Technische Universität München
Max-Planck-Institut für Quantenoptik

Algorithms
for
Classical and Quantum Simulations

Michael Lubasch

Vollständiger Abdruck der von der Fakultät für Physik
der Technischen Universität München
zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. R. Gross
Prüfer der Dissertation:	1. Hon.-Prof. J. I. Cirac, Ph.D.
	2. Univ.-Prof. Dr. M. Zacharias

Die Dissertation wurde am 27.11.2014
bei der Technischen Universität München eingereicht
und durch die Fakultät für Physik am 22.01.2015 angenommen.

Abstract

This thesis contributes to the development of algorithms for the simulation of quantum systems.

The first part of this work considers the classical simulation of quantum many-body systems with the help of Tensor Network States (TNS). TNS denotes a variational class of states that, during the last years, has led to completely new algorithms to tackle many-body problems. Firstly, we focus on a particular subset of TNS called Projected Entangled Pair States (PEPS) in two dimensions. We analyze and compare several recently proposed PEPS algorithms, and we introduce our own algorithmic concept which unifies previously independent methods. Furthermore, we present new procedures that improve the precision and efficiency of PEPS computations, and we provide the currently best PEPS results for the ground state approximation of Heisenberg and quantum Ising Hamiltonians on finite square lattices of sizes 10×10 to 21×21. Secondly, we address the question wether TNS concepts could be useful for density functional theory (DFT). Within DFT, extremely successful numerical procedures have been devised for the analysis of realistic many-electron problems such as, e.g., real molecules or solids. However, there exists no general DFT scheme that provides an error estimate for the final result of the DFT computation or a systematic way of improving it. With the aim of constructing such a highly desirable DFT algorithm, we study different approaches motivated by TNS, identify the necessary characteristics that our approach must have, and present first promising results for one-dimensional Fermi-Hubbard Hamiltonians as well as fermions with a Coulomb interaction.

The second part of this work deals with the quantum simulation of quantum many-body systems by means of ultracold atoms in optical lattices. An unprecedented degree of control over this system allows current experiments to prepare it in such a way that it is described by a Hubbard Hamiltonian whose parameters can be tuned with high accuracy. Firstly, we consider a realization of the strongly interacting Fermi-Hubbard model, i.e., of the $t - J$ Hamiltonian, whose phase diagram is still discussed controversially and might give insight into the mechanism underlying high-T_c superconducting cuprates. A key challenge for the experimental preparation of this Hamiltonian is the low entropy required for the emergence of magnetic order. We propose an adiabatic protocol that starts from a state that can be created experimentally with low entropy. Our calculations for one- and two-dimensional systems demonstrate the feasibility of our protocol for reasonable system sizes and time scales. Secondly, we consider a realization of the Bose-Hubbard model, and we investigate the entanglement properties between the left and right halves of finite one-dimensional systems after the local particle number in one half has been measured. We observe that ground and thermal states are only weakly entangled, and we propose an experimentally realizable protocol that enhances this amount of entanglement by several orders of magnitude and turns out

to be robust against experimental imperfections.

The third part of this work contains an additional algorithm for classical simulation. We demonstrate that the fidelity measure can be applied to the analysis of quantum chaotic energy spectra.

In the appendix, we present an algorithm for the diagonalization of complex symmetric matrices that is particularly suited for the computation of all eigenvalues and eigenvectors of dense matrices. Although this work is not directly related to the rest of the thesis, it was done during this Ph.D. and we include it for completeness.

Zusammenfassung

Diese Doktorarbeit trägt zu der Entwicklung von Algorithmen zur Simulation von Quantensystemen bei.

Der erste Teil dieser Arbeit betrachtet die klassische Simulation von Quanten-Vielteilchensystemen mit Hilfe von Tensornetzwerkzuständen (TNZ). TNZ bezeichnet eine variationelle Klasse von Zuständen welche, während der letzten Jahre, zu komplett neuen Algorithmen geführt hat um Vielteilchenprobleme anzugehen. Erstens konzentrieren wir uns auf eine bestimmte Untermenge von TNZ genannt Projected Entangled Pair States (PEPS) in zwei Dimensionen. Wir analysieren und vergleichen verschiedene kürzlich vorgeschlagene PEPS Algorithmen, und wir führen unser eigenes algorithmisches Konzept ein welches zuvor unabhängige Methoden vereinheitlicht. Außerdem präsentieren wir neue Verfahren welche die Genauigkeit und Effizienz von PEPS Berechnungen verbessern, und wir stellen die momentan besten PEPS Ergebnisse zur Grundzustandsapproximation von Heisenberg und Quanten-Ising Hamiltonians auf endlichen Quadratgittern der Größen 10×10 bis 21×21 zur Verfügung. Zweitens gehen wir der Frage nach ob TNZ Konzepte für Dichtefunktionaltheorie (DFT) nützlich sein könnten. Innerhalb von DFT wurden ausserordentlich erfolgreiche Verfahren zur Analyse von realistischen Vielelektronproblemen wie, z. B., echten Molekülen oder Festkörpern entwickelt. Jedoch gibt es kein allgemeines DFT Schema welches eine Fehlerabschätzung für das Endergebnis einer DFT Berechnung oder einen systematischen Weg dieses zu verbessern zur Verfügung stellt. Mit dem Ziel solch einen sehr wünschenswerten DFT Algorithmus zu konstruieren, untersuchen wir unterschiedliche, von TNZ motivierte Ansätze, identifizieren die notwendigen Charakteristiken die unser Ansatz haben muss, und präsentieren erste vielversprechende Ergebnisse für eindimensionale Fermi-Hubbard Hamiltonians als auch Fermionen mit Coulomb-Wechselwirkung

Der zweite Teil dieser Arbeit handelt von der Quantensimulation von Quanten-Vielteilchensystemen mittels ultrakalter Atome in optischen Gittern. Ein zuvorniedagewesenes Maß an Kontrolle über dieses System erlaubt heutigen Experimenten es in solcher Weise zu präparieren dass es durch einen Hubbard Hamiltonian beschrieben wird wessen Parameter mit hoher Genauigkeit eingestellt werden können. Erstens betrachten wir die Realisierung des stark wechselwirkenden Fermi-Hubbard Modells, d. h., des $t - J$ Hamiltonians, dessen Phasendiagramm noch kontrovers diskutiert wird und in den Mechanismus Einsicht geben könnte welcher Hochtemperatur-supraleitenden Kupraten unterliegt. Eine besondere Herausforderung für die experimentelle Präparierung dieses Hamiltonians ist die niedrige Entropie die zur Entstehung von magnetischer Ordnung benötigt wird. Wir schlagen ein adiabatisches Protokoll vor, das von einem Zustand aus startet, der experimentell mit niedriger Entropie erzeugt werden kann. Unsere Berechnungen für ein- und zweidimensionale Systeme demon-

strieren die Machbarkeit unseres Protokolls für vernünftige Systemgrößen und Zeitskalen. Zweitens betrachten wir eine Realisierung des Bose-Hubbard Modells, und wir untersuchen die Verschränkungseigenschaften zwischen den linken und rechten Hälften endlicher eindimensionaler Systeme nachdem die lokale Teilchenzahl in einer Hälfte gemessen wurde. Wir beobachten dass Grundzustände und thermische Zustände nur schwach verschränkt sind, und wir schlagen ein experimentell realisierbares Protokoll vor welches diese Menge an Verschränkung um mehrere Größenordnungen erhöht und sich als robust gegen experimentelle Ungenauigkeiten herausstellt.

Der dritte Teil dieser Arbeit beinhaltet einen zusätzlichen Algorithmus zur klassischen Simulation. Wir zeigen dass das Fidelity-Maß zur Analyse von quantenchaotischen Energiespektren verwendet werden kann.

Im Appendix stellen wir einen Algorithmus zur Diagonalisierung von komplexen symmetrischen Matrizen vor, welcher besonders für die Berechnung aller Eigenwerte und Eigenvektoren von dichten Matrizen geeignet ist. Obwohl diese Arbeit nicht direkt mit dem Rest der Doktorarbeit zusammenhängt, wurde sie während dieser Doktorarbeit fertiggestellt, und wir fügen sie der Vollständigkeit halber ein.

Publications

1. Chapter 1:
   Unifying projected entangled pair state contractions
   M. Lubasch, J. I. Cirac, and M.-C. Bañuls
   New Journal of Physics **16**, 033014 (2014).

2. Chapter 2:
   Algorithms for finite projected entangled pair states
   M. Lubasch, J. I. Cirac, and M.-C. Bañuls
   Physical Review B **90**, 064425 (2014).

3. Chapter 3:
   Density functional theory beyond the local density approximation
   M. Lubasch, J. I. Fuks, H. Appel, A. Rubio, J. I. Cirac, and M.-C. Bañuls
   In preparation.

4. Chapter 4:
   Adiabatic Preparation of a Heisenberg Antiferromagnet Using an Optical Superlattice
   M. Lubasch, V. Murg, U. Schneider, J. I. Cirac, and M.-C. Bañuls
   Physical Review Letters **107**, 165301 (2011).

5. Chapter 5:
   Dynamical enhancement of spatial entanglement in massive particles
   M. Lubasch, F. Mintert, and S. Wimberger
   Physical Review A **84**, 063615 (2011).

6. Chapter 6:
   Detection of avoided crossings by fidelity
   P. Plötz, M. Lubasch, and S. Wimberger
   Physica A: Statistical Mechanics and its Applications **390** (2011) 1363.

7. Appendix A:
   Generalized Householder transformations for the complex symmetric eigenvalue problem
   J. H. Noble, M. Lubasch, and U. D. Jentschura
   European Physical Journal Plus **128** (2013) 93.

Acknowledgements

First of all, I would like to thank my supervisor Ignacio Cirac. Thanks to him, I could work on the most exciting physics problems, in an outstanding scientific environment. I am deeply grateful for his generous support, his ingenious ideas in our projects, and all discussions with him. He has taught me what the scientific method is about.

I am also particularly thankful to my advisor Mari-Carmen Bañuls. She always had time for me and all my problems. Her positive attitude and her deep understanding of physics as well as of numerics, which she always communicates very clearly, were absolutely crucial for this work. I am deeply indebted to her for the countless discussions we had.

This work would not have been possible without my other collaborators. Valentin Murg was a great help when I started to program my PEPS code. Johanna Fuks, Heiko Appel, and Angel Rubio thoroughly introduced me to density functional theory. Ulrich Schneider patiently explained me numerous details of optical lattice experiments. I could keep up my interest in entanglement measures thanks to Florian Mintert, in quantum chaos thanks to Sandro Wimberger, and in high-precision numerics thanks to Ulrich Jentschura. I am additionally indebted to Patrick Plötz and Jonathan Noble for their important contributions to our projects.

I would also like to use this opportunity to thank Rudolf Gross and Martin Zacharias for being part of my Ph.D. commission.

The atmosphere in our group has always been great and I simply thank the complete group for that, since thanking all the wonderful characters of our group in detail would fill many pages. However, I have to explicitly acknowledge invaluable discussions about physics (and also many other things) with Román Orús, Hong-Hao Tu, Matteo Rizzi, Géza Giedke, Norbert Schuch, Gemma De las Cuevas, Maarten van den Nest, and Manuel Endres. And I also have to mention my previous and current office mates, Christine Muschik, Heike Schwager, Anika Pflanzer, Oliver Buerschaper, Martin Schütz, and András Molnár, who have always made office hours (and also other hours) so enjoyable. And furthermore, I must thank Oriol Romero-Isart, Fernando Pastawski, Leonardo Mazza, Birger Horstmann, Eric Kessler, and Stefan Kühn, for all the fun we had together.

Finally, I would like to express my deep gratitude to my parents, who have always supported me unconditionally.

Funding from the Max Planck Institute of Quantum Optics and the Nanosystems Initiative Munich is gratefully acknowledged.

Contents

Chapter 0

Introduction

Programming is understanding.
- Kristen Nygaard

Quantum mechanics explains how the equations describing quantum systems are written down but not how they are solved. For some problems exact analytical solutions exist, and then also problems in the vicinity of such problems might be solvable by means of perturbation theory. However, many interesting quantum systems can only be theoretically analyzed with the help of clever algorithms running on powerful technical devices.

Often quantum many-body systems fall into this category. Although classical computers can, in principle, handle such systems exactly, the exact quantum mechanical description of a system comprising N constituents requires the storage of a number of parameters that grows exponentially with N. Then any calculation performed for such a system, as, e.g., the evaluation of an observable, requires the execution of a number of operations that depends exponentially on N. As a consequence, exact calculations with classical computers are limited to relatively small system sizes N.

Much larger values of N can be treated by classical computers approximately, thanks to smart algorithmic approaches. Prominent algorithms of that kind are the quantum Monte Carlo method [1, 2], density functional theory [3, 4], and the density matrix renormalization group [5]. Each of these algorithms approximates a different aspect of the quantum many-body problem and, therefore, functions particularly well for a specific class of systems, namely, for which that particular aspect is not of crucial importance. Because these methods are already known for some time, their strengths have been identified as well as their fundamental limitations. The quantum Monte Carlo method is fundamentally hindered by the sign problem, which shows up, e.g., for frustrated or fermionic systems. Density functional theory does, typically, not provide an error estimate for a final result, which then has to be checked with an experiment or with another numerical method. The density matrix renormalization group is, in principle, limited to one-dimensional systems with not too much entanglement.

In recent years, a new class of algorithms has been introduced that generalizes the density matrix renormalization group to higher-dimensional systems with more entanglement, can provide error estimates for their computations, and can be applied to frustrated or fermionic systems. These new algorithms are based on

a description of quantum many-body states as Tensor Network States (TNS). The concept of TNS is very general and has led to significant algorithmic progress in many different areas, ranging from the computation of higher-dimensional classical partition functions to the general simulation of ground, thermal, time evolved, or higher-dimensional quantum systems, their variational optimization and evaluation, and their adaptation to, e.g., fermionic degrees of freedom [6–23]. First results for problems like the Kagome Heisenberg model [24] or the $t - J$ Hamiltonian on a two-dimensional square lattice [25, 26] compare well to the best results of longer existing and further developed alternative algorithms. These results demonstrate the potential of TNS, and, thus, of classical simulations, to contribute to the solution of truly outstanding problems in condensed matter theory.

Another technical device, different from a classical computer, for the simulation of large quantum many-body systems, is a quantum simulator [27–29]. This is a quantum system that can be very well controlled and that is described by a known Hamiltonian whose parameters can be adjusted accurately. The idea is, that this Hamiltonian shall also describe other interesting quantum many-body problems, which are hard to solve by other means, and can then be investigated by analyzing the properties of the quantum simulator. Numerous quantum simulators have been realized and their potential for the simulation of interesting models ranging from condensed matter to quantum field theory has been demonstrated. They were based on different physical systems such as, e.g., photons, ultracold atoms in optical lattices, arrays of trapped ions, quantum dots, or Josephson junctions, each one having different advantages and disadvantages.

During the last years, an unprecedented degree of control has been achieved over ultracold atoms in optical lattices [30–33]. This system is realized by loading a cold gas of atoms into an optical lattice, i.e., a periodic potential resulting from counter-propagating laser fields. Since the atom number as well as the lattice size can, in principle, be increased easily without loss of coherence, an advantage of this quantum simulator is its easy scalability. When the atoms are cold enough and the lattice is deep enough, this system is described by a one-band Hubbard Hamiltonian [34]. That this regime can really be reached has been demonstrated by experiments which successfully prepared quantum phases of both the Bose-Hubbard [35] and the Fermi-Hubbard model [36, 37]. Nowadays, these models can be analyzed experimentally with a remarkable resolution of single lattice sites [38, 39]. The tremendous progress of this research field within few years and the many possibilities of this quantum simulator to outperform current classical computers, e.g., by simulating higher-dimensional Fermi-Hubbard Hamiltonians [37] or long time evolutions of quantum many-body systems [40], suggest that this technical device might one day allow to understand quantum phenomena that are not accessible by any other means.

In this thesis, we develop algorithms for classical computers, i.e., classical simulations, and for quantum simulators, i.e., quantum simulations. The major part of this work is concerned with the quantum many-body problem, and presents new approaches for classical simulations based on TNS and for quantum simulations based on ultracold atoms in optical lattices. These two approaches are not disconnected: The classical TNS tools turn out to be of fundamental utility for the design and validation of quantum simulations. A minor part of this work contains additional algorithmic proposals that are useful for classical simulations not necessarily of many-body systems. We have therefore structured this thesis in three parts: parts I and II for classical and quantum simulations of quantum many-body systems respectively, and part III, and the appendix, for other classical simulations.

Part I contains contributions to two different fields of TNS: to a specific subset of TNS called Projected Entangled Pair States (PEPS), in chapters 1 and 2, and to the use of TNS concepts in density functional theory (DFT), in chapter 3.

While PEPS have very promising properties for the description of strongly correlated quantum many-body systems, in general, their corresponding tensor networks can only be evaluated approximately and numerical calculations with them are characterized by a high computational cost [41]. During the last years, many PEPS algorithms have come out that propose different strategies for the evaluation of PEPS tensor networks and are characterized by different computational costs [22, 42, 43]. In chapter 1, after a brief introduction to PEPS, we analyze and compare several of these proposals, and we find that those with low computational cost, typically, also lead to low precision in the result of the PEPS computation. By introducing our own algorithmic concept, we are able to unify previously independent methods, and obtain a PEPS algorithm in which the desired precision can be adjusted together with the computational cost. We have published this analysis in reference [44]. Chapter 2 is based on the published article [45], in which we study other numerical aspects of PEPS algorithms, present new procedures that increase their precision and decrease their computational cost, and demonstrate the validity of the PEPS ansatz for the ground state approximation of Heisenberg and quantum Ising Hamiltonians on finite square lattices of sizes 10×10 to 21×21. Although this chapter is relatively technical, the aspects we discuss are essential, and can not be ignored by any implementation of PEPS that aims to obtain significant results in a practical application.

Chapter 3 focuses on a novel application of TNS techniques to a field which in principle does not seem connected to the more common cases of use of these techniques, namely DFT. DFT is a general theory for the classical simulation of quantum many-body systems that has proven extremely powerful for the study of real many-electron systems. However, typical DFT algorithms provide neither an error estimate for the final result nor a way of systematically improving it. This holds, in particular, for the widely used DFT routines based on the local density approximation (LDA). The development of general extensions beyond the LDA, that would allow error estimates or systematic improvements of general DFT results, is a topic of intensive research [46–50] but has only obtained moderate success. In chapter 3, we present first promising results for the construction of such a general DFT algorithm with the help of TNS concepts.

Part II contributes to two different kinds of quantum simulations that can be performed with ultracold atoms in optical lattices: of a $t - J$ Hamiltonian resulting from a strongly interacting Fermi-Hubbard system, in chapter 4, and of a Bose-Hubbard Hamiltonian, in chapter 5.

The phase diagram of the $t - J$ Hamiltonian on a two-dimensional square lattice might contain the answer to a "holy grail" question of condensed matter physics, namely, what the mechanism is that leads to high-temperature superconductivity in cuprates [51]. So far, all attempts, in particular also by classical computers, of characterizing the relevant parts of this phase diagram have encountered serious obstacles and have not led to definite conclusions. The first successful quantum simulation of this problem might, therefore, not only provide crucial insight into an important physical phenomenon but would, furthermore, outperform current classical simulations. Ultracold atoms in optical lattices are a very natural choice for the construction of a quantum simulator to approach this problem, since two species of sufficiently cold and strongly interacting fermions in a sufficiently deep optical lattice are described by the $t - J$ Hamiltonian. However, the key chal-

lenge of current experiments is the low temperature, or entropy, required for the emergence of order within the two species of fermions. In chapter 4, we focus on a particular ground state of the $t - J$ model, namely, the Heisenberg antiferromagnet, and we propose its creation via an adiabatic protocol, starting from a state that can be experimentally prepared with low entropy and then adiabatically raising a second optical lattice. Our analysis of this protocol in one and two dimensions concludes that emergence of the desired order can be expected for feasible system sizes within feasible time scales, and is published in reference [52].

Chapter 5 discusses the entanglement properties of the Bose-Hubbard model when the entangled entities are the spatial Fock states corresponding to the left and right half of a finite optical lattice. The choice of these entangled entities has the advantage, that, by experimentally raising the intermediate barrier between the two lattice halves, two spatially separated quantum systems are obtained, that each can be locally addressed, in the spirit of the Alice and Bob paradigm in quantum information theory [53]. We furthermore demand that each entangled entity has a fixed total particle number, i.e., we measure the total particle number of each lattice half before we compute the entanglement between the two halves. Our results are published in reference [54], in which we show that ground and thermal states feature only low amounts of entanglement that, however, can be increased by several orders of magnitude when the lattice height is periodically modulated. This analysis demonstrates a way of achieving a high degree of control over the entanglement produced in an experimental setup.

Part III presents a different work, which departs from the common theme of the previous chapters. It reflects, nevertheless, aspects that are of relevance for the study of quantum systems: namely the characterization of quantum chaos, in chapter 6. Since the procedures introduced in this chapter are mostly relevant for the numerical understanding of quantum problems, we have included them under the name of classical simulations.

For the identification of a quantum critical point, a fidelity measure defined via the overlap of two nearby ground states has proven successful [55]. This can be understood from the fact that a quantum critical point separates two different quantum phases, i.e., ground states of a certain quantum mechanical Hamiltonian, and, while the overlap between nearby ground states should be close to 1 within a quantum phase, it should differ from 1 when a quantum phase transition takes place in a thermodynamic system. In a finite system, a quantum critical point appears as an avoided crossing in the energy spectrum between the ground and the first excited state, and one might ask the question wether a natural fidelity can be defined to characterize such avoided crossings. In chapter 6, we propose such a fidelity measure and lift its applicability to many eigenstates in the context of quantum chaos. Our study is published in reference [56] and demonstrates, e.g., that energy level-spacing statistics, which are the basis of standard procedures for the analysis of quantum chaos, can be obtained from our fidelity. However, while the standard quantum chaos procedures require many energy levels in order to make a clear statement about their statistics, our fidelity measure can equally be applied to very few energy levels and can thus characterize quantum chaos also only in a small part of the spectrum.

Finally, in the appendix, we include another work that can be useful for the numerical analysis of quantum systems: on the diagonalization of complex symmetric matrices, in appendix A.

If a complex symmetric matrix A, i.e., with complex entries and the property $A = A^{\mathrm{t}}$, is diagonalizable, then it can be diagonalized by a complex orthogonal

matrix Z, i.e., having the inverse Z^t. Although this is a well-known fact, common numerical libraries, such as LAPACK or Intel's MKL, do not provide routines that make use of this fact. In appendix A, we develop our own algorithm, that makes use of complex orthogonality, and we observe a speedup factor 2 compared with the best available LAPACK routine. This numerical analysis is carried out in the context of $\mathcal{PT}$-symmetric quantum mechanics where complex symmetric matrices arise and is published in reference [57].

Part I

Classical Simulations of Quantum Many-Body Systems

Chapter 1

Unifying Projected Entangled Pair State Contractions

Projected Entangled Pair States (PEPS) are a promising ansatz for the study of strongly correlated quantum many-body systems in two dimensions. However, due to their high computational cost, developing and improving PEPS algorithms is necessary to make the ansatz widely usable in practice. The approximate contraction of a PEPS tensor network is a fundamental ingredient of any PEPS algorithm, required for the optimization of the tensors in ground state search or time evolution, as well as for the evaluation of expectation values. An exact contraction is in general impossible, and the choice of the approximating procedure determines the efficiency and accuracy of the algorithm. We analyze different previous proposals for this approximation, and show that they can be understood via the form of their environment, i.e. the operator that results from contracting part of the network. This provides physical insight into the limitation of various approaches, and allows us to introduce a new strategy, based on the idea of clusters, that unifies previous methods. The resulting contraction algorithm interpolates naturally between the cheapest and most imprecise and the most costly and most precise method. We benchmark the different algorithms with finite PEPS, and show how the cluster strategy can be used for both the tensor optimization and the calculation of expectation values. Additionally, we discuss its applicability to the parallelization of PEPS and to infinite systems (iPEPS). This chapter is published in reference [44].

1.1 Introduction

In the last years, Tensor Network States (TNS) have revealed as a very promising choice for the numerical simulation of strongly correlated quantum many-body systems. A sustained effort has led to significant conceptual and technical advancement of these methods, e.g. [7–13, 15–17, 41, 42, 58–68].

In the case of one-dimensional systems, Matrix Product States (MPS) are the variational class of TNS underlying the density matrix renormalization group (DMRG) [69]. Insight gained from quantum information theory has allowed the understanding of DMRG's enormous success at approximating ground states of spin chains, and the extension of the technique to dynamical problems [7–10,12,58] and lattices of more complex geometry [11, 15, 16, 59].

Projected Entangled Pair States (PEPS) [11] constitute a family of TNS that naturally generalizes MPS to spatial dimensions larger than one and arbitrary lattice geometry. As MPS, PEPS incorporate the area law by construction, what makes them a very promising variational ansatz for strongly correlated systems which might not be tractable by other means, e.g. frustrated or fermionic states where quantum Monte Carlo methods suffer from the sign problem. Although originally defined for spin systems, PEPS have been subsequently formulated for fermions [18–20, 22], and their potential in the numerical simulation of fermionic phases has been demonstrated [21–23, 25]. First PEPS results for such problems in condensed matter as the $t - J$ model or the Kagome antiferromagnet compare well to the best currently known results achieved by other means [24–26]. But in contrast to MPS, even local expectation values cannot be computed exactly in the case of PEPS. This is because the exact evaluation of the TN that represents the observables has an exponential cost in the system size. The same difficulty affects the contraction of the TN that surrounds a given tensor, the so-called *environment*, required for the local update operations in the course of optimization algorithms. It is nevertheless possible to perform an approximate TN contraction with controlled error, albeit involving a much higher computational cost than in the case of MPS. This limits the feasible PEPS simulations to relatively small tensor dimensions.

Lately, several algorithmic proposals have come out [42, 43, 70, 71] that make larger tensors accessible by using new approximations in the PEPS contraction. Although these approaches allow the manipulation of a larger set of PEPS, their assumptions have an impact on the accuracy of the ground state approximation, and this accuracy is not always directly related to the maximum bond dimension the algorithm can accommodate. It is nevertheless possible to analyze the various approximations from the unifying point of view of how they treat the environment contraction, which in turn has a physical meaning. This allows us to understand how a given strategy may attain only a limited precision approximation to the ground state, even when its computational cost allows for large bond dimensions.

Contraction strategies proposed in the literature include the original PEPS method [11], the simple update [42],[1] and the single-layer [43] algorithm. In this chapter we investigate these algorithms from the unifying perspective introduced above, and present a new contraction scheme that naturally interpolates between the cheapest and most imprecise method and the most expensive and precise one. We illustrate our findings with finite-size PEPS with open boundary conditions. A finite PEPS, in which each tensor contributes independent variational parameters, is less biased than its infinite counterpart iPEPS [17], in which a unit cell of variational tensors is replicated infinitely often and the form of that unit cell can have an effect on the observed order. However, all our results apply also to iPEPS, and, as we will argue, provide the basis for a new promising approach in that context.

1.1.1 Reader's Guide

This chapter is structured as follows. In section 1.2 we briefly introduce the basic PEPS concepts and original algorithms. Section 1.3 reviews the simple update method introduced in [42], and analyzes its performance with finite PEPS. We find that the resulting ground state energies can be less accurate than those of the original algorithm when the environment form assumed in the method is far from the

[1] In the original proposal [42] the simple update does not denote a contraction strategy but a tensor update procedure for imaginary time evolution. However, the environment used in this update corresponds also to a certain contraction method, as we will show later.

true one. The single-layer algorithm proposed in [43] can be seen as a first, conceptual generalization of the simple update, and we investigate it in section 1.4. We show that the error introduced by this method exhibits a strong system size dependence, in contrast to the original algorithm. With the gained insight, in section 1.5 we formulate and investigate a new strategy for the environment approximation based on the idea of clusters, that is applicable to both the tensor update and the computation of expectation values. Furthermore, we discuss how this cluster strategy is also beneficial to the parallelization of PEPS as well as to the infinite case, i.e. iPEPS. Finally in section 1.6 we briefly summarize our results.

1.2 PEPS: Basic Concepts and Algorithms

We consider a system of N quantum particles with Hilbert spaces of dimensions d_l, for $l = 1, \ldots, N$, spanned by individual bases $\{|s_l\rangle\}$, with $s_l = 1, \ldots, d_l$. Projected Entangled Pair States (PEPS) [11] are states for which the coefficients in the product basis are given by the contraction of a tensor network,

$$|\psi^{PEPS}\rangle := \sum_{s_1, s_2, \ldots, s_N} \mathcal{F}(A_1^{s_1} A_2^{s_2} \ldots A_N^{s_N}) |s_1 s_2 \ldots s_N\rangle \quad ,$$

with one tensor A_l per physical site. The tensors A_l are arranged in a certain lattice geometry and connected to neighboring sites by shared indices, such that the coordination number, c, of a certain lattice site coincides with the number of connecting indices. The latter are called virtual, and apart from them, each tensor A_l possesses one physical index s_l, standing for the physical degree of freedom of the quantum particle on lattice site l. The function $\mathcal{F}$ represents the contraction of all virtual indices. Each of them ranges up to the parameter D which is named *bond dimension*. D determines the number of variational parameters of each tensor, namely dD^c.[2] The bond dimension sets an upper bound to the entanglement entropy of the state, in fulfillment of the area law. In particular, if we consider a subsystem delimited by a regular shape of side length ℓ, the entropy of its reduced density matrix, ρ_ℓ, is upper-bounded by $S(\rho_\ell)_{\max} \propto \ell^{dim-1} \log(D)$, where dim denotes the system's dimensionality.

In one dimension, PEPS are called Matrix Product States (MPS). For periodic boundary conditions, every MPS tensor has two virtual indices and is thus, for fixed physical index, a matrix of dimension $D \times D$. In that case, the function $\mathcal{F}$ becomes a trace over a product of matrices. For open boundary conditions, almost all MPS tensors have two virtual indices, only the two boundary tensors have just one virtual index. In that case, the boundary tensors, for fixed physical index, are vectors and the function $\mathcal{F}$ is not needed. The MPS family constitutes a very good approximation to thermal states of local Hamiltonians as well as to ground states of gapped local Hamiltonians in one dimension [72, 73]. The computational cost of MPS algorithms, including ground state search and time evolution, scales as $\mathcal{O}(ND^3)$, for systems with open boundary conditions. The one-dimensional TN structure of the ansatz allows its exact contraction and therefore, e.g., the exact computation of (local) expectation values and the norm. All these aspects make MPS a very powerful tool for the analysis of one-dimensional quantum many-body problems, and we refer the interested reader to the review articles [41, 74–76] for more details on MPS algorithms and applications.

[2] In the case of open boundary conditions the tensors on the boundaries have fewer virtual indices and variational parameters.

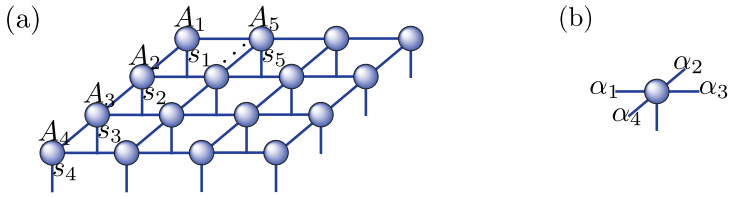


Figure 1.1: (a) A 4×4 PEPS $|\psi^{PEPS}\rangle := \sum_{s_1,s_2,\ldots,s_{16}} \mathcal{F}(A_1^{s_1} A_2^{s_2} \ldots A_{16}^{s_{16}})|s_1 s_2 \ldots s_{16}\rangle$ on a square lattice. (b) A tensor from the interior with four virtual indices α_1 to α_4. The function $\mathcal{F}$ performs the contraction of the tensor network by summing over connected virtual indices.

In any dimension, the PEPS family is known to represent a very good approximation for thermal states of local Hamiltonians and it is believed to represent a very good approximation for ground states of gapped local Hamiltonians [77]. The computational cost of higher-dimensional PEPS algorithms is larger than for MPS and depends strongly on the considered algorithm as well as the considered geometry. Throughout this chapter, we consider PEPS on two-dimensional square lattices of size $N = L \times L$ with side length L and open boundary conditions. An example is shown in figure 1.1.

For general PEPS, the computation of an expectation value or even the norm is known to be hard [78], like the evaluation of a two-dimensional classical partition function [6]. Hence only an approximate contraction is possible for already moderate lattice sizes. The originally proposed algorithm [11, 41] approximates the two-dimensional TN of an expectation value $\langle\psi|\hat{O}|\psi\rangle$ or the norm $\langle\psi|\psi\rangle$ by means of a succession of one-dimensional MPS contractions, as sketched in figure 1.2 (a) for the norm. The procedure starts by identifying two opposite sides of the TN, e.g. the upper- and bottommost rows, with MPS, and each of the intermediate rows with a Matrix Product Operator (MPO) [79]. Beginning from one of the edges, the contraction of the last row with the immediately neighboring one is then a MPS-MPO product, which can be optimally approximated by a MPS of fixed bond dimension, D'. By repeating the procedure from both opposite sides, successive MPS-MPO approximations lead to a representation of both halves of the TN by MPS. Finally the row in the center is contracted between both MPS to give the approximate expectation value or norm.

At each point of this procedure, the obtained MPS approximates the boundary between the contracted part of the network and the rest. This MPS can be interpreted as an operator that maps the virtual indices of the ket boundary to the bra and thus we will refer to it as the *boundary MPO*, shown in figure 1.2 (b).[3] The approximate contraction of the norm has the leading cost $\mathcal{O}(D^4 D'^3) + \mathcal{O}(dD^6 D'^2)$, and thus both cost and error are determined solely by the bond dimension D' of the boundary MPO. Although in principle D' could scale exponentially with the number of rows, in practice it typically scales as $D' \propto D^2$ independent of the system size, such that the original contraction [11,41] has the total computational cost $\mathcal{O}(D^{10})$.

For certain problems, this observed mild scaling can be given a more rigorous ground. Indeed, the boundary MPO can be interpreted as the thermal state of a Hamiltonian defined on the virtual degrees of freedom of the boundary. This

[3] If the contraction was exact, this boundary MPO would always be positive in the case of the norm computation, due to the bra-ket structure of all the rows involved, but this positive character is in general lost in the truncation.

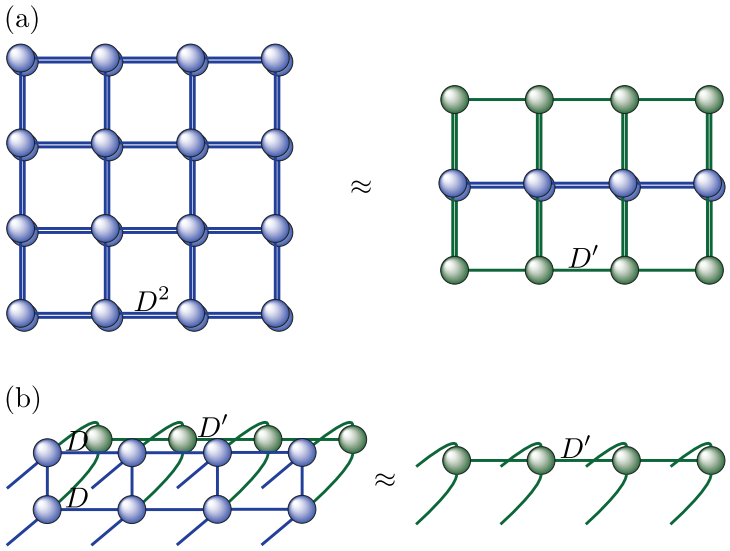


Figure 1.2: Original contraction [11,41] of the norm $\langle\psi|\psi\rangle$. (a) The TN corresponding to $\langle\psi|\psi\rangle$ results from figure 1.1 (a) as ket and as bra and contraction over the physical indices (left). The hard computation of this two-dimensional PEPS TN is approximated by an efficiently contractible one-dimensional MPS expectation value (right). (b) This is done by successively approximating the action of a bulk row of the PEPS TN, of bond dimension D, on a boundary MPO, of bond dimension D', (left) by a new boundary MPO (right). The latter can be determined with computational cost $\mathcal{O}(D^4D'^3) + \mathcal{O}(dD^6D'^2)$.

boundary Hamiltonian is obtained by identifying its excitation spectrum with the entanglement spectrum of the state [80]. Such a construction is very natural in the framework of PEPS and establishes a holographic principle [81]. While PEPS are expected to represent the low energy sector of local Hamiltonians well, it has not been proven when expectation values can be computed efficiently with them. However, if the boundary Hamiltonian is local, as evidence suggests for gapped models [81], the corresponding thermal state will be efficiently approximated by a MPO [77].

In the following, we obtain the PEPS approximation to the ground state of a certain Hamiltonian by means of imaginary time evolution. It is based on the idea that $e^{-t\hat{H}}|\psi_0\rangle$ converges to the ground state of H exponentially fast with t, as long as the ground state is not degenerate and has non-vanishing overlap with the initial state, $|\psi_0\rangle$. In the context of TNS [8], the initial state is chosen within the appropriate TNS family, and a Suzuki-Trotter decomposition of the evolution operator $U(t) = e^{-t\hat{H}} = (e^{-\tau\hat{H}})^n$ is applied to local Hamiltonians, such that each step of the evolution, $\tau = t/n$, is approximated by a product of local *Trotter gates*. The resulting state after each gate or set of gates, is again approximated by an adequate TNS. In particular, the action of a certain operator $\hat{O}$ on a PEPS $|\phi\rangle$ can be approximated by a new PEPS $|\psi\rangle$ by minimizing the cost function $\mathrm{d}(|\psi\rangle) = |||\psi\rangle - \hat{O}|\phi\rangle||^2$. We perform this minimization for each gate via an alternating least squares (ALS) scheme, optimizing one tensor at a time while the others are fixed, and sweeping only over the tensors on which the Trotter gate acts. The optimal tensor at position l is the solution of a system of linear equations $N_l\vec{A}_l = \vec{b}_l$, where the norm matrix N_l is defined from the tensor network $\langle\psi|\psi\rangle$ by leaving out the tensor A_l in the ket

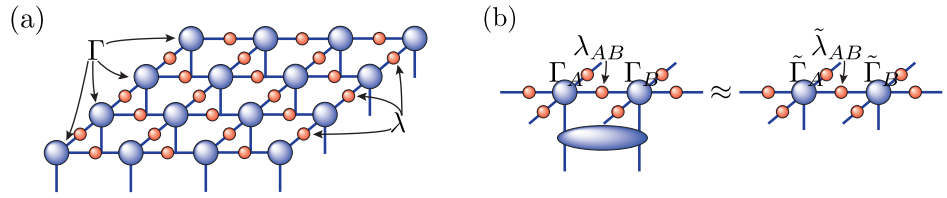


Figure 1.3: (a) A 4×4 PEPS $|\psi^{PEPS,SU}\rangle := \sum_{s_1,s_2,\ldots,s_{16}} \mathcal{F}(\Gamma_1^{s_1}\lambda_1 \ldots \Gamma_{16}^{s_{16}})|s_1 s_2 \ldots s_{16}\rangle$ of the simple update (SU) form, composed of Γ tensors and λ matrices. (b) Assuming nearest-neighbor Trotter gates, the 6 λ matrices surrounding a tensor pair Γ_A and Γ_B are sufficient for the update of this pair and its λ_{AB}.

and A_l^* in the bra, and the vector $\vec{b}_l$ results from the tensor network $\langle\psi|\hat{O}|\phi\rangle$ by removing A_l^* from the bra.

The environment of a tensor at site l is the open TN that results when this tensor and its adjoint are removed from the norm of the state. Contracting the environment is necessary to evaluate N_l and b_l, which determine the local equation for A_l. Such contraction can only be carried out approximately, and the approximation strategy is decisive both for the accuracy and for the computational cost of the algorithm.[4] Because we process the Trotter gates one after another and modify only the tensors on which the gate directly acts, in the following we focus on the contraction of the norm TN around a Trotter gate. The importance of the environment approximation has been recognized also in other works, e.g. [82], or in the different context of tensor renormalization group algorithms [67] where a more precise environment representation lead to significant improvements [68].

1.3 Simple Update

The simple update method (SU) [42] directly generalizes the one-dimensional TEBD [7–10, 14] and proposes for the PEPS tensors the decomposition

$$|\psi^{PEPS,SU}\rangle \quad := \sum_{s_1,s_2,\ldots,s_N} \mathcal{F}(\Gamma_1^{s_1}\lambda_1\Gamma_2^{s_2}\lambda_2 \ldots \Gamma_N^{s_N})|s_1 s_2 \ldots s_N\rangle \qquad ,$$

formally analogous to the canonical form for MPS [7], where the Γ_l are tensors with the same dimensions as the original A_l, and the λ_l are $D \times D$ diagonal and positive matrices, see figure 1.3. Although in the case of PEPS, the λ matrices do not have the clear meaning of their one-dimensional counterparts, the SU has proven a successful strategy in the context of iPEPS, starting with [42]. This success can be attributed, on the one hand, to the low computational cost of the tensor update, which is why large values of D can be reached easily. Indeed, the SU rule requires only the λ matrices that are closest to a tensor pair and as a consequence has the computational cost $\mathcal{O}(D^5)$. On the other hand, all parts of the algorithm are well-conditioned. These positive aspects arise at the expense of an oversimplified representation of the environment as separable and local, that, in general, can only be a rough approximation of the true environment.

In order to illustrate its performance, we employ the SU to find ground states

[4] In general, also the tensor update operations contribute to the final cost. If we restrict the variational parameters to the reduced tensor [22, 70], the linear equations can be solved with $\mathcal{O}(D^6)$ operations.

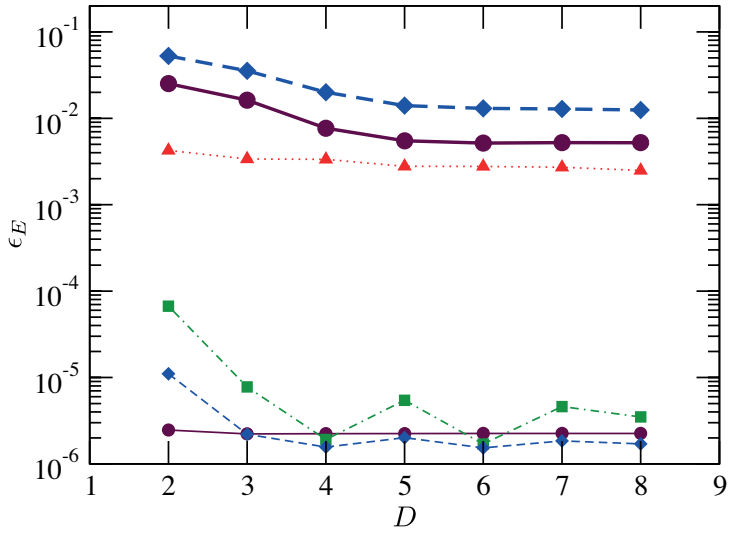


Figure 1.4: Relative energy error $\epsilon_E := |E(D) - E_0|/|E_0|$ of the SU, $E(D)$, with regard to the exact ground state energy, E_0. We consider the Ising model (thin lines) on a 4×4 lattice with $B = 1.0$ (solid), 2.0 (dotted), 3.0 (dash-dotted), 4.0 (dashed), and the Heisenberg model (thick lines) on a 4×4 (dashed) and 10×10 (solid) lattice.

of the quantum Ising Hamiltonian with transverse field

$$\hat{H} = -\sum_{\langle l,m \rangle} \sigma_l^z \otimes \sigma_m^z - B \sum_l \sigma_l^x \tag{1.1}$$

and of the Heisenberg model

$$\hat{H} = \sum_{\langle l,m \rangle} \vec{S}_l \cdot \vec{S}_m \quad , \tag{1.2}$$

where $\langle l, m \rangle$ denotes pairs of neighboring sites l and m. In the context of finite PEPS considered here, to the best of our knowledge, the SU had not been yet used. We determine the ground state of a particular problem by evolving an initial state long enough in imaginary time and successively decreasing the time step τ until convergence.

Figure 1.4 compares SU results to exact ground state energies. The scheme performs remarkably well for the Ising model at $B = 1.0$, 3.0, and 4.0, where the relative energy error is below 10^{-5} already with $D = 3$. But at $B = 2.0$ and for the Heisenberg Hamiltonian, we observe that the energy does not improve significantly beyond a certain value of the bond dimension, and remains far from the exact value. We identify this as a limitation, not of the ansatz, but of the update procedure, since the original PEPS algorithm [11] achieves for the Heisenberg model on a 4×4 lattice with $D = 3$ already lower energy than any of the SU values from the figure, and with $D = 4$ it attains an energy per site -0.5739 already very close to the exact value -0.5743. Although we observed that the SU result can depend on the initial state, in particular for the larger bond dimensions,[5] this dependence

[5]This became evident by running the algorithm with various values of D separately, each run starting from a product state in which the tensors' zeroes were replaced by random numbers from the interval $[-0.01, 0.01]$. In this setting we observed that the SU can lead to a final $D = 8$ energy above the $D = 7$ value.

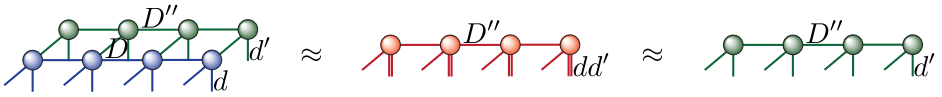


Figure 1.5: SL contraction of the norm $\langle\psi|\psi\rangle$. As explained in the text, the approximation of figure 1.2 is achieved by means of operations in the ket alone.

appeared not so strongly when we increased the bond dimension successively during imaginary time evolution.

1.4 Single-Layer

The single-layer (SL) algorithm for the computation of the norm $\langle\psi|\psi\rangle$ was presented in [43]. This method takes into account the bra-ket structure of the norm TN, and maintains it and hence the positive character of the environment while the contraction of the network progresses from one edge. To achieve this structure, the approximate contraction proceeds by successive MPO-MPS operations, like the original algorithm, but this time performed on a single layer of the norm TN. Then the boundary, i.e. the already contracted part of the network, is always approximated by a purification MPO [12], namely the result of tracing out a part of the physical indices at every site of a MPS. This MPS is assumed to have some maximum bond dimension, D'', and physical dimension $D \times d'$, where d' is the dimension of the traced out degrees of freedom, what we call *purification bond*. In this way, the local and separable environment defined by the λ matrices in the SU is generalized by means of purification MPS that can better capture non-local and non-separable boundary correlations. Moreover, the boundary purification MPO is always a positive operator, and it allows to devise a stable tensor update procedure for imaginary time evolution [43].

The SL operations take place in the two steps shown in figure 1.5. First, the ket part of a PEPS row is applied as a MPO to the MPS of the boundary purification. The result is truncated to a MPS with bond dimension D'' and increased purification bond, dd'. Then the purification bond is reduced from dd' to d', by imposing the canonical form [7] and projecting the reduced density matrix of each site onto the space spanned by its d' largest eigenvectors. The computational cost of the first step, which proceeds via the standard ALS scheme, scales as $\mathcal{O}(dd'D^4D''^2) + \mathcal{O}(dd'D^2D''^3)$, while the leading cost of the second step is $\mathcal{O}(d^2d'^2DD''^2)$, negligible only when d' is small. Because the purification bond satisfies $d' \leq DD''^2$, the maximum cost can at most grow as $\mathcal{O}(dD^5D''^4) + \mathcal{O}(dD^3D''^6)$ when d' takes on its largest possible value. In [43,71] it was proposed to set $d' = D = D''$, in which case the number of operations scales only like $\mathcal{O}(D^7)$, and a clear computational gain compared to the original contraction can be expected.

In order to analyze the performance of the SL procedure, we study the accuracy of the norm contraction, $\mathcal{N} \equiv \langle\psi|\psi\rangle$, as a function of the truncation parameters, D'' and d', for a set of different PEPS, and compare the results to those of the original algorithm. In particular, we consider $D = 2-4$ PEPS ground state approximations from the SU for the Ising model (1.1) with transverse fields $B = 1.0$, 2.0, 3.0, and 4.0, on lattices with side lengths $L = 11$ and 21. In all cases, the exact norm was estimated by means of the original contraction [11,41] with bond dimension $D' = 100$, large enough to make the error negligible.

In the case of the original algorithm, the relative error always decreases expo-

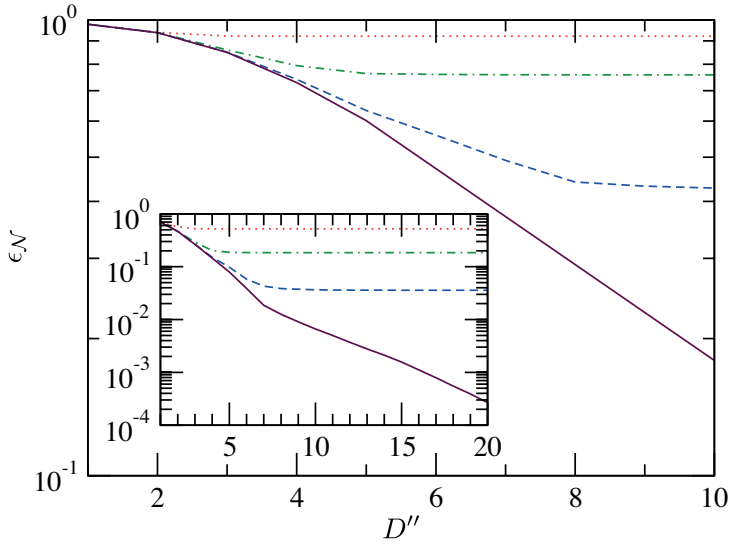


Figure 1.6: Relative norm error $\epsilon_{\mathcal{N}}$ in the SL contraction of a $D = 2$ SU ground state approximation of the Ising model with $B = 3.0$ on a 11×11 (inset) and 21×21 (main) lattice. We consider $d' = 2$ (dotted), $d' = 4$ (dash-dotted), $d' = 8$ (dashed), and the maximum possible $d' = DD''^2$ (solid). We defined $\epsilon_{\mathcal{N}} := |\mathcal{N}(D'') - \mathcal{N}_{ref}|/|\mathcal{N}_{ref}|$ where the reference value $\mathcal{N}_{ref}$ was computed with the original contraction using $D' = 100$.

nentially with the bond dimension of the general boundary MPO, D'. Moreover, for a fixed bond dimension of the PEPS, D, this error shows no system size dependence. In the SL algorithm, for fixed purification bond d', the contraction error converges quickly as function of D'' to a final value that is entirely determined by d'. Even when that purification bond dimension takes on its maximum value, $d' = DD''^2$, this error lies many orders of magnitude above the one from the original contraction [11, 41] with the same $D' = D''$. It is worth noticing that for large $D' = D'' \gg D$ the computational cost of the original method is actually lower than the one of the SL algorithm with maximum $d' = DD''^2$.

The differences between the original and the SL contraction become even more apparent when the lattice size is increased to $N = 21 \times 21$, because the SL algorithm depends strongly on the system size as can be gathered from Fig. 1.6. In that case, given $d' = DD''^2$ and $D'' = 10$, the norm error grows from $\epsilon_{\mathcal{N}} \approx 0.007$ in the 11×11 to $\epsilon_{\mathcal{N}} \approx 0.1$ in the 21×21 lattice, in marked contrast to $\epsilon_{\mathcal{N}} \approx 10^{-11}$ in the original contraction with $D' = 10$ obtained for both lattice sizes. And we observe a similar scaling for PEPS with larger bond dimensions. For instance, the SL value to $D = 4$, $d' = 8$, and $D'' = 10$ grows from $\epsilon_{\mathcal{N}} \approx 0.06$ in the 11×11 to $\epsilon_{\mathcal{N}} \approx 0.6$ in the 21×21 lattice, which has to be compared to $\epsilon_{\mathcal{N}} \approx 10^{-5}$ in the original contraction with $D' = 10$ achieved for both lattice sizes.

From this analysis we conclude that the choice $d' = D = D''$, which ensures the advantageous computational cost $\mathcal{O}(D^7)$, is in general too restrictive in order to get a comparable precision to that of the original algorithm. Moreover, because the required values of the parameters d' and D'' for a certain contraction precision depend strongly on the system size, one cannot make a general statement about the cost scaling of the SL algorithm. This is different from the situation in the original algorithm, where the parameter D' controlling the cost can typically be chosen as

$D' \propto D^2$ with a prefactor that seems not to depend on the system size but only on the state.

The environment approximation in the SL scheme, despite being positive, does not correspond to the most general boundary purification, a fact that provides some insight into the limitations of the method. The purification of a mixed state is only defined up to an isometry on the traced-out degrees of freedom. But the optimization in the SL algorithm does merely allow for local isometries, i.e. for tensor products of isometries each acting on a single site of the boundary only. It is possible to devise an ALS algorithm that searches the optimal general boundary purification, at the expense of a cost function for each site which is no longer quadratic but quartic in the local tensor variables. The minimum of such a cost function is no longer the solution of linear but of nonlinear equations, which are numerically much more demanding than the simple QR decomposition that gives the optimal general boundary MPO in the original algorithm. Therefore such a strategy may result in an undesirable slowing down of the algorithm. Notice also that, while a given purification can always be written efficiently as a positive MPO, namely via contraction of the tensors at each site over their purification bond, the reverse statement is not true, since there exist positive MPO that cannot be written efficiently as purifications [83]. We conclude that, for the problems considered here, it is more advisable to work with a general boundary MPO upon which positivity is not explicitly imposed,[6] and based on it formulate contraction algorithms where cost and precision grow simultaneously, as we shall do in the following section.

1.5 Clusters

The most precise environment approximation is achieved by the original algorithm, in the form of a MPO with sufficiently large bond dimension D'. On the opposite extreme of the spectrum, the lowest computational cost corresponds to the SU, where the environment is represented by a tensor product of matrices each acting on a single virtual bond only. Here, we aim at a contraction scheme that allows to systematically tune the environment precision together with the cost and that interpolates between the SU and the original algorithm.

This goal is achieved with the help of *clusters*. In a state with short-range correlations, we expect that the major contribution to the environment of a given tensor comes from the closest sites. If such sites are not correlated with further ones, or among themselves, the environment will actually be a product, similar to the SU approximation. Correlations in the state cause the environment to be nonseparable in general, and to incorporate relevant contributions from faraway sites. Hence, by progressively taking into account the contribution of sites at longer distances, we would improve the environment description.

In our PEPS algorithm, we are interested in the environment of a row (respectively column), which is required for the update of all the tensors in it. We therefore define a *cluster* of size δ around a certain row as all the surrounding rows which are separated from it by a distance smaller or equal to δ, and similarly for columns. The environment contribution from sites outside the cluster can be roughly approximated by a product in the spirit of the SU, while the contribution from sites inside the cluster is taken into account with more precision, as in the original algo-

[6] Although, when D' is chosen too small, the negative eigenvalues of the environment can lead to instabilities in the tensor update, when D' is large enough, the tensor update is stable and then more accurate.

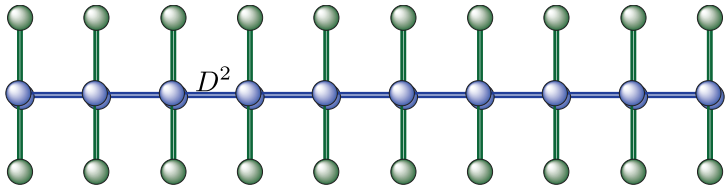


Figure 1.7: Separable Environment with $D' = 1$.

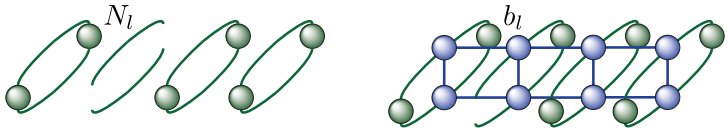


Figure 1.8: Tensor contractions during the ALS sweeping in the computation of the environment for CU_0. For each local update, the norm matrix (left) is the identity times a positive constant, such that the solution is simply proportional to b_l (right), which is given by a positive TN if each of the local tensors is positive.

rithm. This defines a new contraction scheme that we call cluster update in analogy to [71], and that we abbreviate as CU_δ for cluster size δ. The limiting cases of this strategy are $\delta = L - 1$, when the environment reverts to the one of the original algorithm, and $\delta = 0$, which is closely related to the SU.

1.5.1 Cluster Size $\delta = 0$: A Generalized Simple Update

The particular case $\delta = 0$ (CU_0) leads to the environment approximation of a certain row, or column, as a product MPO, illustrated in figure 1.7. This can be found by optimizing the boundary MPO with $D' = 1$, where the standard MPO-MPS ALS scheme can now yield a positive MPO. Indeed, if each of the local tensors of the MPO is positive, this positivity is maintained during the update procedure, since for each local optimization the norm matrix in $N_l \vec{A}_l = \vec{b}_l$ is proportional to the identity, and the TN to b_l is positive, as explained in figure 1.8. Starting the ALS sweeping from an initial positive MPO, which can be trivially constructed from positive local tensors (e.g. of the form $X^\dagger X$ with random X), ensures then a positive environment. Moreover, all contractions can be performed with $\mathcal{O}(dD^5)$ operations, so that the computation of the optimal separable environment does not exceed the leading cost of the SU.

Imaginary time evolution based on a positive separable boundary MPO leads to an algorithm which is very similar to the SU. Both schemes are characterized by the same computational cost,[7] and make use of a separable environment, but the CU_0 method proposed here optimizes the approximate boundary over all possible separable MPO, and hence can be interpreted as a generalized simple update.

In order to elucidate the connection between both algorithms, we study how imaginary time evolution with CU_0 changes PEPS ground state approximations from the SU for the Ising model (1.1) with various magnetic fields. In our quantitative comparison we consider a specific virtual bond between two neighboring sites in the center of the lattice, and focus on the corresponding λ matrix generated by the SU after convergence, λ^{SU}. The diagonal of that matrix can directly

[7] Just like the environment approximation, the tensor update in a separable environment can be performed with $\mathcal{O}(D^5)$ operations.

	$B = 1.0$	$B = 2.0$	$B = 3.0$	$B = 4.0$
λ_2^{SU}	0.006007	0.026032	0.078572	0.071486
$\lambda_2^{\mathrm{CU}_0}$	0.006022	0.026252	0.078953	0.071210
$\Sigma_2^{\mathrm{CU}_0}$(i)	0.006008	0.026049	0.078399	0.071216
$\Sigma_2^{\mathrm{CU}_0}$(ii)	0.006008	0.026048	0.078421	0.071216
$\Sigma_2^{\mathrm{CU}_0}$(iii)	0.005784	0.025236	0.077678	0.071391
$\Sigma_2^{\mathrm{CU}_0}$(iv)	0.005784	0.025237	0.077828	0.071391
$\Sigma_2^{\mathrm{CU}_0}$(v)	0.005567	0.024434	0.076890	0.071566
$\Sigma_2^{\mathrm{CU}_0}$(vi)	0.005567	0.024434	0.077094	0.071566

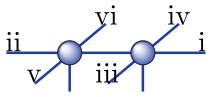


Table 1.1: We apply the SU with $D = 2$ to a 11×11 Ising model with different magnetic fields B. All λ matrices have converged to machine precision and we report the final second entry λ_2^{SU} on the vertical virtual bond at row 5 and column 6. The resulting PEPS is further evolved with the CU_0 until convergence. We show the second singular value $\lambda_2^{\mathrm{CU}_0}$ emerging in the tensor update on the considered bond and the second eigenvalue $\Sigma_2^{\mathrm{CU}_0}$ of the boundary MPO matrix at that place whenever it enters a tensor update. This happens on the six different positions relative to a tensor pair defined in the figure on the right, during the approximation of the four sets of Trotter gates in one time evolution step. We adopt the normalization in which each first λ entry, singular value and eigenvalue is always 1.

be compared to the converged singular values emerging in the CU_0 every time a gate is applied to this particular pair of sites, λ^{CU_0}. As shown in table 1.1 for a 11×11 PEPS with $D = 2$, the relative difference between the entries of these two λ matrices is below $\approx 10^{-2}$.

We can analyze the similarities between both algorithms in more detail by looking at the role of the λ matrices in the environment for the update operations. In the SU, the entries in λ^{SU} are determined after applying one gate to the relevant pair of tensors, but (in the here considered case of nearest neighbor interactions) they are not affected by gates which involve only one member of the pair. For the latter tensor updates, the λ^{SU} matrix enters the environment unchanged, even after the Γ tensors of the pair have been modified. In contrast, in the CU, the environment for a given update operation depends on the surrounding tensors, and changes every time they are updated. In the case of CU_0, a similar role to that of λ^{SU} is played by the eigenvalues of the local tensor in the boundary MPO at the site corresponding to this particular virtual bond, Σ^{CU_0}. For nearest neighbor interactions and the bond we are considering, there are six Trotter gates in each time step (see the figure in table 1.1) that involve only one of the tensors of the pair. The Σ^{CU_0} entries change only slightly, $\approx 10^{-2}$, for each such tensor update, as can be appreciated in table 1.1. And their difference to the corresponding λ^{SU} is of the same order. Additionally, we computed the separable boundary MPO for the SU PEPS and compared the eigenvalues of the local tensors to the corresponding Σ^{CU_0}, to find a similar agreement. We observed the same behavior in larger lattices, with larger bond dimensions, as well as on different virtual bonds.

Our observations provide an explanation for the functioning of the SU: because the latter scheme applies the same λ^{SU} matrix to the tensor updates of all four sets of Trotter gates, this λ^{SU} can be seen as a mean value for the six Σ^{CU_0} from the optimal positive separable environment, and the SU indeed converges it to that mean value.

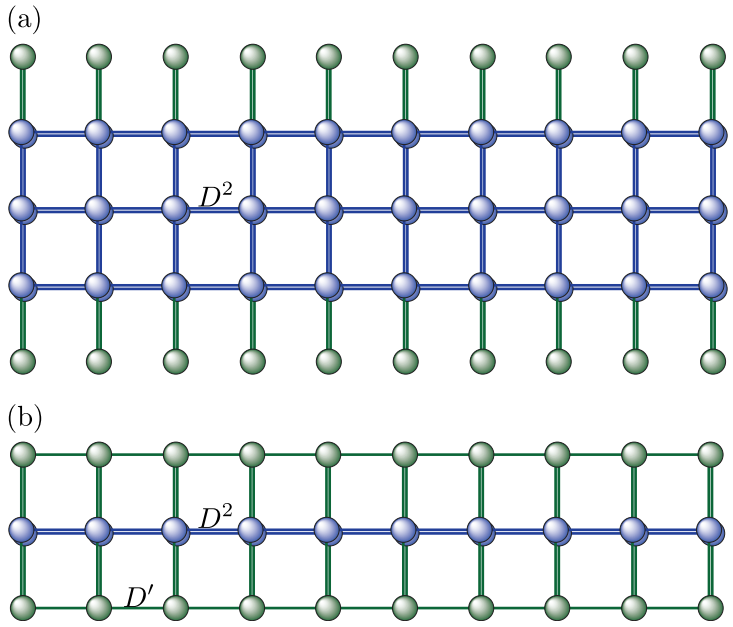


Figure 1.9: Environment for the CU_1 of a middle row. (a) Outside the cluster, the approximate contraction of the TN is performed via a positive separable boundary MPO. (b) The contraction continues inside the cluster via a general boundary MPO with bond dimension D'.

The CU with $\delta = 0$ always uses the best separable environment in each tensor update, and therefore depends less on the initial state and can produce energies slightly below the ones from the SU. However, the final energies of both methods are very close to each other (compare figures 1.4 and 1.10).

Although the SU is an algorithm completely formulated in the SL, our study in the double layer picture provided crucial insight into it, thus reinforcing the idea that the boundary should be described as a general MPO.

1.5.2 From Simple to Full Update

By considering clusters of size $\delta \geq 1$ we can systematically take into account more of the correlations in the environment approximation. Outside the cluster, the environment is represented by a separable MPO and determined as in the CU_0 described above. Then the cluster tensors are contracted row by row with this boundary, as in the original algorithm, to produce a new boundary MPO with larger bond dimension. The approximation is controlled by the cluster size and the bond dimension D' used in the contractions within the cluster.

Figure 1.9 shows the smallest non trivial cluster, for CU_1, in which only the two rows adjacent to the one to be updated enter the cluster contraction. In this case, the optimal boundary MPO with bond dimension D' for the update of a row is computed from the action of a bulk row on a separable boundary MPO with $\mathcal{O}(dD^5D'^2)$ operations. This is the dominant cost in the environment approximation of CU_1, given the fact that the separable MPO outside the cluster is obtained with only $\mathcal{O}(dD^5)$ operations. Hence, the environment approximation for clusters of size $\delta = 1$ is computationally cheaper than the full contraction with cost $\mathcal{O}(D^4D'^3) + \mathcal{O}(dD^6D'^2)$.

To examine the usefulness of this cluster strategy, we compare its performance to that of the SU via their ground state accuracies for the Heisenberg model (1.2). Starting from converged SU PEPS, we ran the CU imaginary time evolution with several cluster sizes for various bond dimensions D and D' on 4×4 and 10×10 lattices. Figure 1.10 contains our cluster results for $\delta = 0$, 1, and $L - 1$, as function of D, such that they can be compared directly to the SU results of figure 1.4. The convergence of the CU with cluster size δ as well as with bond dimension D' can be gathered from figure 1.11 for 10×10 PEPS with $D = 2$ and 4. We refer to the CU with maximum cluster size $\delta = L - 1$ as *full update* (FU), a notion taken from iPEPS (see e.g. [22]). The FU is not identical to the original PEPS algorithm [11], because in the CU the action of single Trotter gates is approximated sequentially, such that for each gate the only tensors updated are those on which the gate directly acts. Thanks to this procedure, the FU requires just the approximate contraction of the norm TN, and is therefore computationally less demanding than the original algorithm, in which, additionally, the PEPS TN with a full set of Trotter gates acting on all the state has to be contracted.

We find that the CU_0 produces very similar energies as the SU, with slight improvement for small systems or for large bond dimensions. The difference between both methods is most apparent in case of the smaller 4×4 lattice where the CU_0 gives lower energies for bond dimensions $D \geq 4$. This can be understood taking into account that the effect of the system boundary, better captured by the environment approximation in CU_0, is more important for smaller systems. We observe then that the CU_1 improves the SU energies considerably. Finally, the FU reduces the energy further significantly when $D \geq 4$, and its effect appears more pronounced with increasing bond dimension D. For $D = 2$ and 3, the FU improves upon the CU_1 only in case of the smaller 4×4 system. Notice that the tables 1.2-1.7 contain the precise energy values that were used in this analysis.

From the arguments above it is apparent that a better representation of the environment is crucial for an improved PEPS approximation of the true ground state. We also infer that larger bond dimensions D require more precise environment representations in the tensor update. Within the CU, this improvement can be achieved systematically by gradually increasing, firstly, the cluster size δ and, secondly, for each fixed δ, the boundary bond dimension D'. Indeed, we can see in figure 1.11 for each fixed cluster size that with growing D' the energy decreases consistently, as the precision of the environment representation in the tensor update increases. The energy converges at a certain value D'_{max} that depends both on the bond dimension D of the considered PEPS and on the cluster size. While for $D = 2$ the lowest energy is already attained with CU_1, for $D = 4$ the energy improves when larger clusters are used and the FU value is obtained with CU_4.

This behavior agrees with our previous observation that larger bond dimensions benefit more from accurate environment representations. We can gain further insight into this feature by looking at the convergence of a boundary MPO as function of D' for different cluster sizes. In figure 1.12 the environment MPO for the leftmost column, computed with different cluster sizes, is compared to the full contraction of the $L - 1$ right columns with large enough D', for PEPS with bond dimensions $D = 2$ and 4 on a 20×20 lattice. We find that for each cluster size δ there exists a maximum value of D' beyond which the distance to the reference boundary MPO does not decrease anymore, and that this value is smaller than the largest possible $D' = D^{2\delta}$. Considering a sufficiently large fixed D', the distance drops exponentially with increasing cluster size until the value of the full contraction is reached. Beyond this, larger clusters have no effect. Finally, we can directly

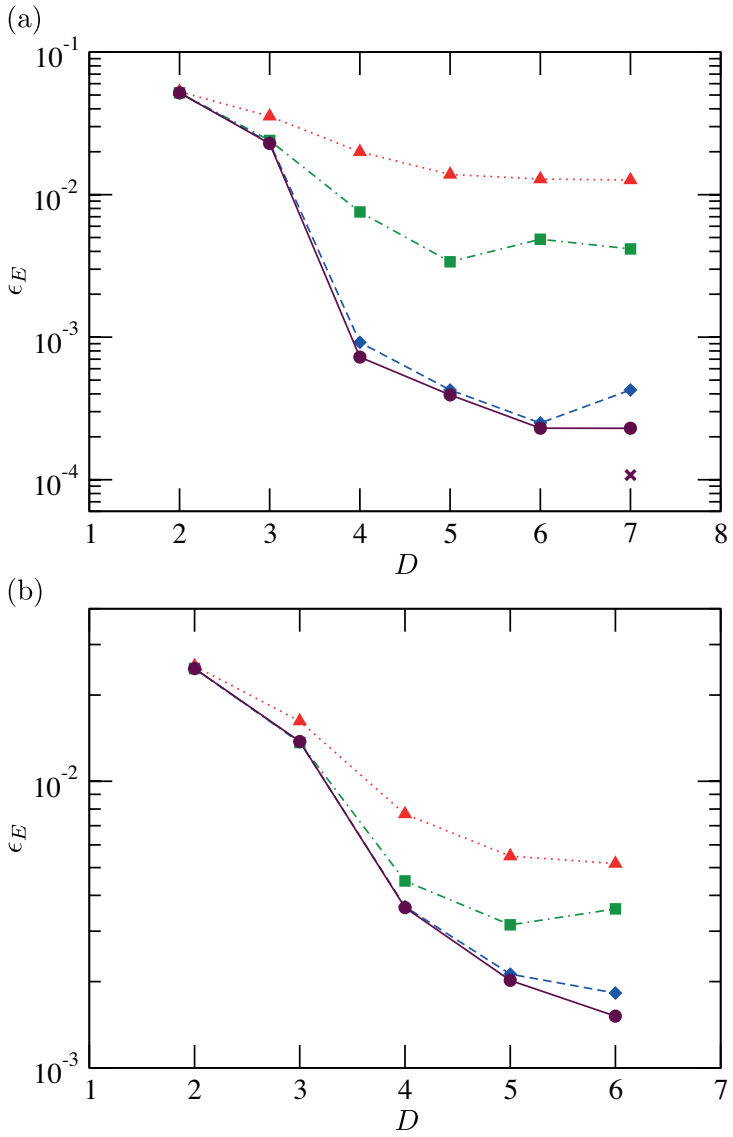


Figure 1.10: Relative energy error ϵ_E as in figure 1.4 for a 4×4 (a) and 10×10 (b) Heisenberg model. We consider the CU_0 (dotted), the CU_1 with $D' = D^2$ (dash-dotted), and the FU with $D' = D^2$ (dashed), $D' = 2D^2$ (solid), and $D' = 130$ (cross).

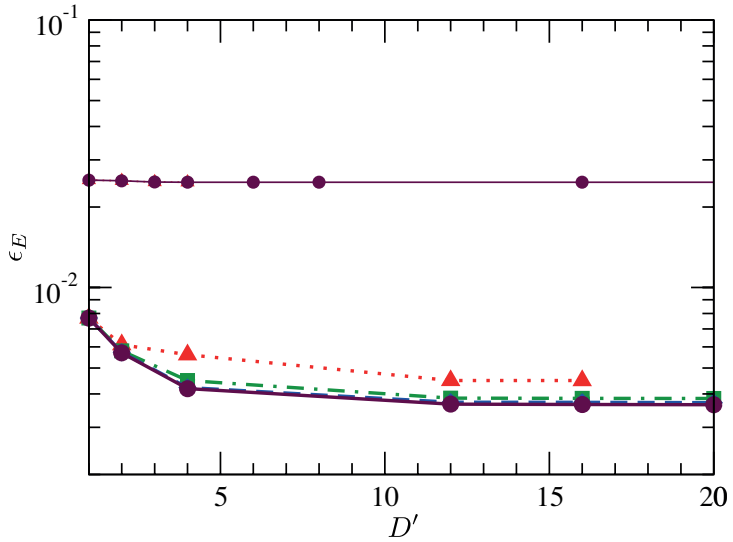


Figure 1.11: Relative energy error ϵ_E as in figure 1.4 for a 10×10 Heisenberg model, obtained with various fixed values of the bond dimension D' of the boundary MPO. We propagated $D = 2$ (upper thin lines) and $D = 4$ (lower thick lines) PEPS with the CU with cluster size $\delta = 1$ (dotted), 2 (dash-dotted), 3 (dashed), and with the FU (solid).

see that, in order to have the same distance, the $D = 4$ PEPS requires larger clusters and larger boundary bond dimensions than the $D = 2$ PEPS, which explains why it responds stronger to a better environment representation.

1.5.3 Computation of Expectation Values

Although we introduced clusters in the specific context of environment approximations for the tensor update, figure 1.12 suggests that, in fact, the reduced density matrix of any part of a PEPS can be accurately approximated by a cluster around that part, with a precision determined by the cluster size. Therefore the cluster strategy could also be applied to the evaluation of (local) expectation values, without the need for an accurate contraction of the full TN. To validate this idea, we computed the energy of PEPS with $D = 2$ and 4 on a 20×20 lattice using clusters of different sizes around the local terms of the Hamiltonian, shown in figure 1.13. We observe, analogously to figure 1.12, that for each cluster size the energy error converges for a certain value of D', and that the larger bond dimensions require larger clusters and larger values of D'. Most remarkably, we find again that for large enough fixed D' the error drops exponentially with the cluster size.

1.5.4 Applicability to a Parallel PEPS Code and to iPEPS

In the context of finite PEPS, considered in this study, the computational cost of the environment approximation for CU_δ is lower than that for the full contraction only when $\delta = 0$ ($\mathcal{O}(dD^5)$) and when $\delta = 1$ ($\mathcal{O}(dD^5D'^2)$). Indeed, if the boundary MPO has bond dimension D'_0, contracting it with a PEPS row and approximating the result by a new boundary with bond dimension D'_1 needs $\mathcal{O}(dD^6D'_0D'_1) + \mathcal{O}(D^4D'^2_0D'_1) + \mathcal{O}(D^4D'_0D'^2_1)$ operations. If $D'_0 = D'_1$, we recover the scaling of the full contraction, so that the CU only results in a cheaper scheme if

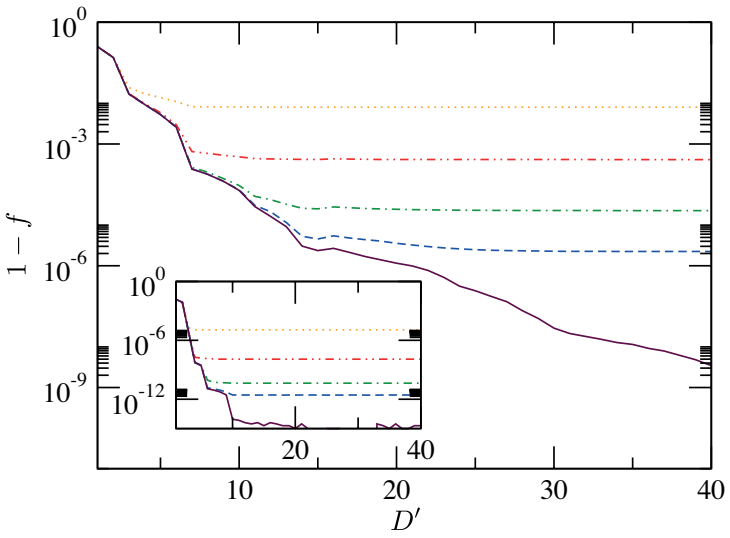


Figure 1.12: Distance to the exact boundary MPO, $1 - f$ with $f := |\mathrm{tr}(\rho_{ref}^{\dagger}\rho(D'))|$, for the left-most column boundary MPO $\rho(D')$ of a $D = 2$ (inset) and $D = 4$ (main) SU Heisenberg PEPS on a 20×20 lattice. We compare the cluster contraction based on clusters of size $\delta = 1$ (dotted), 2 (dash-double-dotted), 3 (dash-dotted), and 4 (dashed), to the full contraction (solid). The reference boundary MPO ρ_{ref} comes from the full contraction with $D' = 100$, and we adopt the normalization $\mathrm{tr}(\rho^{\dagger}\rho) = 1$.

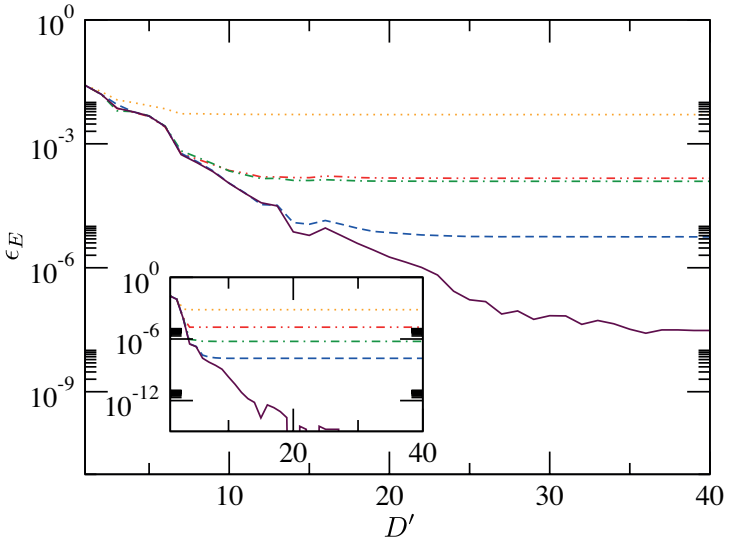


Figure 1.13: Relative energy error $\epsilon_E := |E(D') - E_{ref}|/|E_{ref}|$ of the $D = 2$ (inset) and $D = 4$ (main) PEPS from figure 1.12, for the same setting. The reference value E_{ref} comes from the full contraction with $D' = 100$, and the clusters are formed around the individual terms of the Hamiltonian independently of each other.

the environment bond dimensions D'_i decrease as we increase the distance i to the row (or column) to be updated. Moreover, after every update of a row, the complete cluster surrounding the next row has to be contracted, without being able to reuse the previously obtained cluster boundary MPO. This situation is different from the FU, where, when moving to the update of a new row, only one new boundary MPO has to be determined, as the previously computed and properly stored boundaries can be reused. In the CU, the only boundary MPO that can be reused are the previously obtained separable ones, and $2\delta + 1$ new MPO have to be computed for the update of the next row. Of those, one is separable and thus determined with computational cost $\mathcal{O}(dD^5)$, and two require $\mathcal{O}(dD^5D'^2)$ operations, which we can neglect, such that $2\delta - 2$ new boundary MPO have to be found with cost $\mathcal{O}(dD^6D'^2) + \mathcal{O}(D^4D'^3)$. On the other hand, the CU takes up less memory than the FU. The separable boundary MPO do not have to be written to hard disk but can be stored in main memory since they take up much less memory than MPO with bond dimensions $D' > 1$, and then the cluster boundary MPO are computed on the fly.

Although the CU_δ with cluster sizes $\delta > 1$ does not reduce the computational cost of a sequential algorithm, in which one tensor is updated after another, it can reduce the cost of a parallel algorithm, in which different rows or columns are updated simultaneously on different processors. Assuming that the time for the optimization of a boundary MPO (for a middle row) is t_B on average, and that the update of all the tensors in a row or column is achieved in the time t_U, then the sequential FU requires $2(L-2) \cdot t_B + L \cdot t_U$ for one set of Trotter gates. In contrast, each row or column update with the CU_δ for $\delta \geq 1$ necessitates the computation of only $2(\delta - 1)$ boundary MPO and thus has the cost $2(\delta - 1) \cdot t_B + t_U$. We conclude that a parallel CU algorithm can attain a L/δ speed-up. Since δ does not depend on the system size, L, but only on the bond dimension, D, it can be chosen much smaller than L, such that this speed-up factor may be large. This estimation neglects all computations with sub-leading costs $\mathcal{O}(dD^5)$ and $\mathcal{O}(dD^5D'^2)$ and the communication between the parallel processors. Although the latter will have an impact on the final performance of the algorithm, we expect the speed-up to be still significant, given the fact that just the small individual tensors of separable boundary MPO have to be exchanged between different processors after each set of Trotter gates.

The success of this parallelization strategy relies heavily on the simultaneous update of tensors in different rows. As described in section 1.2, each tensor update is based on solving a system of linear equations that arises from the minimization of a cost function for the whole PEPS by utilizing an ALS scheme. In this scheme one sweeps over the tensors and for each one minimizes the cost function under the assumption that all the others are fixed. This guarantees a non-increasing cost function only when the tensors are updated sequentially. An important question is then whether the convergence of the energy in imaginary time is as fast with the independent updates as with the sequential ones. That this is indeed the case can be gathered from figure 1.14. The plot demonstrates an impressive agreement, which can be attributed to a minor modification of the tensor when the action of a time evolution gate is approximated in a sequential update. We conclude that imaginary time evolution with the CU constitutes a natural basis for a parallel ground state search algorithm based on PEPS. A similar agreement as in figure 1.14 cannot be expected in direct energy minimization, where a tensor is changed significantly during an update.[8]

[8] In direct energy minimization, the energy $\langle\psi|\hat{H}|\psi\rangle / \langle\psi|\psi\rangle$ is minimized directly by sweeping over

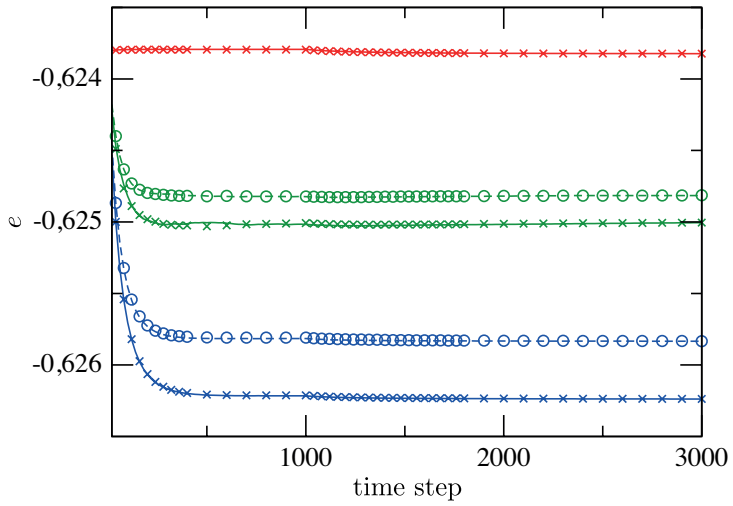


Figure 1.14: Energy per site e of the Heisenberg model on a 10×10 lattice during imaginary time evolution of a $D = 4$ PEPS via the CU with parallel (lines) and sequential (symbols) tensor updates. The boundary MPO have fixed bond dimension $D' = 1$ (top red lines and symbols), $D' = 2$ (middle green lines and symbols), and $D' = 16$ (bottom blue lines and symbols), and we further distinguish cluster size 1 (dashed lines and circles) and 2 (solid lines and crosses). The state propagates 1000 time steps with $\tau = 0.01$ and then 2000 time steps with $\tau = 0.001$.

Although we carried out our analysis in the framework of finite PEPS, it is clear that the CU_δ can also be applied to iPEPS, to replace the costly computation of the environment via the dominant boundary eigenvector with $D' > 1$ [17], fixpoint corner transfer matrices [66], or second renormalized environment [68]. A CU procedure would only require the search for the dominant boundary eigenvector with $D' = 1$, which needs $\mathcal{O}(dD^5)$ operations, followed by a cluster contraction as in the finite case. Then, the cost and precision of both tensor update and expectation value computation would be determined by the cluster size and the bond dimension D' employed in the contraction of the cluster. Furthermore, it is always possible to evaluate clusters by means of Monte Carlo sampling [64, 65]. This method requires only the contraction of the PEPS coefficients, computationally less costly than the contraction of the PEPS norm TN, for different sampled values of the physical indices $s_1, s_2, \ldots$, depicted in figure 1.1. While a full infinite PEPS cannot be sampled, since this would necessitate determining infinitely many classical spin values, clusters open the door to variational Monte Carlo in the realm of iPEPS. For the sampling of an observable as well as of an energy gradient, a cluster would be formed around the considered tensors and then only the physical indices of that cluster would have to be sampled. Larger clusters necessitate longer sampling times such that the error of a finite cluster could be adjusted together with the Monte Carlo error according to the available computational resources.

the tensors with an ALS scheme. This algorithm converges typically within much fewer sweeps over the PEPS than imaginary time evolution, when all sweeps for all time steps are taken into account, and therefore modifies a tensor considerably in an update.

1.6 Conclusions

In this chapter, we have analyzed the environment representation in previous proposals, namely the simple update and the single-layer algorithm. We have shown how the different approximations applied to the environment explain the limitations of each method in the achievable ground state accuracy, an issue that we have studied quantitatively in the context of finite PEPS. Based on this deeper understanding, we have formulated a new proposal, the cluster strategy, that allows a systematic increase of the environment precision from the simplest possible representation, in the SU, to the most accurate full contraction, in the FU.

In its simplest form, CU_0 provides an explanation for the simple update in terms of a separable boundary approximation, and constitutes a slightly improved version of the latter for the models analyzed here, characterized by the same computational cost. The first non trivial cluster update, CU_1, produces significantly better ground state energies than the SU, and has a lower computational cost than the FU. In general, CU_δ interpolates naturally between the SU and the FU. We have shown that increasing the cluster size improves the precision of the environment approximation exponentially. This improvement applies directly to the computation of local observables, which can always be accelerated with the help of clusters.

Our analysis of the computational costs of the CU revealed that in the sequential update of finite PEPS any cluster size $\delta > 1$ exceeds the cost of the FU, which can reuse intermediate calculations more efficiently. However, the CU_δ forms the basis of a very promising parallel PEPS algorithm, with a prospective large gain in computational time also for larger clusters. Although our numerical studies have all been carried out in the framework of finite PEPS, we have also argued how the CU_δ is straightforwardly useful for the infinite iPEPS ansatz.

In summary, we have shown that the environment approximation is a key ingredient to the precision of any PEPS contraction, whether we are interested in the norm, or in some expectation value. The CU_δ provides the means to control this approximation accuracy and can be used in any contraction. It is then reasonable to think of its potential applicability to other PEPS algorithms.

1.7 Appendix

1.7.1 Heisenberg Reference and PEPS Energies

Here, we list explicitly a selection of precise values, as they are used in the main text.

Regarding the reference energies, for $L = 4$ the exact energy values come from exact diagonalization. The exact Heisenberg ground state energy per site on a 4×4 lattice reads -0.57432544. For $L = 10$, we computed the exact values with the quantum Monte Carlo loop algorithm from the ALPS library [84–86], and we use the result for temperature $T = 0.0001$, where we have checked consistency with $T = 0.01$ and 0.001. That energy per site for a 10×10 system reads $-0.628655(2)$.

1.7.2 Numerical Details

We used the following setup for time evolution and energy computation, if not explicitly stated otherwise.

We initialize imaginary time evolution with a separable $D = 2$ PEPS in which the zeroes are replaced by noise as uniformly distributed random numbers from

D	4×4	10×10
2	-0.54404(1)	-0.61281(1)
3	-0.55396(2)	-0.61846(2)
4	-0.56281(1)	-0.62382(1)
5	-0.56628(2)	-0.62520(2)
6	-0.56684(3)	-0.62541(2)
7	-0.56696(2)	-0.62537(2)
8	-0.56715(3)	-0.62538(2)

Table 1.2: Energy per site e of the Heisenberg model on a 4×4 and 10×10 lattice, obtained by means of the SU, presented in figure 1.4.

D	4×4	10×10
2	-0.54404(3)	-0.61280(2)
3	-0.55397(2)	-0.61846(2)
4	-0.56287(5)	-0.62382(2)
5	-0.56637(2)	-0.62521(2)
6	-0.56694(2)	-0.62541(2)
7	-0.56706(3)	

Table 1.3: Energy per site e of the Heisenberg model on a 4×4 and 10×10 lattice, obtained by means of the CU_0, presented in figure 1.10.

D	4×4	10×10
2	-0.54458(2)	-0.61310(2)
3	-0.5605(3)	-0.62007(1)
4	-0.56999(3)	-0.62583(2)
5	-0.57238(7)	-0.62667(2)
6	-0.57153(7)	-0.6264(2)
7	-0.57194(1)	

Table 1.4: Energy per site e of the Heisenberg model on a 4×4 and 10×10 lattice, obtained by means of the CU_1 with $D' = D^2$, presented in figure 1.10.

D	D'	4×4	10×10
2	4	-0.54458(2)	-0.61310(2)
	8	-0.54458(2)	-0.61310(2)
3	9	-0.56101(2)	-0.62002(2)
	18	-0.5612(1)	-0.62000(2)
4	16	-0.5738(3)	-0.62636(3)
	32	-0.5739(2)	-0.62637(2)
5	25	-0.57408(1)	-0.62732(4)
	50	-0.57410(3)	-0.62739(1)
6	36	-0.57418(2)	-0.62751(2)
	72	-0.57419(1)	-0.62770(7)
7	49	-0.57408(1)	
	98	-0.57419(1)	
	130	-0.57426(1)	

Table 1.5: Energy per site e of the Heisenberg model on a 4×4 and 10×10 lattice, obtained by means of the FU, presented in figure 1.10.

D'	CU_1	FU
1	-0.61280(2)	-0.61280(2)
2	-0.61290(2)	-0.61289(1)
3	-0.61307(2)	-0.61307(1)
4	-0.61310(2)	-0.61310(2)
100		-0.61310(2)

Table 1.6: Energy per site e of the Heisenberg model on a 10×10 lattice, from $D = 2$, presented in figure 1.11.

D'	CU_1	CU_2	CU_3	CU_4	FU
1	-0.62382(2)	-0.62382(2)	-0.62382(2)	-0.62382(2)	-0.62382(2)
2	-0.62481(1)	-0.62501(1)	-0.62506(5)	-0.62504(1)	-0.62508(4)
4	-0.62513(3)	-0.62583(4)	-0.62600(1)	-0.62607(2)	-0.62602(1)
12	-0.62583(2)	-0.62623(2)	-0.62631(2)	-0.62634(3)	-0.62635(2)
16	-0.62583(2)	-0.62623(2)	-0.62632(3)	-0.62635(3)	-0.62636(3)
20		-0.62624(2)	-0.62632(2)	-0.62635(2)	-0.62636(2)
32					-0.62637(3)

Table 1.7: Energy per site e of the Heisenberg model on a 10×10 lattice, from $D = 4$, presented in figure 1.11.

$[-0.01, 0.01]$. This state is evolved for N_1 steps with τ_1, followed by N_2 steps with τ_2, and so on, what we abbreviate to the short notation $(N_1 \times \tau_1, N_2 \times \tau_2, \ldots)$ for fixed bond dimension D. In order to specify a successively growing value of D, we introduce the recursive notation $(D_{i+1} = D_i^{\tau} + 1, N_1 \times \tau_1, N_2 \times \tau_2, \ldots)$. It defines the next PEPS for the propagation with bond dimension D_{i+1} as the final state of the previous evolution with bond dimension D_i and time step τ_i with a by 1 incremented bond dimension. In the case of the cluster and full update, the additional parameter D' is typically chosen as $D' = 1, 2$, and, related to D, as $D' = D, D^2$, and so on. The final PEPS obtained with a certain value of D' is always the initial state for increased D with that D'.

Regarding the energy computation, all energies are evaluated with $D' = 100$ for the final PEPS corresponding to the smallest time step. We define the energy error as the difference between the energy of this final state and the energy of an intermediate state. The latter is either the PEPS obtained after half of the evolution or the final PEPS corresponding to the immediately larger time step, depending on wether or not the propagation was also performed with this larger time step.

In the following, we list the precise configurations that were used for the results which are shown in the figures:

Figure 1.4:
We propagate the initial $D = 2$ PEPS 1000 time steps with $\tau = 0.1$, then 2000 time steps with $\tau = 0.01$, then 8000 time steps with $\tau = 0.001$, and then according to the configuration $(D_{i+1} = D_i^{\tau=0.01} + 1, 2000 \times \tau = 0.01, 8000 \times \tau = 0.001)$.

Figure 1.6:
We propagate the initial $D = 2$ PEPS each time 10000 steps first with $\tau = 0.1$, then with $\tau = 0.01$, then with $\tau = 0.001$, then 20000 steps with $\tau = 0.0001$, and finally 50000 steps with $\tau = 0.00001$. The $D = 3$ and 4 results, used in the analysis, were

obtained by evolving the final $D = 2$ PEPS from $\tau = 0.0001$ further according to the configuration $(D_{i+1} = D_i^{\tau=0.0001} + 1, 20000 \times \tau = 0.0001)$.

Table 1.1:
We apply the SU to a 11×11 Ising model with different magnetic fields B, and propagate the initial $D = 2$ PEPS 10000 time evolution steps with $\tau = 0.1$ and then 10000 steps with $\tau = 0.01$. The resulting PEPS is further evolved with the CU_0 for 10000 time steps with $\tau = 0.01$. All shown numbers are converged to machine precision.

Figure 1.10 (a) ($N = 4 \times 4$):
The initial state of the imaginary time evolution is the converged $D = 2$ SU ground state approximation to time step $\tau = 0.01$. We propagate this state 1000 time steps with $\tau = 0.01$, then 2000 time steps with $\tau = 0.001$, and then according to the configuration $(D_{i+1} = D_i^{\tau=0.01} + 1, 1000 \times \tau = 0.01, 2000 \times \tau = 0.001)$ up to bond dimension $D = 5$. Then, we continue with $(D_{i+1} = D_i^{\tau=0.01} + 1, 500 \times \tau = 0.01, 1000 \times \tau = 0.001)$. In the case of the FU with $D = 7$ and $D' = 130$, the state propagates 100 time steps with $\tau = 0.01$.

Figure 1.10 (b) ($N = 10 \times 10$):
The initial state of the imaginary time evolution is the converged $D = 2$ SU ground state approximation to time step $\tau = 0.01$. We propagate this state 1000 time steps with $\tau = 0.01$, then 2000 time steps with $\tau = 0.001$, and then according to the configuration $(D_{i+1} = D_i^{\tau=0.01} + 1, 1000 \times \tau = 0.01, 2000 \times \tau = 0.001)$. In the cases of the CU_0 and the CU_1 we use this time evolution configuration up to $D = 5$, and for $D = 6$ propagate the states 500 time steps with $\tau = 0.01$ and then 500 time steps with $\tau = 0.001$. In the case of the FU we use this configuration up to $D = 4$, and for $D = 5$ evolve the states 500 time steps with $\tau = 0.01$ and then 1000 time steps with $\tau = 0.001$, and for $D = 6$ we propagate the states 500 time steps with $\tau = 0.01$.

Figure 1.11:
The initial state of the imaginary time evolution is the converged SU ground state approximation to time step $\tau = 0.01$ with the considered bond dimension. In the case of $D = 2$, we propagate this state 1000 time steps with $\tau = 0.01$, then 2000 time steps with $\tau = 0.001$. In the case of $D = 4$, for CU_1 we use $(1000 \times \tau = 0.01, 2000 \times \tau = 0.001)$, for CU_2 $(1000 \times \tau = 0.01, 1000 \times \tau = 0.001)$, and for CU_3, CU_4, and FU we use $(500 \times \tau = 0.01, 1000 \times \tau = 0.001)$.

Chapter 2

Algorithms for Finite Projected Entangled Pair States

In the preceding chapter, we focused on a particular part of PEPS, namely, their environment approximation. Here, we analyze several algorithmic aspects of the method in order to achieve a more complete understanding of PEPS computations. On the one hand, we quantify the connection between the correlation length of the PEPS and the accuracy of its approximate contraction, and discuss how purifications can be used in the latter. On the other, we present algorithmic improvements for the update of the tensor that introduce drastic gains in the numerical conditioning and the efficiency of the algorithms. Finally, the state-of-the-art general PEPS code is benchmarked with the Heisenberg and quantum Ising models on lattices of up to 21 $\times$ 21 sites. This chapter is published in reference [45].

2.1 Introduction

As explained in the previous chapter, during the last years, significant conceptual and algorithmic progress has been made in the context of PEPS, e.g. references [17, 22, 42, 43, 66–68, 70, 87]. Many of the numerical studies have focused on systems in the thermodynamic limit, for which the iPEPS [17] ansatz can be used. In such case, the translational invariance of the system is exploited to reduce the number of variational parameters to the few tensors in a small unit cell. But the non translationally invariant finite PEPS ansatz is also of great importance. On the one hand, by avoiding a predefined unit cell it allows a more unbiased approach to the thermodynamic limit, when combined with finite size scaling (although, also, a systematically increased unit cell in iPEPS can be expected to produce more and more unbiased results [24, 26]). On the other, it is the proper ansatz for problems that are intrinsically non translationally invariant, such as the simulation of current optical lattice experiments that are being carried out in inhomogeneous traps. In exchange, the price to pay is a more involved implementation and longer running times that scale with the system size.

The original PEPS algorithms [11, 41] can cope with the non translationally invariant situations [52,61,62,88], but a straightforward implementation attains only small tensor dimensions, and is not enough to explore the power of the ansatz. In order to reach larger dimensions (i.e. comparable to those used in present iPEPS

calculations) and to approach the optimal ground state approximations for them, it is necessary to take into account and to optimize the cost and stability of every stage of the algorithms, which is only feasible through a thorough understanding of the various possibilities. Only then it will be possible to adopt the optimal strategies for the particular problem at hand.

In this chapter we aim at a global understanding of the algorithmic aspects of finite PEPS, both at the physical and technical level. We address the two fundamental ingredients of PEPS algorithms, namely the environment approximation, i.e. the approximate contraction of the tensor network (TN), and the tensor update. In the previous chapter, we focused on the environment approximation. We studied the physical significance and limitations of various contraction strategies, and introduced the cluster scheme, which unifies previous methods and gives rise to a new contraction algorithm with a trade-off between precision and computational cost. Here, we extend the analysis of the environment approximation to provide new insight into the convergence of the cluster strategy by relating it to the correlation length of the system. Additionally, we show how the environment approximation can be kept exactly positive with the help of purifications. Regarding the tensor update, we investigate the effect of restricting the variational parameters to a reduced tensor [22], a technique used often in the case of iPEPS and characterized by a lower computational cost. Via the reduced tensor we derive new numerical methods, namely suitable gauge choices, which significantly enhance the stability of the update algorithm. These gauge choices admit also a generalization to cases where the full tensor needs to be updated.

Furthermore, we benchmark the state-of-the-art finite PEPS algorithms using the Heisenberg Hamiltonian and the quantum Ising model with transverse field. By presenting converged finite PEPS results for lattice sizes typically considered in the context of finite size scaling, we not only assess the validity of the ansatz, but enable a systematic comparison to other methods and implementations.

2.1.1 Reader's Guide

The rest of the chapter is structured as follows. In section 2.2 we briefly present the basic notation and concepts common to PEPS algorithms. The algorithmic details regarding the convergence of the cluster scheme, the use of positive environments and the strategies to improve the tensor update are discussed in section 2.3. Section 2.4 collects the numerical results corresponding to our best PEPS ground state approximations for the benchmark models. Finally in section 2.5 we summarize our conclusions. Supplemental material is provided in the appendix in section 2.6.

2.2 Notation and Preliminary Concepts

PEPS algorithms for finding ground states can be classified in two types, namely variational minimization of energy and imaginary time evolution. With only minor changes, the second one allows also the simulation of real time evolution. Both kinds of algorithm can be formulated in terms of the minimization of a certain cost function by varying the tensor parameters.[1] This minimization is realized in practice by means of an alternating least squares (ALS) scheme, in which one sweeps

[1] We remark that the initial time evolution method [7,8,10] works slightly different, and the basis of this chapter is rather given by reference [12].

over the tensors and updates them one after another, each time choosing the components that minimize the cost function under the constraint that all the other tensors are fixed.

Throughout this chapter we focus (almost exclusively) on the imaginary time evolution. In this case, it is customary to use a Suzuki-Trotter approximation of the evolution operator where the Hamiltonian is split into parts containing only mutually commuting terms. The cost function to be minimized is then the distance $\mathrm{d}(|\psi\rangle) = |||\psi\rangle - \hat{O}|\phi\rangle||^2$, where $|\phi\rangle$ is the initial PEPS, $\hat{O}$ is an operator representing one (or more) *Trotter gates*, i.e. the exponential of one (or several) such Hamiltonian terms [41], and $|\psi\rangle$ is the resulting PEPS. During the ALS sweeping, the tensor for site l is the one that minimizes

$$\mathrm{d}(A_l) \quad =: \quad \vec{A}_l^\dagger N_l \vec{A}_l - \vec{A}_l^\dagger \vec{b}_l - \vec{b}_l^\dagger \vec{A}_l + \text{const.} \tag{2.1}$$

It is given by the solution of the linear system of equations $N_l \vec{A}_l = \vec{b}_l$, i.e. $\vec{A}_l = N_l^{-1}\vec{b}_l$, where the norm matrix N_l results from the norm TN $\langle\psi|\psi\rangle$ by leaving out the tensor A_l^* in the bra and A_l in the ket, and $\vec{b}_l$ results from the TN $\langle\psi|\hat{O}|\phi\rangle$ by leaving out A_l^* in the bra. This procedure can be iterated for the necessary number of steps to reach the desired total (real or imaginary) time.[2]

Two main parts, namely the *environment approximation* and the *tensor update*, constitute the building blocks of this algorithm and will be often referred to in the rest of this chapter. The first notion corresponds to the exact or approximate evaluation of the effective matrix (N_l) and vector ($\vec{b}_l$) that determine the local equation to be solved for the tensor at a given site. The second term denotes the solution of the vector equation and the corresponding change of the PEPS with the updated tensor.

Some strategies developed in the context of iPEPS can also be applied to the finite case, and we will do that in the following. The most widely used iPEPS method, due to its efficiency and stability, is the simple update (SU) [42], in which the environment is assumed to be separable and then the tensors are updated via simple SVD. As we showed in the previous chapter, the SU works equally with finite PEPS but produces results with limited accuracy. The full update (FU) [17, 22] is based on a more accurate approximation of the environment, in closer analogy to the original finite PEPS algorithm [11, 41], but differing from it in the fact that Trotter gates are not applied simultaneously, so that the environment for the update of one gate does only require the norm contraction around that gate. We will in the following use the term FU in the context of finite PEPS to denote the sequential application of Trotter gates together with the full contraction of the norm TN, as in the previous chapter.

2.3 Algorithmic Aspects

In this section we analyze several distinct aspects of finite PEPS algorithms, regarding both the environment approximation and the tensor update.

In particular, for the environment approximation we show how the success of the cluster scheme introduced in the previous chapter is deeply connected to

[2] In the direct minimization of the energy, the cost function to be minimized is $E(|\psi\rangle) = \langle\psi|\hat{H}|\psi\rangle / \langle\psi|\psi\rangle$ over the PEPS $|\psi\rangle$. For an update during the ALS sweeping, the tensor at position l is set to A_l minimizing $E(\vec{A}_l) = \vec{A}_l^\dagger H_l \vec{A}_l / \vec{A}_l^\dagger N_l \vec{A}_l$, which can be found as the lowest eigenstate of the generalized eigenproblem $H_l \vec{A}_l = \lambda N_l \vec{A}_l$. Here, the matrix H_l is defined from the TN $\langle\psi|\hat{H}|\psi\rangle$, in which one leaves out the tensor A_l in the ket and A_l^* in the bra, and N_l is the norm matrix.

the correlation length of the state. We also discuss the feasibility and the cost of explicitly keeping a positive environment by making use of purification MPO.

For the tensor update we propose gauge choices for each possible update scheme, and show how they improve the numerical stability of the algorithms. We additionally discuss how the reduced tensor, originally introduced in the context of iPEPS [22], can similarly be used in the finite case to speed up the computations. The normalization of the tensors is another factor that can improve the stability of the method.

While part of this section is significantly technical, the considerations exposed here are relevant for the implementation of any (finite) PEPS algorithm. Furthermore they have also clear physical implications, especially in the case of the environment contraction.

2.3.1 Environment Approximation

In the imaginary time algorithm, the update of one tensor at lattice site l involves the contraction around that tensor of the norm TN $\langle\psi|\psi\rangle$ and of the TN $\langle\psi|e^{-\tau\hat{H}_x}|\phi\rangle$ for a certain subset of (mutually commuting) Hamiltonian parts $\hat{H}_x$. The first contraction leads to the norm matrix N_l and the second to the vector $\vec{b}_l$ from which the new tensor for that lattice site follows as $\vec{A}_l = N_l^{-1}\vec{b}_l$. The original algorithm [11, 41] includes the complete $e^{-\tau\hat{H}_x}$, i.e. all (mutually commuting) Trotter gates, in the TN $\langle\psi|e^{-\tau\hat{H}_x}|\phi\rangle$ and thus requires two independent environment approximations, one for $\langle\psi|\psi\rangle$ and one for $\langle\psi|e^{-\tau\hat{H}_x}|\phi\rangle$. However, in the following we adopt the strategy from the previous chapter: If Trotter gates are applied one by one, and only the tensors on which a given gate acts are modified, then it suffices to consider the environment approximation of the norm TN alone, and, starting from this environment, the vector $\vec{b}_l$ is constructed from the exact contraction of a single Trotter gate.[3]

Accuracy of the Cluster Contraction

The cluster update (CU) introduced in the previous chapter allows a trade-off between precision and efficiency in the environment approximation. The SU and the FU are special cases of this procedure, which naturally interpolates between them in both accuracy and computational cost. Because clusters are not only useful for the tensor update but equally for the computation of expectation values, they realize a unifying framework for PEPS contractions.

We observed in the previous chapter that the contraction error decreases exponentially with the cluster size for PEPS ground state approximations of the Heisenberg model. This property, which justifies the usability of clusters, is ultimately related to a finite correlation length of the system, as we appreciate here with the help of the quantum Ising model. This model becomes critical in the thermodynamic limit at transverse field $B \approx 3.044$, and, thus, by varying B we can create states with different correlation lengths.

We have analyzed the cluster contraction error of a local observable acting on the center of the lattice, $\epsilon_\alpha(\delta) := |\langle\sigma^\alpha\rangle_\delta - \langle\sigma^\alpha\rangle|/|\langle\sigma^\alpha\rangle|$, for $\alpha = X, Z$, where $\langle\sigma^\alpha\rangle_\delta$ is the approximated contraction using cluster size δ, and $\langle\sigma^\alpha\rangle$ the result of contracting the full TN. The behavior of this quantity can be compared to the correlation

[3] In an efficient implementation of this algorithm we only store and update the boundary MPO for the rows and columns of the norm TN, and the contraction for the vector $\vec{b}_l$ is performed on the fly.

function, $G_\alpha(x) := \langle \sigma_l^\alpha \sigma_{l+x}^\alpha \rangle - \langle \sigma_l^\alpha \rangle \langle \sigma_{l+x}^\alpha \rangle$, for two sites separated by a distance x along the central column of the lattice. All contractions were performed with large enough $D' = 100$ such that the contraction error was independent of D'.[4] We observe in figure 2.1 (a) that the decrease of the contraction error is always steeper for a faster decaying correlation function. In order to make this statement more precise, we can fit the decay of the error to an exponential function of the cluster size, $\epsilon_\alpha(\delta) \propto \exp(-\delta/\delta_0)$, and obtain a characteristic cluster size δ_0. Correspondingly, we can extract a correlation length ζ from a similar fit of the correlation function $G_\alpha(x) \propto \exp(-x/\zeta)$. After having calculated δ_0 and ζ for several PEPS,[5] we plot δ_0 as a function of ζ in figure 2.1 (b) and conclude that $\delta_0 \approx \zeta$.[6] This demonstrates an extremely clear quantitative connection between the cluster contraction error for a given cluster size and the correlations in the state.

Positive Environment

The exact norm environment, resulting from an exact contraction of $\langle \psi | \psi \rangle$ around one (or several) site(s), is positive by construction, as can be seen in figure 2.2. Although this positive characteristic is considered a desirable property for the environment approximation, in general it is not respected by the approximated contractions. Nevertheless, it is possible to use schemes that maintain it. In particular, the single-layer (SL) algorithm was introduced in reference [43] as a way to improve the environment approximation of the SU while preserving its efficiency and numerical stability. The SL method performs the norm contraction by means of transformations in the ket alone [43]. Then the boundary is described by a *purification* MPO [12], defined via a MPS of virtual bond dimension D'' and physical dimension $D \times d'$ in such a way that the MPO results from tracing over the *purification bonds* of dimension d'. Approximating the environment in the SL way and then updating the tensors as explained in reference [43] ensures a stable algorithm, but as seen in the previous chapter the error in the environment approximation can be several orders of magnitude above that of the original contraction [11, 41], and it can depend strongly on the system size.

Several factors can cause these accuracy limitations. Even if there exists a good positive MPO approximation for the boundary with moderate bond dimension D' (as observed for gapped systems [81]), it does not necessarily follow that D'' is small [83, 89] and hence it is not clear a priori that fixing the maximum D'' produces an accurate approximation for the environment. Moreover, as argued in the previous chapter, by operating on a single layer, the scheme does not find the most general purification with given bond D''. Here we want to address the question wether the accuracy limitations of the SL algorithm are due to the description of the boundary as purification or wether they are due to the specific operations proposed in reference [43] to determine that boundary purification.

One way to allow for a more general purification is to formulate an algorithm in the double-layer picture in the following way. Given the MPS bond dimension, D'', and the purification bond, d', we write a purification MPO [12] by using in the lower layer the complex conjugated tensors from the upper layer. The problem of approximating the boundary after the contraction of one further row of the

[4] The dependence of the cluster contraction on D' was already investigated in the previous chapter.

[5] We determine δ_0 via the two values of $\epsilon_\alpha(\delta)$ at $\delta = 2$ and 4, and ζ via the two values of $G_\alpha(x)$ at $x = 4$ and 8.

[6] The slight deviation of figure 2.1 (b) from the exact diagonal for larger ζ is due to the fact that they correspond to $B \approx 3.0$ where both $\epsilon_\alpha(\delta)$ and $G_\alpha(x)$ decrease rather polynomially and thus δ_0 as well as ζ depend more strongly on the two values from which they were determined.

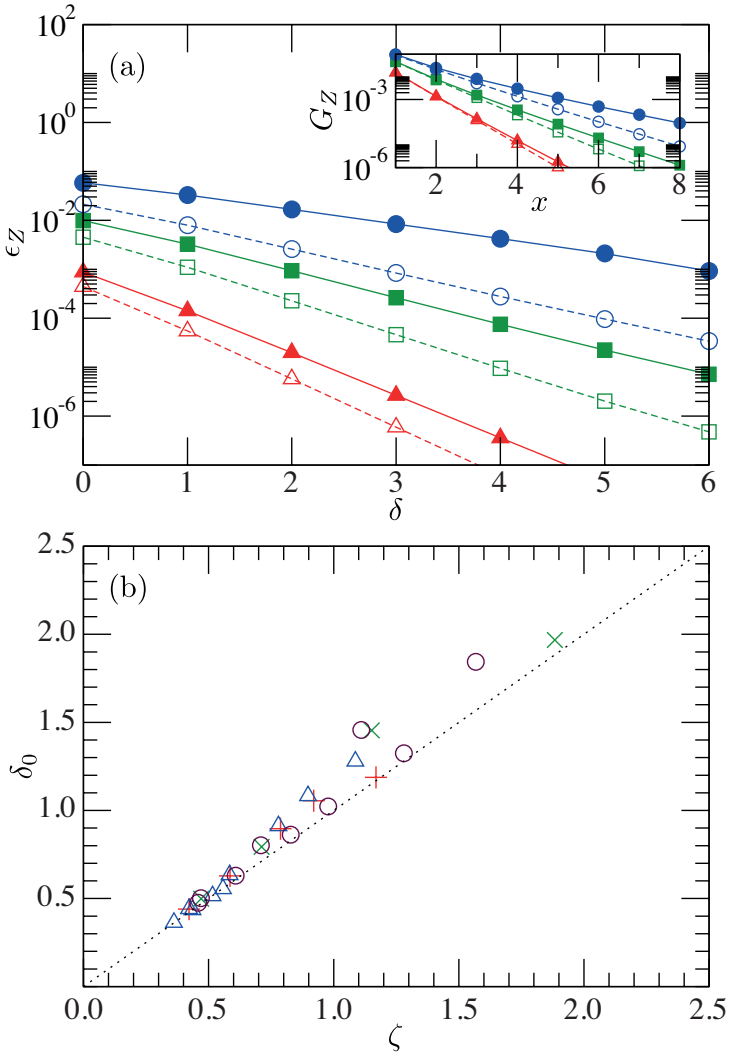


Figure 2.1: Relation between cluster error and correlation function in the Ising model on a 21×21 lattice. (a) Cluster error (main plot) and correlation function (inset) for observable σ^Z, for $D = 2$ (open symbols), 3 (filled symbols), and $B = 2.0$ (triangles), 2.5 (squares), 2.8 (circles). (b) Characteristic cluster size, δ_0, versus correlation length, ζ, for several values of $B \in [2, 4]$, for observable σ^Z with $D = 2$ (plusses), 3 (crosses), and for σ^X with $D = 2$ (triangles), 3 (circles).

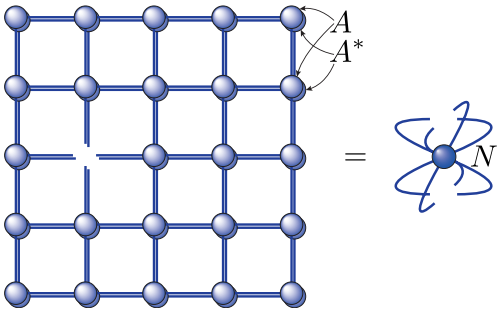


Figure 2.2: Norm environment for a single site of a 5×5 PEPS. Because each PEPS tensor A (see figure 1.1) is contracted with its complex conjugate A^* over the physical index, an exact contraction of this norm TN would give a positive Hermitian norm matrix N. For large PEPS, an exact contraction of N is not feasible, and the original contraction approximation [11,41] based on general boundary MPO (shown in figure 1.2) does not keep the positivity.

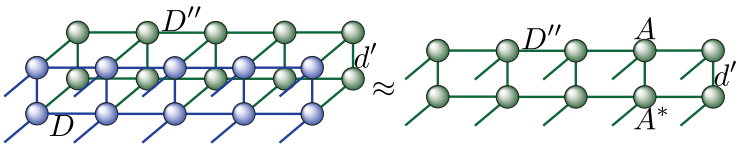


Figure 2.3: Positive contraction of the norm TN for a PEPS with bond dimension D. The product of a bulk row with a boundary purification of virtual bond dimension D'' and purification bond dimension d' is approximated by a new boundary purification of the same dimensions. Each update of a tensor A constitutes a nonlinear problem, and solving the linearized equations costs $\mathcal{O}(dD^6D''^4) + \mathcal{O}(D^4D''^6) + \mathcal{O}(d'^3D^3D''^6)$ where $d' \leq DD''^2$.

PEPS norm TN is then formulated for this structure instead of the general MPO, as sketched in figure 2.3. The local equations result from replacing the single tensor of the general MPO by the structure consisting of A_l and A_l^*. Following the standard ALS procedure, we sweep over the sites l, and for each site solve the corresponding optimization problem for A_l. However, in this case the cost function to be minimized is no longer quadratic, but quartic in the variables of a tensor at site l, and its minimum corresponds to the solution of nonlinear equations, in contrast to the linear equations encountered in the original contraction [11, 41] of figure 1.2. The nonlinear equations for A_l have to be solved iteratively. We describe and benchmark several options in section 2.6.1.

We compare this scheme to the original [11, 41] and the SL algorithm based on the norm contraction of the same PEPS used in the analysis of the previous chapter. The results are shown in figure 2.4. For a fixed purification bond d', we observe that the relative error of the norm decreases fast as a function of D''. The comparison of the 11×11 to the 21×21 lattice (figure 2.4 (a)) shows that, similar to the original algorithm [11, 41], the error does not have a strong dependence on the system size. On the other hand, with growing d', the curves tend to converge to the error of the original contraction [11, 41]. This effect can be observed already with small purification bonds for $B = 1.0$ (figure 2.4 (b)). These results suggest that the error in the SL method is mainly due to the restricted class of purifications it can attain, and not to the description of the boundary as a purification with small

bond D''.[7]

From the discussion above we conclude that it is possible to efficiently find a (close to) optimal general purification by means of the solution of nonlinear equations. In the context of PEPS contractions, this technique improves the SL scheme significantly, but given its higher computational cost compared to the original contraction [11, 41], resulting from the iterative routines (see section 2.6.1), it is not a practical option. Hence, in the following, all our cluster and full contractions will be based upon the original contraction algorithm [11,41] and thus make use of general boundary MPO as shown in figure 1.2. Nevertheless, the procedures analyzed here may be useful for other problems where the question of numerically optimizing a purification MPO appears, such as for the description of one-dimensional thermal states or open systems.

2.3.2 Tensor Update

Once the environment is computed, the actual update of the tensors takes place by solving the appropriate local equations. It is also possible to use simplifications of this step which render a more efficient and stable algorithm.

For the update of a pair of neighboring tensors, the environment can be approximated in general by a MPO with periodic boundary conditions, as illustrated in figure 2.5. A first simplification of the tensor update procedure comes from sequentially processing the Trotter gates, as described above, and changing only the tensors on which each gate acts. Then all the update operations on the pair are performed with a fixed environment.

The computational cost of the tensor update can be greatly decreased by restricting the update to the *reduced tensor* [22]. This reduced tensor update minimizes the cost function (2.1) for the full tensors exactly only in the case of a separable environment (see previous chapter), as e.g. in the SU, but it is worth studying its performance in a more general situation. In any case, it allows to work with larger bond dimensions, which might compensate for the smaller number of variational parameters.

Another major difference between MPS and PEPS concerns the conditioning of the effective norm matrix N_l. For MPS with open boundary conditions, a *gauge transformation*[8] can be chosen such that $N_l = 1$, which guarantees the stability of the tensor update. Although this is impossible for PEPS, we will show how a proper gauge choice and tensor normalization drastically improve the stability of the algorithm.

Reduced Tensor

Before performing the update under a nearest-neighbor Trotter gate, the tensor for a lattice site can be decomposed into the contraction of two tensors, in such a way that one of them carries the physical index and the virtual bond corresponding to

[7] Although our conclusions are based on PEPS from the SU, because we had thoroughly analyzed exactly the same PEPS with the SL algorithm in the previous chapter, we clearly expect similar improvements for PEPS from the FU (in particular the ground state of the quantum Ising model at $B = 1.0$, away from criticality, should be well approximated by the SU and thus figure 2.4 (b) should only change slightly for the corresponding FU PEPS).

[8] In a PEPS, for any pair of neighboring tensors that are connected via a virtual index, an arbitrary matrix M can be contracted with one tensor and the matrix M^{-1} with the other in such a way that the state does not change. This establishes the analogue of a (local) gauge freedom. Throughout this chapter, the matrix M is called *gauge matrix* or *gauge transformation* in accordance with reference [41], and the term *gauge fixing* or *gauge choice* refers to the process of choosing a specific matrix M.

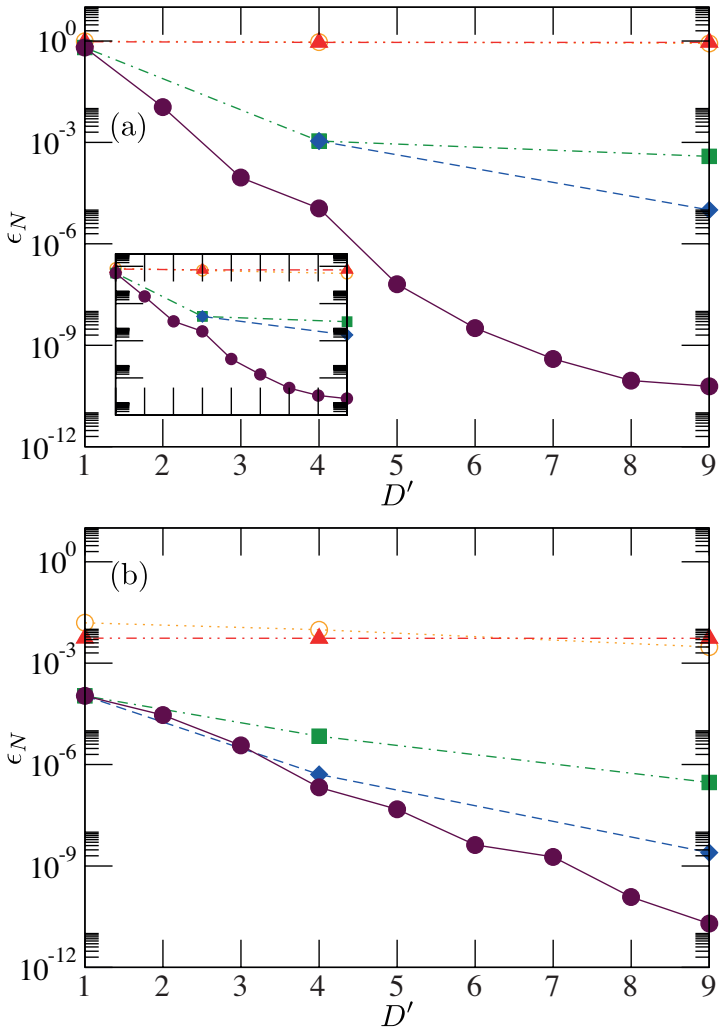


Figure 2.4: Relative error of the norm contraction using boundary purification MPO, for SU Ising ground state approximations. (a) $B = 3.0$ and $D = 2$ on lattices of size 21×21 (main plot) and 11×11 (inset). (b) $B = 1.0$ and $D = 4$ on a 11×11 lattice. For reference, we show the error of the SL method with maximum $d' = DD''^2$ (open circles) and of the original algorithm [11, 41] (filled circles). Our purification contraction was performed with $d' = 1$ (triangles), 2 (squares), and 3 (diamonds), and $D'' = \sqrt{D'}$.

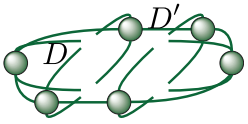


Figure 2.5: The 6 environment tensors of a nearest-neighbor tensor pair. They form a periodic boundary MPO with virtual bond dimension D' and physical dimension D. It is constructed with $\mathcal{O}(dD^6D'^2) + \mathcal{O}(D^4D'^3)$ operations, resulting from the optimal search for a boundary MPO and the contraction of the environment up to the tensor pair.

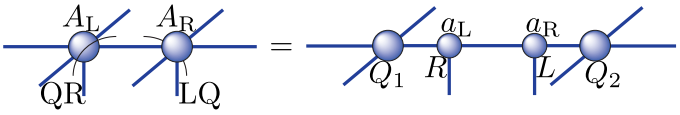


Figure 2.6: A QR decomposition of the left full tensor A_{L} generates the left reduced tensor a_{L} as the R. Similarly, a LQ decomposition of the right full tensor A_{R} gives the right reduced tensor a_{R} as the L. The initial dD^4 variational parameters of the full tensor are decreased to the d^2D^2 variational parameters of the reduced tensor.

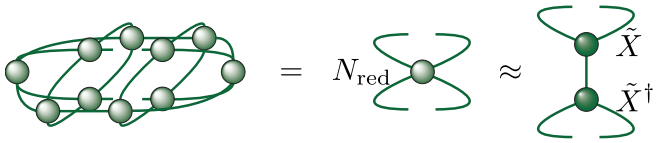


Figure 2.7: Environment tensor N_{red} of a reduced nearest-neighbor tensor pair and its closest positive semidefinite approximant $\tilde{X}\tilde{X}^\dagger$ constructed as explained in the text. The contractions are characterized by the leading computational cost $\mathcal{O}(d^4D^4D'^2) + \mathcal{O}(d^2D^6D'^2) + \mathcal{O}(d^2D^4D'^3)$ and the computation of the positive approximant requires additionally $\mathcal{O}(d^6D^6)$ operations.

the link on which the two-site gate acts, i.e. all the indices directly affected by the gate. This tensor is called the *reduced tensor* [22], and can be obtained from the full tensor by means of a QR decomposition, as sketched in figure 2.6.

In the reduced tensor update, only the components of such reduced tensor are modified during the update procedure, while the remaining part of the full tensor is left unchanged. These remaining parts of both tensors in the pair are contracted with the periodic MPO of figure 2.5 to get the environment for the reduced tensor pair, N_{red}, shown in figure 2.7. Due to the approximate contractions, this reduced environment is in general not positive, neither is it Hermitian, but its positive approximant can be constructed in two steps [90]. First, we compute the optimal Hermitian approximant $\tilde{N}_{\mathrm{red}} := (N_{\mathrm{red}} + N_{\mathrm{red}}^\dagger)/2$. Second, from its eigendecomposition $\tilde{N}_{\mathrm{red}} = U\Sigma U^\dagger$ we obtain the positive approximant as $U\Sigma_+U^\dagger$ where Σ_+ results from Σ by setting all negative eigenvalues to zero. Finally, the environment is written as $\tilde{X}\tilde{X}^\dagger$ in terms of its square root $\tilde{X} := U\sqrt{\Sigma_+}$.

The computational cost of the contractions for the periodic boundary MPO (figure 2.5), needed in both the reduced and the full tensor update, reads $\mathcal{O}(dD^6D'^2) + \mathcal{O}(D^4D'^3)$. The construction of N_{red} (figure 2.6 and 2.7) is only slightly more expensive with $\mathcal{O}(d^4D^4D'^2) + \mathcal{O}(d^2D^6D'^2) + \mathcal{O}(d^2D^4D'^3)$ operations. Its eigendecomposition requires $\mathcal{O}(d^6D^6)$. In the complete update of the reduced tensors via the sweeping of the ALS scheme, all further operations have lower computational cost. Notice that in the case of the full tensors the contraction of the norm environment for a single tensor needs $\mathcal{O}(D^8D'^2) + \mathcal{O}(D^4D'^3)$ operations while the eigendecomposition of the norm matrix has the cost $\mathcal{O}(d^3D^{12})$.

In order to study wether the reduced tensor limits the accuracy of the method, we considered imaginary time evolution of the Heisenberg model on 4×4 and 10×10 lattices, and compared the final energies from the reduced tensor update to the ones from the full tensor update. We found that, while for the small bond dimensions $D = 2$ and 3 the full tensor update produced better energies, for $D = 4$ the energies of both approaches were already very similar. This can be appreciated by comparison of the results in section 2.6.3 (obtained with the reduced tensor) to

the full tensor results published in references [11, 62].

Because the reduced tensor update is less costly, we could reach larger bond dimensions than with the full tensor update, and, in the end, obtained the lowest energies with the reduced tensors. Therefore the reduced tensor update was used for the results presented in this chapter.

Gauge Fixing

In the case of MPS, it is possible to keep up a canonical form of the tensors during their updates with the help of a local gauge fixing, and that ensures the stability of the algorithm and optimizes its performance [41]. In the case of PEPS, there exists neither such a canonical form nor any means to locally gauge away the norm matrix. Nevertheless, using the gauge freedom, it is possible to improve the conditioning of the norm matrix and positively affect the precision and stability of the method, as we describe in the following.[9]

We propose a gauge fixing that is inspired by the one-dimensional case with open boundary conditions. In that case, the norm tensor can be reduced to the identity by (partially) imposing the canonical form of the MPS, achieved by QR (or LQ) decomposition of each tensor after its update [41]. Alternatively, for an arbitrary MPS it is always possible to reduce the norm matrix to the identity by taking the square roots of the unconnected left and right environment halves and absorbing part of their QR (LQ) decompositions in the tensor to be updated.

In the case of PEPS, it is not possible to ensure an identity norm matrix by means of QR or LQ decompositions after the tensor update. Hence, we adapt the second possibility and obtain the gauge transformations from the environment before the tensor update, namely from the norm tensor itself, such that the norm matrix is better-conditioned. Because this gauge fixing can be combined with any of the environment approximations described previously, we propose a precise scheme for each case.

When the environment of the tensor pair is separable, i.e. $D' = 1$ in figure 2.5, it decomposes into six positive semidefinite matrices, which can be determined by the algorithm from section 1.5.1. We compute the square roots of these matrices and absorb them in the tensor pair. After contraction of the tensor pair with the Trotter gate, a SVD is performed to find the new tensors, and finally these are multiplied by the inverses of the previous square roots. This procedure coincides with the SU [42] in which the λ matrices surrounding the tensor pair are substituted here by the square roots of the environment matrices corresponding to each link. Since the positive separable environment of the tensor pair is obtained with $\mathcal{O}(dD^5)$ operations, the leading cost of the complete update is $\mathcal{O}(d^6D^3) + \mathcal{O}(d^2D^5)$, under the assumption $d \leq D^2$.

When the environment of the tensor pair is non-separable, and we restrict the update to the reduced tensor, we propose the gauge fixing from figure 2.8. By taking R and L from independent QR and LQ decompositions of the same $\tilde{X}$ from figure 2.7, we treat both virtual bonds of the environment equally, such that both reduced tensors will experience similar condition numbers in the linear equations of the following sweeping.[10] After we have obtained the desired better-conditioned

[9] A recent alternative approach for an improved imaginary time evolution is presented in reference [91], in which the authors propose a "quasicanonical form" that arises as fix point of the SU performed with nearest-neighbor identity gates, and in which they demonstrate how that form allows an efficient and stable imaginary time evolution with Projected Entangled Pair Operators (PEPO) in the context of iPEPS.

[10] We also explicitly checked the resulting condition numbers when first a QR and then a LQ, or the

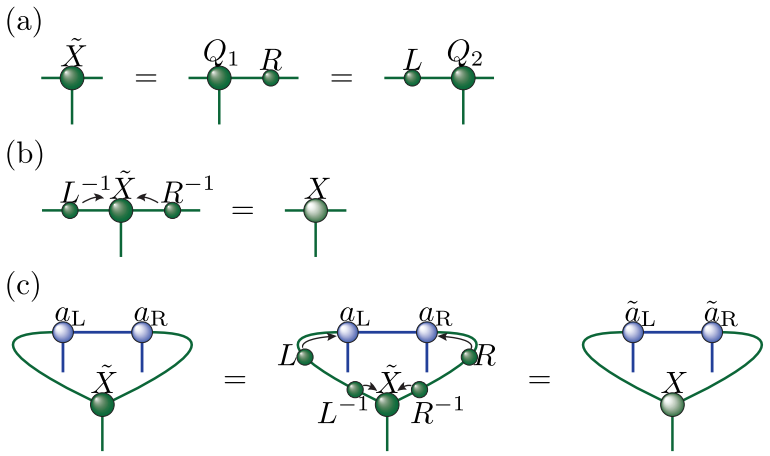


Figure 2.8: Gauge fixing on the environment tensor of the reduced tensor pair, when the environment is non-separable. (a) We perform a QR and LQ decomposition on $\tilde{X}$ from figure 2.7 independently of each other (notice that we have shortened here the horizontal open indices of $\tilde{X}$ compared to figure 2.7). (b) Contraction of $\tilde{X}$ with L^{-1} and R^{-1} gives the final square root of the environment tensor, X. (c) In order to leave the state unchanged, the left and right reduced tensors a_L and a_R from figure 2.6 have to be contracted with the gauge transformations L and R as shown here, which gives the starting tensors $\tilde{a}_L$ and $\tilde{a}_R$ for the update explained in figure 2.9.

square root of the environment tensor, X, in order to leave the state unchanged, the left (right) reduced tensor has to be contracted with L (R) over its left (right) virtual index.

After our gauge fixing has been applied, the actual update takes place in three steps. First, the tensors are initialized using a SVD as shown in figure 2.9 (a). This step coincides with the SU. If the environment is separable, the cost function is already minimal. In any other case, we can anticipate good starting tensors that are closer to the minimum of the cost function. Second, we optimize the tensors by means of the standard ALS sweeping, in which each tensor update is followed by the standard gauge fixing [41], i.e. the left (right) tensor is QR (LQ) decomposed along its right (left) virtual bond. Third, a gauge choice is made on the internal link of the converged pair as shown in figure 2.9 (b).

We have observed that our gauge choices improve the condition number of the norm matrix by several orders of magnitude in all studied cases. This statement is quantified by the results in table 2.1, which compares typical condition numbers found in the simulation of the Ising and Heisenberg models with and without our gauge fixing. Strictly speaking, the condition number of the norm matrix N_l provides only an upper bound for the final error of the solution $\vec{A}_l$ to the linear system of equations [92]: Therefore, a large condition number does not imply low accuracy, but a small condition number implies high accuracy of the solution. In practical computations with finite PEPS, when our gauge transformations are not used instabilities can occur (e.g. as reported in reference [43]) that we have never encountered after our gauge fixing.

We can also investigate the effect of our gauge fixing on the convergence of

other way round, or only a single decomposition is applied. In these cases, one tensor always encounters better condition numbers than the other in the linear equations of the following sweeping.

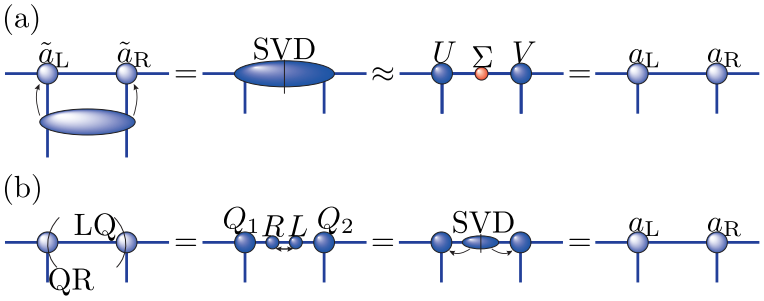


Figure 2.9: Initial and final step of the reduced tensor update. (a) Initialization: Before we update the reduced tensors by sweeping, we apply a SVD to their joint contraction with the Trotter gate, and keep only the D largest singular values in Σ. Splitting Σ gives the initial tensors $a_L := U\sqrt{\Sigma}$ and $a_R := \sqrt{\Sigma}V$ for the ALS procedure. (b) Final form: After convergence of the ALS sweeping, we put the two tensors on an equal footing.

(a)	Model	Positive approximant	Gauge fixing
	$B = 1.0$ Ising	$(2 \pm 3) \cdot 10^7$	1.1 ± 0.1
	$B = 3.0$ Ising	$(2 \pm 3) \cdot 10^3$	1.6 ± 0.1
	Heisenberg	$(8 \pm 5) \cdot 10$	1.08 ± 0.02
(b)	Model	Positive approximant	Gauge fixing
	$B = 1.0$ Ising	$(9 \pm 205) \cdot 10^{13}$	$(1 \pm 3) \cdot 10^4$
	$B = 3.0$ Ising	$(4 \pm 158) \cdot 10^{13}$	$(5 \pm 6) \cdot 10^2$
	Heisenberg	$(3 \pm 2) \cdot 10^4$	5 ± 3

Table 2.1: We show the mean condition number of the norm matrix with its standard deviation in the reduced tensor update without our gauge fixing, using only the positive approximant, and with our gauge fixing during the FU imaginary time evolution of $D = 2$ (a) and $D = 4$ (b) PEPS of size $N = 11 \times 11$ for the Ising model and of size $N = 10 \times 10$ for the Heisenberg model. The values were obtained averaging over 10 time steps and all tensors in the lattice.

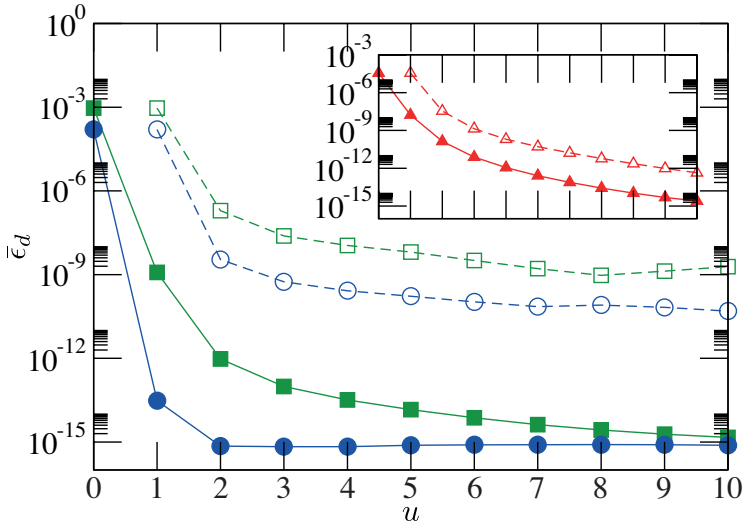


Figure 2.10: Mean value of the relative change $\epsilon_d(u) := |d(u) - d(u-1)|/|d_{\text{init}}|$ of the cost function d, equation (2.1), after consecutive update sweeps u over a tensor pair computed with respect to the initial value of the cost function d_{init}, for the $D = 4$ reduced tensor update setting of table 2.1. We compare the FU evolution without our gauge fixing using only the positive approximant (open symbols) to the same propagation with our gauge fixing (filled symbols), for a 11×11 Ising model at $B = 1.0$ (circles) and 3.0 (squares), and for a 10×10 Heisenberg model (triangles).

the ALS sweeping, which can be gathered from figure 2.10 for the update of the reduced tensor. Most remarkably, in the presence of the gauge transformations, already the initial SVD drastically reduces the cost function equation (2.1), by a value that in all considered cases is larger than the one attained after one sweep without the gauge transformations. Furthermore, the final total reduction of the cost function is also larger with our gauge fixing than without. Because the relative change of the cost function in figure 2.10 always decreases faster in the presence of our gauge transformations, we conclude that the latter accelerate the convergence of the ALS scheme. Our results indicate that a simplified tensor update consisting of the combination gauge fixing and SVD only, without the ALS sweeps, might be successful. Indeed, for the Ising model, the cost function after our gauge fixing and SVD is already smaller than after 10 sweeps without our gauge fixing. However, the sweeping can further decrease the cost function, and this is revealed most evidently for the Heisenberg model.

So far, we assumed that the update is performed on two directly neighboring tensors, after applying on them one of the Trotter gates of a nearest-neighbor Hamiltonian. The discussion can be extended to the update of more distant tensors, as would appear in the case of Hamiltonians with long-range interactions. However, for n non-adjacent reduced tensors the dimension of the norm matrix is typically $(d^2D^2)^n \times (d^2D^2)^n$, and its diagonalization, even for two tensors, is not desirable. A further simplification is to choose the gauge transformations from the local environment of a single tensor, since the norm matrix in this case has size $d^2D^2 \times d^2D^2$. By means of a numerical simulation we confirmed that such a local gauge choice can produce condition numbers comparable to the ones obtained

from the gauging of the environment of the pair.[11] Moreover, we found that the gauge matrices L and R computed from the local norm tensors of each separate tensor in the pair can also be applied to the pair environment, and then we can follow the update procedure of figure 2.9. Thus, we expect that local gauge choices similarly improve the tensor update in the case of general long-range interactions, where the tensor initialization and final form in figure 2.9 will be given by their analogues from TEBD [7].

While the discussion here is focussed on the reduced tensors, in section 2.6.2 we derive an efficient gauge fixing for the full tensors. This gauge fixing equally improves the condition number of the norm matrix and the convergence of the ALS sweeping in the tensor update of the full tensors. Because all our gauge transformations are derived from and applied to the norm TN alone, they do not explicitly depend on the operator whose action on the PEPS is approximated. We would therefore expect that our gauge choices similarly improve the original time evolution algorithm [11, 41] in which the action of Projected Entangled Pair Operators (PEPO) on PEPS is approximated.[12]

Stability Issues

The previously described gauge choices guarantee a better conditioned norm matrix. But for the stability, precision and efficiency of the algorithms, especially when the environment approximation is very rough (e.g. by using small clusters or boundary bond dimensions), also the following factors need to be taken into account.

- For PEPS, the matrices N_l are not exactly Hermitian and positive semidefinite. The advisable strategy is to replace them by their closest Hermitian approximants $(N_l + N_l^\dagger)/2$, and additionally set to zero any negative eigenvalues in order to get the closest positive semidefinite approximant of N_l, as described above for the environment of the reduced tensor pair.[13]

- In general, some eigenvalues of N_l are zero and its positive subspace is ill-conditioned. That is why N_l^{-1} must be a pseudoinverse. A cutoff is set such that only the subspace of N_l with eigenvalues larger than a certain value is considered in the construction of the pseudoinverse.

- Finally, the correct tensor normalization has a decisive impact. Imaginary time evolution steadily modifies the norm of the state. Thus we impose the normalization of the PEPS, $\langle\psi|\psi\rangle = 1$, after each set of Trotter gates, and, in order to avoid the existence of very small or very large tensors, we additionally scale all PEPS tensors to have the same largest element absolute.

[11] We propagated, using the FU, several 11×11 $D = 2$ PEPS, obtained from the SU for the Ising model with different fields, and we monitored eigenvalues and singular values of norm and gauge matrices throughout the evolution.

[12] We can also expect that the direct variational minimization of the energy benefits from our gauge transformations because they improve the condition number of the norm matrix N_l that enters the generalized eigenproblem $H_l\vec{A}_l = \lambda N_l \vec{A}_l$ which has to be solved for the tensor update at lattice site l.

[13] In the direct variational minimization of the energy, also the matrices H_l are not exactly Hermitian and hence should be replaced by their closest Hermitian approximants $(H_l + H_l^\dagger)/2$. Furthermore, the eigendecomposition of $(N_l + N_l^\dagger)/2$ enables us to replace the generalized eigenproblem by a standard one, $\sqrt{N_l}^{-1} H_l \sqrt{N_l}^{-1} \vec{B}_l = \lambda \vec{B}_l$. Its lowest eigenvector, $\vec{B}_l$, yields the desired new variational parameters via $\vec{A}_l = \sqrt{N_l}^{-1} \vec{B}_l$.

2.4 Performance of Finite PEPS

With the aim of analyzing its performance in terms of system size and bond dimension, we have applied the generic finite PEPS code to the ground state search for the Heisenberg and quantum Ising model, and compared the results to those obtained by other numerical methods, when available. Our best PEPS results were obtained with the FU, i.e. updating the reduced tensors, applying the Trotter gates sequentially, and approximating the full contraction of the environment by means of general boundary MPO. This combination of techniques allowed us to push the simulations to lattices of size up to 21×21. On the same systems, we ran also the SU for finite PEPS.

2.4.1 Convergence Procedure

In each case, the PEPS ground state approximation was found by means of imaginary time evolution. The initial state was always a $D = 2$ PEPS which was constructed by embedding in it a separable PEPS and replacing the zero entries by small random numbers. Beginning with the time step $\tau = 0.01$, the propagation was performed long enough for the energy to converge, and then the procedure was repeated for smaller time step(s). After convergence was attained for the minimum time step, the scheme was iterated for a larger bond dimension, starting from a previously converged PEPS as initial state.

We observed that the converged SU PEPS of a certain bond dimension was always a good initial state for further propagation with the FU for this bond dimension. On the one hand, in general, the SU PEPS can already be a good ground state approximation and then only few further steps with the FU are required. On the other hand, we have found, that such state required smaller values of D' when the evolution was continued with the FU.

Energies and correlators reported here for a certain value of D correspond to the final PEPS for the smallest time step. The error in the corresponding observable was estimated via the difference to the expectation value calculated with the converged PEPS for the previous time step. All contractions were performed with boundary bond dimension $D' = 100$, big enough to neglect contraction errors, as we explicitly checked by comparison to results from $D' = 200$.

2.4.2 Heisenberg Model

We considered imaginary time evolution with an antiferromagnetic Heisenberg Hamiltonian $\hat{H} = \sum_{\langle l,m \rangle} \vec{S}_l \cdot \vec{S}_m$. This model on a two-dimensional square lattice is a paradigmatic benchmark Hamiltonian because quantum Monte Carlo methods provide quasi exact results for very large system sizes [93], and thus we can directly compare our results to quantum Monte Carlo.[14] In the context of PEPS, the ground state order parameter of this model, i.e. the squared staggered magnetization $M^2_{\text{stag}} := \frac{1}{N^2} \sum_{l,m=1}^{N} (-1)^{l+m} \langle \vec{S}_l \cdot \vec{S}_m \rangle$, is particularly challenging [94] (also on a honeycomb lattice [68]) and a precise determination has so far only been possible with very large bond dimension $D = 16$ in reference [70]. Here we want to find out what our improved algorithmic procedures can do.

To our Heisenberg Hamiltonian we added a small staggered magnetic field $B_Z \sum_l (-1)^l S^Z_l$ which we slowly switched off during the evolution, starting from $B_Z = 10^{-3}$. In the presence of this staggered field the SU(2) symmetry of the

[14] We obtain our quantum Monte Carlo reference values from the ALPS library [84–86].

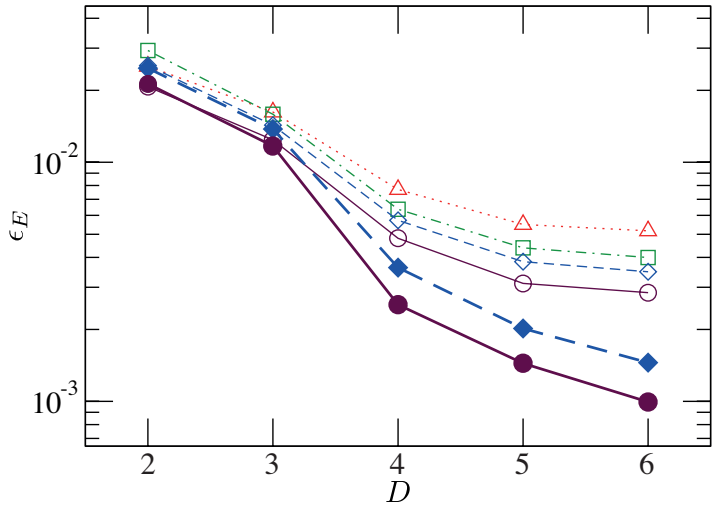


Figure 2.11: Relative energy error $\epsilon_E := |E(D) - E_0|/|E_0|$, where E_0 denotes the exact ground state energy from the ALPS library [84–86], of the SU (open symbols) and the FU (filled symbols) for different lattice sizes. In the case of the SU, we consider $N = 10 \times 10$ (triangles), 14×14 (squares), 16×16 (diamonds), and 20×20 (circles). In the case of the FU, we consider $N = 10 \times 10$ (diamonds) and 14×14 (circles).

Heisenberg model is explicitly broken and smaller values of D' suffice. This procedure improved the convergence of all our algorithms significantly. In the case of the SU, it helped to avoid local minima and reach lower final energies, in particular on the largest 20×20 lattice. And, in the case of the FU, already when the staggered field was still switched on, low values of the energy were attained while smaller values of D' were required. All propagations were performed for time steps $\tau = 10^{-2}$ and 10^{-3}.

Figure 2.11 shows the convergence of the energy with increasing bond dimension. We observe that, while the FU energy error decreases rapidly with D, the SU energies saturate, and for bond dimensions up to $D = 6$, the lowest SU energies lie between the values for $D = 3$ and 4 obtained with the FU. This is consistent with our earlier observations in the previous chapter based on smaller lattices. Both the SU and the FU produce better energies when the lattice size increases.

We can now compare our energy accuracies to the existing literature. The original finite PEPS algorithm [11, 41] obtained a lowest energy per site -0.62515 on a 10×10 lattice, using time step $\tau = 0.001$ and bond dimension $D = 4$ (this PEPS result is given in references [63, 95]). For this system size and the same values of τ and D we now achieve the slightly lower energy per site $-0.62637(2)$, and we can also provide the converged $D = 6$ result $-0.62774(1)$. All our energies as well as our quantum Monte Carlo reference values corresponding to figure 2.11 are collected in section 2.6.3. Our $D = 4$ energy per site is already lower than the best values reported for the wave function ansatzes (block) sequentially generated states (-0.61713) [63], entangled-plaquette states ($-0.6258(1)$) [95], and string bond states (-0.6225) [96], to which we can directly compare because they also considered finite systems with open boundary conditions. For infinite systems, the iPEPS ansatz attains slightly better energy precisions between 10^{-3} and 10^{-4} for $D = 4$ to 6, as reported in reference [97]. And for large finite cylinders, the best DMRG results are also more accurate: Reference [98] analyzes the Heisenberg

model on a cylinder with a constant staggered magnetic field on the boundaries and, by making use of S^Z symmetry in the algorithm, reaches an energy accuracy of 10^{-4} on a 20×10 lattice.

In order to check the accuracy of the ground state approximation, we evaluated also non-local observables. In particular, we computed the correlator $\langle \vec{S}_l \cdot \vec{S}_{l+x} \rangle$, in the center of the lattice for two sites separated by a distance x, either along the diagonal or along the same column. We checked explicitly that the correlators of the converged PEPS along the diagonal and vertical direction are quantitatively very similar. This feature is obviously due to the PEPS ansatz and would be harder to reproduce e.g. with MPS in two dimensions. The precision of our considered spin-spin correlator $\langle \vec{S}_l \cdot \vec{S}_{l+x} \rangle$ also indicates the precision that can be expected for the order parameter $M^2_{\text{stag}} := \frac{1}{N^2} \sum_{l,m=1}^{N} (-1)^{l+m} \langle \vec{S}_l \cdot \vec{S}_m \rangle$. Since the former quantity, being dependent on the distance x, provides more information than the latter quantity, being just a single number, we focus here on the spin-spin correlator.

The results for the diagonal correlators in 10×10 PEPS are shown in figure 2.12 (a), and figure 2.12 (b) displays the vertical correlators for 14×14 PEPS. We observe that the FU converges quickly to the true correlator with increasing bond dimension. Although for fixed D the error grows with the distance x, for fixed x it decreases fast with D. In particular, if we consider the correlator at distance $x = L/2$, as commonly done for the construction of the thermodynamic value via finite size scaling, we read off $\epsilon_C^{D=6} \approx 0.01$ and $\epsilon_C^{D=7} \approx 0.003$ on the 10×10 lattice, and we find $\epsilon_C^{D=5} \approx 0.07$ and $\epsilon_C^{D=6} \approx 0.01$ on the 14×14 lattice. As for the energy, the SU results saturate, and they get better when the system size is larger.

We want to compare our results for the spin-spin correlator to previous works. The widely used iPEPS algorithms achieve a remarkably low relative energy error in the thermodynamic limit [97] while their relative correlator error ≈ 0.1 reported in reference [94] for $D = 5$ is still rather high (although larger values of D are accessible within iPEPS algorithms nowadays [26,99] by making use of symmetries [100,101]). In reference [70] the SU was used together with Monte Carlo sampling to reach much larger bond dimensions, and their best accuracies obtained with $D = 16$ were $0.003(2)$ on a 8×8 lattice and $0.013(2)$ on a 16×16 lattice, assuming periodic boundary conditions. We now attain the same precisions here on 10×10 and 14×14 lattices already with much smaller bond dimensions $D = 6$ and 7. Again, the best DMRG results are still more accurate: Reference [98] reports an uncertainty of 0.0007 for the observable $|\langle S^Z \rangle|$ in the center of a 20×10 cylinder with constant staggered magnetic fields on the boundaries.

We can try to understand the characteristics of the SU and the FU results with the help of the environment approximation used in their tensor updates. As we have argued in the previous chapter, SU and FU represent special cases of a unifying CU_δ: the SU is equivalent to clusters of size $\delta = 0$ in the tensor update, while the FU corresponds to the largest possible cluster size $\delta = L - 1$. Here we showed in section 2.3, that the cluster contraction error as a function of the cluster size behaves like the correlation function of the considered PEPS, such that states with short correlation lengths can be accurately contracted by means of small clusters. It is then reasonable to expect that the cluster size δ used in CU_δ limits the finally achievable correlation length. We address this question on a 10×10 lattice with $D = 4$ in the main part of figure 2.13. Indeed, the correlation function decays slower when larger clusters are used in the CU.

Moreover, we can gather from figure 2.13 that the correlation functions for system size 14×14 from SU as well as FU decay faster than the corresponding ones for system size 10×10, while figure 2.11 shows that a higher energy accuracy is

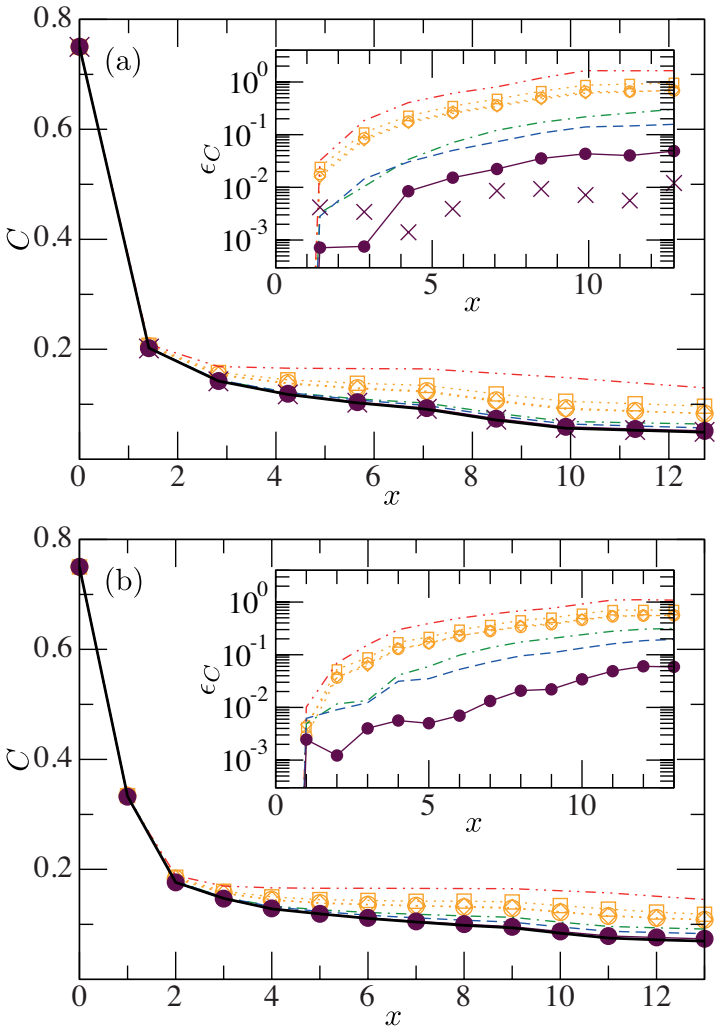


Figure 2.12: Spin correlations $C(x) := |\langle \vec{S}_l \cdot \vec{S}_{l+x} \rangle|$ (main) and relative error $\epsilon_C(x) := |C(x) - C_0(x)|/|C_0(x)|$ (inset), with the exact values $C_0(x)$ (thick line) from the ALPS library [84–86], for two sites separated by distance x along the diagonal in the center of 10×10 PEPS (a) and along the vertical in the center of 14×14 PEPS (b). We consider PEPS Heisenberg ground state approximations from the FU with $D = 2$ (dash-double-dotted), 4 (dash-dotted), 5 (dashed), 6 (filled circles), and 7 (crosses), and from the SU with $D = 4$ (squares), 6 (diamonds), and 8 (open circles).

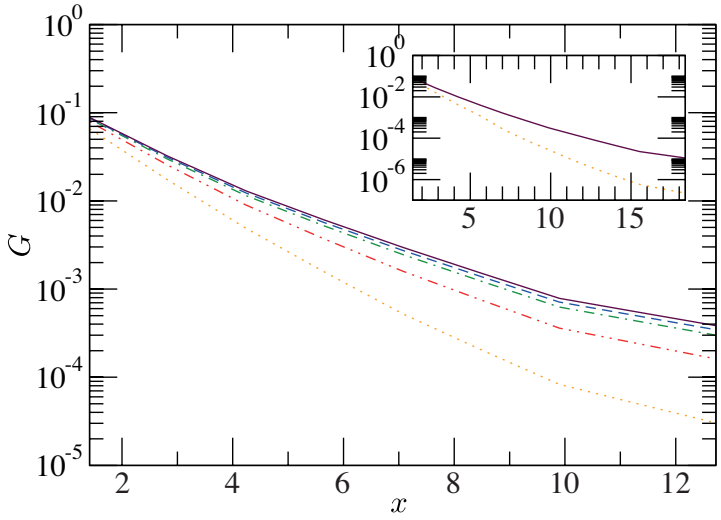


Figure 2.13: Correlation function $G(x) := \langle \vec{S}_l \cdot \vec{S}_{l+x} \rangle - \langle \vec{S}_l \rangle \cdot \langle \vec{S}_{l+x} \rangle$ for two sites separated by distance x along the diagonal in the center of 10×10 (main) and 14×14 (inset) PEPS Heisenberg ground state approximations of bond dimension $D = 4$. We compare SU (dotted), CU_1 (dash-double-dotted), CU_2 (dash-dotted), CU_3 (dashed), and FU (solid).

attained on the larger lattice. This indicates that, for the finite systems with open boundary conditions considered here, the true correlation length of the Heisenberg model slightly decreases with growing lattice size. In the context of the SU, this would explain why the SU results of figures 2.11 and 2.12 are better on larger lattices. And in the context of the FU, this would explain our numerical observation that the convergence of energies and spin-spin correlators required smaller values of D' for larger systems:[15] A smaller correlation length can be captured with a smaller cluster size δ in the CU_δ and the contraction precision achieved with such δ can equally be obtained by the full contraction, used in the FU, with correspondingly smaller value of D' (see figures 1.12 and 1.13 in the previous chapter).

2.4.3 Quantum Ising Model

We have also applied our finite PEPS algorithms to the quantum Ising model with transverse field, $\hat{H} = -\sum_{\langle l,m \rangle} \sigma_l^Z \sigma_m^Z - B \sum_l \sigma_l^X$. This Hamiltonian features a quantum phase transition in the thermodynamic limit, and its critical point $B_c \approx 3.044$ and exponent $\beta \approx 0.327$ are known very accurately thanks to finite size scaling with quantum Monte Carlo [102]. Since iPEPS have already very successfully demonstrated the adequacy of the PEPS ansatz for the quantum Ising model even at criticality [17,66] (reference [66] reports $B_c \approx 3.04$ and $\beta \approx 0.328$), we present our results here and in section 2.6.3 just for benchmark purposes, e.g. to enable a comparison with another PEPS implementation or with another wave function ansatz. We thus consider here only few different values of the magnetic field around $B = 3$ and run our computations only for the two system sizes 11×11 and 21×21. For

[15] To be precise, our FU energies and spin-spin correlators for the Heisenberg model were converged with $D' = 2D^2$ for all $D = 2$ to 6 on the 14×14 lattice, i.e. our results did not change anymore when we further ran the FU with larger D'. However, this convergence occurred only with $D' = 75$ for $D = 5$ and $D' = 126$ for $D = 6$ on the 10×10 lattice.

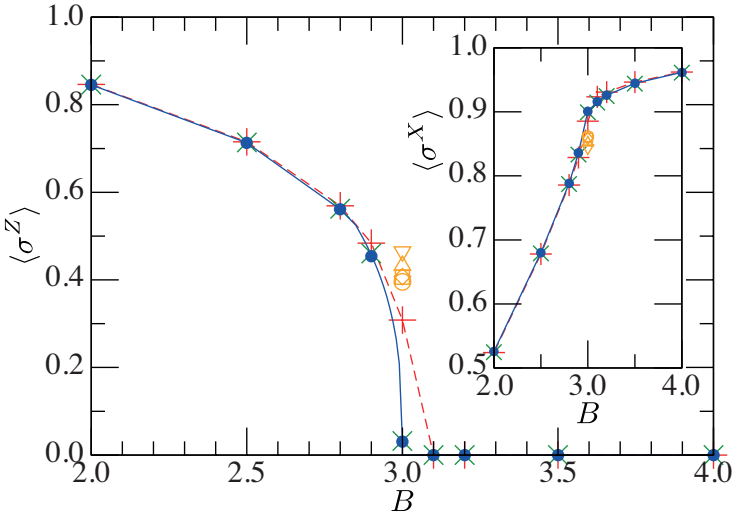


Figure 2.14: Observables $\langle \sigma^Z \rangle$ (main) and $\langle \sigma^X \rangle$ (inset) evaluated in the center of 21×21 PEPS Ising ground state approximations from the FU with $D = 2$ (plusses), 3 (crosses), and 4 (filled circles), where the $D = 4$ are basically on top of the $D = 3$ results. Open symbols show the SU at $B = 3.0$, for $D = 2$ (down-triangles), 4 (up-triangles), 5 (squares), 6 (diamonds), and 7 (open circles). We interpolate the FU $D = 4$ $\langle \sigma^Z \rangle$ results between $B = 2.85$ and $B_0 := 3.000035$ with $|B - B_0|^{0.34}$.

each value of B, we converge the imaginary time evolution independently, using time steps $\tau = 10^{-2}$, 10^{-3}, and 10^{-4}.

Figure 2.14 shows the order parameter evaluated in the center of 21×21 PEPS from the FU for several points in the phase diagram. Without performing a finite size scaling, we can already extract estimates of the critical point $B_c \approx 3.0$ and exponent $\beta \approx 0.34$ from this finite system, which are close to the iPEPS results [17, 66]. We conclude that this lattice is already large enough to display features similar to iPEPS.

For comparison, we also present SU results at $B = 3$. As expected from our previous analysis, the SU does not work well there, where the correlation length should be large. Figure 2.15 shows that the FU can indeed generate PEPS with larger correlation lengths. While a least squares fit gives a correlation length for the FU $D = 4$ PEPS $\zeta_{FU}^{D=4} \approx 2.6$, it reveals for the SU $D = 7$ PEPS only $\zeta_{SU}^{D=7} \approx 1.2$. The inset of figure 2.15 demonstrates the largest correlation length $\zeta_{FU}^{D=4} \approx 4.3$ for the 11×11 lattice. Notice that, here, we have not performed such an extensive convergence analysis with D' as we have done before for the Heisenberg Hamiltonian.[16] Nevertheless, we want to emphasize that our correlation functions for the 21×21 lattice are in perfect agreement with the best iPEPS results [66].

Remarkably, long correlation lengths can be analyzed, i.e. large clusters can be contracted, with very high accuracy in the framework of PEPS. This constitutes clear evidence for the power of general boundary MPO. They can capture the correlations of a large cluster size δ with a boundary bond dimension much smaller than the one needed for the exact contraction, $D' = D^{2\delta}$.

[16] For the quantum Ising model and all considered transverse fields B, on the 11×11 lattice, we ran the FU with $D' = 8$ for $D = 2$, $D' = 54$ for $D = 3$, and $D' = 128$ for $D = 4$, while on the 21×21 lattice, we ran the FU with $D' = 8$ for $D = 2$, $D' = 36$ for $D = 3$, and $D' = 128$ for $D = 4$.

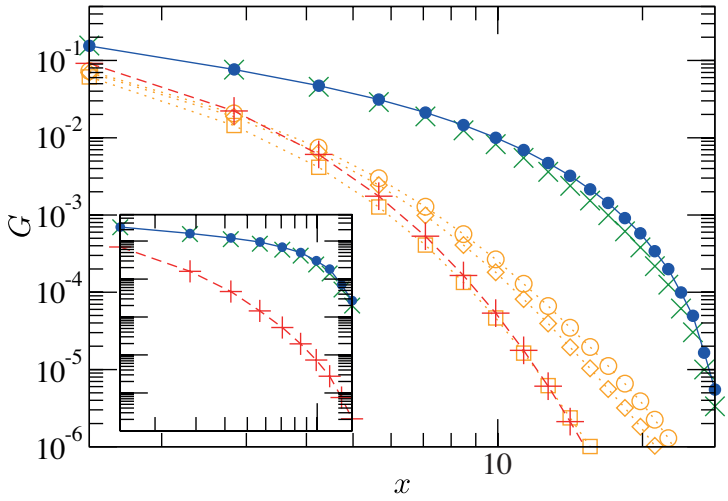


Figure 2.15: Correlation function $G(x) := \langle \sigma_l^Z \sigma_{l+x}^Z \rangle - \langle \sigma_l^Z \rangle \langle \sigma_{l+x}^Z \rangle$ for two sites separated by distance x along the diagonal in the center of PEPS ground state approximations of the Ising model on a 21×21 lattice with transverse field $B = 3.0$ (main) and on a 11×11 lattice with $B = 2.8$ (inset). The results were obtained with the FU using bond dimension $D = 2$ (plusses), 3 (crosses), and 4 (filled circles), and with the SU using $D = 4$ (squares), 6 (diamonds), and 7 (open circles).

2.5 Conclusions

In this chapter we have reviewed various aspects that need to be taken into account in the implementation of efficient state-of-the-art finite PEPS algorithms. Within the two main parts of PEPS algorithms, namely the environment approximation and the tensor update, we have analyzed algorithmic strategies that improve the efficiency and stability of the procedures, and the physical properties of the solution.

The environment approximation has decisive influence on the precision of the final PEPS of an imaginary time evolution, and is equally crucial for the computation of expectation values. We have shown how the accuracy of the cluster strategy, which allows for a natural trade-off between precision and computational cost of the environment, is fundamentally connected to the correlation length of the state. Additionally, we have demonstrated that it is possible to make use of purification MPO in order to ensure a positive environment approximation, and that this overcomes the limitations of the single-layer algorithm [43]. The numerical techniques analyzed in this problem can straightforwardly be applied to the cluster update and to the full update, but also to other scenarios where a positive MPO is required, e.g. to describe the mixed state of a one-dimensional system.

Not only the environment approximation, but also the method chosen for the tensor update affects the cost and stability of the routines. We have proposed an update scheme that is more efficient and better conditioned than the one from the original algorithm [11,41]. By restricting the variational parameters to the reduced tensor, the update is drastically accelerated. For both the reduced and the full tensor, we have formulated gauge fixings that significantly improve the conditioning. These gauge fixings, additionally, when combined with a cheap SVD, constitute a promising simplified but fast tensor update procedure.

Finally, we have combined the ingredients discussed above in an efficient im-

plementation of finite PEPS imaginary time evolution, capable of dealing with large systems and bond dimensions. In particular we have opted for the sequential application of Trotter gates, using general boundary MPO in the contraction of the full environment,[17] and restricting the update to the reduced tensors. To benchmark the performance of finite PEPS and to quantitatively assess the algorithmic properties, we have applied the code to the ground state search for the Heisenberg and the quantum Ising model.

We have presented ground state calculations for system sizes up to 21×21 and bond dimensions up to $D \approx 7, 8$. Our results demonstrate the adequacy of the PEPS ansatz for the description of strongly correlated quantum many-body systems, with energy and order parameter converging fast with increasing bond dimension, when they were obtained with the full update. In that case, thanks to the algorithmic improvements developed in this chapter, we have been able to achieve precisions of the spin-spin correlator in the Heisenberg model using bond dimensions $D = 6$ and 7 that previously had only been attained using a much larger $D = 16$ in reference [70]. Our analysis of a 21×21 quantum Ising model gave, already without finite size scaling, critical point, critical exponent and correlation functions in good agreement with the iPEPS results [17, 66].

The simple update [42] and the cluster update using small cluster sizes, while ensuring a less costly environment and thus being able to deal with larger bond dimensions, do not produce the best ground state approximation for a certain value of D, and, in particular, give rise to PEPS with limited correlation length, which is especially relevant for strongly correlated systems as e.g. the Ising model close to criticality. This makes clear that the largest bond dimension attained is not the significant measure of the power of a PEPS algorithm.

By reaching system sizes typically considered for finite size scaling, we have given evidence that finite PEPS, when all algorithmic details are taken into account, offer a feasible unbiased alternative to their infinite counterpart iPEPS [17]. On the other hand, the algorithmic methods proposed here can also be applied to iPEPS, and the feasibility of large finite PEPS demonstrated here suggests that large unit cells are possible in iPEPS, such that their potential bias due to a finite unit cell can be well analyzed by systematically increasing the unit cell from small to very large size.

Our analysis has been carried out with a generic implementation of PEPS algorithms, so that one can expect that adapting the methods to the specific properties of a certain problem will further enhance the performance. A particularly promising next step is to incorporate the symmetries of the considered Hamiltonian in the tensors [100, 101], a key element of ground-breaking two-dimensional DMRG studies, such as references [103, 104], and of seminal iPEPS calculations, such as references [26, 99].

2.6 Appendix

2.6.1 Purification Approximations

Approximating the boundary by a purification MPO, as described in section 2.3.1, requires the solution of nonlinear equations for each tensor A_l. Different algorithms can be used for this purpose, and we have tried and compared three methods.

[17] Notice that clusters could be beneficial in a parallel implementation as discussed in the previous chapter.

- Linearization: Instead of solving the equations for the product $A_l A_l^*$, we solve them for $A_l B_l$, treating A_l and A_l^* as independent tensors. In order to achieve convergence, the change of the tensor in each iteration needs to be small, and hence we construct the solution of the ith iteration according to $A_l^{(i)} = (1-\alpha)A_l^{(i-1)} + \alpha A_l$, where A_l solves the linearized equations of the previous iteration and is added to the previous solution $A_l^{(i-1)}$ with a weight α. The latter parameter must be chosen small enough to guarantee a decreasing cost function, and large enough to avoid unnecessarily long convergence times.[18] The construction of the individual parts of the linear equations has the leading cost $\mathcal{O}(dD^6D''^4) + \mathcal{O}(D^4D''^6)$. Because we cannot impose a kind of canonical form that gives a trivial norm matrix $N_l = 1$, we have to explicitly contract its tensor network, which contributes a cost $\mathcal{O}(d'^2D^2D''^6)$. Finally, computing the pseudoinverse to solve the linear equations requires $\mathcal{O}(d'^3D^3D''^6)$ operations, which typically represent the dominant cost when $d' \geq D$.

- Conjugate gradient: We employ a canned routine[19] that comprises a conjugate gradient method with line minimization. It has the lowest computational cost, as it only requires the computation of the cost function and its gradient with respect to a single tensor, which can be obtained with $\mathcal{O}(dD^6D''^4) + \mathcal{O}(D^4D''^6)$ operations.

- Newton method: It approaches a root of the gradient by iterating $\mathcal{H}_l^{(i-1)}(\vec{A}_l^{(i)} - \vec{A}_l^{(i-1)}) = -\vec{\mathcal{G}}_l^{(i-1)}$, where $\mathcal{H}_l^{(i-1)}$ denotes the Hessian matrix and $\vec{\mathcal{G}}_l^{(i-1)}$ the gradient of the cost function with respect to the tensor components at site l, evaluated with the solution $A_l^{(i-1)}$ of the previous iteration.[20] The Newton method has the advantage that the step width is naturally given, in contrast to the linearized equations where α needs to be chosen heuristically, and in contrast to the conjugate gradient routine where it is determined via line search. In addition to the parts of the conjugate gradient algorithm, the Newton method needs the Hessian matrix of the cost function, which contributes $\mathcal{O}(d'^2D^2D''^6)$ to the cost, and its pseudoinverse, determined by $\mathcal{O}(d'^3D^3D''^6)$ operations, such that the leading cost $\mathcal{O}(dD^6D''^4) + \mathcal{O}(D^4D''^6) + \mathcal{O}(d'^3D^3D''^6)$ is the same as for the linearized problem.

To compare the different alternatives, we benchmarked their performance in the search for an optimal purification with fixed $D'' = 2$ and varying d', given a reference purification with $D'' = 4$ and $d' = 4$. The latter was constructed by taking two rows from one edge of a PEPS norm TN, for several 11×11 $D = 2$ SU ground state approximations of the Ising model at various magnetic fields. As a general rule, the initial tensors for the search with incremented purification bond $d'+1$ were chosen as the previous solution for d' where the extra elements were filled with uniformly distributed random numbers. From the three considered algorithms, the Newton method performed best. It converged reliably for all d' within few local updates per tensor.

[18] We found that the optimization worked well if the initial value was $\alpha = 0.01$ and was multiplied by 0.8 whenever the cost function increased.

[19] For conjugate gradient minimization we use nag_opt_conj_grad from the NAG library [105].

[20] We observed that, occasionally, the Hessian matrix had many negative eigenvalues and then it was crucial to include only the positive ones (above a certain cutoff) in the construction of its pseudoinverse.

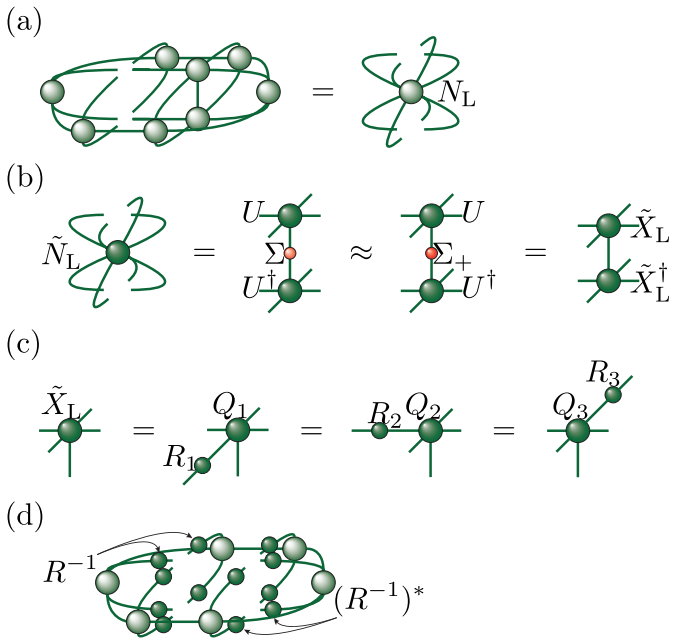


Figure 2.16: Gauge fixing on the environment tensors of the full tensor pair, when the environment is non-separable. (a) The environment tensor N_{L} of the left full tensor. (b) We determine the positive approximant for the environment of the left full tensor via a diagonalization of the Hermitian approximant $\tilde{N}_{\mathrm{L}} := (N_{\mathrm{L}} + N_{\mathrm{L}}^\dagger)/2 = U\Sigma U^\dagger$, in which, then, the negative eigenvalues are discarded in Σ_+, and, finally, the environment is written as $\tilde{X}_{\mathrm{L}}\tilde{X}_{\mathrm{L}}^\dagger$ in terms of its square root $\tilde{X}_{\mathrm{L}} := U\sqrt{\Sigma_+}$. (c) We perform three independent QR decompositions on $\tilde{X}_{\mathrm{L}}$. (d) After equally having carried out the previous steps (a) to (c) with the right full tensor, we have six different matrices R. Their inverses are contracted with the corresponding tensors of the boundary MPO.

All the methods benefit from initial variables that are already close to the final solution. A sensible numerical approach to purification approximations can then be implemented in two steps: firstly, the computation of the optimal purification via the SL algorithm [43], and secondly, the further optimization of that purification via the Newton method.

2.6.2 Gauge Fixing for the Full Tensors

When the environment is non-separable, and the update of the full tensors is considered, gauge transformations can be efficiently computed in such a way that an eigendecomposition of the $D^6 \times D^6$ dimensional norm environment of the pair is not necessary. Instead, the $D^4 \times D^4$ dimensional environments of the left (N_{L}) and right (N_{R}) tensor are independently computed (figure 2.16 (a)) and replaced by their positive approximants (figure 2.16 (b)) like in previous cases. Their square roots are used to obtain the desired gauge transformations for each of the virtual bonds of the pair (figure 2.16 (c)). On each virtual bond we then insert the corresponding product $R^{-1}R$ and absorb the R matrices in the full tensors and their inverses in the environment (figure 2.16 (d)).

The update of the full tensor pair proceeds in the way explained in figure 2.17, analogously to the reduced tensor update. As in the latter context, if the environ-

Model	Positive approximant	Gauge fixing
$B = 1.0$ Ising	$(2 \pm 3) \cdot 10^9$	1.3 ± 0.2
$B = 3.0$ Ising	$(2 \pm 2) \cdot 10^4$	2.9 ± 0.7
Heisenberg	$(1.3 \pm 0.8) \cdot 10^3$	1.15 ± 0.05

Table 2.2: We show the mean condition number of the norm matrix with its standard deviation in the full tensor update, for $D = 2$ and the setting of table 2.1.

ment is separable, the tensor initialization figure 2.17 (a) to (c) already minimizes the cost function, while, if the environment is close to separable, we can expect a significant decrease of the cost function. In general, we can anticipate good starting tensors for the following ALS sweeping.

Table 2.2 contains typical condition numbers of the norm matrix in the full tensor update without our gauge fixing, using only the positive approximant, and with our gauge fixing. Our gauge fixing improves the condition number drastically.

2.6.3 Finite PEPS Energies

Here we collect some precise energy values obtained with the PEPS ground state approximations considered in this chapter.

In the case of the Heisenberg model, we compare our results to energies from the quantum Monte Carlo loop algorithm of the ALPS library [84–86], summarized in table 2.3. The presented values and errors correspond to temperature $T = 10^{-4}$, and they agree with the ones corresponding to $T = 10^{-3}$ within the error bars.

10×10	14×14	16×16	20×20
-0.628656(2)	-0.639939(2)	-0.643531(2)	-0.648607(1)

Table 2.3: Energy per site of the Heisenberg model on square lattices of various sizes from quantum Monte Carlo, computed with the ALPS library [84–86].

D	10×10	14×14
2	-0.61310(2)	-0.62631(1)
3	-0.61999(1)	-0.63246(1)
4	-0.62637(2)	-0.63832(3)
5	-0.62739(1)	-0.63901(1)
6	-0.62774(1)	-0.63930(1)

Table 2.4: Energy per site of PEPS Heisenberg ground state approximations from the FU.

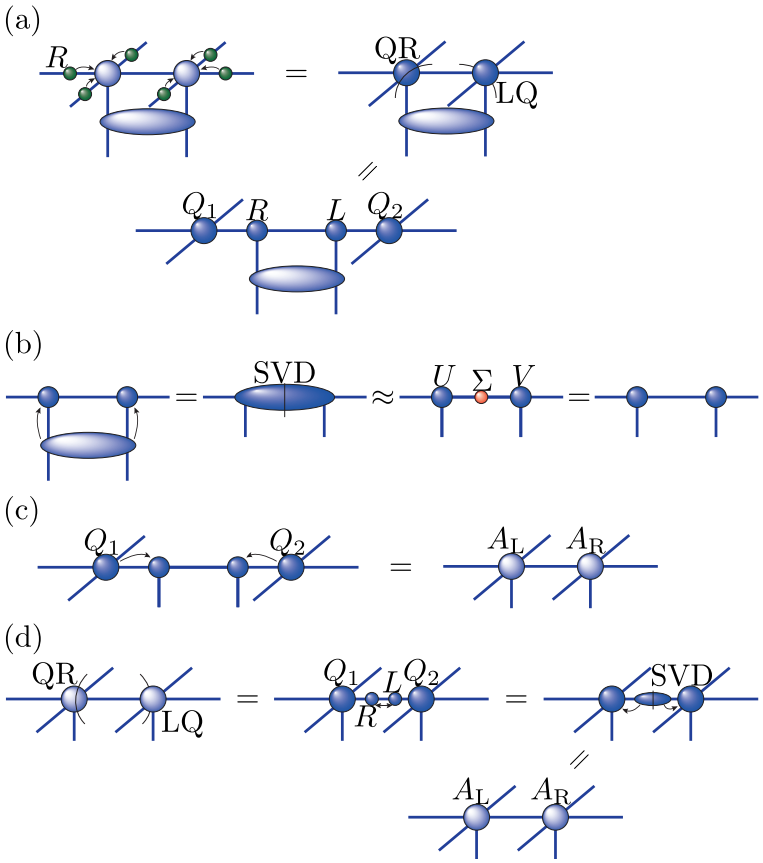


Figure 2.17: Like in the reduced case of figure 2.9, the update of the full tensor pair also consists of the three stages initialization, optimization, and final form. The optimization is the standard ALS sweeping, in which each full tensor is gauged after its update in the standard way, i.e. the left tensor is QR decomposed along its right virtual bond and the right tensor is LQ decomposed along its left virtual bond. (a) Initialization I: Firstly, we contract the gauge transformations from figure 2.16 with the full tensors, and split off their reduced parts. (b) Initialization II: Secondly, we construct new reduced tensors from a SVD on the tensor pair and the Trotter gate, equally sharing the D largest singular values between the left and right tensor. (c) Initialization III: We recover the full tensors A_{L} and A_{R}, which are now the initial tensors for the optimization via ALS sweeping. (d) Final form: After convergence of the sweeping, we put the two tensors on the same footing.

D	10×10	14×14	16×16	20×20
2	-0.61281(1)	-0.62115(1)	-0.62719(2)	-0.63519(2)
3	-0.61846(2)	-0.62977(1)	-0.63433(1)	-0.64056(2)
4	-0.62382(1)	-0.63587(1)	-0.63985(1)	-0.64549(2)
5	-0.62520(2)	-0.63713(2)	-0.64106(1)	-0.64659(2)
6	-0.62541(2)	-0.63738(2)	-0.64129(2)	-0.64676(2)

Table 2.5: Energy per site of PEPS Heisenberg ground state approximations from the SU.

D	2.0	2.5	2.8
2	-2.40075(1)	-2.74230(2)	-2.98947(5)
3	-2.40076(1)	-2.74243(1)	-2.99094(2)
4	-2.40076(1)	-2.74243(1)	-2.99099(1)
D	2.9	3.0	3.1
2	-3.07945(5)	-3.17128(4)	-3.26400(4)
3	-3.08071(1)	-3.17210(1)	-3.26457(1)
4	-3.08073(1)	-3.17210(1)	-3.26457(1)
D	3.2	3.5	4.0
2	-3.35744(4)	-3.64097(3)	-4.12064(2)
3	-3.35785(1)	-3.64116(1)	-4.12071(1)
4	-3.35785(1)	-3.64116(1)	-4.12071(1)

Table 2.6: Energy per site of 11×11 PEPS Ising ground state approximations from the FU for different transverse fields B.

D	2.0	2.5	2.8
2	-2.45219(1)	-2.77340(2)	-3.00705(4)
3	-2.45219(1)	-2.77346(1)	-3.00737(1)
4	-2.45219(1)	-2.77346(1)	-3.00737(1)
D	2.9	3.0	3.1
2	-3.09228(5)	-3.18128(6)	-3.27326(4)
3	-3.09287(1)	-3.18242(1)	-3.27406(1)
4	-3.09287(1)	-3.18243(1)	-3.27406(1)
D	3.2	3.5	4.0
2	-3.36617(4)	-3.64849(3)	-4.12685(2)
3	-3.36672(1)	-3.64873(1)	-4.12694(1)
4	-3.36672(1)	-3.64873(1)	-4.12694(1)

Table 2.7: Energy per site of 21×21 PEPS Ising ground state approximations from the FU for different transverse fields B.

D	3.0
2	-3.1792(4)
3	-3.1806(4)
4	-3.1807(4)
5	-3.1812(5)
6	-3.1812(4)
7	-3.1814(5)

Table 2.8: Energy per site of 21×21 PEPS Ising ground state approximations from the SU for transverse field $B = 3.0$.

Chapter 3

Density Functional Theory Beyond the Local Density Approximation

This chapter describes a completely different problem in which TNS techniques can serve as a fundamental tool. In the context of density functional theory (DFT), we develop a novel ansatz for the universal functional of Hohenberg and Kohn [3] that can go beyond the local density approximation (LDA) in a systematic way. This is achieved by parametrizing the functional via increasingly non-local terms. We describe a method that allows the determination of the optimal parameters for a set of ground state densities, and discuss its performance for several Hubbard and Coulomb models in one dimension. The usage of MPS enables us to exactly solve the chosen training set of problems before performing the optimization. While our analysis mainly aims at assessing the capability of our ansatz to approximate the universal functional in different scenarios and therefore mostly assumes already known exact ground state densities, we also move one step further and demonstrate how our ansatz can be applied within DFT to compute such densities. This chapter presents work in progress [106].

3.1 Introduction

Density functional theory (DFT) tackles interacting electron problems via the densities instead of the wave functions. This is possible thanks to the fundamental Hohenberg-Kohn theorem [3] which proves that for ground states of a large class of interesting Hamiltonians the density determines the wave function. Ground state observables that are naturally functions of the wave function are then expressed as functions of the density, i.e., as density functionals. This concept has led to algorithms that bridge the gap between the microscopic and the macroscopic and allow the quantum mechanical analysis of realistic systems ranging from atoms and molecules to solids and even plasmas [107, 108]. The Nobel Prize in Chemistry in 1998 was awarded for the development of DFT and related computational methods [109], and the two initial publications [3] and [4] are the two most cited in the history of Physical Review from 1893 to 2003 [110], each of them having over 15000 citations at the time of this writing in June 2014.

A central ingredient of DFT and its algorithms is the *energy functional* whose

minimization gives the ground state energy and density [3]. It consists of two parts: one that it is universal for all problems (with the same kinetic and interaction term in their Hamiltonians) and another that is special for the particular problem at hand (and defined by the external potential term in its Hamiltonian). While the second part takes on a simple form that is known, the first part captures the entire complexity of the considered physical systems. In general, this first part, called *universal functional*, is non-local as it depends on the density at all positions in space, its precise form is not known and finding it is QMA-hard [111].

Therefore, in practical calculations efficient approximations of the universal functional have to be used. The most famous of these is the *local density approximation* (LDA), in which the functional depends only on the density at the position coordinate where it is evaluated. DFT has been enormously successful with the LDA derived from the *homogeneous electron gas*. This system is defined in the thermodynamic limit as the ground state of a uniform density of electrons with Coulomb interactions whose negative charge is neutralized by a positively charged background. Because the density is the same at each point in space, the LDA leads to an exact energy functional for this system. Such a LDA functional can also be applied to an inhomogeneous density, by evaluating its energy pointwise in space, and this procedure works extremely well in many cases, what justifies the method a posteriori. The final error of a LDA calculation, however, cannot be estimated within the algorithm but necessitates a comparison with other methods or with an experiment. Although there are many approaches that improve upon the LDA (s., e.g., references [46–49, 112] and the review article [50] and references therein), to the best of our knowledge, there exists no general scheme with a systematic improvement and error estimates.

This chapter aims at such a scheme and is motivated by our knowledge about Tensor Network States (TNS), which, typically, allow to systematically improve results and estimate errors. In TNS methods, the accuracy of the approximation can be increased by including more correlations in the ansatz, which is achieved by increasing the dimensions of tensors. We want to investigate if TNS concepts can be useful for DFT, and if a similar scheme can be devised where the non-local contributions to the functional, that escape the LDA, can be gradually included.

Our analysis is carried out in a discretized one-dimensional space where exact ground state energies and densities can be computed with Matrix Product State (MPS) algorithms. Throughout this study we focus on the universal functional. We first investigate its exact form for shorter Hubbard chains and, in that context, we study the non-locality of the functional via a MPS approximation. Having larger system sizes in mind for which the exact universal functional is not feasible, we propose an efficient ansatz for its approximation that includes and goes beyond the LDA. We first explain how our ansatz can be optimally determined when the exact functional is known for discrete density values on a predefined density grid and demonstrate its performance for short Hubbard chains. We then discuss how our ansatz has to be modified for continuous density values by means of both general polynomials as well as interpolating splines, and we show how both can be optimally fitted to an arbitrary set of training densities. This constitutes the basis of our general algorithm that we then apply to several paradigmatic larger Coulomb problems. Finally, we show how our ansatz can be used to obtain ground state densities, and we discuss potential improvements.

3.1.1 Reader's Guide

This chapter is organized as follows. Section 3.2 presents the necessary background for this analysis, namely the DFT basics and the considered Hamiltonians. We begin our discussion of different approximations for the universal functional in section 3.3, first for shorter Hubbard chains and then for larger Coulomb problems. The applicability of our functional ansatz to typical DFT problems is demonstrated in section 3.4. Finally, section 3.5 summarizes our results and discusses possible refinements of our algorithm.

3.2 Background

3.2.1 DFT on a Lattice

In this section, the necessary DFT background for this chapter is presented in a concise way. More comprehensive presentations can be found in reference [113] for DFT in general and in reference [114] for DFT on a lattice.

Although DFT is widely used for realistic three-dimensional systems in the continuum, we restrict ourselves here to a one-dimensional setting on a discrete lattice, where we can obtain accurate ground state energies and densities for large strongly correlated quantum many-body systems via our MPS algorithms. In this scenario, we consider a system of fermions on a lattice of size L. The density is then a function of the lattice site l, where $l \in \{1, 2, \ldots, L\}$, and we simply write it as vector $\vec{n} := (n_1, n_2, \ldots, n_L)^t$ such that all density functionals are simply functions of this vector.

Key functionals in DFT are the energy and the universal functional [3]. For their derivation we assume a Hamiltonian

$$\hat{H} \quad = \quad \hat{H}_0 + \sum_l v_l \hat{n}_l \tag{3.1}$$

where $\hat{H}_0$ comprises the kinetic and interaction energy and the v_l denote an external local potential, and thus we have separated a part $\hat{H}_0$ that is the same for all problems, with the same kinetic and interaction term, from a part that specifies the considered problem. The precise form of $\hat{H}_0$ as well as of the occupation number operator $\hat{n}_l$ is not relevant for the general discussion here and will be defined for the problems of this chapter in the next section. Denoting by $|\psi_{\vec{n}}\rangle$ a wave function $|\psi\rangle$ having density $\vec{n}$ such that $n_l = \langle\psi|\hat{n}_l|\psi\rangle$ holds for all components of the density vector, the ground state energy E can be found via a minimization over densities:

$$\begin{aligned}
E &= \min_{|\psi\rangle} \langle\psi|\hat{H}|\psi\rangle \\
&= \min_{\vec{n}} \Big(\min_{|\psi_{\vec{n}}\rangle} \langle\psi_{\vec{n}}|\hat{H}|\psi_{\vec{n}}\rangle \Big) \\
&= \min_{\vec{n}} \Big(\min_{|\psi_{\vec{n}}\rangle} \langle\psi_{\vec{n}}|\hat{H}_0 + \sum_l v_l \hat{n}_l|\psi_{\vec{n}}\rangle \Big) \\
&= \min_{\vec{n}} \Big(\min_{|\psi_{\vec{n}}\rangle} \langle\psi_{\vec{n}}|\hat{H}_0|\psi_{\vec{n}}\rangle + \sum_l v_l n_l \Big) \\
&=: \min_{\vec{n}} \Big(F(\vec{n}) + \sum_l v_l n_l \Big) \quad .
\end{aligned}$$

Throughout this chapter, we assume that all wave functions are normalized to 1 (in particular, the preceding minimization was performed under this constraint). So the energy functional

$$E(\vec{n}) \quad := \quad F(\vec{n}) + \sum_l v_l n_l \tag{3.2}$$

is defined in terms of the universal functional

$$F(\vec{n}) \quad := \quad \min_{|\psi_{\vec{n}}\rangle} \langle \psi_{\vec{n}} | \hat{H}_0 | \psi_{\vec{n}} \rangle \tag{3.3}$$

in such a way that its minimization for a specific external potential $\vec{v} := (v_1, v_2, \ldots, v_L)^{\mathrm{t}}$ gives the corresponding ground state density $\vec{n}_0$ and energy $E(\vec{n}_0)$, and

$$\frac{\partial F}{\partial n_l}\Big|_{\vec{n}_0} \quad = \quad -v_l(\vec{n}_0) \tag{3.4}$$

holds ($\forall l \in \{1, 2, \ldots, L\}$).

An alternative to the direct minimization of the energy functional (3.2) is provided by the *Kohn-Sham* method [4]. The Kohn-Sham method maps the intractable interacting quantum many-body problem of equation (3.1) to a tractable non-interacting Kohn-Sham system:

$$\hat{H}^{\mathrm{KS}} \quad = \quad \hat{T} + \sum_l v_l^{\mathrm{KS}} \hat{n}_l \qquad , \tag{3.5}$$

where the Kohn-Sham potential v_l^{KS} is chosen to produce a ground state density $\vec{n}_0$ that coincides with the one of the interacting system. Assuming that such a potential exists [113, 115], both the energy functional of the interacting, $E^{\mathrm{I}}(\vec{n}) = F^{\mathrm{I}}(\vec{n}) + \sum_l v_l n_l$, and of the non-interacting problem, $E^{\mathrm{NI}}(\vec{n}) = F^{\mathrm{NI}}(\vec{n}) + \sum_l v_l^{\mathrm{KS}} n_l$, are minimal at the same density $\vec{n}_0$, i.e., $(\partial E^{\mathrm{I}}/\partial n_l)|_{\vec{n}_0} = 0 = (\partial E^{\mathrm{NI}}/\partial n_l)|_{\vec{n}_0}$, and thus:

$$\frac{\partial F^{\mathrm{I}}}{\partial n_l}\Big|_{\vec{n}_0} + v_l(\vec{n}_0) = 0 = \frac{\partial F^{\mathrm{NI}}}{\partial n_l}\Big|_{\vec{n}_0} + v_l^{\mathrm{KS}}(\vec{n}_0) \qquad .$$

We now introduce the functional

$$W(\vec{n}) \quad := \quad F^{\mathrm{I}}(\vec{n}) - F^{\mathrm{NI}}(\vec{n}) \tag{3.6}$$

and conclude that the Kohn-Sham potential fulfills:

$$v_l^{\mathrm{KS}}(\vec{n}_0) \quad = \quad \frac{\partial W}{\partial n_l}\Big|_{\vec{n}_0} + v_l(\vec{n}_0) \qquad . \tag{3.7}$$

This potential generates the ground state density of the interacting problem, while the ground state energy follows from:

$$E^{\mathrm{I}}(\vec{n}) \quad = \quad E^{\mathrm{NI}}(\vec{n}) + W(\vec{n}) - \sum_l \frac{\partial W}{\partial n_l} n_l \qquad . \tag{3.8}$$

So far, the developed DFT formalism is impractical because, in general, neither $F(\vec{n})$ nor $W(\vec{n})$ is known exactly. Practical DFT algorithms are based on good approximations of these quantities. Because for many interesting applications better approximations could be found for $W(\vec{n})$ than for $F(\vec{n})$, the Kohn-Sham algorithms

are currently more widely used than the so-called *orbital free* methods whose goal is a direct minimization of the energy functional (3.2). It is also important to realize that when a problem is non-interacting and thus efficiently solvable, a Kohn-Sham procedure immediately becomes exact, as can be appreciated from equations (3.7) and (3.8), while an orbital free method unnecessarily complicates this situation as it still requires a proper ansatz for $F^{\mathrm{NI}}(\vec{n})$. Nevertheless, we will henceforth study approximations of both $F(\vec{n})$ and $W(\vec{n})$.

The famous LDA actually approximates the *exchange-correlation energy* $E_{\mathrm{xc}}(\vec{n}) := W(\vec{n}) - E_{\mathrm{H}}(\vec{n})$ where $E_{\mathrm{H}}(\vec{n})$ denotes the known Hartree contribution to the interaction. Here one hopes that a subtraction of all known terms finally leads to a small remainder to be approximated such that its approximation error has only a small effect on the final result. While the mean-field part E_{H} can be expected to play an important role in higher dimensions, this is not so clear for the one-dimensional systems studied here, and thus we will not subtract it in our approximations. The term $W(\vec{n})$ is called *Hartree-exchange-correlation energy* in the literature and we adopt this notion in the following. The LDA ansatz takes on the form

$$W_{\mathrm{LDA}}(\vec{n}) \quad = \quad \sum_l w(n_l) \tag{3.9}$$

where $w(n)$ comes from the exact ground state solution of the homogeneous electron gas in the thermodynamic limit for which this ansatz is exact.

3.2.2 Hamiltonians

In this analysis, we consider two paradigmatic models for interacting electrons, namely a Hubbard Hamiltonian

$$\hat{H} \quad = \quad -t\sum_{l,\sigma}(c^\dagger_{l,\sigma}c_{l+1,\sigma} + c^\dagger_{l+1,\sigma}c_{l,\sigma}) + U\sum_l \hat{n}_{l,\uparrow}\hat{n}_{l,\downarrow} + \sum_l v_l\hat{n}_l \tag{3.10}$$

and a Coulomb Hamiltonian

$$\hat{H} \quad = \quad -t\sum_{l,\sigma}(c^\dagger_{l,\sigma}c_{l+1,\sigma} + c^\dagger_{l+1,\sigma}c_{l,\sigma}) + U\sum_{l\leq m}\frac{\hat{n}_l\hat{n}_m}{\sqrt{(l-m)^2+1}} + \sum_l v_l\hat{n}_l\ . \tag{3.11}$$

They shall describe two species of fermions, $\sigma \in \{\uparrow,\downarrow\}$, on a lattice of length L, with lattice site index $l \in \{1,2,\ldots,L\}$, and with operators for annihilation, $c_{l,\sigma}$, creation, $c^\dagger_{l,\sigma}$, and number, $\hat{n}_{l,\sigma} = c^\dagger_{l,\sigma}c_{l,\sigma}$ and $\hat{n}_l = \hat{n}_{l,\uparrow} + \hat{n}_{l,\downarrow}$. The fermions can hop from one lattice site to a neighboring one via the tunneling element t, interact via interaction U, and they experience an external potential via the v_l.

These Hamiltonians are very often used for the theoretical analysis of realistic systems of Coulomb interacting electrons. The Hubbard model assumes local interactions, which might be justified for materials in which the screening effect is very strong. In the Coulomb model, a soft-Coulomb, $\propto 1/\sqrt{(l-m)^2+1}$, replaces the true Coulomb interaction, $\propto 1/\sqrt{(l-m)^2}$, in order to circumvent any problems related to the singularity at $l = m$.

We will compute ground states with the help of MPS algorithms and thus restrict ourselves to one dimension and open boundary conditions. In our computations, the desired total particle number is obtained by adjusting a chemical potential term $-\mu\sum_l \hat{n}_l$ that is also part of the above Hamiltonians and that we implicitly included in the external potential. All the following discussion will be based upon the density operator $\hat{n}_l = \hat{n}_{l,\uparrow} + \hat{n}_{l,\downarrow}$ corresponding to the total density, and we will

not resolve its individual components $\hat{n}_{l,\uparrow}$ and $\hat{n}_{l,\downarrow}$ (s., e.g., reference [116] on the role of spin in DFT).

3.3 Searching for Approximations Beyond the LDA

The universal functional $F(\vec{n})$ (3.3) and the Hartree-exchange-correlation energy $W(\vec{n})$ (3.6) are functions of the density vector $\vec{n}$ which has L entries n_1, n_2, ..., n_L where n_l is the density at lattice site l. We can see both density functionals as tensors with L indices that take on continuous values, i.e., as objects of the form $F_{\vec{n}}$ and $W_{\vec{n}}$.

We will first look at shorter Hubbard chains which allow an exact computation of $F_{\vec{n}}$ and $W_{\vec{n}}$ for discretized density values in a wide range. This enables us to investigate an approximation by MPS. We then propose a functional ansatz for continuous density values and large systems. After showing how the optimal form of this ansatz is fitted when the exact density functional tensors are known, we move on to larger Coulomb problems and demonstrate how that optimal fitting is performed when only an arbitrary set of training densities is provided.

3.3.1 Shorter Hubbard Chains

In this section, we first construct the exact density functional tensors $F_{\vec{n}}$ and $W_{\vec{n}}$ for equations (3.3) and (3.6), and then study different approximations for them. We focus on shorter Hubbard chains, Hamiltonian (3.10), for which we compute exact ground states by means of Lanczos. Because we approach this problem and compute ground states numerically, we can only take into account a finite total number of ground state densities in the construction of $F_{\vec{n}}$ and $W_{\vec{n}}$. We thus define a discrete density by values $n_{\min}$ and $n_{\max}$ and R such that the interval $[n_{\min}, n_{\max}]$ is divided into $R-1$ intervals.

In this section, we choose these R points such that they lie equidistantly to each other. Imposing the same density grid at each lattice site l yields R^L possible density vectors $\vec{n}$ of which, however, only those leading to integer total particle number $N = \sum_l n_l$ are relevant here. We store the corresponding $< R^L$ values of the universal functional and the Hartree-exchange-correlation energy in the R^L dimensional tensors $F_{\vec{n}}$ and $W_{\vec{n}}$.

Inversion

For a given density vector $\vec{n}$, the corresponding value $W(\vec{n})$, equation (3.6), follows from the two independent values of the universal functional for the interacting, $F^{\mathrm{I}}(\vec{n})$, and non-interacting system, $F^{\mathrm{NI}}(\vec{n})$. For either system, we extract the corresponding value $F(\vec{n})$ from the energy functional (3.2) via subtraction of the potential energy:

$$F(\vec{n}) \quad = \quad E(\vec{n}) - \sum_l v_l n_l \qquad . \tag{3.12}$$

For given $\vec{n}$ we have to find the external potential $\vec{v}(\vec{n})$ that gives rise to this ground state density. The corresponding ground state energy is the value $E(\vec{n})$. Since the tools we have at hand (exact diagonalization, MPS) allow us to solve the direct problem, i.e., given an external potential, to find the density and energy of the corresponding ground state, we need to perform an inversion [117]. The existence of such a potential $\vec{v}(\vec{n})$, although not guaranteed for all $\vec{n}$, can be expected for the

physically motivated densities considered here. In fact, for each considered target density, we checked explicitly that the obtained corresponding potential leads to a ground state density close to the target within machine precision.

We use an inversion procedure similar to the one proposed in [118]. Given a target density $\vec{n}^{\text{target}}$, we find the corresponding potential iteratively, such that in iteration i the ground state density $\vec{n}(i)$ of the potential $\vec{v}(i)$ determines the next potential

$$\vec{v}(i+1) = \vec{v}(i) + \alpha(i)(\vec{n}(i) - \vec{n}^{\text{target}})$$

where $\alpha(i) > 0$ is a function that accelerates the convergence in the following way: In each iteration i we compute the ground state for $\alpha(i) = q \cdot \alpha(i-1)$ and for $\alpha(i) = \alpha(i-1)/q$ for an initially specified $q > 0$, and then we take the $\alpha(i)$ and following density $\vec{n}(i)$ that give the smaller density error $||\vec{n}(i) - \vec{n}^{\text{target}}||$. We tried several values of q and of the initial $\alpha(0)$, and the choice $q = 0.9$ and $\alpha(0) = 0.5$ performed best. In all results discussed here, the density error is guaranteed to be below 10^{-14}, and convergence to this value was typically achieved within ≈ 100 iterations.

In the following, we determine the exact tensors $F_{\vec{n}}$ and $W_{\vec{n}}$ for half-filled Hubbard chains of on-site interaction $U = 0.0$, 0.1, and 1.0. To access a possible dependence on lattice size and density grid, we consider two different system sizes and for each system size we analyze two different density grids. We study a system of length $L = 4$ and particle number $N = 4$ with $R = 13$ on the intervals $n_l \in [0.8, 1.2]$ and $\in [0.4, 1.6]$, and a $L = 6 = N$ system with $R = 5$ on the intervals $n_l \in [0.8, 1.2]$ and $\in [0.6, 1.4]$. These intervals are chosen to mimic both the situation of very smooth potentials, i.e. $n_l \in [0.8, 1.2]$, and the situation of less smooth potentials, i.e. $n_l \in [0.4, 1.6]$ and $n_l \in [0.6, 1.4]$. After having obtained the exact tensors $F_{\vec{n}}$ and $W_{\vec{n}}$, we investigate several approximations for them that we denote by $G_{\vec{n}}$ for $F_{\vec{n}}$ and $H_{\vec{n}}$ for $W_{\vec{n}}$. The respective relative approximation errors read $\epsilon_F := \sum_{\vec{n}} |G(\vec{n}) - F(\vec{n})| / \sum_{\vec{n}} |F(\vec{n})|$ and $\epsilon_W := \sum_{\vec{n}} |H(\vec{n}) - W(\vec{n})| / \sum_{\vec{n}} |W(\vec{n})|$.

MPS Approximation

A MPS with a bond dimension D growing at most polynomially with the system size represents an efficiently storable tensor network, in which the complexity corresponds to increasing non-locality. The exact density functional tensors $F_{\vec{n}}$ and $W_{\vec{n}}$ are not efficiently storable since their number of entries R^L grows exponentially with the lattice size. But we can ask ourselves if these tensors could be well approximated by MPS of small bond dimension, i.e., if the non-locality in $F_{\vec{n}}$ and $W_{\vec{n}}$ has the right pattern so that their information has an efficient approximated encoding.

We obtain the optimal MPS approximation of bond dimension D for the tensors $F_{\vec{n}}$ and $W_{\vec{n}}$ by means of successive truncated singular value decompositions [119, 120]. Figure 3.1 shows the relative errors ϵ_F and ϵ_W for MPS approximations as a function of their bond dimension, for a selection of analyzed cases. In all analyzed cases, we could always make out two separate regimes for D: For small D, the approximation error is independent of the interaction strength U and decreases slower than exponentially, and for large D, the error depends on U and decreases exponentially. Moreover, for large D, the relative error is always greater for greater values of U, and, for a given value of U, ϵ_W is always smaller than ϵ_F. The two regimes are separated at a specific value of the bond dimension that seems to depend only on R. For fixed $R = 5$, we can compare figure 3.1(c) for $L = 4 = N$

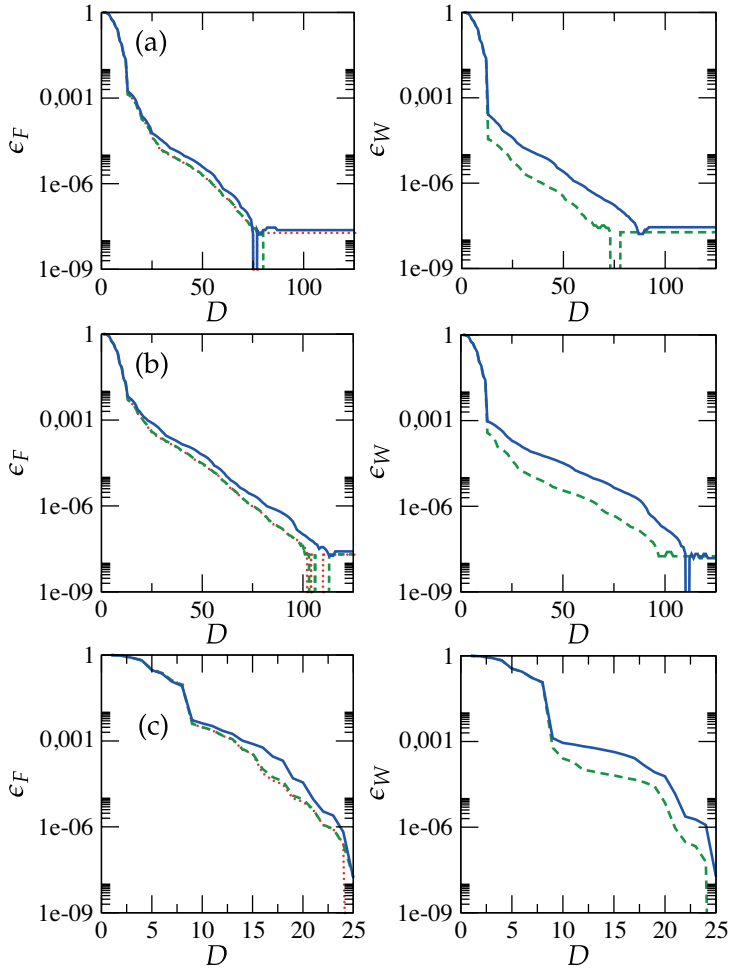


Figure 3.1: Relative errors ϵ_F and ϵ_W for MPS approximations as a function of the MPS bond dimension D. We consider a grid of $R = 5$ values of $n_l \in [0.8, 1.2]$ for $L = 6 = N$ (a), $n_l \in [0.6, 1.4]$ for $L = 6 = N$ (b), and $n_l \in [0.6, 1.4]$ for $L = 4 = N$ (c). The interaction strengths are $U = 10.0$ (solid line), $U = 0.1$ (dashed line), and $U = 0$ (dotted line).

with figure 3.1(b) for $L = 6 = N$ and conclude that a given value of D leads to very similar approximation errors for both system sizes. These results have to be compared to the largest bond dimensions possible for these cases, $R^{L/2}$, namely 125 and 25. We observe that a relative error below 10^{-3} is attained from $D \geq 15$ in both cases. This points out the possibility of an efficient MPS approximation for the exact functionals, although a scaling study with larger chains would be needed for a rigorous statement.

Our construction of the MPS approximation for $F_{\vec{n}}$ and $W_{\vec{n}}$ via successive truncated singular value decompositions becomes more difficult when the system size increases. In fact, the computational cost of the initial singular value decompositions in this procedure grows exponentially with the number of lattice sites L. Reference [120] presents an algorithm that constructs a MPS approximation of a black-box tensor, i.e., which can be efficiently evaluated for any values of its possibly large number of indices. Because both $F_{\vec{n}}$ and $W_{\vec{n}}$ are such black-box tensors which can be calculated efficiently for any density vector $\vec{n}$ by means of the in-

version procedure described above, we tried an open-source Matlab implementation of the scheme [120] which is accessible online [121]. We observed that the code [121] performed extremely well when we searched for a MPS approximation of a black-box tensor having MPS form. However, when our considered black-box tensor deviated from an exact MPS, the computations ran much longer and we were not able to obtain MPS approximations for large L within reasonable amounts of time.

Our Ansatz

Therefore, we now propose an alternative approximation ansatz that allows us to include non-local correlations in a controlled way, namely

$$G(\vec{n}) = \sum_{x=0}^{X} \sum_{l=1}^{L-x} g_{x,l}(n_l, n_{l+x}) \tag{3.13}$$

for the functional $F(\vec{n})$ and similarly $H(\vec{n})$ in terms of $h_{x,l}(n_l, n_{l+x})$ for $W(\vec{n})$. Our ansatz in equation (3.13) comprises terms of different range x, starting at $x = 0$ and stopping at the maximal range $x = X$ which can be at most $X = L - 1$. The local terms for $x = 0$, $g_{x=0,l}(n_l)$, are in the spirit of the LDA while the non-local terms for $x > 0$, $g_{x>0,l}(n_l, n_{l+x})$, extend beyond the LDA and shall capture more non-locality of the true density functional. At the same time, our ansatz is efficient, as its defining tensor $g_{x,l}(n_l, n_{l+x}) = g_{x,l}^{n_l,n_{l+x}}$ always has less than R^2L^2 entries.

We want to analyze the performance of our ansatz as a function of the maximal range X. For that we have to find the optimal tensor $g_{x,l}(n_l, n_{l+x}) = g_{x,l}^{n_l,n_{l+x}}$ that minimizes the distance $\sum_{\vec{n}} |F_{\vec{n}} - \sum_{x,l} g_{x,l}^{n_l,n_{l+x}}|^2$. This least squares problem can be recast into a system of linear equations $C\vec{g} = \vec{F}$, where $\vec{g}$ results from vectorizing all $\mathcal{N} < R^2L^2$ components of $g_{x,l}^{n_l,n_{l+x}}$, $\vec{F}$ from vectorizing all $\mathcal{M} < R^L$ non-zero elements of $F_{\vec{n}}$, and C is a $\mathcal{M} \times \mathcal{N}$ dimensional matrix whose non-vanishing entries all have the value 1 and establish the correct correspondence between the left- and right-hand side of the linear equations. Under the assumption $\mathcal{M} \geq \mathcal{N}$ our problem is solved by $\vec{g} = C^{-1}\vec{F}$ with the help of the pseudoinverse C^{-1}.

We first focus on the smooth training densities for $n_l \in [0.8, 1.2]$. Tables 3.1 and 3.2 show the approximation errors of our ansatz with different values of the maximal range X for different interaction strengths U. For given U, both ϵ_F and ϵ_W decrease rapidly with increasing X until $X = 2$ is reached, beyond which no substantial improvement occurs anymore. For given X, both relative errors increase when the interaction strength U increases. For fixed values of X and U, ϵ_W is always smaller than ϵ_F. The local terms alone, i.e., $X = 0$, lead to a relative error $\epsilon_F \approx 10^{-2}$ for $G(\vec{n})$ that varies only slightly for different U, while they give a relative error $\epsilon_W \approx 10^{-3}$ for $H(\vec{n})$ when $U = 1.0$ and $\epsilon_W \approx 10^{-4}$ when $U = 0.1$. Including the long-range terms, i.e., $X > 0$, improves all approximation errors by at least an order of magnitude when $X = 2$. Since these statements hold equally for both system sizes $L = 4 = N$ and $L = 6 = N$, we can hope that our ansatz in equation (3.13) provides an efficient approximation for the true tensors $F_{\vec{n}}$ and $W_{\vec{n}}$ when the system size increases.

We can additionally look at the functional form of our solution by plotting the obtained tensor elements, $g_{x,l}^{n_l,n_{l+x}}$, for different values of x and l as functions of n_l and n_{l+x}, $g_{x,l}(n_l, n_{l+x})$. Here, we consider the tensor $g_{x,l}^{n_l,n_{l+x}}$ resulting from a fit of our ansatz in equation (3.13) with $X = L - 1$, i.e., including all terms of all possible ranges x. The local terms for $x = 0$ are shown in figure 3.2 for $L = 4 = N$. We

(a)	X	$U = 0.0$	$U = 0.1$	$U = 1.0$	(b)	X	$U = 0.1$	$U = 1.0$
	0	$8.8 \cdot 10^{-3}$	$9.0 \cdot 10^{-3}$	$1.2 \cdot 10^{-2}$		0	$9.8 \cdot 10^{-5}$	$1.8 \cdot 10^{-3}$
	1	$3.7 \cdot 10^{-5}$	$3.8 \cdot 10^{-5}$	$5.8 \cdot 10^{-5}$		1	$4.0 \cdot 10^{-6}$	$4.3 \cdot 10^{-5}$
	2	$6.3 \cdot 10^{-6}$	$6.4 \cdot 10^{-6}$	$9.7 \cdot 10^{-6}$		2	$6.3 \cdot 10^{-7}$	$6.6 \cdot 10^{-6}$
	3	$6.1 \cdot 10^{-6}$	$6.3 \cdot 10^{-6}$	$9.4 \cdot 10^{-6}$		3	$6.1 \cdot 10^{-7}$	$6.4 \cdot 10^{-6}$

Table 3.1: Relative error ϵ_F (a) and ϵ_W (b) for $L = 4 = N$ and our ansatz with different values of the maximal range X on a grid of $R = 13$ values of $n_l \in [0.8, 1.2]$.

(a)	X	$U = 0.0$	$U = 0.1$	$U = 1.0$	(b)	X	$U = 0.1$	$U = 1.0$
	0	$7.0 \cdot 10^{-3}$	$7.2 \cdot 10^{-3}$	$9.2 \cdot 10^{-3}$		0	$1.3 \cdot 10^{-4}$	$2.0 \cdot 10^{-3}$
	1	$1.8 \cdot 10^{-3}$	$1.8 \cdot 10^{-3}$	$2.1 \cdot 10^{-3}$		1	$5.5 \cdot 10^{-5}$	$5.5 \cdot 10^{-4}$
	2	$4.4 \cdot 10^{-4}$	$4.5 \cdot 10^{-4}$	$5.5 \cdot 10^{-4}$		2	$1.8 \cdot 10^{-5}$	$1.9 \cdot 10^{-4}$
	5	$4.3 \cdot 10^{-4}$	$4.4 \cdot 10^{-4}$	$5.5 \cdot 10^{-4}$		5	$1.8 \cdot 10^{-5}$	$1.9 \cdot 10^{-4}$

Table 3.2: Relative error ϵ_F (a) and ϵ_W (b) for $L = 6 = N$ and our ansatz with different values of the maximal range X on a grid of $R = 5$ values of $n_l \in [0.8, 1.2]$.

observe that they all feature the same functional behavior $\propto (n_l - 1)^2$. Figure 3.3 shows examples of non-local terms for $x > 0$ when $L = 4 = N$. While the non-local terms of $G(\vec{n})$ behave as $g_{x,l}(n_l, n_{l+x}) \propto (n_l - 1)(n_{l+x} - 1)$, the ones of $H(\vec{n})$ behave as $h_{x,l}(n_l, n_{l+x}) \propto -(n_l - 1)(n_{l+x} - 1)$. This holds for all non-local terms, as we checked explicitly. We similarly carried out this analysis for $L = 6 = N$ and obtained exactly the same functional forms as for $L = 4 = N$.

The observed functional behavior can be very well understood taking into account that, so far, we have only considered smooth training densities for $L = N$ where $n_l \in [0.8, 1.2]$. In this case, every training density departs only slightly from the homogeneous situation for $L = N$ where $n_l = 1$ on every site l. If we denote the latter homogeneous density by $\vec{n}_0 = (1, 1, \ldots, 1)^t$, a second-order Taylor series expansion of the universal functional $F(\vec{n})$ around $\vec{n}_0$ gives

$$\begin{aligned} F(\vec{n}) &\approx F(\vec{n}_0) + \sum_l \frac{\partial F}{\partial n_l}\Big|_{\vec{n}_0} (n_l - 1) + \frac{1}{2} \sum_{l,m} \frac{\partial^2 F}{\partial n_l \partial n_m}\Big|_{\vec{n}_0} (n_l - 1)(n_m - 1) \\ &\approx F(\vec{n}_0) - \sum_l v_l(\vec{n}_0)(n_l - 1) + \frac{1}{2} \sum_{l,m} \frac{\partial^2 F}{\partial n_l \partial n_m}\Big|_{\vec{n}_0} (n_l - 1)(n_m - 1) \end{aligned}$$

where we have used the fact that $(\partial F / \partial n_l)|_{\vec{n}_0} = -v_l(\vec{n}_0)$ from equation (3.4). Since the dependence $\propto (n_l - 1)$ is not visible in our solutions, we conjecture that $v_l(\vec{n}_0) \approx 0$ for the Hubbard model with $L = N$.

We now consider the less smooth training densities for $n_l \in [0.4, 1.6]$ in the $L = 4 = N$ case and for $n_l \in [0.6, 1.4]$ in the $L = 6 = N$ case. The approximation errors of our ansatz are collected in the tables 3.3 and 3.4. Compared to the smooth training densities, qualitatively, our results here lead to the same conclusions, while quantitatively, the errors here are slightly larger. In particular, the significant improvement with increasing X, and the low values $\epsilon_F \approx 0.004$ and $\epsilon_W \approx 0.001$ attained for $X = 2$ in all cases at least, support the validity of our ansatz.

The functional form of our solution for the less smooth training densities reveals deviations from the second-order Taylor series expansion that well describes our solution for the smooth training densities. This can be seen in figure 3.4 for

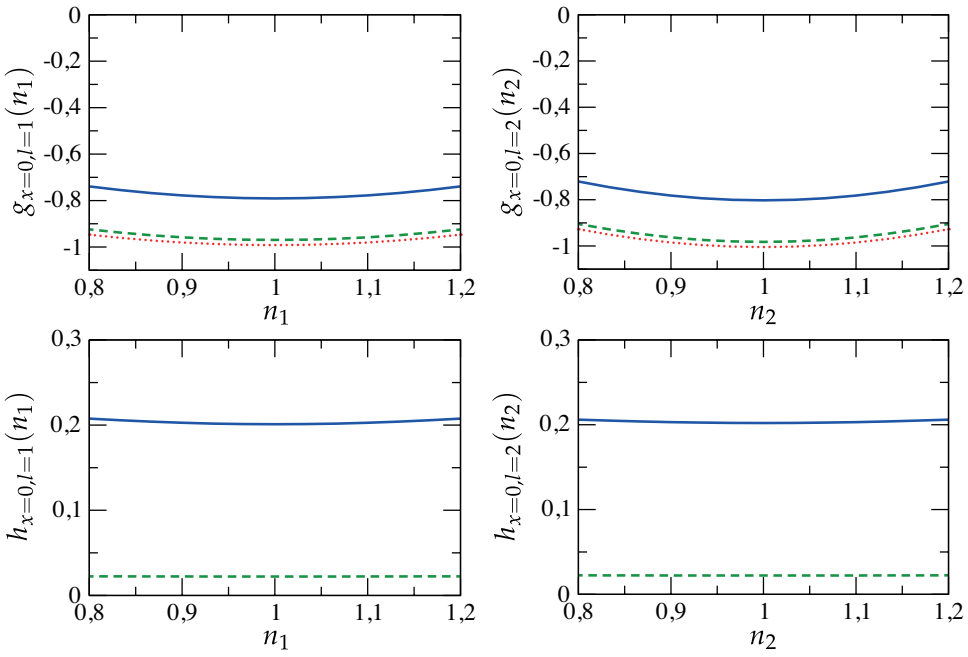


Figure 3.2: $g_{x=0,l}(n_l)$ and $h_{x=0,l}(n_l)$ for lattice sites $l = 1$, 2, and $L = 4 = N$, obtained from a grid of $R = 13$ values of $n_l \in [0.8, 1.2]$. The considered interaction strengths are $U = 1.0$ (solid line), $U = 0.1$ (dashed line), and $U = 0$ (dotted line). The curves for lattice sites $l = 3$ and 4 (not shown) coincide with the ones for $l = 2$ and 1 respectively.

(a)	X	$U = 0.0$	$U = 0.1$	$U = 1.0$
	0	$4.7 \cdot 10^{-2}$	$4.8 \cdot 10^{-2}$	$6.2 \cdot 10^{-2}$
	1	$1.9 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$	$2.8 \cdot 10^{-3}$
	2	$5.1 \cdot 10^{-4}$	$5.2 \cdot 10^{-4}$	$7.1 \cdot 10^{-4}$
	3	$5.0 \cdot 10^{-4}$	$5.2 \cdot 10^{-4}$	$7.0 \cdot 10^{-4}$

(b)	X	$U = 0.1$	$U = 1.0$
	0	$1.1 \cdot 10^{-3}$	$7.2 \cdot 10^{-3}$
	1	$1.6 \cdot 10^{-4}$	$9.2 \cdot 10^{-4}$
	2	$1.9 \cdot 10^{-5}$	$1.9 \cdot 10^{-4}$
	3	$1.3 \cdot 10^{-5}$	$1.4 \cdot 10^{-4}$

Table 3.3: Relative error ϵ_F (a) and ϵ_W (b) for $L = 4 = N$ and our ansatz on a grid of $R = 13$ values of $n_l \in [0.4, 1.6]$.

(a)	X	$U = 0.0$	$U = 0.1$	$U = 1.0$
	0	$2.1 \cdot 10^{-2}$	$2.1 \cdot 10^{-2}$	$2.6 \cdot 10^{-2}$
	1	$5.8 \cdot 10^{-3}$	$6.0 \cdot 10^{-3}$	$7.2 \cdot 10^{-3}$
	2	$3.5 \cdot 10^{-3}$	$3.6 \cdot 10^{-3}$	$4.4 \cdot 10^{-3}$
	5	$3.5 \cdot 10^{-3}$	$3.6 \cdot 10^{-3}$	$4.4 \cdot 10^{-3}$

(b)	X	$U = 0.1$	$U = 1.0$
	0	$4.0 \cdot 10^{-4}$	$2.9 \cdot 10^{-3}$
	1	$2.4 \cdot 10^{-4}$	$1.8 \cdot 10^{-3}$
	2	$1.4 \cdot 10^{-4}$	$1.3 \cdot 10^{-3}$
	5	$1.4 \cdot 10^{-4}$	$1.2 \cdot 10^{-3}$

Table 3.4: Relative error ϵ_F (a) and ϵ_W (b) for $L = 6 = N$ and our ansatz on a grid of $R = 5$ values of $n_l \in [0.6, 1.4]$.

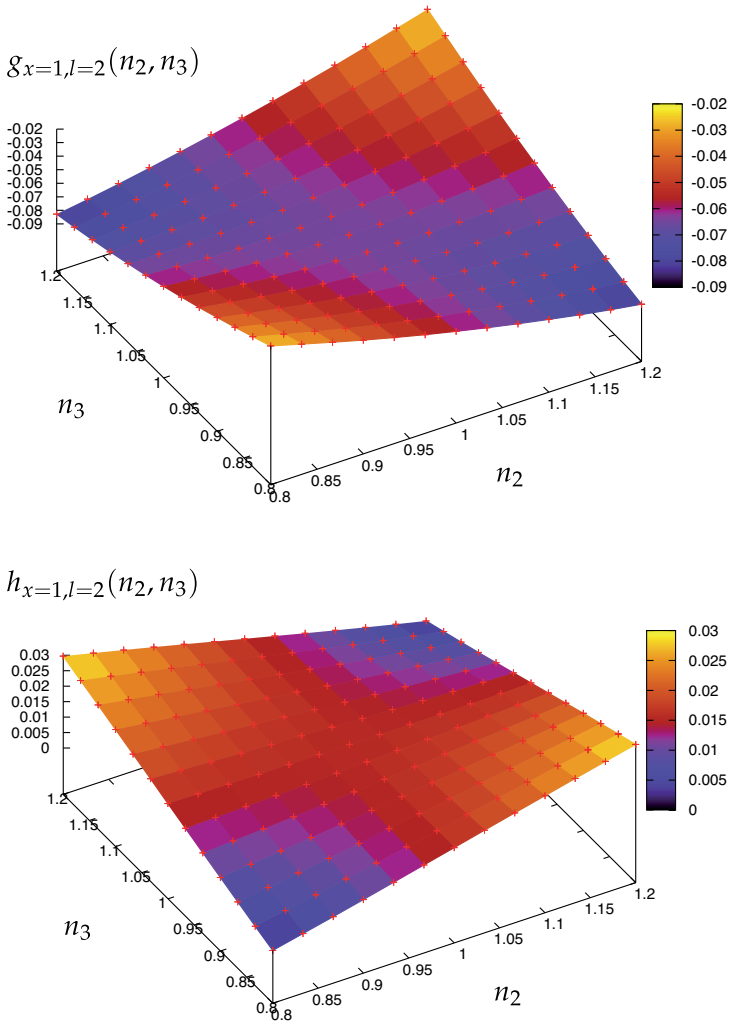


Figure 3.3: $g_{x=1,l=2}(n_2, n_3)$ and $h_{x=1,l=2}(n_2, n_3)$ for $L = 4 = N$ and $U = 1.0$, obtained from a grid of $R = 13$ values of $n_l \in [0.8, 1.2]$.

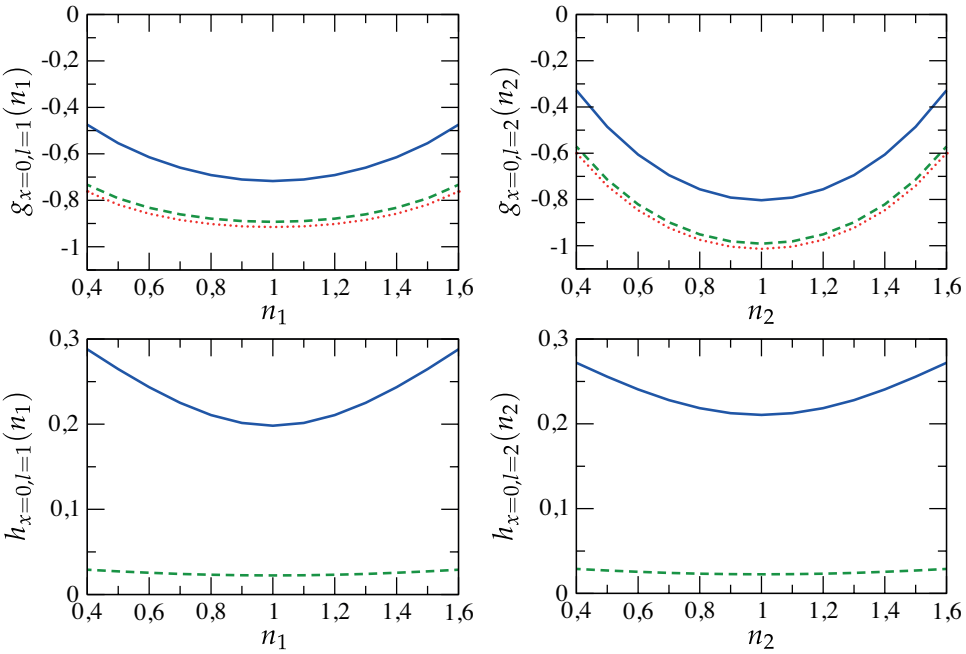


Figure 3.4: $g_{x=0,l}(n_l)$ and $h_{x=0,l}(n_l)$ for lattice sites $l = 1, 2$, and $L = 4 = N$, obtained from a grid of $R = 13$ values of $n_l \in [0.4, 1.6]$. The considered interaction strengths are $U = 1.0$ (solid line), $U = 0.1$ (dashed line), and $U = 0$ (dotted line). The curves for lattice sites $l = 3$ and 4 (not shown) coincide with the ones for $l = 2$ and 1 respectively.

$x = 0$ and in figure 3.5 for $x > 0$, for a selection of $L = 4 = N$ results. We observe that in the central region of n_l close to 1, all terms behave like the ones from the smooth training densities, while in the boundary region of n_l close to $n_{\min}$ or $n_{\max}$, the terms deviate from the second-order Taylor series expansion. Again, we equally analyzed the $L = 6 = N$ case which lead to the same conclusions as the $L = 4 = N$ case.

Polynomial Interpolation

Up to now, we considered our ansatz in equation (3.13) to be a tensor $g_{x,l}^{n_l,n_{l+x}}$ with four indices labelled by x, l, n_l, and n_{l+x}. Since such indices can only take on discrete values, our tensor ansatz can only be optimally determined for training densities $\vec{n}$ on a predefined density grid and applied to target densities on the same density grid. Obviously, it would be desirable to have a functional form of our ansatz that would enable us to evaluate our ansatz at target densities having a continuous range of values. Moreover, a differentiable form of our ansatz would be especially helpful, because both orbital free and Kohn-Sham methods benefit from well-defined and computable derivatives.

These desired properties are fulfilled by a polynomial ansatz in equation (3.13), in which the local terms are general polynomials of degree γ_0,

$$g_{x=0,l}(n_l) \quad := \quad \sum_{s_l=0}^{\gamma_0} g_{x=0,l}^{s_l} n_l^{s_l} \qquad , \tag{3.14}$$

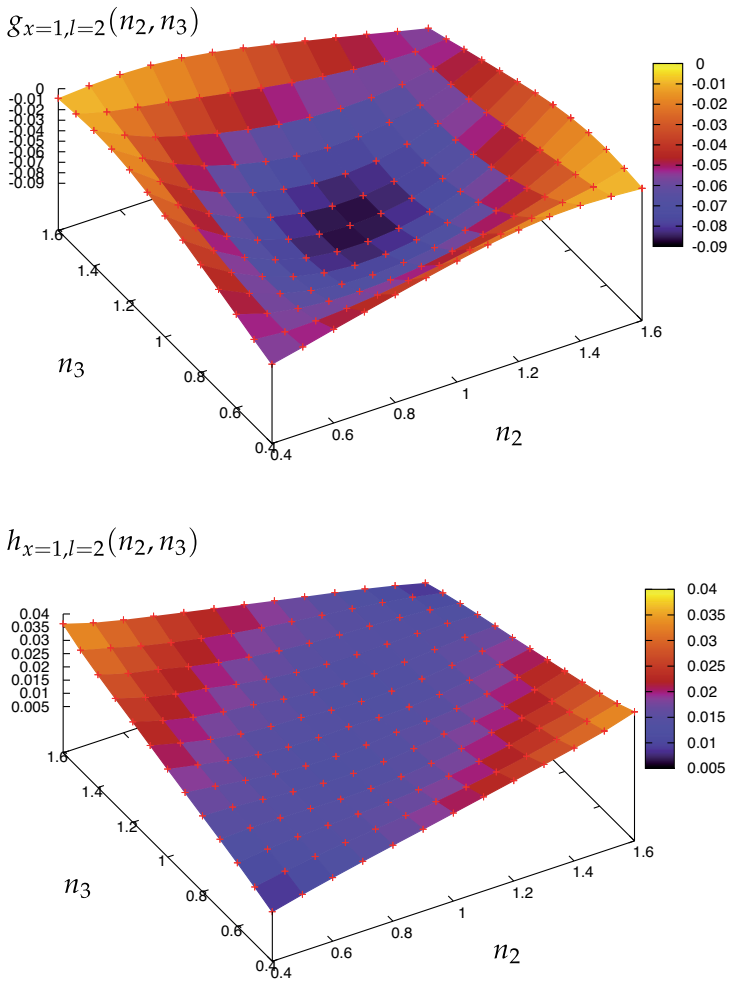


Figure 3.5: $g_{x=1,l=2}(n_2, n_3)$ and $h_{x=1,l=2}(n_2, n_3)$ for $L = 4 = N$ and $U = 1.0$, obtained from a grid of $R = 13$ values of $n_l \in [0.4, 1.6]$.

and the nonlocal terms are general polynomials of degree γ_1,

$$g_{x>0,l}(n_l, n_{l+x}) \quad := \quad \sum_{s_l, s_{l+x}=0}^{\gamma_1} g_{x>0,l}^{s_l, s_{l+x}} n_l^{s_l} n_{l+x}^{s_{l+x}} \qquad . \tag{3.15}$$

Assuming this ansatz, we have to determine the optimal tensors $g_{x=0,l}^{s_l}$ and $g_{x>0,l}^{s_l,s_{l+x}}$ such that the distance $\sum_{\vec{n}} |F_{\vec{n}} - \sum_l (\sum_{s_l} g_{x=0,l}^{s_l} n_l^{s_l} + \sum_{x>0} \sum_{s_l,s_{l+x}} g_{x>0,l}^{s_l,s_{l+x}} n_l^{s_l} n_{l+x}^{s_{l+x}}|^2$ is minimal. Again, we reformulate this least squares problem as a system of linear equations $C\vec{g} = \vec{F}$, where $\vec{g}$ is a vector of all $\mathcal{N}$ possible values of the coefficient tensors $g_{x=0,l}^{s_l}$ and $g_{x>0,l}^{s_l,s_{l+x}}$, $\vec{F}$ is a vector of all $\mathcal{M}$ values of $F_{\vec{n}}$ corresponding to the considered $\mathcal{M}$ training densities, and C is a $\mathcal{M} \times \mathcal{N}$ dimensional matrix whose non-vanishing entries are the values of $n_l^{s_l}$ and of $n_l^{s_l} n_{l+x}^{s_{l+x}}$ evaluated for the training densities.

In order to assess the usefulness of such a polynomial ansatz, we now analyze its approximation of the exact tensor $F_{\vec{n}}$ corresponding to $L = 4 = N$, $U = 0.1$, and a grid of $R = 13$ values of $n_l \in [0.8, 1.2]$. We can understand the restrictions of our polynomial ansatz if we compare our results here to the ones that we previously obtained when our ansatz was a general tensor. We find that, for a fixed value of X, when γ_0 and γ_1 increase, the relative error ϵ_F always quickly approaches the minimum error reached for our ansatz without polynomial constraints. E.g., already with $\gamma_0 = 4$ and $\gamma_1 = 4$, we get $\epsilon_F(X = 0) = 8.9673335 \cdot 10^{-3}$, $\epsilon_F(X = 1) = 6 \cdot 10^{-5}$, $\epsilon_F(X = 3) = 4 \cdot 10^{-5}$, where the corresponding results obtained without polynomial constraints read $\epsilon_F(X = 0) = 8.9673328 \cdot 10^{-3}$, $\epsilon_F(X = 1) = 4 \cdot 10^{-5}$, $\epsilon_F(X = 3) = 6 \cdot 10^{-6}$. However, the functional form does not converge to the one obtained without polynomial constraints.

The latter aspect can be fixed if we restrict the general polynomial ansatz of equations (3.14) and (3.15) to the form of the above Taylor series expansion, such that the local terms satisfy

$$g_{x=0,l}(n_l) \quad := \quad \sum_{s_l=0}^{\gamma_0} g_{x=0,l}^{s_l} (n_l - 1)^{s_l} \tag{3.16}$$

with $s_l \in \{0, 2, 4, \ldots, \gamma_0\}$ and the nonlocal terms satisfy

$$g_{x>0,l}(n_l, n_{l+x}) \quad := \quad \sum_{s_l, s_{l+x}=1}^{\gamma_1} g_{x>0,l}^{s_l,s_{l+x}} (n_l - 1)^{s_l} (n_{l+x} - 1)^{s_{l+x}} \tag{3.17}$$

with $s_l + s_{l+x} \in \{2, 4, \ldots, 2\gamma_1\}$. Running s_l and $s_l + s_{l+x}$ only over even numbers guarantees that we recover the above second-order Taylor series expansion for $\gamma_0 = 2$ and $\gamma_1 = 1$. Given a set of training densities, the optimal tensors $g_{x=0,l}^{s_l}$ and $g_{x>0,l}^{s_l,s_{l+x}}$ follow from the same procedure as used above for the general polynomial ansatz. As in this case, we consider an approximation of the exact tensor $F_{\vec{n}}$ corresponding to $L = 4 = N$, $U = 0.1$, and a grid of $R = 13$ values of $n_l \in [0.8, 1.2]$. We observe that the approximation errors basically coincide with the ones attained above for the general polynomial ansatz. However, now the functional form converges to the desired best one obtained without polynomial constraints, s. figure 3.6 for a $L = 4 = N$ example.

3.3.2 Larger Coulomb Problems

In order to move on to more realistic problems, this section departs from the scenario of the previous section in two respects. On the one hand, we now investigate fermions with a Coulomb interaction, Hamiltonian (3.11), which represents

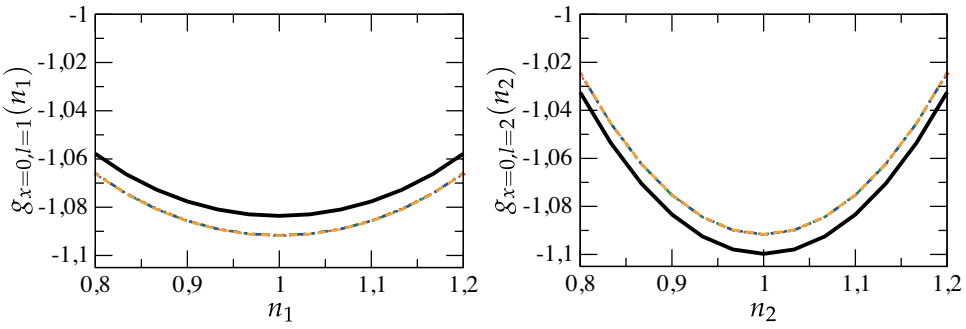


Figure 3.6: $g_{x=0,l}(n_l)$ for lattice sites $l = 1$ and 2. We compare the exact solution (thick solid line), obtained without polynomial constraints, to the polynomial solutions for $X = 0$ and $\gamma_0 = 2$ (dotted line), 4 (dashed-doubledotted line), 6 (dashed-dotted line), and 100 (dashed line). Our polynomial solutions all lie on top of each other. The curves for lattice sites $l = 3$ and 4 (not shown) coincide with the ones for $l = 2$ and 1 respectively.

the typical situation of DFT calculations. On the other hand, we now aim at larger lattices and compute (quasi) exact ground states with our MPS algorithms. The ultimate motivation for our analysis of larger system sizes comes from the fact that DFT commonly treats continuous space, which we can approximate via discretization with a fine grid, leading to a Hamiltonian (3.11) defined on a large lattice. Consequently, in the following, we do not have access to the exact tensors $F_{\vec{n}}$ and $W_{\vec{n}}$ for all possible training densities $\vec{n}$ anymore, but, rather, a limited set of solutions $F(\vec{n})$ and $W(\vec{n})$ for training densities $\vec{n}$ arising from a certain class of physically relevant problems. For such given training set, we recover the optimal parameters for an ansatz of the form (3.13). As in the previous section, the functionals are required to provide energies for continuous values of the density in the allowed range and hence an interpolation scheme is needed. In the following, we describe the splines method we have used to this end, which is more general than the polynomial ansatz.

Spline Interpolation

In the preceding section, we argued that our ansatz (3.13) should be defined and differentiable for a continuous range of density values. We have seen that a polynomial interpolation satisfies these criteria, but, in general, does not lead to the correct functional form. While we could correct that for smooth training densities, for which we knew the correct polynomial form from a Taylor series expansion, such kind of correction is not feasible when the correct polynomial form is not known a priori, which is the case, in general, for less smooth training densities. It is well-known that an interpolation by a polynomial of high degree often does not recover the true form of the fitted function and that, in this regard, an interpolation by splines can perform better. Thus, from now on, we formulate and analyze our ansatz (3.13) in terms of splines.

The standard spline approach. In order to introduce the standard definition of cubic splines considered here as well as the corresponding fitting and evaluation procedure [122], we first describe the procedure for a univariate function $\tilde{g}(n)$, which could correspond, e.g., to each of the local terms in our ansatz, $g_{x=0,l}(n_l)$.

Assuming that the value of $\tilde{g}(n)$ is known at R grid points $n_{\min} := n_1 < n_2 < \ldots < n_R =: n_{\max}$, i.e., $\tilde{g}(n)$ takes on known values y_i at the grid points n_i, then a spline interpolation has the form:

$$g(n) \quad = \quad \begin{cases} g_1(n), & \text{if} \quad n_1 \leq n < n_2 \\ g_2(n), & \text{if} \quad n_2 \leq n < n_3 \\ \quad \vdots & \\ g_{R-1}(n), & \text{if} \quad n_{R-1} \leq n \leq n_R \end{cases} \quad , \tag{3.18}$$

where each $g_i(n) = a_{i,0} + a_{i,1}(n - n_i) + a_{i,2}(n - n_i)^2 + a_{i,3}(n - n_i)^3$ is a polynomial of degree three for the cubic splines considered here. The parameters of the spline, $a_{i,j}$, have to be such that $g(n_i) = y_i$ at the grid points n_i and the function, $g(n)$, and its first and second derivative, $g'(n)$ and $g''(n)$, are continuous in the complete range of allowed values of $n \in [n_{\min}, n_{\max}]$. It turns out that a cubic splines interpolation is uniquely defined when all values of y_i and additionally the two values of $g''(n)$ at n_1 and n_R are specified. A common choice is to set $g''(n_1) = 0 = g''(n_R)$, in which case the function behaves like a straight line at its endpoints n_1 and n_R, and we will only discuss this so-called *natural spline* in the following.[1]

After vectorizing the desired parameters $a_{i,j}$ in $\vec{a} := (a_{1,0}, a_{1,1}, a_{1,2}, a_{1,3}, a_{2,0}, a_{2,1}, \ldots, a_{R-1,3})^{\mathrm{t}}$ and the given values y_i in $\vec{y} := (y_1, y_2, \ldots, y_R)^{\mathrm{t}}$, we can express the solution of our splines interpolation as a matrix vector multiplication:

$$\vec{a} \quad = \quad (C_1 + C_2 C_3^{-1} C_4)\vec{y} \quad . \tag{3.19}$$

The non-vanishing entries of the four matrices C_1 to C_4 depend only on the grid spacings $h_i := n_{i+1} - n_i$ in the following way:

$$\begin{aligned} (C_1)_{(i,j)} &:= \delta_{i,4(j-1)+1} - \frac{1}{h_j}\delta_{i,4(j-1)+2} + \frac{1}{h_{j-1}}\delta_{i,4(j-2)+2} \\ (C_2)_{(i,j)} &:= -\frac{h_{j+1}}{3}\delta_{i,4j+2} - \frac{h_j}{6}\delta_{i,4(j-1)+2} + \frac{1}{2}\delta_{i,4j+3} - \\ &\quad \frac{1}{6h_{j+1}}\delta_{i,4j+4} + \frac{1}{6h_j}\delta_{i,4(j-1)+4} \\ (C_3)_{(i,j)} &:= h_i\delta_{i,j+1} + 2(h_i + h_{i+1})\delta_{i,j} + h_{i+1}\delta_{i,j-1} \\ (C_4)_{(i,j)} &:= \frac{6}{h_i}\delta_{i,j} - 6(\frac{1}{h_i} + \frac{1}{h_{i+1}})\delta_{i,j-1} + \frac{6}{h_{i+1}}\delta_{i,j-2} \quad , \end{aligned}$$

where the dimension of C_1 is $4(R-1) \times R$, C_2 is $4(R-1) \times (R-2)$, C_3 is $(R-2) \times (R-2)$, and C_4 is $(R-2) \times R$.

Having calculated the spline coefficients $\vec{a}$, we can evaluate our solution $g(n)$ of equation (3.18) at an arbitrary density value n in $[n_{\min}, n_{\max}]$. For that, we have to find the interval i enclosed by grid points n_i and n_{i+1} such that $n_i \leq n < n_{i+1}$. The cubic polynomial corresponding to this interval i is defined by the coefficients $a_{i,0}$, $a_{i,1}$, $a_{i,2}$, and $a_{i,3}$, which are listed in $\vec{a}$ on positions $4(i-1)+1$ to $4(i-1)+4$ in this consecutive order. If we introduce a vector $\vec{n}_i$ of the same dimension $4(R-1)$ as $\vec{a}$ and with zero entries everywhere but on positions $4(i-1)+1$ to $4(i-1)+4$ where its elements are 1, $n - n_i$, $(n - n_i)^2$, and $(n - n_i)^3$ in this sequence, then the value $g(n)$ of our spline solution at density n takes on the form

$$g(n) \quad = \quad \vec{n}_i^{\mathrm{t}} \cdot \vec{a} \quad . \tag{3.20}$$

[1] An adaptation of our discussion to the two alternative choices [122] that we have also implemented, namely, $g''(n_1) = g''(n_2)$ and $g''(n_R) = g''(n_{R-1})$ (*parabolic runout spline*), and $g''(n_1) = 2g''(n_2) - g''(n_3)$ and $g''(n_R) = 2g''(n_{R-1}) - g''(n_{R-2})$ (*cubic runout spline*), is straightforward.

Our spline approach for the local terms. The form of equation (3.20) provides an alternative approach for fitting a splines function that will become the basis of our fit when the ansatz (3.13) is formulated in terms of splines. In this alternative approach, we consider an arbitrary function $f(n)$ of which we know the values $f(n_k)$ for $\mathcal{M}$ training densities n_k, and we want to determine the optimal splines function $g(n)$ of equation (3.18) that minimizes the distance $\sum_{k=1}^{\mathcal{M}} |f(n_k) - g(n_k)|^2$. Here, for $g(n)$, both the number of grid points, R, and their positions, $n_1 < n_2 < \ldots < n_R$, are parameters of choice in our ansatz, independent of the training densities n_k. For a choice of the density grid, we will be searching the $\mathcal{N} = 4(R-1)$ parameters $\vec{a}$ of the optimal $g(n)$. As in our previous fits (s. discussion after equation (3.13) as well as after equations (3.14) and (3.15)), we vectorize the known function values $\vec{f} := (f(n_1), f(n_2), \ldots, f(n_{\mathcal{M}}))^t$ and replace the least squares problem by a system of linear equations $C\vec{a} = \vec{f}$, where the $\mathcal{M} \times \mathcal{N}$ dimensional matrix C holds in row k the transposed vector $\vec{n}_k^t$ of equation (3.20) corresponding to training density n_k. For $\mathcal{M} \geq \mathcal{N}$, the minimum of the least squares problem is attained when $\vec{a} = C^{-1}\vec{f}$, by means of the pseudoinverse C^{-1}.

This procedure can be directly generalized to fit our ansatz (3.13) for $X = 0$, i.e., consisting of the sum of L local terms $g_{x=0,l}(n_l)$ only, and we can drop the index x here. In that case, different vectors $\vec{a}_l$ of spline coefficients define the different functions $g_l(n_l)$, and, for simplicity, we impose the same grid on each lattice site l. Notice, nevertheless, that the density values do not need to be uniformly distributed in the given interval. We concatenate all the vectors $\vec{a}_l$ for $l = 1, 2, \ldots, L$ in the $4(R-1)L$ dimensional vector $\vec{a} := (\vec{a}_1^t, \vec{a}_2^t, \ldots, \vec{a}_L^t)^t$. Our goal is to determine the elements of $\vec{a}$ such that, given $\mathcal{M}$ training densities $\vec{n}$ for which $F(\vec{n})$ is known, the distance $\sum_{i=1}^{\mathcal{M}} |F(\vec{n}) - \sum_{l=0}^{L} g_l(n_l)|^2$ is minimized. As before, we vectorize the $\mathcal{M}$ known values $F(\vec{n})$ in $\vec{F}$ and substitute the least squares problem by a system of linear equations $C\vec{a} = \vec{F}$. The $\mathcal{M} \times \mathcal{N}$ dimensional matrix C now stores in row i the $4(R-1)L$ dimensional transposed vector $\vec{n}_{i,l}^t$, which is a concatenation of the L transposed vectors $\vec{n}_l^t$ of the form (3.20) for $l = 1, 2, \ldots, L$, where $\vec{n}_l^t$ corresponds to the lth entry n_l of the ith training density $\vec{n}$. By providing enough training densities, we ensure that $\mathcal{M} \geq \mathcal{N} = 4(R-1)L$ and then obtain the desired solution $\vec{a} = C^{-1}\vec{F}$ via the pseudoinverse C^{-1}.

Our spline approach for the nonlocal terms. In the previous paragraphs, we dealt with the interpolation and fitting of the local terms, which are univariate functions of the local density values. The non-local ansatz of equation (3.13) includes also functions of two density values, for which a recursive spline scheme can be used, which is described in the following.

Let us consider a function of two continuous values, $g(n_x, n_y)$. For simplicity, we impose the same grid $n_1 < n_2 < \ldots < n_R$ on both density axes n_x and n_y, such that $g(n_x, n_y)$ is defined for $n_x \in [n_1, n_R]$ and $n_y \in [n_1, n_R]$. To evaluate the function at a point (n_l, n_m) that does not belong to the grid, we proceed in three steps. First, we determine the R values $g(n_x = n_l, n_y = n_i)$ at the grid points n_i of n_y and the final $n_x = n_l$ coordinate. These are given by the known R splines $g_i(n_x)$ at the position $n_x = n_l$. Second, we interpolate the R values $g_i(n_x = n_l)$ by a spline function $g_y(n_y)$. Third, we compute the spline function $g_y(n_y)$ at $n_y = n_m$. Figure 3.7 illustrates this method when $R = 5$.

The evaluation of our two-dimensional spline $g(n_x, n_y)$ can be written in a compact form similar to equation (3.20). Since $g(n_x, n_y)$ is uniquely defined by the R vectors $\vec{a}_i$ each of dimension $4(R-1)$, we summarize these in a $4(R-1)R$ dimen-

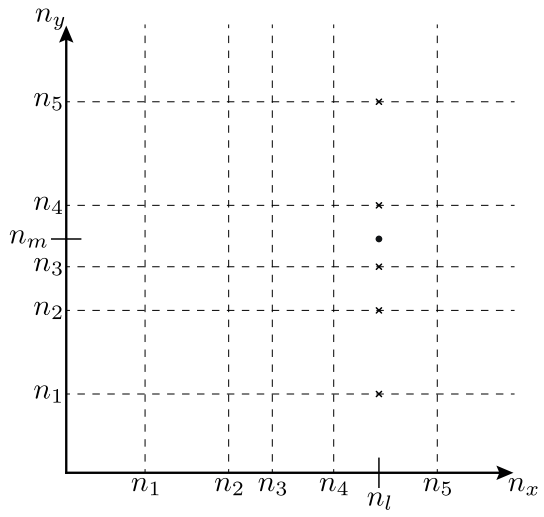


Figure 3.7: As explained in the text, we evaluate a single nonlocal term $g(n_x, n_y)$ at $n_x = n_l$ and $n_y = n_m$ (filled circle) via a spline interpolation of the R values $g(n_x = n_l, n_y = n_i)$ at the grid points n_i of n_y (crosses).

sional vector $\vec{A} := (\vec{a}_1^{\,t}, \vec{a}_2^{\,t}, \ldots, \vec{a}_R^{\,t})^t$. Assuming $n_x = n_l$ is located in interval l, we can construct a $R \times 4(R-1)R$ dimensional matrix C_{n_l}, such that $g_i(n_x = n_l)$ is the ith entry of a vector $\vec{g}_y := C_{n_l}\vec{A}$. For that, the ith row of the matrix C_{n_l} contains the $4(R-1)R$ dimensional transposed vector $\vec{n}_{i,l}^{\,t}$ which is zero everywhere but on the positions $4(R-1)\cdot(i-1)+1$ to $4(R-1)\cdot i$ where the elements of $\vec{n}_l^{\,t}$ from equation (3.20) are stored. Multiplication of the matrix $C_y := C_1 + C_2 C_3^{-1} C_4$ from equation (3.19) with the vector $\vec{g}_y$ gives the coefficients $\vec{a}_y := C_y \vec{g}_y$ of the spline $g_y(n_y)$. Assuming $n_y = n_m$ is located in the interval m, multiplying the transposed vector $\vec{n}_m^{\,t}$ of the form (3.20) with $\vec{a}_y$ leads to the solution $g(n_x = n_l, n_y = n_m) = \vec{n}_m^{\,t}\vec{a}_y$. In summary, the analog of equation (3.20) for our nonlocal term takes on the form:

$$g(n_l, n_m) \quad = \quad \vec{n}_m^{\,t} C_y C_{n_l} \vec{A} \qquad . \tag{3.21}$$

Equation (3.21) enables us to fit an arbitrary two-dimensional function $f(n_x, n_y)$, known at $\mathcal{M}$ training densities $\vec{n}_k$, by our two-dimensional spline $g(n_x, n_y)$, defined via $\mathcal{N} = 4(R-1)R$ coefficients $\vec{A}$, by minimizing the distance $\sum_{k=1}^{\mathcal{M}} |f(n_x, n_y) - g(n_x, n_y)|^2$. The spline function $g(n_x, n_y)$ may be based upon an arbitrary density grid of R grid points $n_1 < n_2 < \ldots < n_R$, but we assume here, for simplicity, that the same density grid is imposed in n_x as well as n_y direction. As before, listing the known values $f(\vec{n}_k)$ in $\vec{f} := (f(\vec{n}_1), f(\vec{n}_2), \ldots, f(\vec{n}_{\mathcal{M}}))^t$, we recast the least squares problem into a system of linear equations $C\vec{A} = \vec{f}$, where the kth row of the matrix C is given by the transposed vector $\vec{n}_m^{\,t} C_y C_{n_l}$ of equation (3.21) corresponding to training density $\vec{n}_k = (n_l, n_m)^t$. Having generated a number of training densities $\mathcal{M}$ no less than there are parameters $\mathcal{N}$, our least squares problem is minimized when $\vec{A} = C^{-1}\vec{f}$ via the pseudoinverse C^{-1}.

Our spline approach for our complete ansatz. Thanks to the above developed methods, it is now straightforward to fit our ansatz (3.13) in its most general form, i.e., for an arbitrary value of the maximal range X. Because a particular term $g_{x,l}(n_l, n_{l+x})$ is uniquely identified by the values of its indices x and l, the corresponding vectors of spline coefficients $\vec{a}_{x,l}$ carry these same indices. Each vector

$\vec{a}_{x,l}$ is defined according to the previous discussion. We concatenate all vectors of spline coefficients for all possible values of x and l in a single vector $\vec{A}$. This allows us, again, to find the minimum of the least squares problem $\sum_k |F(\vec{n}) - \sum_{x,l} g_{x,l}(n_l, n_{l+x})|^2$ by solving the linear equations $C\vec{A} = \vec{F}$ where $\vec{F}$ comprises the $\mathcal{M}$ known values $F(\vec{n}_k)$ at training densities $\vec{n}_k$ of the function F to be fitted. The matrix C is $\mathcal{M} \times \mathcal{N}$ dimensional, where $\mathcal{N} = 4(R-1)L + 4(R-1)R\sum_{x=1}^{X}(L-x)$, and its ith row follows from proper concatenation of the vectors $\vec{n}_l^{\,\mathrm{t}}$ of the form (3.20) and $\vec{n}_m^{\,\mathrm{t}} C_y C_{n_l}$ of the form (3.21) which are derived from the entries n_l and n_m of the ith training density $\vec{n}$. As before, we have to guarantee $\mathcal{M} \geq \mathcal{N}$ and, in general, use the pseudoinverse C^{-1} in order to get the solution $\vec{A} = C^{-1}\vec{F}$.

Particular Details of Our Spline Fit. Before we can conclude this discussion on the spline interpolation of our ansatz (3.13), final technical comments regarding our implementation are still in order:

- Although the above discussion explains the computation of the optimal spline coefficients, in our code, we instead find the function values of the terms $g_{x,l}(n_l, n_{l+x})$ at all grid points. The spline coefficients can always be derived uniquely from these function values at the grid points, s., e.g., equation (3.19). We make use of this fact because it reduces the number of variational parameters $\mathcal{N}$ such that fewer training densities $\mathcal{M} \geq \mathcal{N}$ suffice.
- We impose a physically motivated constraint on our ansatz (3.13). Since the considered Hamiltonian (3.11) is reflection-symmetric, our ansatz also has that property. Notice that this additionally reduces the number of variational parameters and, thus, of necessary training densities.
- In order to better assess the magnitude of nonlocality of the function $F(\vec{n})$ to be fitted, we determine the $g_{x,l}(n_l, n_{l+x})$ with increasing x one after another. This means that, given a certain maximal range X, we first fit the known function values $\vec{F}$ by our ansatz (3.13) including only the $x = 0$ terms ($C_0\vec{A}_0 = \vec{F}$), then the remainder $\vec{F} - C_0\vec{A}_0$ by our ansatz including only the $x = 1$ terms ($C_1\vec{A}_1 = \vec{F} - C_0\vec{A}_0$), then $\vec{F} - C_0\vec{A}_0 - C_1\vec{A}_1$ by only the $x = 2$ terms, etc., until we are done with the fit of the $x = X$ terms.

Performance of Splines

Having thoroughly described how our ansatz (3.13) is formulated and fitted in terms of spline functions, we are ready to analyze its performance. Equation (3.13) allows a free choice of the maximal range X as well as of the number of grid points R and their positions. We want to find out how these parameters of our ansatz affect its performance. Similar to the preceding performance analyses, carried out for tensors and polynomials, we are interested in both the approximation error and the functional form of the optimal spline interpolation.

Notice that now our training densities are not obtained by means of inversion anymore. Instead, we choose a set of training scenarios, corresponding to certain types of external potentials. Here, we consider three scenarios that we call *kink*, *barrier*, and *well*, and whose parameters are illustrated in figure 3.8. Our training densities are the various ground state densities emerging in each training scenario when its parameters are varied in small steps within a wide range. The corresponding external potentials shall model, as simply as possible, physically relevant situations.[2]

[2]We remark that, for the considered Hamiltonian (3.11), the ground state density in a certain training

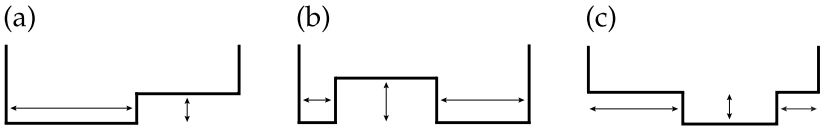


Figure 3.8: External potentials of training scenario *kink* (a), *barrier* (b), and *well* (c). Each scenario is defined by the set of potentials that arise when the corresponding free parameters, indicated by arrows, are varied in small steps within a wide range.

Clearly, for each training potential, we still have to specify the total number of particles. We are able to compute the ground states of Hamiltonian (3.11) for large numbers of fermions on long lattices: in the noninteracting case, $U = 0$, via exact diagonalization of $2L \times 2L$ dimensional matrices and in the interacting case, $U \neq 0$, via our MPS algorithms. The spline fit might certainly depend on the total particle numbers of the considered training densities. Therefore we distinguish here the two simplest extremes, namely, firstly, training with exactly two particles and, secondly, training with all possible particle numbers within a certain range.

Before we can begin to investigate these two cases, some technical details still need to be discussed. From the various parameter choices possible in our analysis, we use the following, if not stated otherwise. In the considered Hamiltonian (3.11), we set the tunneling element to $t = 1/2$, and we refer to the noninteracting case when $U = 0$, and we speak of the interacting case when $U = 1$. The standard lattice size we are concerned with is $L = 21$. This choice of $t = 1/2$ and $L = 21$ allows us to understand Hamiltonian (3.11) as arising from discretizing continuous space of length 20 via a discretization of $\Delta = 1$. We compared results from lattice sizes $L = 11$, 21, and 41, and found that $L = 21$ captures the essential physics we are interested in here. At the same time, this rather small lattice size enables us to quickly compute many training densities even in the interacting case using our MPS code, with which we always checked convergence with bond dimension D. Regarding our spline fit, for the positions of the R density grid points, we always choose $n_1 = n_{\min} = 0$ and $n_R = n_{\max} = 2$ such that our spline interpolations are defined for all possible density values $n \in [0, 2]$ on each lattice site. Furthermore, for each value of R, we always perform a fit for two density grids, one with equal spacing between the R grid points and another with unequal spacing between them, where the unequal spacing is chosen from several random trials as the spacing leading to the smallest approximation error. If we do not explicitly mention the spacing of the density grid, equal spacing is assumed. We use natural splines throughout this work and solve all encountered linear equations by means of pseudoinverses calculated with a cutoff 10^{-6}, which turned out to be the best cutoff value between 10^{-2} and 10^{-14} as we have checked for a wide range of problems.

Fixed Total Particle Number N = 2. For a fixed total particle number $N = 2$, we consider all three training scenarios kink, barrier, and well, and obtain a set of 58302 training densities for $U = 0$ and of 65298 training densities for $U = 1$. These two training sets are the basis of the following analysis.

We first fix the maximal range to $X = 0$, such that our ansatz (3.13) consists of the local terms only. In this case, we can study the effect of the number of grid

potential has the same value of the Hohenberg-Kohn functional as the reflected density, corresponding to the ground state of the reflected potential, s. equation (3.12). Therefore, if a certain training potential is not reflection-symmetric already, we always include both the ground state density and its reflected counterpart with the same value of F in our training set.

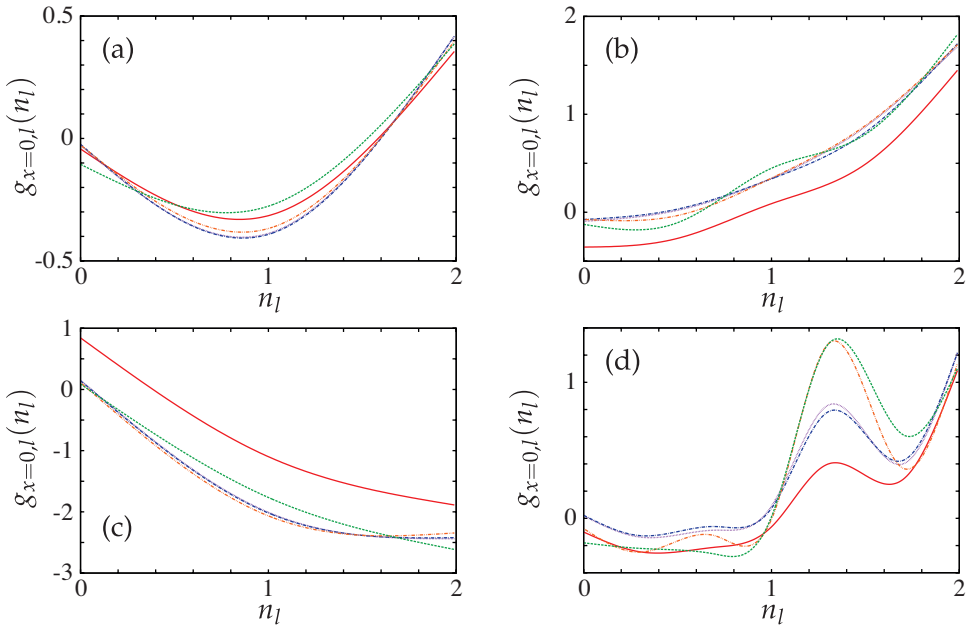


Figure 3.9: Local terms $g_{x=0,l}(n_l)$ on lattice sites $l = 1$ (dashed), $l = 4$ (dash-doubledotted), $l = 7$ (dash-dotted), $l = 8$ (dotted), and $l = 11$ (solid), for $L = 21$, $U = 0$ and $R = 3$ grid points $\{0, 1, 2\}$ (a), $U = 0$ and $R = 5$ grid points $\{0, 0.5, 1, 1.5, 2\}$ (b), $U = 1$ and $R = 3$ grid points $\{0, 1, 2\}$ (c), and $U = 1$ and $R = 7$ grid points $\{0, 0.3, 0.7, 1, 1.3, 1.7, 2\}$ (d).

points R as well as of their positions. For $U = 0$, our mean relative error reads ≈ 0.03 when $R = 3$ and decreases to ≈ 0.01 when $R = 7$. The effect of different grid spacing can be significant, e.g., for $R = 5$, the equally spaced density grid gives an error ≈ 0.017 while the unequally spaced grid gives ≈ 0.014. For $U = 1$, our mean relative error reads ≈ 0.4 when $R = 3$ and reduces to ≈ 0.09 when $R = 7$. Again, the grid spacing affects the error, e.g., for $R = 7$, the grid with unequal spacing gives ≈ 0.08.

We observe that both the number of grid points R and the grid spacing affect the functional form of the local terms. With growing value of R, the spline interpolations always feature stronger oscillations. Figure 3.9 illustrates this behavior for the non-interacting and interacting cases.

When we take into account the first nonlocal terms in our ansatz, i.e., $X = 1$, the mean relative error for $R = 3$ reads ≈ 0.015 when $U = 0$ and ≈ 0.05 when $U = 1$. For $R = 5$, we obtain a mean relative error ≈ 0.005 when $U = 0$ and ≈ 0.04 when $U = 1$. The convergence of the error as a function of the number of training densities occurs very fast, namely, within the first 10000 to 20000 training densities for all cases considered here.

Many Different Particle Numbers. We extend our training based on a single total particle number now to include many different particle numbers. Our training set comprises all training densities resulting from the kink scenario with total particle numbers $N = 1, 2, \ldots, 30$. This gives a total number of training densities 25200 for $U = 0$ and 25186 for $U = 1$.

With these new training sets, we continue the performance analysis of our ansatz (3.13) for larger values of the maximal range X. For $U = 0$, our results are

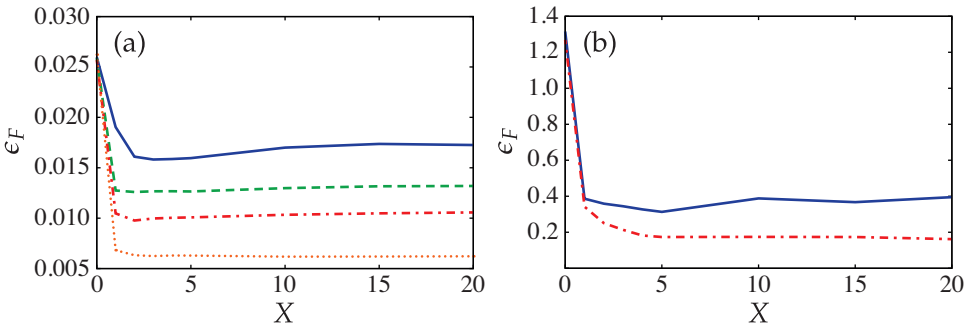


Figure 3.10: Mean relative error ϵ_F as a function of the maximal range X, for $L = 21$, $U = 0$ (a), $U = 1$ (b), and several values of R density grid points, namely $R = 4$ (solid), $R = 5$ (dashed), $R = 6$ (dash-dotted), and $R = 8$ (dotted). The density grids are given by $\{0, 0.65, 1.35, 2\}$ when $R = 4$ (a), $\{0, 0.8, 1.2, 2\}$ when $R = 4$ (b), $\{0, 0.8, 1, 1.2, 2\}$ when $R = 5$, $\{0, 0.7, 0.9, 1.1, 1.3, 2\}$ when $R = 6$, and $\{0, 0.2, 0.8, 0.95, 1.05, 1.2, 1.8, 2\}$ when $R = 8$.

shown in figure 3.10(a). Initially, when $X = 0$, the mean relative error is ≈ 0.026 independent of the number of grid points $R = 3$, 4, 5, 6, and 8. This is different from the previous training sets where R had a clear effect on the approximation error. The mean relative error improves substantially when $X = 1$ and it saturates beyond $X = 1$, which indicates that the long-range terms $g_{x,l}(n_l, n_{l+x})$ with $x > 1$ are not particularly helpful for this training set of noninteracting fermions. The improvement for $X = 1$ depends on R again, such that with a finer grid we get smaller errors. Our results for $U = 1$ are shown in figure 3.10(b). In this case, we consider the two values $R = 4$ and 6. When $X = 0$, the mean relative error is roughly the same for both R and reads ≈ 1.3. A substantial improvement with increasing X is visible until $X = 4$. Beyond $X = 4$, the error still decreases slightly for $R = 6$ and oscillates for $R = 4$, which can only be due to numerical inaccuracies.

Just like the corresponding approximation error, the functional form of our ansatz (3.13) with $X = 0$ features no strong dependence on R. This is true both in the noninteracting case, $U = 0$, and in the interacting case, $U = 1$. In particular, the terms $g_{x=0,l}(n_l)$ on the central lattice sites l can all be described by the same function of n_l, possibly shifted by a constant value, s. figure 3.11(a) for $U = 0$ and figure 3.11(b) for $U = 1$. For $U = 0$, this function has a form similar to the one obtained previously when our ansatz (3.13) was formulated in terms of general tensors. The functional form of the non-local terms, in general, depends on R, similar to the mean relative approximation error when $X \geq 1$. The various spline functions $g_{x>0,l}(n_l, n_{l+x})$ for fixed range $x > 0$ obtained on different central lattice sites l can have the same form, and we have observed that for $U = 0$ as well as for $U = 1$ and for both $x = 1$ and $x = 2$.

When fitted with the previous training set of varying total particle numbers, our ansatz (3.13) with $X = 0$ produces local terms that converge on the central lattice sites to a specific functional form, which appears also in the calculation of the exact LDA. The latter is derived from the ground state energies of a homogeneous system, i.e., having the same density everywhere, in the thermodynamic limit [113]. We can take a look at this quantity for finite systems. Because the density is the same on each lattice site, the Hohenberg-Kohn functional, $F(\vec{n})$, and the exchange-correlation energy functional, $W(\vec{n})$, are simply functions of only a single parameter $n = N/L$ that is determined by the total particle number N for a given

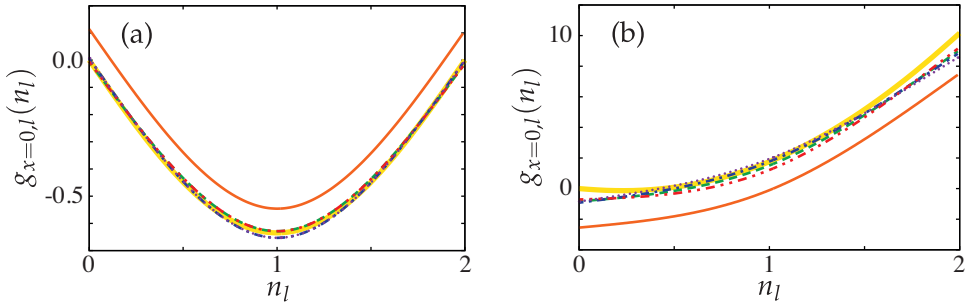


Figure 3.11: Local terms $g_{x=0,l}(n_l)$ on central lattice sites $l = 7$ (dotted), $l = 8$ (dash-doubledotted), $l = 9$ (dash-dotted) $l = 10$ (dashed), and $l = 11$ (solid), for $L = 21$, $R = 3$, $U = 0$ (a), and $U = 1$ (b). We also show the exact values of F for exactly homogeneous densities (thick solid yellow) (see figure 3.12 and corresponding discussion).

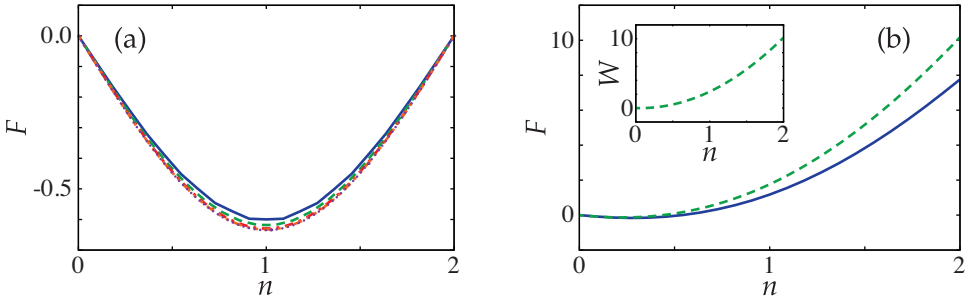


Figure 3.12: Exact $F(n)$ for $U = 0$ (a), for $U = 1$ (b), and exact $W(n)$ for $U = 1$ (inset in (b)), for exactly homogeneous densities $n = N/L$ on lattices of sizes $L = 11$ (solid), $L = 21$ (dashed), $L = 51$ (dash-dotted), $L = 101$ (dash-doubledotted), and $L = 201$ (dotted).

system size L. For a fixed lattice size L, we define these functions $F(n)$ and $W(n)$ by tabulating the values of F, equations (3.3) and (3.12), and W, equation (3.6), for all possible particle numbers N. For each N, we make use of the inversion procedure to find the external potential that gives rise to the homogeneous ground state density $n = N/L$. Our results are shown in figure 3.12. The solution corresponding to $U = 0$ (figure 3.12(a)) converges faster to the thermodynamic limit with increasing system size than the one corresponding to $U = 1$ (figure 3.12(b)), which suggests that the long-range Coulomb interaction amplifies the finite-size effects. According to equation (3.4), $F(n)$ must have a minimum at the particular ground state density n that is obtained without external potential and, as expected, this minimum is located at $n = 1$ for the noninteracting fermions and moves to $n < 1$ in the presence of interactions. In the following, we refer to the exact $F(n)$ of figure 3.12 as our exact LDA. Notice that the widely used LDA is not defined for F but for the slightly different quantity E_{xc}, see section 3.2.1.

3.4 Applications Beyond the LDA

In the previous section, we investigated different approximations of the universal functional $F(\vec{n})$ (3.3) as well as of the Hartree-exchange-correlation energy $W(\vec{n})$

(3.6). We proposed the ansatz (3.13) that contains local terms depending on local density values only, in the spirit of the LDA, and nonlocal terms depending on distant density values, extending beyond the LDA. We found that a formulation of equation (3.13) in terms of spline functions simplifies both its fit and its evaluation.

We have developed a comprehensive software library that determines the optimal form of our ansatz (3.13) for an arbitrary set of training densities. The next step is to study the performance of our ansatz in scenarios where DFT is applied.

In a reasonable scenario the target problem to be solved will be different from the set of training problems used to fit the ansatz. This scenario is the basis of the first part of this section, where we will study the performance of our algorithms in two settings, namely, a simpler setting of fermions without external potential and a more demanding setting of the dissociation of H_2. For these two problems, we will again focus on the question how well the exact $F(\vec{n})$, i.e., $E(\vec{n})$, gets reproduced by our approximation (3.13), and we will again assume that the exact ground state densities are known.

A typical DFT calculation goes one step further and produces also approximate ground state densities. In the second part of this section, we will therefore concentrate on the question how well exact densities are obtained by our approximation (3.13) when it is applied in standard self-consistent Kohn-Sham cycles. We will fit (3.13) with the various training sets that were used in the previous section and benchmark the resulting (3.13) with the simplest possible target densities.

The following analysis is also carried out with the Coulomb Hamiltonian (3.11). We will use the same parameters as before, i.e., system size $L = 21$, tunneling element $t = 1/2$, etc.

3.4.1 Energies

In this section, we analyze the performance of our ansatz (3.13) when it is applied to a set of target densities that is different from the set of training densities used in the fit. We will discuss two scenarios: a simpler one of fermions without external potential, in which we choose the training closer to the target set, and a more demanding one of H_2 dissociation, in which we compare several different training sets. As in the preceding analysis, we assume that all considered densities $\vec{n}$ are known exactly and we study the approximation error of our ansatz $G(\vec{n})$ for the universal functional $F(\vec{n})$ under this assumption. This assumption allows to assess the limitations of the specific form (3.13) for the approximation of the generally nonlocal $F(\vec{n})$ in the clearest possible way.

Fermions Without External Potential

The target densities for the first setting of our analysis are given by a system of possibly interacting fermions that are otherwise free in the sense that there is no external potential present. Since our lattice of $L = 21$ sites can be filled with total particle numbers $N = 1, 2, \ldots, 42$, our target set contains 42 different ground state densities and values of the universal functional. We want to train our ansatz with densities that are similar to the target densities but still differ from them. This is accomplished by a training set comprising exactly homogeneous densities and densities that vary slightly from being exactly homogeneous. We construct these ground state densities by means of the above explained inversion procedure, in which we aim at the exactly homogeneous densities $\vec{n}^{\mathrm{hom}}$ and, during the iteration of the inversion, include all those densities $\vec{n}$ in our training set for which the

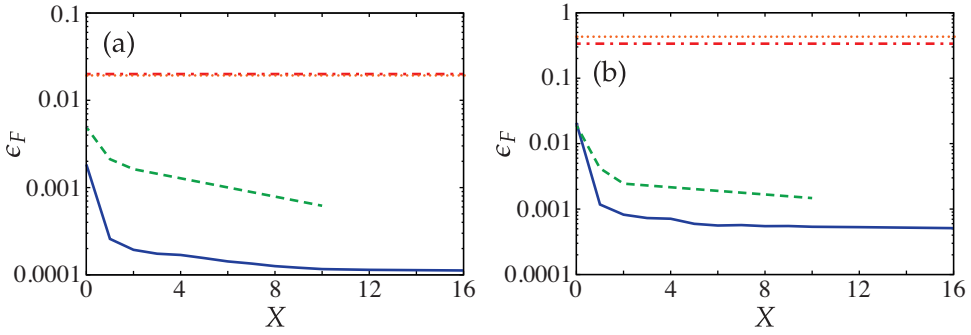


Figure 3.13: Mean relative error ϵ_F as a function of the maximal range X, for our ansatz (3.13) fitted with the homogeneous training set and applied to the target set of fermions without external potential (dashed), for $U = 0$ (a), and for $U = 1$ (b). We also show the error of our exact LDA (from figure 3.12) for this target set (dotted). Additionally, we show the error obtained for the training set, by our ansatz (solid), and by our exact LDA (dash-dotted). The fit of our ansatz was performed for a density grid of $R = 3$ equidistant points $\{0, 1, 2\}$.

distance $||\vec{n} - \vec{n}^{\text{hom}}||$ is smaller than a certain threshold.

Our results are shown in figure 3.13(a) for noninteracting fermions and in figure 3.13(b) for interacting fermions. We plot the relative error both of the exact LDA (figures 3.12(a) and 3.12(b)) and of our ansatz (3.13) when they are applied to the target set as well as to the training set. Our spline fit is performed for a density grid of $R = 3$ equidistant points $\{0, 1, 2\}$. The exact LDA achieves roughly the same error for both sets of densities, which reads ≈ 0.02 when $U = 0$ and ≈ 0.4 when $U = 1$. Our ansatz with $X = 0$ improves this error more or less by an order of magnitude, as it gives ≈ 0.002 for the training set and ≈ 0.005 for the target set when $U = 0$, and it gives ≈ 0.02 for both sets of densities when $U = 1$. In all cases, our ansatz with $X \geq 1$ shows a drastic improvement when $X = 1$ and then further improves our results with increasing X. While this is the expected behavior for the error of the training set, this behavior is not obvious for the target set. More precisely, when $U = 0$, the error improves from ≈ 0.0002 for $X = 1$ to ≈ 0.0001 for $X = 10$ for the training set, and it improves from ≈ 0.002 for $X = 1$ to ≈ 0.0006 for $X = 10$ for the target set. When $U = 1$, the error improves from ≈ 0.001 for $X = 1$ to ≈ 0.0005 for $X = 10$ for the training set, and it improves from ≈ 0.004 for $X = 1$ to ≈ 0.001 for $X = 10$ for the target set.

H_2 Dissociation

DFT calculations based on the LDA alone are not able to correctly describe the dissociation of the H_2 molecule, neither in three spatial dimensions [123] nor in one spatial dimension [124, 125]. Due to this fact, molecular dissociation has become a widely used testbed for novel DFT approaches that go beyond the LDA.

We consider here a simplified one-dimensional H_2 dissociation problem: two electrons in the external potential of two fixed nuclei separated by a distance $\mathcal{R}$, where each nucleus has the positive charge equivalent to one electron, we assume soft-Coulomb interaction between all particles, and we approximate continuous space via discretization. These several simplifications still keep the essential dissociation physics [124, 125] and, at the same time, make the dissociation problem tractable for our numerical algorithms. For discretization $\Delta = 1$, the corresponding

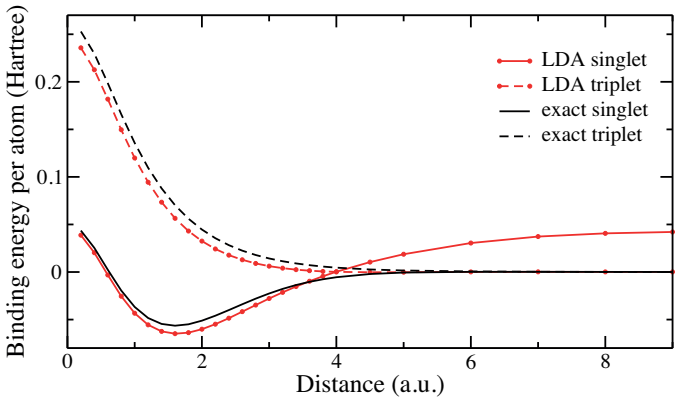


Figure 3.14: Dissociation curves taken from reference [125]. The considered system size is large enough and the spatial discretization is small enough such that these parameters do not affect the shown results anymore. Only the singlet curves are relevant here: The LDA solution (solid with dots) is compared to the one from exact diagonalization (solid).

model is described by the Hamiltonian (3.11) with $t = 1/2$, $U = 1$, and external potential

$$v_l = -\frac{1}{\sqrt{(l+\mathcal{R}/2)^2+1}} - \frac{1}{\sqrt{(l-\mathcal{R}/2)^2+1}} \quad .$$

To the ground state energy of this Hamiltonian, for a given separation $\mathcal{R}$, we still have to add the constant term $1/\sqrt{\mathcal{R}^2+1}$ from the nucleus-nucleus repulsion in order to obtain the total energy of the system.

This total energy, as a function of the separation $\mathcal{R}$, is shown in figure 3.14, after convergence was achieved with increasing system size and decreasing discretization width. Figure 3.14 compares the exact dissociation curve, from exact diagonalization, to the one computed via the one-dimensional LDA of reference [125]. For small separations $\mathcal{R}$, the LDA curve is close to the exact solution and also predicts the position of the energy minimum well (and thus can roughly predict the equilibrium bond length of the one-dimensional H_2 molecule). However, for large separations $\mathcal{R}$, the LDA curve departs from the exact solution (and features a finite slope, indicating an attractive force between the two atoms, while no slope is present in the exact case, and thus the atoms are already free). Our simplified model of system size $L = 21$ and discretization $\Delta = 1$ has the exact dissociation curve shown in figure 3.15. The oscillations, visible for larger $\mathcal{R}$, are due to the finite size and discretization.

First, we train our ansatz (3.13) with the dissociation densities. When we now apply our ansatz to the dissociation problem, the mean relative error reads $\approx 10^{-4}$ for $X = 0$, $\approx 8 \cdot 10^{-8}$ for $X = 1$, and $\approx 6 \cdot 10^{-8}$ for $X = 2$. As we can gather from figure 3.15, all resulting dissociation curves lie on top of each other. The excellent agreement achieved with $X = 0$ even for large separations $\mathcal{R}$ is remarkable, given that exact LDA calculations have difficulties with large $\mathcal{R}$ as described previously. Nevertheless, the accuracy achieved in the energy value is not enough to conclude that the functional can be used in a more general setting. On the one hand, we do not find convergence of the functional form of the local and non-local terms with the chosen density grid. On the other hand, when the training set is different, we do not get such a precision.

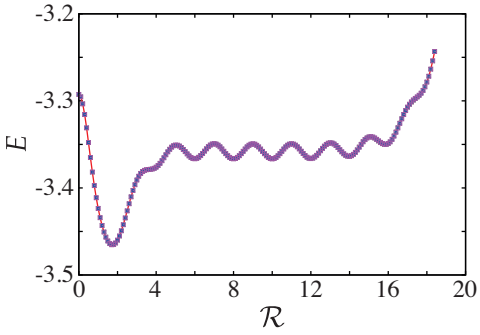


Figure 3.15: Exact dissociation curve for our model defined in the text (solid), and the dissociation curves obtained from our ansatz (3.13) with $X = 0$ (crosses), 1 (stars), and 2 (squares), when it is trained with the dissociation densities. All curves lie on top of each other. The mean relative errors ϵ_F obtained with our ansatz are $\epsilon_F \approx 10^{-4}$ for $X = 0$, $\epsilon_F \approx 8 \cdot 10^{-8}$ for $X = 1$, and $\epsilon_F \approx 6 \cdot 10^{-8}$ for $X = 2$.

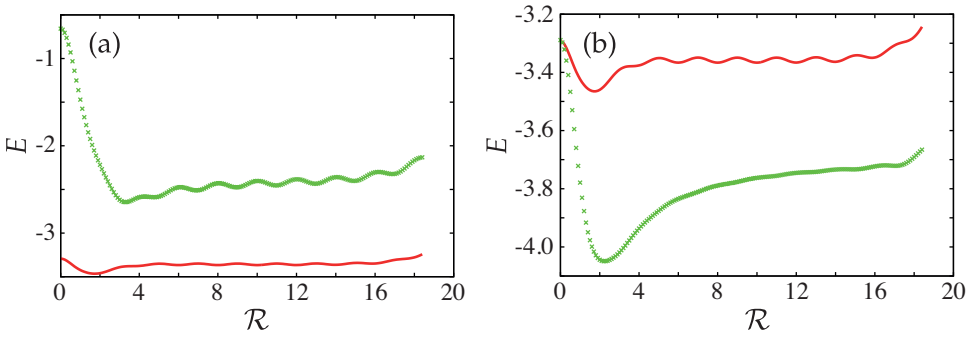


Figure 3.16: We compare the exact dissociation curve of the model defined in the text (solid) to the one obtained from our exact LDA (crosses in (a)) and to the one obtained from an average of several central local terms of our ansatz trained with the dissociation densities (crosses in (b)).

We have observed that this situation improves if we include more physical information in the fit. Indeed, because the different terms enter the functional (and hence the fit) in given combinations, they are not uniquely determined by the training sets, independent of how large these are. This is partly fixed in the case of local terms by imposing that each term acquires the correct known values at the extreme densities $n = 0$ and $n = 2$. Another way of removing this indeterminacy of the individual terms is to impose some invariance, e.g., by taking the average of several local terms. We can compare the result obtained in this way, in figure 3.16(b), with that of our exact LDA for this particular system, in figure 3.16(a). These results indicate that the training with exact energies and densities needs to be complemented with all the physical constraints we can impose in the ansatz.

3.4.2 Densities From Kohn-Sham

The complete preceding analysis dealt with the question how well our ansatz G of equation (3.13) approximates the universal functional F. After we had fitted G with a certain set of training densities, for a given target density $\vec{n}$ with known $F(\vec{n})$, we always compared the value $G(\vec{n})$ to the value $F(\vec{n})$, where both functionals were

evaluated with the same exact ground state density $\vec{n}$. In this section, we want to go one step further and use our functional G also for the computation of approximate ground state densities. This is actually the normal use of a DFT functional in a self-consistent Kohn-Sham cycle. In our case, we can test the usability of our ansatz in this fashion by obtaining an approximate density $\vec{n}'$ from G and we can then compare $G(\vec{n}')$ with $F(\vec{n})$. Clearly, in general, this error should be larger than the previously considered one because, now, additionally the density $\vec{n}'$ in $G(\vec{n}')$ has an error.

If the ground state of an interacting system in a certain external potential is sought, then the Kohn-Sham method finds the ground state of a noninteracting system in the Kohn-Sham potential (3.7) that produces the same density as the interacting ground state. Having obtained the desired density, the ground state energy of the interacting system can be calculated with the help of equation (3.8). This procedure requires the Hartree exchange-correlation energy (3.6), $W(\vec{n}) := F^{\mathrm{I}}(\vec{n}) - F^{\mathrm{NI}}(\vec{n})$, and its derivatives. In the following analysis, we will fit our approximation G^{I} of F^{I} independently from our approximation G^{NI} of F^{NI}, and we will then construct our approximation $\tilde{W}$ of W from the difference of these two functionals, i.e., $\tilde{W} := G^{\mathrm{I}} - G^{\mathrm{NI}}$. Thanks to the formulation of our ansatz (3.13) for G^{I} and for G^{NI} in terms of spline interpolations, the construction of $\tilde{W}$ can proceed via subtraction of the individual spline functions $g^{\mathrm{I}}_{x,l}$ and $g^{\mathrm{NI}}_{x,l}$ for all possible index pairs $\{x, l\}$ one after another. In this way, for each index pair $\{x, l\}$, the function $w_{x,l} := g^{\mathrm{I}}_{x,l} - g^{\mathrm{NI}}_{x,l}$ has the same spline form as $g_{x,l}$, and, thus, our complete approximation $\tilde{W}$ takes on the same form of our ansatz (3.13), i.e., $\tilde{W}(\vec{n}) = \sum_{x,l} w_{x,l}(n_l, n_{l+x})$, in terms of the same kind of spline interpolations as G. Due to the spline form, the computation of derivatives is straightforward.

We will distinguish two different tests for the computation of the Kohn-Sham density. Our first test consists in computing the Kohn-Sham potential produced by our ansatz for the exact ground state density. The corresponding non-interacting problem can then be solved, and we evaluate the error in the density. Our second test regards the more realistic scenario, in which the exact ground state density is not known. Hence we also apply our functional to the self-consistent solution of equations (3.7) and (3.5), as in standard Kohn-Sham calculations.

For each of the two tests, we will consider the two training sets from the second part of section 3.3.2. The first training set contains all training densities corresponding to scenarios kink, barrier, and well, for a total number of particles $N = 2$. And the second training set comprises all training densities corresponding to scenario kink for total particle numbers $N = 1, 2, \ldots, 30$. The resulting approximation $\tilde{W}$ will then be applied to the computation of ground state densities for $N = 2, 5$, and 10 interacting fermions, for simplicity, without any external potential.

Direct Calculation of the Kohn-Sham Density

If we know the exact ground state density for a given interacting problem, we can compute the corresponding Kohn-Sham potential (3.7) via our approximate $\tilde{W}$ directly. The noninteracting ground state of the Kohn-Sham Hamiltonian (3.5) with that potential would have the same density as the considered interacting system if $\tilde{W}$ would be exact.

We fit our ansatz with the $N = 2$ training densities from scenarios kink, barrier, and well, and apply it to a $N = 2$ target density of interacting fermions without external potential. Figure 3.17 shows the exact target density as well as the non-interacting ground state densities of the Kohn-Sham potentials obtained from our

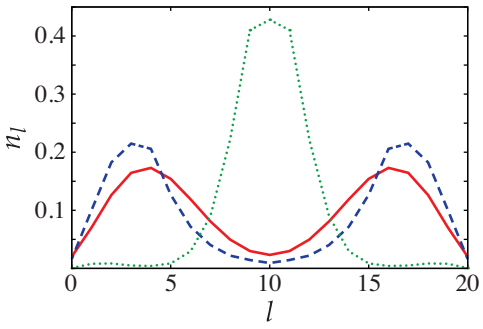


Figure 3.17: Exact density (solid) and corresponding Kohn-Sham densities from our ansatz with $X = 0$ (dashed) and $X = 1$ (dotted), for a a density grid of $R = 3$ values $\{0, 1, 2\}$.

ansatz, for $R = 3$, $X = 0$ and $X = 1$. While the $X = 0$ solution is close to the target density, the $X = 1$ solution is quite different. With increasing R or X, we could not observe a systematic improvement of our Kohn-Sham density. Such a systematic improvement does also not appear for the other target densities of $N = 5$ and $N = 10$ fermions. We obtain the same conclusions when our ansatz is fitted with the second training set.

Self-Consistent Calculation of the Kohn-Sham Density

We now assume that the exact interacting ground state density is not known and we solve equations (3.7) and (3.5) via our approximate $\tilde{W}$ self-consistently. For that, we will always perform the following iteration. We start from a homogeneous density $\vec{n}(0) := (n, n, \ldots, n)^t$ where $n := N/L$ is given by the considered total particle number N and system size L. That density $\vec{n}(0)$ enters equation (3.7) with our approximation $\tilde{W}$ and gives the first Kohn-Sham potential $\vec{v}^{\mathrm{KS}}(1)$. This potential $\vec{v}^{\mathrm{KS}}(1)$ enters the Kohn-Sham Hamiltonian (3.5) whose ground state gives the first Kohn-Sham density $\vec{n}(1)$. The density $\vec{n}(1)$ gives the second Kohn-Sham potential $\vec{v}^{\mathrm{KS}}(2)$ via equation (3.7), which gives the second Kohn-Sham density $\vec{n}(2)$ via the ground state of Hamiltonian (3.5), which gives the third Kohn-Sham potential $\vec{v}^{\mathrm{KS}}(3)$ via equation (3.7), and so on. Convergence of this scheme can be defined when the distance $d(i) := ||\vec{n}(i) - \vec{n}(i-1)||$ is below a small enough threshold. In general, for an approximate W, convergence is not guaranteed[3] and, in order to improve convergence, several density mixing algorithms have been proposed, e.g., in references [127–130], of which we will implement the simplest: In each iteration i, we modify our previous solution $\vec{n}(i-1)$ by adding the Kohn-Sham density $\vec{n}^{\mathrm{KS}}(i)$ with a weight α, i.e., $\vec{n}(i) := (1-\alpha)\vec{n}(i-1) + \alpha\vec{n}^{\mathrm{KS}}(i)$. We will run all our calculations both with $\alpha = 1$ and with $\alpha = 0.1$, and we always present our results for the value of α corresponding to the better final density.

For a target density of $N = 2$ interacting fermions without external potential, when the training set contains our training densities with total particle number $N = 2$ from scenarios kink, barrier, and well, a relatively good approximation of the true ground state density is achieved with $X = 0$ and $R = 3$, shown in figure 3.18(a). That approximation, however, does not improve when we increase X to $X = 1$, as we can gather from figure 3.18(b). As in the preceding section, we do

[3] When the exact functional is known, guaranteed convergence of the Kohn-Sham method can be proven under certain circumstances [126].

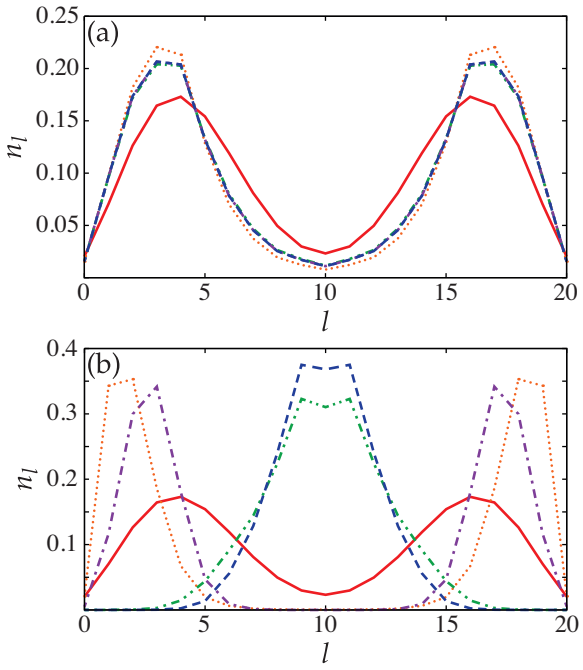


Figure 3.18: Target density (solid) and Kohn-Sham densities from our ansatz with $X = 0$ (a) and $X = 1$ (b) during the self-consistency cycle after iteration 1 (dotted), 2 (dash-doubledotted), 5 (dash-dotted), and 10 (dashed). The self-consistency cycle was run with $\alpha = 1$ and started from a homogeneous initial density, with $n_l = N/L$ on all lattice sites l. We used a density grid of $R = 3$ equidistant values $\{0, 1, 2\}$.

not observe a systematic improvement with increasing X or R. This holds also for the other target densities as well as for the second training set. Interestingly, we observed that our exact LDA typically produced good approximations to the true ground state densities, and a particularly good match can be seen in figure 3.19.

3.5 Conclusions and Outlook

In summary, in the first half of this chapter, we compared approximations of the universal functional F by several ansatzes, that were inspired by TNS concepts and had a parameter to systematically increase their nonlocality, modifying the ansatz from a local form in the spirit of the LDA to a nonlocal form going beyond the LDA. The two key questions in the comparison of the various ansatzes were, firstly, how the ansatz is optimally determined for a set of training densities and, secondly, how the ansatz is applied to a set of target densities. Having realistic DFT problems as a goal and, therefore, requiring an optimal fit of our ansatz for training densities on large lattices, we concluded with a specific ansatz for the approximation of the universal functional. Our ansatz consists of local terms, that, in the spirit of the LDA, each depend only on the density value of a single lattice site, and of nonlocal terms, that, extending beyond the LDA, each depend on the density values of two lattice sites separated by a distance x. We define our ansatz by a maximal range X such that it contains all terms with $x \leq X$, and then the idea underlying our ansatz is that its nonlocality can be systematically increased by increasing X, hopefully, leading to a systematically improved approximation of

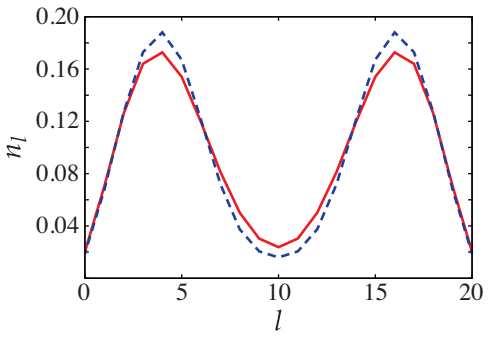


Figure 3.19: Target density (solid) and Kohn-Sham density from our exact LDA (dashed) after convergence of the self-consistency cycle. As in figure 3.18, the cycle was initialized with a homogeneous density and then run with $\alpha = 1$. The density converged within 10 iterations.

F. The necessity of being able to apply our ansatz to arbitrary target densities led to a continuous formulation in terms of spline functions.

We then extensively analyzed the performance of our ansatz as a function of its parameters for different training densities. In this analysis, for each set of parameters, we always determined the optimal form of our ansatz for certain training densities and then calculated the mean relative error of F for the same densities, i.e., the training densities were always the same as the target densities. Because our optimal fit minimizes a quantity very similar to this mean relative error of F, the errors discussed in the first half of this chapter are close to the lowest possible that can be achieved with our ansatz for the considered parameters and training densities.

In particular, this implies that one can only expect larger errors when the target densities differ from the training densities. We convinced ourselves of the correctness of this expectation in the first part of the second half of this chapter, where we considered target densities different from the training densities. When the target densities were chosen close to the training densities, qualitatively the target errors behaved like the training errors, although quantitatively the target errors were larger than the training errors. When the target densities were chosen different from the training densities, as in our considered H_2 dissociation problem, we were unable to get good target results.

So far, in our analysis, we always evaluated our ansatz for the exact ground state densities, i.e., we did not yet use our ansatz to compute approximate ground state densities, as done in typical DFT calculations. We want to emphasize that we can only expect larger errors when not only our ansatz is an approximation of F but also the density with which it is evaluated is only approximate. This situation was analyzed in the second part of the second half of this chapter, where we considered the approximate computation of ground state densities with our ansatz. We found that, often our ansatz with $X = 0$ and with few spline grid points gave a good ground state density approximation, that, however, could not be improved systematically by increasing the number of spline grid points or the maximal range X.

This chapter contains a comprehensive performance analysis of our ansatz, when it is used as an approximation for the universal functional F, and when it is optimally determined by minimization of its error in F. We conclude this chapter with a discussion of alternative ideas as well as of possible refinements of our

algorithms that might improve our results.

- The nonlocal terms in our proposed ansatz depend on the density of two lattice sites separated at most by a distance X. We saw that our achieved approximation error saturates already for relatively small values of X. A further improvement should therefore be gained by including nonlocal terms depending on the density of more than two separated lattice sites. The optimal fit as well as the evaluation of such an ansatz is quickly implemented when its formulation is done in terms of general polynomials (although it is also feasible, but requires more implementation effort, for spline functions).

- Our exact LDA, constructed by tabulating the values of F for exactly homogeneous ground state densities, performed remarkably well for many different target densities, regarding both its approximation error in F and its resulting Kohn-Sham densities. A generalization of this construction by tabulating the values of F for exactly two-periodic ground state densities gives a function that depends on two density values. Then, three-periodic ground state densities give an exact function of three density values, and so on. Because the exactly homogeneous densities are contained in the set of higher-periodic ground state densities, the generalized constructions include our exact LDA. Therefore, it would be interesting to find out if these functions improve the results obtained previously with our exact LDA.

- In general, for an arbitrary training set, the fit of our ansatz, via minimization of its error in the values of F, leaves a lot of freedom for the resulting functional form of our local and nonlocal terms. This can be understood by looking at a training set of exactly homogeneous densities, i.e., having the same density value $n = N/L$ on every lattice site. The corresponding values of the universal functional define a function $F(n)$ and a minimization of the error of our $X = 0$ ansatz $G_{X=0}(\vec{n}) = \sum_l g_{x=0,l}(n_l)$ with respect to $F(n)$ clearly gives a vanishing approximation error. However, for each value of n, the local terms $g_{x=0,l}(n)$ on the different lattice sites l can take on arbitrary values as long as their sum adds up correctly to $F(n) = \sum_l g_{x=0,l}(n)$. We also checked that, when our ansatz was trained with many more homogeneous densities, namely, exactly homogeneous and slight deviations thereof, the resulting local and nonlocal terms were still different on all the lattice sites. As a consequence, it makes sense to average all terms with the same x over all lattice sites and apply that average to the considered target densities.

- Related to the previous point are the difficulties we encountered when our ansatz was used for the computation of approximate ground state densities. For that, we employed the Kohn-Sham method, in which the desired density follows from the noninteracting ground state in a Kohn-Sham potential, which we obtained from derivatives of our ansatz. However, since our fit leaves a lot of freedom to the functional form of our ansatz, the same is true for its derivatives. One way out might be to impose our ansatz as an approximation for the Kohn-Sham potentials and determine its optimal form for these potentials instead of for the universal functional. In this case, we would have to enforce the constraint $\partial v_l/\partial n_m = \partial v_m/\partial n_l$ (equivalent to $\partial W/(\partial n_l \partial n_m) = \partial W/(\partial n_m \partial n_l)$) in our fit. Additionally, the energy, see equation (3.8), would have to be calculated via numerical integration of the approximate potentials, leading to an approximate W, or via an independent fit of our ansatz as approximation of $W = F^{\mathrm{I}} - F^{\mathrm{NI}}$ as before.

- Alternative to the previous point, we could keep our ansatz as an approximation of F and additionally include derivative information of F in our fit, since we know our training potentials and thus the derivatives of F for the corresponding training densities. Then the constraint $\partial v_l/\partial n_m = \partial v_m/\partial n_l$ would be naturally fulfilled as well as the energy would be directly calculated from our approximate F as before.

- A difficulty, in this and the preceding point, might arise from the fact that each training potential is well-defined only up to a constant offset, i.e., up to an arbitrary value of the chemical potential. This means that arbitrarily similar training densities can have arbitrarily different training potentials and, thus, derivatives entering the fit of our ansatz. Therefore, on at least one lattice site, the value of the external potential has to be fixed a priori, and then kept for all training densities. However, a fixed value of the external potential implies a fixed value of the derivative on that lattice site, independent of the density, which might then lead to a trivial and inconsistent local term on that lattice site in the fit of our ansatz. The possible inconsistency of such a local term is most easily seen for a training set of exactly homogeneous densities in a system with periodic boundary conditions. Because each training density could be computed in the same zero external potential, it would contribute the same zero derivative to our fit, for all density values, which would have to give constant local terms resulting from our fit. But we can tabulate the exact values of F for these densities in $F(n)$ and then know the optimal local term exactly, and that term in general depends on the density. It might make sense to tackle this problem by specifying lattice sites, e.g., the ones on the boundaries, where both the external potential and the density has to be zero for all training densities. Then, the optimal local terms on these boundary lattice sites can be constantly zero, consistent with all training densities.

- Additional to derivative information, we can include known values of F as constraints in our fit, e.g., its value for no density at all and its value for maximal density.

- When we directly approximate F by our ansatz, we could also replace the Kohn-Sham procedure by an orbital-free method. In such an orbital-free method, we would minimize our energy functional, resulting from our approximation of F, directly, and this minimization might be more efficient than the self-consistent Kohn-Sham calculations. A particularly simple orbital-free algorithm arises if we formulate our ansatz G in terms of low-order polynomials in such a way that the local terms depend at most quadratically and the nonlocal terms depend at most linearly on the density. In this case, the derivative of G with respect to the density on a specific lattice site is a linear function of the density and, thus, the energy minimization becomes an easily solvable linear problem. For higher-order polynomials as well as for our spline interpolations, the energy minimization becomes a nonlinear problem in general, that we might be able to tackle efficiently with the help of Conjugate Gradient or Newton methods.

- It might make sense to impose our ansatz as an approximation of the exchange-correlation energy $E_{\mathrm{xc}} = W - E_{\mathrm{H}} = F^{\mathrm{I}} - F^{\mathrm{NI}} - E_{\mathrm{H}}$, such that both the efficiently computable F^{NI} (via noninteracting inversion) and the known Hartree term E_{H} (meanfield-like contribution) are subtracted before the fit. In fact, the standard LDA of DFT approximates E_{xc}. The Hartree contribution

E_{H} is defined as $U \sum_l \langle \hat{n}_{l,\uparrow} \rangle \langle \hat{n}_{l,\downarrow} \rangle$ for the Hubbard Hamiltonian (3.10) and as $U \sum_{l \leq m} \langle \hat{n}_l \rangle \langle \hat{n}_m \rangle / \sqrt{(l-m)^2 + 1}$ for the soft-Coulomb Hamiltonian (3.11). A first attempt of fitting $F^{\mathrm{I}} - E_{\mathrm{H}}$ instead of F^{I} gave promising results for the dissociation curve.

Part II

Quantum Simulations of Quantum Many-Body Systems

Chapter 4

Adiabatic Preparation of a Heisenberg Antiferromagnet Using an Optical Superlattice

Our MPS and PEPS algorithms allow us to simulate and investigate large quantum many-body systems. One scenario particularly apt for their application is that of ultracold atoms in optical lattices. In this context, TNS techniques can play a significant role for the design and validation of new quantum simulation algorithms. This chapter presents an example of such a successful application. We analyze the possibility to prepare a Heisenberg antiferromagnet with cold fermions in optical lattices, starting from a band insulator and adiabatically changing the lattice potential. The numerical simulation of the dynamics in 1D allows us to identify the conditions for success, and to study the influence that the presence of holes in the initial state may have on the protocol. We also extend our results to two-dimensional systems. This chapter is published in reference [52].

4.1 Introduction

Ultracold atoms trapped in optical lattices offer a unique possibility to experimentally explore strongly correlated states of quantum matter. Currently, one of the main experimental challenges in this field is the preparation of a Heisenberg antiferromagnet (AFM), which represents the necessary next experimental step towards a true quantum simulator of the fermionic Hubbard model [32].

Although the creation of a fermionic Mott insulator (MI) has recently been reported [36,37], the realization of antiferromagnetic order requires temperature and entropy significantly lower than presently achieved [131, 132], despite many existing proposals for direct cooling within the lattice [133]. An alternative to the direct generation is to use an adiabatic protocol [134, 135]. In such a scheme, it is desirable to tune interactions initially to give a ground state with very low entropy. Then, they are changed slowly, until the Heisenberg Hamiltonian is realized at the end. If the process is adiabatic, the entropy will stay low and the final state will be the desired AFM. The following questions immediately arise: What are the conditions to achieve adiabaticity? What occurs if these conditions are too restrictive and cannot be met? And, how will the protocol be affected by a finite temperature

and the presence of a harmonic trap?

In this chapter, we propose a specific adiabatic scheme and analyze these issues. Our adiabatic protocol is the first to attain an AFM with ultracold fermions within feasible timescales, even in the presence of experimental imperfections. Additionally, we show that it is possible to realize antiferromagnetic order on a part of the sample in a shorter time than required for the whole system. Finally, we simulate the dynamics of holes to demonstrate their destructive effect on the AFM and devise a strategy to control them.

The initial ground state of our protocol is a band insulator (BI), which is transformed first to an array of decoupled singlets and finally to the AFM, by adiabatically changing the depth of two superimposed optical lattices. A BI is easier to prepare with low entropy than a MI for two reasons. On the one hand, its energy gap is given by the band gap, which is much larger than the interaction energy (MI gap) and favors a redistribution of the entropy towards the surrounding metallic shell [136]. On the other hand, the preparation can be done using weakly or non-interacting atoms, thereby avoiding the long timescales associated with mass and entropy transport at higher interactions [137, 138].

For the one-dimensional case, we simulate the fermionic $t - J$ model with Matrix Product States (MPS).[1] We first identify the adiabatic conditions that allow the preparation of the antiferromagnetic state in an ideal case with no defects. Second, we study how these conditions are relaxed if one imposes that antiferromagnetic order is only obtained on a subset of fermions around the center of the sample. We observe that, when restricted to a middle sublattice, adiabaticity is determined by an effective gap related to this sublattice and not by the gap of the total system. Third, we include the presence of holes in the initial state, expected to occur in real experiments due to the finite temperature. The large initial energy of the holes can in principle destroy the AFM as they delocalize inside the sample. We find that, if the holes are initially located at the outer part of the sample, as expected in an experiment, a tradeoff can be reached between the degree of adiabaticity of the process and the distance the holes travel inside the chain, so that the antiferromagnetic order is still produced in the center. Moreover, we show that a harmonic trap can prevent the destructive effect of holes by confining them to the outside of the sample. Finally, via Projected Entangled Pair States (PEPS) [11, 61], we complement our analysis with a simulation of the two-dimensional $t - J$ model of hardcore bosons with antiferromagnetic interaction. This setting is easier to investigate numerically than the corresponding 2D fermionic system and provides evidence that the physics studied in the one-dimensional case can be extrapolated to understand the conditions of an equivalent scheme in 2D.

4.1.1 Reader's Guide

This chapter is organized as follows. We specify the Hamiltonian, our adiabatic protocol, and the considered observables in section 4.2. The one-dimensional analysis is carried out in section 4.3, first for an idealized scenario without holes, then with holes, and then additionally with the harmonic trap. Section 4.4 contains our results for two dimensions. Finally, a brief discussion of our conclusions is presented in section 4.5. The appendix in section 4.6 provides supplemental material.

[1] Instead of the initial method [8], we use the scheme introduced in reference [12].

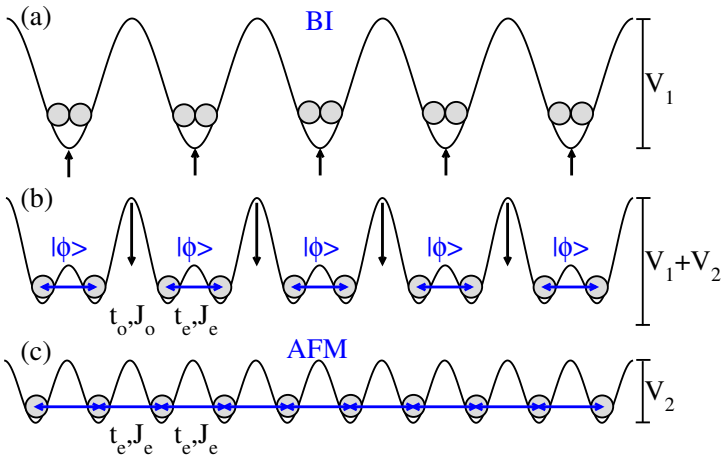


Figure 4.1: The proposed adiabatic protocol. First, a BI in a lattice with depth V_1 (a) is transformed to a product of singlets $|\phi\rangle$ (b) by slowly switching on a second lattice with depth V_2 and half the original wave length. Then, by lowering the barrier $V_1 \to 0$, the system turns into an AFM (c).

4.2 Model

In the following, unless stated otherwise, we will focus on the one-dimensional case and consider a two component Fermi gas in an optical lattice. The physical setting consists of two adiabatic stages, depicted in figure 4.1. The first transition has already been realized experimentally [139–141] and can be straightforwardly described, so that we can focus on the second one and take figure 4.1(b) as the initial state for our theoretical study. In this situation, the system is governed by a one-band $t-J$ model. This model emerges in the limit of strong interactions from the Hubbard model, which describes ultracold atoms in optical lattices [34]. We consider a bipartite $t-J$ Hamiltonian with different couplings for even and odd links, $\hat{H} = \hat{H}_e + \hat{H}_o$, where

$$\begin{aligned} \hat{H}_\ell &= -t_\ell \sum_{k\in\ell,\sigma=\uparrow,\downarrow} (c^\dagger_{k,\sigma} c_{k+1,\sigma} + \text{H.c.}) \\ &\quad + J_\ell \sum_{k\in\ell} \left(\vec{S}_k \cdot \vec{S}_{k+1} - \frac{\hat{n}_k \hat{n}_{k+1}}{4} \right), \quad \ell = e, o. \end{aligned} \tag{4.1}$$

The superexchange interaction J_ℓ and the tunneling parameter t_ℓ are related through the on-site interaction U as $J_\ell = 4t_\ell^2/U$. We fix the couplings on the even links, $t_e = t$ and $J_e = J$, and choose a linear ramping of the superexchange interaction on the odd links, over total ramping time T so that $J_o(\tau) = J \cdot \tau/T$ and $t_o(\tau) = t \cdot \sqrt{\tau/T}$, for $0 \le \tau \le T$.[2] In the following, we set $J = 1$. A harmonic trap is included by adding a term $V_t \sum_k (k-k_0)^2 \hat{n}_k$ to equation (4.1).

The ramping time from figure 4.1(b) to 4.1(c) needs to be long enough such that the final state is close to the true AFM. The required time to ensure a certain degree of adiabaticity can be seen to scale as $T \propto 1/\Delta^2$ [142], where Δ is the minimum gap between the ground and the first excited state during the evolution. A closer look at the relevant energy levels reveals that in this adiabatic transition the gap decreases monotonically from J to the Heisenberg gap, which vanishes in the thermodynamic limit, and there is no phase transition occurring in between [143].

[2] Other ramping schemes $J_o(\tau) = J(\tau/T)^x$ for different values of x give qualitatively similar results.

The adiabaticity of the evolution and the final antiferromagnetic order can be probed by two experimentally accessible observables. The first one, the squared staggered magnetization, $M^2_{\text{stag}} = \frac{1}{N^2}\sum_{l,m=1}^{N}(-1)^{l+m}\langle \vec{S}_l \cdot \vec{S}_m \rangle$, is the antiferromagnetic order parameter and can be determined by noise correlations [144]. The second one is the double well singlet fraction, $P_0 = \frac{2}{N}\sum_{k\in e}\left(\frac{1}{4} - \langle \vec{S}_k \cdot \vec{S}_{k+1}\rangle\right)$. Since the initial state, figure 4.1(b), has a pure singlet in each double well, measuring $P_0 \neq 1$ at the end indicates a change in the state. The generation and detection of singlet and triplet dimers in double well lattices has been recently reported [141]. Whereas the squared staggered magnetization is experimentally detected over the whole sample and captures information on long-range correlations, the singlet fraction can be determined *in situ* and hence allows us to probe parts of the sample. For the Heisenberg antiferromagnetic chain in the thermodynamic limit, these observables take on the values $M^2_{\text{stag,TD}} = 0$ and $P_{0,\text{TD}} \approx 0.693$ [145], while in 2D $M^2_{\text{stag,TD}} \approx 0.0945$ and $P_{0,\text{TD}} \approx 0.585$ [146]. Notice that for the finite 1D systems considered in this work, M^2_{stag} does not vanish, but has a sizable value, comparable to the 2D thermodynamic limit, s. section 4.6.

4.3 One-Dimensional Case

4.3.1 Absence of Holes

Using the numerical simulation of the chain dynamics with MPS, we investigate the state at the end of the protocol for varying ramping time. To characterize the antiferromagnetic order independently of the system size, we define the relative quantities $m^2(T) := M^2_{\text{stag}}(T)/M^2_{\text{stag,AFM}}$, $p_0(T) := P_0(T)/P_{0,\text{AFM}}$, and $e_{\text{spin}}(T) := E_{\text{spin}}(T)/E_{\text{spin,AFM}}$, where the denominator is the expectation value of the observable in the true AFM for a given lattice, s. section 4.6. For the last quantity, E_{spin} is the expectation value of the spin term in the total Hamiltonian of equation (4.1).

Figure 4.2 shows our results for an ideal case with no holes in the initial state, with all relative quantities converging to 1, as expected, in the limit of large T [figures 4.2(a),(b)]. The ramping time necessary to reach a certain relative magnetization m^2 grows with the system size. If we study, given T, which is the largest system for which a fixed value m^2 can be achieved [figure 4.2(a) inset], we find $N^2 \propto T$, which is consistent with the adiabaticity condition for a gap closing like $\Delta \propto 1/N$ [147].

For very long chains, the required T might not be experimentally feasible. Remarkably, this does not exclude the preparation of antiferromagnetic order on large systems. We may evaluate the magnetization over a sublattice in the center of the sample. If, given T and N, we ask for the largest sublattice size L for which the magnetization reaches a fixed value [figure 4.2(c)], we find a scaling $T \propto L^2$, as governed by an effective local gap, and not by the gap of the total system. In contrast to m^2, the observables p_0 and e_{spin} do not depend on L [figure 4.2(d)]. This can be understood from the fact that p_0 and e_{spin} are determined by a two-site observable $\vec{S}_k \cdot \vec{S}_{k+1}$ averaged over a sublattice of length L, whereas m^2 is a true L-site observable, and thus effectively probes the adiabaticity on the sublattice. The experimental consequence is that with a large sample, high values of m^2 can be obtained in short ramping times T on small parts of the system, $L \propto \sqrt{T}$.

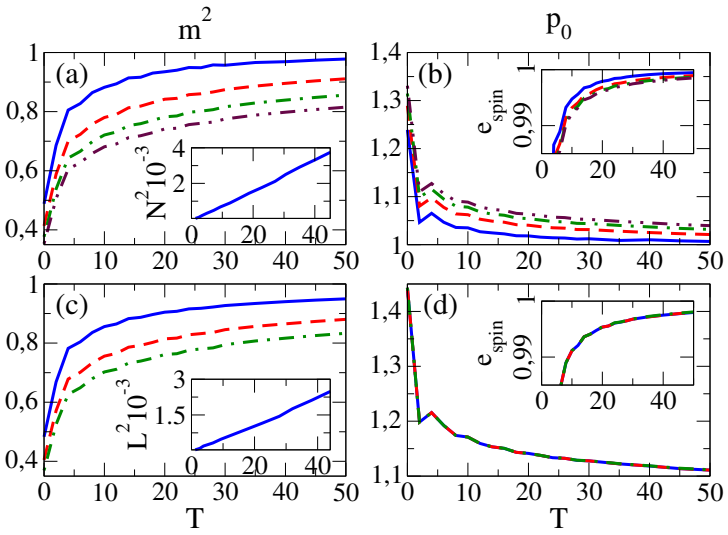


Figure 4.2: Absence of holes. (a), (b) m^2, p_0, and e_{spin} as functions of the ramping time T, for chain length $N = 22$ (solid blue lines), 42 (dashed red lines), 62 (dash-dotted green lines), and 82 (dash double-dotted brown lines). The inset of (a) shows the squared size N^2 of the longest chain reaching a fixed $m^2 = 0.85$ at ramping time T, and reveals the scaling $T \propto N^2$. (c), (d) Same quantities as above, evaluated on sublattices of length $L = 22$ (solid blue lines), 42 (dashed red lines), and 62 (dash-dotted green lines) for $N = 82$. Now, the inset of (c) shows the squared size L^2 of the largest sublattice reaching $m^2 = 0.85$ at ramping time T, and reveals the scaling $T \propto L^2$. All results were obtained with MPS of bond dimension $D = 60$ and Trotter step $\delta t = 0.02$ (s. section 4.6).

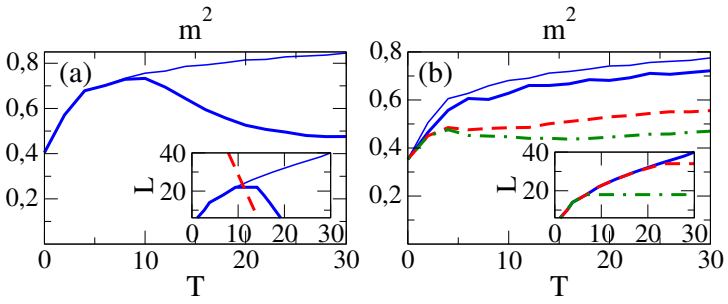


Figure 4.3: Effect of holes and harmonic trap. (a) m^2 as a function of the ramping time T, evaluated on a sublattice of length $L = 42$ for $N = 86$, without holes (thin solid line), and with initially 2 holes at each boundary (thick solid line), and $t = 2$. The inset shows the largest sublattice size L reaching $m^2 = 0.85$ at ramping time T, for the two cases of the main plot, and the size of the hole-free region (dashed red line). (b) m^2 evaluated on a middle sublattice of length $L = 82$ for $N = 102$, with no holes (thin solid line), and with initially 10 holes at each boundary and a harmonic trap of strength $V_t = 0.004$ (dash-dotted green line), 0.006 (dashed red line), and 0.02 (solid blue line), and $t = 3$. For $V_t = 0.2$ (not shown) the exact behavior of the ideal case is recovered. Again, the inset shows the largest sublattice size L reaching $m^2 = 0.85$ at ramping time T, for the cases of the main plot. All results were obtained with MPS of bond dimension $D = 60$ and Trotter step $\delta t = 0.02$ (s. section 4.6).

4.3.2 Effect of Holes

In a real experiment, the finite temperature causes the sample to be in a thermal mixture. As a consequence, localized holes will be present in the initial state, figure 4.1(b). Since the double wells are decoupled, the wave function of a hole will be an equal superposition of being in the left and in the right side of a single double well. Our simulation reveals that holes have a highly destructive effect on magnetic order. As seen in figure 4.3(a), a few holes initially located on the boundary of the sample are enough to cause a dramatic reduction of the final staggered magnetization.

We observe that the dynamics of holes can be qualitatively well understood using a simplified picture, in which the spreading of an initially localized hole, propagating in an antiferromagnetic background, is modeled by a free particle. This picture is accurate in the limit $t \gg J$, when the spin term in equation (4.1) is negligible. We checked that it is also valid in the whole experimentally reasonable parameter regime by comparing the behavior of single holes to that of free particles with the same initial wave function: for the relevant range of times and tunneling values, $2 \leq t \leq 8$, the hole spreads like a free particle with a maximal velocity $v = 2t$. In the case of small $t = 2$, a hole with higher initial energy causes a higher spin excitation, while for a large $t = 8$, the hole excites the spin background by $\Delta E_{\text{spin}} \approx 0.5$ in the beginning, independent of its initial kinetic energy. This can be understood by assuming a simple classical Néel background, where the delocalization of a hole, initially positioned at a boundary, breaks up exactly one antiferromagnetic bond. This assumption should become a good approximation for the hole dynamics in the regime $t \gg J$, where the timescale of the delocalization is much faster than the timescale of the reacting spin background. In all cases, the squared staggered magnetization is reduced substantially during propagation of

the hole until it reaches a minimum after the hole has travelled once over the whole sample. We found that the magnetization reduction depends only weakly on the initial kinetic energy, and on t, but it depends strongly on the number of holes in the sample.

The simplified free particle picture allows us to interpret the results from figure 4.3(a). In particular, the strong magnetization drop around $T \approx 8$ indicates the arrival of holes at the middle sublattice. We can roughly predict this arrival time from the spreading of a free particle wave function, which, after time T, will have covered $\frac{4}{3}t \cdot T$ sites, taking the ramping of the lattice into account. We observe that the magnetization for a given sublattice L behaves like in figure 4.2(c) only for short ramping times, while the region is hole-free, until the holes reach the sublattice [figure 4.3(a) inset].

4.3.3 Harmonic Trap

The simplified picture described above points out that the negative effect of holes can be controlled by the presence of a trap. An external potential changes sign for a hole and effectively turns into an inverse trap, capable of confining the holes to the outer parts of the chain. The trap strength should be chosen as large as possible without exceeding the on-site interaction U, what would destroy the MI figure 4.1(c).

The results of the dynamics within the harmonic trap are shown in figure 4.3(b). From energy considerations, a hole delocalizes at most by $\pm 2t$, s. section 4.6. As the trap strength is increased, the holes get more localized on the outside. As a consequence, the magnetization of the total sample increases, and the behavior of the hole-free case is recovered.

4.4 Two-Dimensional Case

For the 2D case, the adiabatic setting consists of an array of initially decoupled chains like figure 4.1(b), connected by a transverse lattice with the time-dependent couplings J_o and t_o. Different to the 1D case, this system exhibits a phase transition in the thermodynamic limit at $J_o \approx 0.5$ [143].

Although the numerical simulation of the 2D setting is much more demanding than the one for chains, we can relatively easily obtain results for hardcore bosons on lattices of moderate size. In the absence of holes, the bosonic and fermionic $t - J$ models are equivalent, and our simulations serve to test our protocol on a 2D system. Similarities between both models under inclusion of holes are a subject of current research [148], but our simulations can still provide a qualitative indication of the controlling effect of the trap. It is worth noticing that the AFM can also be realized with ultracold bosons.[3]

In the ideal case of no holes, we observe [figure 4.4(a)] that, whereas the energy converges quickly, the magnetization does not, and thus we cannot claim convergence to the true AFM within the numerically accessible ramping times studied here. Remarkably enough, we find that for a 10×10 lattice, a significant magnetization value m^2 is obtained in times of the same order of magnitude as for a chain of length 10, suggesting that the generation of antiferromagnetic order on much larger 2D lattices will be experimentally possible within reasonable timescales.

[3]The hardcore bosonic system has a ferromagnetic interaction, but the antiferromagnet can be obtained as the highest excited state [135, 149].

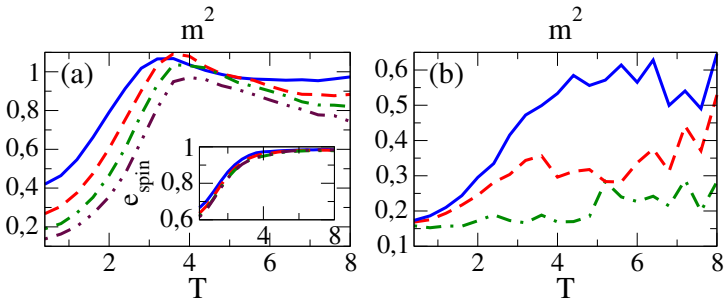


Figure 4.4: Performance in 2D. (a) m^2 and e_{spin} as functions of the ramping time T, for systems of size $N = 4 \times 4$ (solid blue line), 6×6 (dashed red line), 8×8 (dash-dotted green line), and 10×10 (dash double-dotted brown line). (b) m^2 for $N = 8 \times 8$, with initially 4 holes at the boundary and no harmonic trap (dash-dotted green line), and with a trap of strength $V_t = 0.25$ (dashed red line), and 2.5 (solid blue line), and $t = 2.5$. The results were obtained with PEPS of bond dimension $D = 4$ (a) and $D = 2$ (b) and Trotter step $\delta t = 0.03$ (s. section 4.6).

Upon hole injection, a similarly dramatic magnetization reduction is observed [figure 4.4(b)] which can be controlled by the presence of a harmonic trap, as in the 1D case.

4.5 Conclusions

We have proposed and analyzed an adiabatic protocol, suitable to prepare an antiferromagnetically ordered state in an optical lattice, even from an initial state containing defects. The timescales for the finite systems studied in this work lie well within the range of current experiments. Furthermore, we have observed that antiferromagnetic order can be produced in a sublattice in times governed only by its size.

This scheme offers several advantages over other proposals. First, starting from a BI simplifies the preparation of a sufficiently low entropy initial state. Additionally, the initial ground state in figure 4.1(b) already features the final $\mathrm{SU}(2)$ symmetry of the AFM, so that the number of excited states to which the evolution couples is minimum, as compared with an alternative proposal [135] with only $\mathrm{U}(1)$ initial symmetry. Moreover, since hole doping is experimentally feasible, the same procedure can possibly be used to prepare the ground state with varying hole densities. This would open the door to the experimental exploration of open questions in condensed matter theory, ultimately the existence of d-wave superconductivity in the $t - J$ model.

4.6 Appendix

In this appendix, we describe the numerical method used in our article, provide absolute values of experimental observables and discuss the numerical errors.

4.6.1 Numerical Method

We use MPS in one dimension and PEPS in two dimensions for the computation of ground state approximations and the simulation of time evolution. For the defini-

tion of MPS and PEPS, we refer to section 1.2. From the many possible algorithms (see chapters 1 and 2 for examples in the context of PEPS), we chose the variational search (explained, e.g., in references [13, 41]) to compute ground state approximations and the algorithm based on references [8, 12] to simulate time evolution.

In the variational search for ground states, the numerical error comes from limiting the bond dimension to a certain maximum value. Comparing the results with those obtained after running the search with a larger D gives an estimation of the magnitude of this truncation error. Typically, the algorithm is run repeatedly with increasing bond dimension until convergence is achieved within the desired numerical precision.

The simulation of time evolution of MPS (or PEPS) is based on a Suzuki-Trotter decomposition of the evolution operator [8]. This introduces another source of numerical error, in addition to the truncation of the bond dimension. The magnitude of this Trotter error can be controlled by decreasing the size of the time step, δt, or using a higher order decomposition of the exponential. In particular, we use a second order Trotter decomposition in the case of one-dimensional simulations, while a first order decomposition is used in 2D. For each time step, the evolution operator for δt is applied and an optimal MPS or PEPS approximation to the evolved state is found as described in [12]. The magnitude of the Trotter error is controlled by comparing results for various values of δt, and the truncation error, as in the ground state search, is estimated from the comparison of results for different D.

4.6.2 Absolute Values of Experimental Observables

The relative quantities presented in the main text result from normalizing the computed expectation values at the end of the evolution to the corresponding values in the true AFM ground state. For completeness, we provide in this section the actual absolute values obtained for each observable, as well as the reference AFM values, and discuss the convergence of the numerical results.

As explained in the main text, we model our system by a bipartite $t-J$ Hamiltonian, that, in 1D, reads:

$$\begin{aligned}\hat{H} \;=\; & -t_e \sum_{k\in e,\,\sigma} (c^\dagger_{k,\sigma} c_{k+1,\sigma} + \text{H.c.}) + J_e \sum_{k\in e} \left(\vec{S}_k \cdot \vec{S}_{k+1} - \frac{\hat{n}_k \hat{n}_{k+1}}{4}\right) \\ & -t_o \sum_{k\in o,\,\sigma} (c^\dagger_{k,\sigma} c_{k+1,\sigma} + \text{H.c.}) + J_o \sum_{k\in o} \left(\vec{S}_k \cdot \vec{S}_{k+1} - \frac{\hat{n}_k \hat{n}_{k+1}}{4}\right) \quad ,\end{aligned}$$

where the subscripts e and o denote even and odd sites and double occupancy of sites is forbidden, as implicitly assumed for the $t-J$ model. The couplings on even links are constant, $t_e = t$ and $J_e = J$, while the time-dependent odd couplings are increased from 0 to their final values, t and J, during a total ramping time T, according to $t_o(\tau) = t \cdot \sqrt{\tau/T}$ and $J_o(\tau) = J \cdot \tau/T$. We set $J = 1$. The two-dimensional system consists of several such chains, coupled in the transverse direction by $t_o(\tau)$, $J_o(\tau)$.

Reference Values in the AFM Ground State

We use the algorithms described above to compute numerically the true AFM ground state for various lattices, as shown in table 4.1 (for chains of lengths $N = 22 - 82$). The numerical convergence is checked by comparing the values obtained using bond dimensions $D = 80$ and 100. As can be seen from the values in the table, the maximum relative error is of the order 10^{-5} for the magnetization.

N	$M^2_{\text{stag}}(D=80)$	$M^2_{\text{stag}}(D=100)$	$P_0(D=80)$	$P_0(D=100)$
22	0.139654861	0.139654869	0.807121935	0.807121935
42	0.086355374	0.086355479	0.775723907	0.775723906
62	0.064239765	0.064240095	0.760657976	0.760657965
82	0.051802233	0.051803243	0.751408114	0.751408046

N	$E_{\text{spin}}(D=80)$	$E_{\text{spin}}(D=100)$
22	-0.434912540	-0.434912540
42	-0.438751679	-0.438751679
62	-0.440148665	-0.440148665
82	-0.440871682	-0.440871683

Table 4.1: Squared staggered magnetization M^2_{stag}, mean singlet fraction per double well P_0, and mean spin energy per site E_{spin}, for the AFM of total length N obtained from ground state computation with MPS of bond dimension $D = 80$ and $D = 100$.

L	$M^2_{\text{stag}}(D=80)$	$M^2_{\text{stag}}(D=100)$
22	0.141157103	0.141157444
42	0.088248385	0.088248984
62	0.065508481	0.065509337

Table 4.2: Squared staggered magnetization M^2_{stag} for middle sublattices of length $L = 22$, 42, and 62, for an AFM of total length $N = 82$ obtained from ground state computation with MPS of bond dimension $D = 80$ and $D = 100$.

Note that the energy is correct up to 10^{-9}. We use these values as the reference for adiabaticity of the total chain of length N.

As discussed in the main text, for long enough chains, we observe that antiferromagnetic order develops in a middle sublattice faster than on the total chain. We find that the timescales for observables measured on the sublattice are controlled by the range of the observable itself, as far as finite size effects can be ignored. Therefore, in order to quantify this observation, we need to compare the observables in the evolved sublattice with the corresponding AFM values for a sublattice of the same size in an infinite chain. The thermodynamic mean singlet fraction $P_{0,\text{TD}} = \ln(2) \approx 0.693$ and mean spin energy $E_{\text{spin,TD}} = 1/4 - \ln(2) \approx -0.44315$ are well known [145]. The squared staggered magnetization on a finite sublattice, however, cannot be computed exactly, so that we approximate the reference value by the numerical estimation in a long chain ($N = 82$), shown in table 4.2. Increasing the chain length, the reference values do not change significantly, as one can see in table 4.3 for $N = 142$.

L	$M^2_{\text{stag}}(D=80)$	$M^2_{\text{stag}}(D=100)$
22	0.141068047	0.141069522
42	0.088395582	0.088397878
62	0.065999995	0.066003287

Table 4.3: Squared staggered magnetization M^2_{stag} for middle sublattices of length $L = 22$, 42, and 62, for an AFM of total length $N = 142$ obtained from ground state computation with MPS of bond dimension $D = 80$ and $D = 100$.

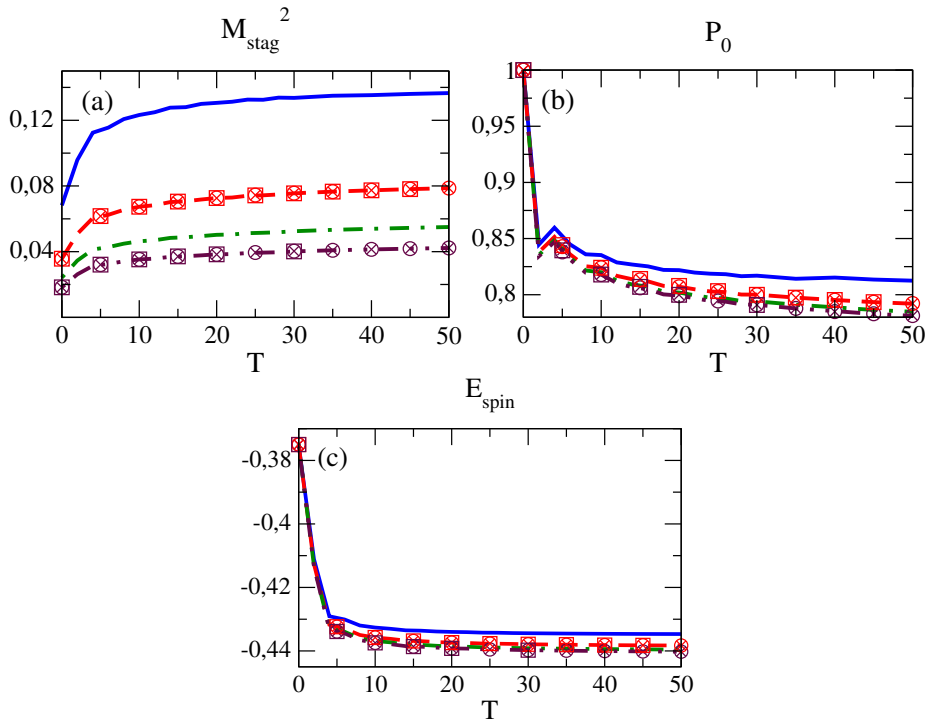


Figure 4.5: M^2_{stag}, P_0, and E_{spin}, as functions of the ramping time T, for chains of length $N = 22$ (solid blue), 42 (dashed red), 62 (dash-dotted green), and 82 (dash double-dotted brown). The results were obtained with MPS of bond dimension $D = 60$ and Trotter step $\delta t = 0.02$ (lines), $D = 40$ and $\delta t = 0.02$ (circles), $D = 40$ and $\delta t = 0.005$ (crosses), and $D = 60$ and $\delta t = 0.005$ (squares).

In 2D, the corresponding values of the AFM were obtained with quantum Monte Carlo by means of the ALPS library [85].

Adiabatically Evolved States

In the following, we present the absolute values of our observables for the state obtained at the end of the adiabatic ramping, in the same sequence as they appear in the main text.

Figures 4.5 and 4.6 show the computed expectation values at the end of the protocol, as a function of the total ramping time, for the case of no holes. Convergence of the numerical results is checked by comparing the results for bond dimension $D = 40$ and 60 and for Trotter steps $\delta t = 0.02$ and 0.005. The largest relative error found (for the largest system and the longest ramping time) is of the order of 10^{-4}, which ensures enough precision for our analysis.

Injecting holes into the sample results in a substantial drop of the squared staggered magnetization, and an increase in the energy (figure 4.7). The latter implies that the system gets excited and the numerical simulation via MPS becomes more demanding. Now, the relative error after the longest ramping time, $T = 30$, for the case of 4 holes on $N = 82$ sites shown in the main text, is of the order of $0.01 - 0.1$, but it becomes significantly smaller for shorter times. This worst-case error does not affect our conclusions, since the main effect we observe, the drop of the magnetization value upon hole arrival, is much larger ($\approx 30\%$) and occurs already at

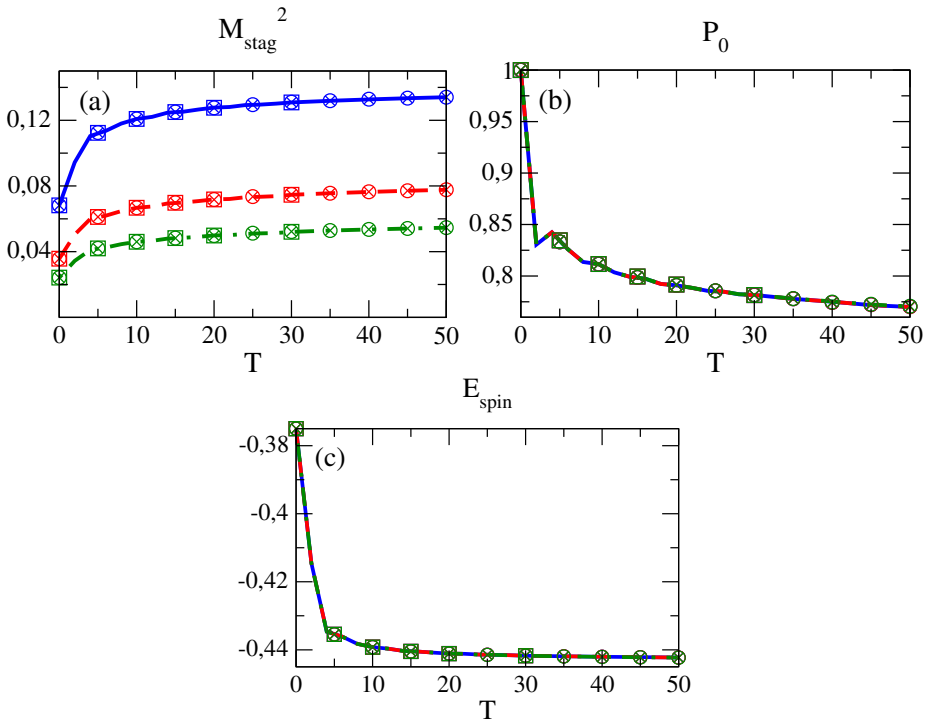


Figure 4.6: Same quantities as in figure 4.5, evaluated on sublattices of length $L = 22$ (solid blue), 42 (dashed red), and 62 (dash-dotted green), of a total lattice of length $N = 82$.

much shorter times ($T \approx 20$), for which the numerical error is only of the order of 10^{-3}. The figure also shows that 2 holes on 42 sites are more dramatic than on 82 sites, and that the negative effect of holes increases with their number.

By including a harmonic trap $V_t \sum_k (k - k_0)^2 \hat{n}_k$, the holes can be confined outside of the sample. We consider 10 holes left and 10 holes right of a sample of size 82, and successively increase the trap strength, as shown in figure 4.8. Consistent with the results in the previous paragraph, our simulation is most demanding for the weakest trap, when holes can still enter the sample and excite the system. The largest relative error in that case is of the order of $0.01 - 0.1$, but it decreases significantly if the trap strength is increased. Again, this worst-case error does not affect any of our conclusions.

In 2D, the time evolution is done with PEPS of bond dimension $D = 2$, $D = 3$ and $D = 4$, and the largest relative error is of the order of 10^{-2}, for the largest system in figure 4.9. Our results suggest that qualitative insight can already be gained from PEPS with $D = 2$. Therefore, figure 4.10 shows the effect of holes and harmonic trap for $D = 2$ without convergence check. Just as in 1D [figure 4.7(a)], the staggered magnetization decreases significantly with increasing number of holes, and the harmonic trap confines the holes on the outside.

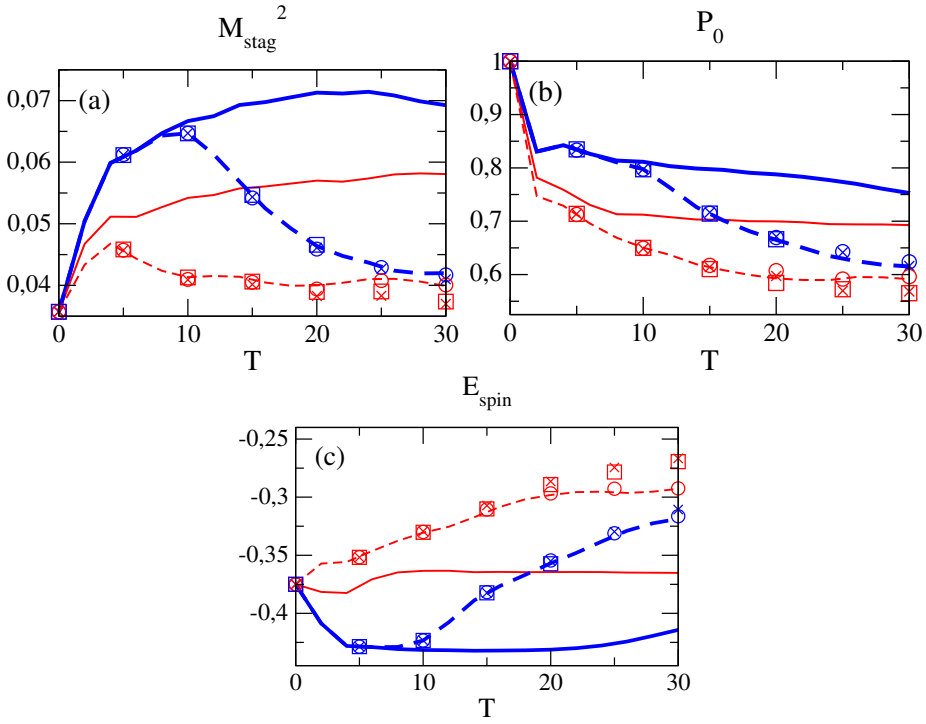


Figure 4.7: $M^2_{\rm stag}$, P_0, and $E_{\rm spin}$, as functions of the ramping time T, evaluated on the middle $L = 42$ site sublattice, for 2 (solid) and 4 holes (dashed) on $N = 82$ sites (thick blue) and $N = 42$ sites (thin red), and the holes are initially located at the boundaries, and $t = 2$. The results correspond to $D = 60$ and Trotter step $\delta t = 0.02$ (lines), $D = 40$ and $\delta t = 0.02$ (circles), $D = 40$ and $\delta t = 0.005$ (crosses), and $D = 60$ and $\delta t = 0.005$ (squares).

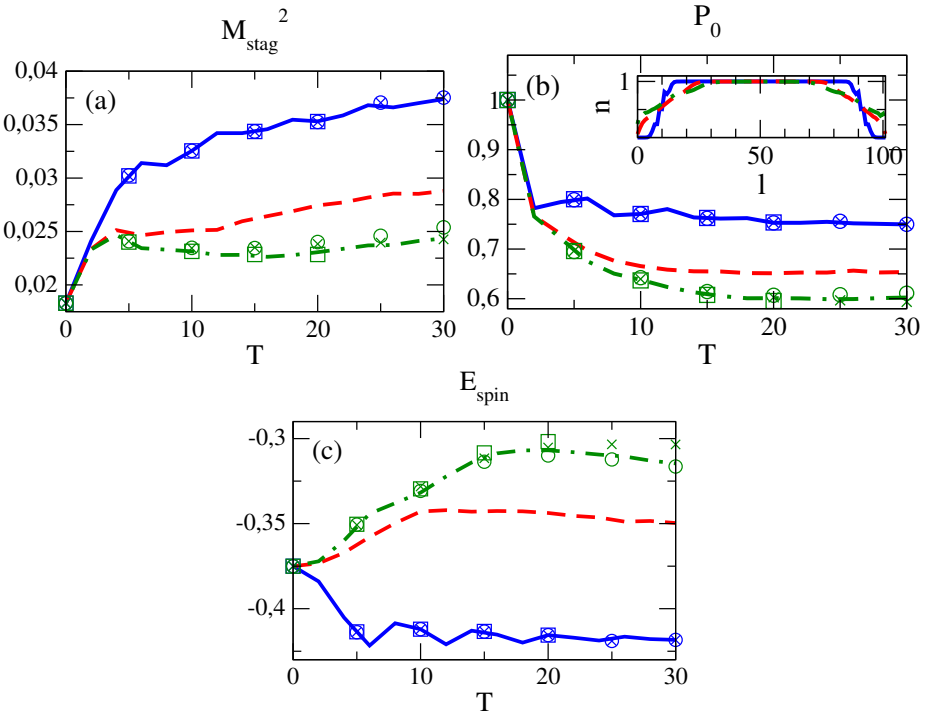


Figure 4.8: M^2_{stag}, P_0, and E_{spin}, as functions of the ramping time T, evaluated on the middle $L = 82$ site sublattice, for 10 holes initially on each boundary of 82 fermions, with a harmonic trap of strength $V_t = 0.004$ (dash-dotted green), $V_t = 0.006$ (dashed red), and $V_t = 0.02$ (solid blue), and $t = 3$. The inset in (b) shows the occupation n of lattice site l after ramping time $T = 30$, and we find that the holes delocalize precisely $\pm 2t$ at the boundaries of the trap. Again, the results correspond to $D = 60$ and Trotter step $\delta t = 0.02$ (lines), $D = 40$ and $\delta t = 0.02$ (circles), $D = 40$ and $\delta t = 0.005$ (crosses), and $D = 60$ and $\delta t = 0.005$ (squares).

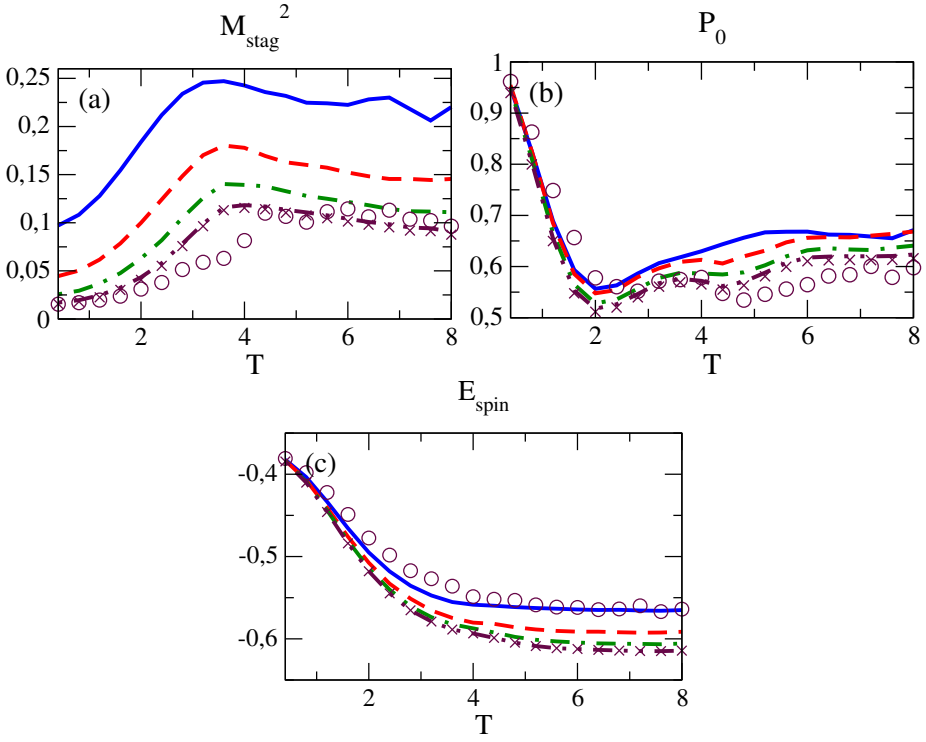


Figure 4.9: M^2_{stag}, P_0, and E_{spin}, as functions of the ramping time T, for $N = 4 \times 4$ (solid blue), 6×6 (dashed red), 8×8 (dash-dotted green), and 10×10 (dash double-dotted brown). The results were obtained with PEPS of bond dimension $D = 4$ and Trotter step $\delta t = 0.03$ (lines), $D = 2$ (circles), and $D = 3$ (crosses).

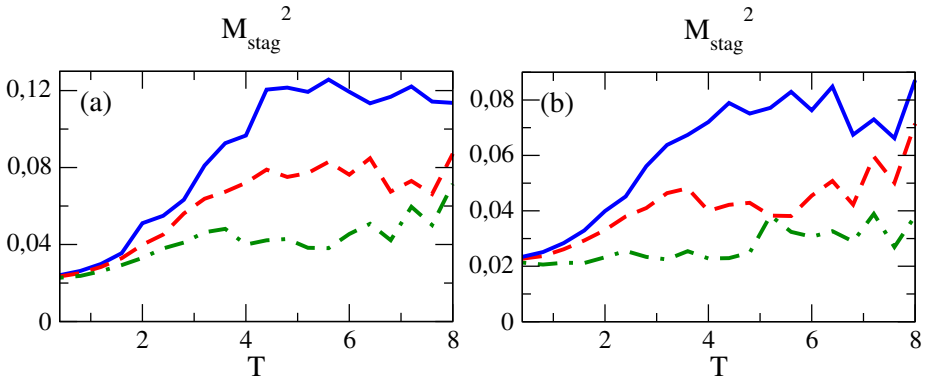


Figure 4.10: (a) M^2_{stag} as a function of the ramping time T, for $N = 8 \times 8$ without holes (solid blue), with 1 hole (dashed red), and 2 holes (dash-dotted green), where the holes are initially localized at the boundary, and $t = 2.5$. (b) $N = 8 \times 8$ with 4 holes initially distributed at the boundary, with no harmonic trap (dash-dotted green), and with a trap of strength $V_t = 0.25$ (dashed red), and $V_t = 2.5$ (solid blue), and again $t = 2.5$. The results correspond to $D = 2$ and Trotter step $\delta t = 0.03$.

Chapter 5

Dynamical Enhancement of Spatial Entanglement in Massive Particles

In the field of quantum many-body simulations, one topic of interest is the generation of highly entangled states. These can be of fundamental interest, but also find applications in quantum information processing. However, entanglement is not always easy to generate in a given setup. In this chapter, we discuss dynamical enhancement of entanglement in a driven Bose-Hubbard model and find an enhancement of two orders of magnitude from the ground state value which is robust against fluctuations in experimental parameters. This chapter is published in reference [54].

5.1 Introduction

Quantum coherence is often thought to be found only in very small systems or under artificial laboratory conditions, since otherwise unavoidable environment coupling results in rapid loss of coherence. Whereas this holds quite generally, it holds in particular for many-body coherence and entanglement. As recent experimental [150, 151] and theoretical [152–157] evidence suggests, however, exceptions to this general rule exist. In particular coherent driving can compensate for the environment-induced loss of coherence and, thereby, stabilize entanglement under conditions under which a static system would be completely separable [158–163].

Such dynamically induced entanglement does not only hold the potential to influence macroscopically observable properties [164], but certainly also opens up new paths towards scalable quantum information processing which otherwise is limited through the unfavorably scaling dephasing times with the system size [53, 165]. However, our current understanding of dynamical enhancement of entanglement is still in its infancy.

In this chapter, we consider ultracold bosonic atoms stored in an optical lattice and investigate the dynamical enhancement of entanglement in the atoms' spatial degree of freedom. The feasibility of the setup we have in mind has already been proven experimentally for spectroscopic measurements [166], however, its potential for the creation of many-body entangled states has not been addressed [167]. The spatial degree of freedom as carrier of entanglement is advantageous since re-

quirements on cooling are much less stringent than in the case of the spin degree of freedom [31]. Since one-dimensional systems ranging from 10 to 100 sites, with one to three atoms per site, and beyond, are experimental routine (see e.g. references [38,39,168,169]), the presently discussed mechanism permits the creation of highly entangled states in a scalable fashion.

5.1.1 Reader's Guide

The structure of this chapter is the following. Section 5.2 introduces the precise model for our problem, and section 5.3 demonstrates that entanglement in its ground and thermal states is very low. Subsequently section 5.4 shows that coherent driving substantially increases entanglement, to a value close to maximal entanglement. In particular, this enhancement is very robust against experimental imperfections, as discussed in section 5.5. Section 5.6 briefly summarizes our conclusions.

5.2 Model

The optical lattice is created by two counterpropagating laser beams of wave length λ and amplitude V_0 in one direction, and a tight perpendicular confinement of strength $V_\perp$ in the other two directions, restricting the motion of the atoms to one dimension. In the deep lattice limit $V_0 \gg E_\mathrm{R}$, where $E_\mathrm{R} = \hbar^2 k^2/2m$ (with $k = 2\pi/\lambda$) is the recoil energy, and at sufficiently low temperatures, this system can be well described [34,35,170] in terms of the Bose-Hubbard Hamiltonian

$$\hat{H} \;=\; -J\sum_{l=1}^{L-1}(a_l^\dagger a_{l+1} + a_{l+1}^\dagger a_l) + \frac{U}{2}\sum_{l=1}^{L}\hat{n}_l(\hat{n}_l - 1) \quad , \tag{5.1}$$

where the creation operator $a_l^\dagger$ creates and the annihilation operator a_l annihilates a boson at lattice site l. The tunneling parameter J and the on-site interaction U depend on the lattice parameters approximately [31,171] via

$$J/E_\mathrm{R} \;=\; \frac{4}{\sqrt{\pi}}(V_0/E_\mathrm{R})^{\frac{3}{4}} e^{-2\sqrt{V_0/E_\mathrm{R}}} \quad , \tag{5.2}$$

$$U/E_\mathrm{R} \;=\; \sqrt{\frac{8}{\pi}} k a_\mathrm{s} (V_0 V_\perp^2/E_\mathrm{R}^3)^{\frac{1}{4}} \quad . \tag{5.3}$$

Whereas the lattice depth V_0 is typically time-independent, we will compare here the dynamics of such an autonomous system with its driven version [166], where V_0 is modulated temporally:

$$V_0(t) \;=\; V\Big(1 + \mathrm{d}V\sin(\omega t)\Big) \quad , \tag{5.4}$$

around its initial height V with relative amplitude $\mathrm{d}V$ and frequency ω. As we will show, the temporal modulation of the tunneling parameter J and the on-site interaction U that results from this lattice depth modulation drives the atoms into a spatially strongly correlated, i.e., entangled state. For the verification of the entanglement properties, we envision a rapid separation of the many-body system into two parts, what can be realized by ramping up a potential barrier as depicted in figure 5.1.

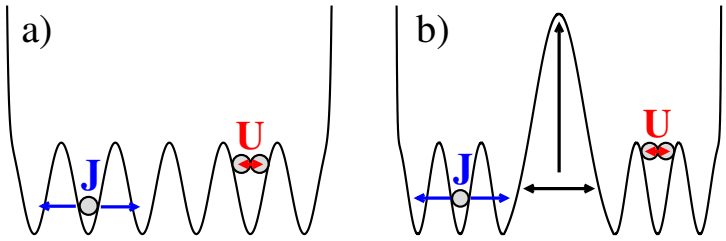


Figure 5.1: a) The system is prepared in an entangled state. b) The system is split into two halves by raising the intermediate barrier. J is the kinetic term and U the on-site interaction in the Bose-Hubbard model (5.1).

This spatial separation effectively switches off the interaction between the two subsystems what freezes the entanglement dynamics. However, it also entails that each subsystem will typically not have a well-defined particle number. Whereas correlations in the particle number formally may imply entanglement, its experimental verification will be technically impossible because this would require the measurement of coherent superpositions of states with different numbers of massive particles [172–175]. We, therefore, consider states after projection onto well-defined local particle number. Experimentally, this can be realized via post-selection, i.e., rejection of data corresponding to undesired local particle numbers, as routinely done in experiments with entangled photons [176–179]; but, given the advances in spatially resolved detection of ultracold atoms in optical lattices [38, 39], it is also becoming feasible to detect the local particle number while leaving spatial coherence within each subsystem intact. We will focus in particular on those cases in which particles are split evenly between the two subsystems, since this is the case that occurs with highest probability, and this is also the case in which the highest entanglement can be achieved. Doing so, we obtain a clean notion of entanglement between the two separated halves of the optical lattice where each half is filled with a fixed number of particles and can be addressed individually.

In the following we quantify the entanglement of the postselected states with the entropy of entanglement [53, 180] in the case of pure states and the negativity [181] in the case of mixed states. The entropy of entanglement is given by the von Neumann entropy of the reduced density matrix ρ_r that is obtained through the partial trace over one subsystem of the entire many-body state, i.e., $E(|\psi\rangle) = -\mathrm{tr}(\rho_\mathrm{r} \log_2(\rho_\mathrm{r}))$. The negativity $N(\rho) = (||\rho^{\mathrm{PT}}||_1 - 1)/2$ of a mixed state ρ is defined in terms of the trace norm of the partially transposed density matrix ρ^{PT}.

5.3 Ground and Thermal State

Figure 5.2a) depicts the ground state entanglement properties for the exemplary case of $L = 6$ wells filled with $N = 6$, 3 and 2 bosons, respectively. The qualitative and quantitative features of the system with $N = 4$ bosons are very similar to the $N = 2$ system, and both $N = 5$ and $N = 7$ are similar to $N = 6$; more generally, we found that the behavior of $N + mL$ bosons (with integer m) is essentially identical to that of N bosons. In all these cases, the amount of entanglement in the ground state does not exceed the value of $E \approx 0.05$. The qualitative dependence of E on the parameter U/J is as expected: If the tunneling dominates the system dynamics,

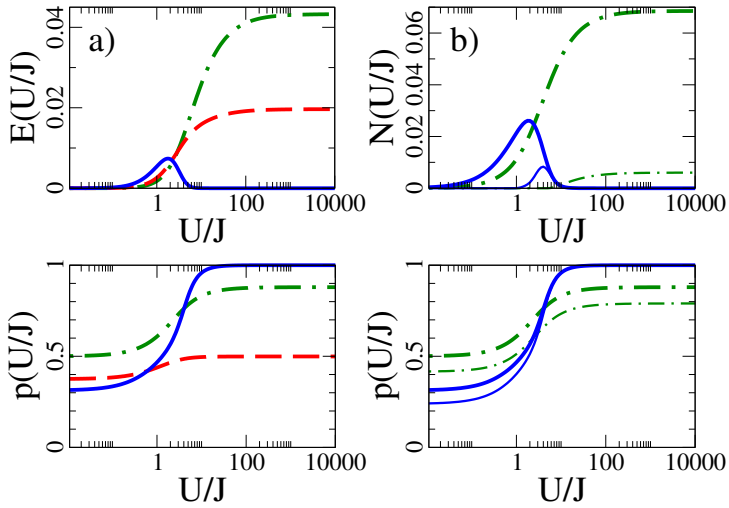


Figure 5.2: a) Entropy of entanglement E and probability for successful postselection of evenly distributed number of bosons between the right and the left part of a $L = 6$ well Bose-Hubbard system. The overall particle number is $N = 6$ (solid blue), $N = 3$ (dashed red) and $N = 2$ (dash-dotted green). b) Negativity N of the thermal state and corresponding probability at temperature $T = 0\,\mathrm{nK}$ (thick line) and $T = 80\,\mathrm{nK}$ (thin line) for some systems of the left panel.

i.e., if $U/J \simeq 0$, the bosons populate the same single-particle states such that after postselection of local particle number the system is separable. For finite interaction U the bosons repel each other and establish correlations; therefore, E typically increases with U/J. There are, however, exceptions, like the case of unit filling (depicted in blue), where a separable, perfect Mott insulator [31, 34, 35, 170, 171] develops for $U/J \to \infty$. This is also reflected in the fact that in this limit the bosons will always be separated in a balanced fashion between the left and right half of the system, whereas typically the probability for this is smaller than unity.

Assuming perfect ground state cooling is certainly a theoretical idealization, but also thermal excitation cannot enhance the entanglement as shown in figure 5.2b), where the negativity of the thermal state $\rho_{\mathrm{th}} = \exp(-\hat{H}/k_{\mathrm{B}}T)/Z$ is shown. The probability of finding the bosons split evenly into left and right half decreases with increasing temperature, and the entanglement is always lower than for $T = 0$, independent of the filling N/L.

5.4 Driven System

As we will see in the following, the ground state entanglement is highly enhanced when the system is coherently driven in resonance. The driving is given by a periodic modulation of the lattice, equation (5.4), around the initial height V by a relative amplitude $\mathrm{d}V$ with frequency ω. The resulting time-dependent lattice height $V_0(t)$ causes time-dependent parameters kinetic term $J(t)$ and on-site interaction $U(t)$ according to equations (5.2) and (3), and thus a time-dependent Bose-Hubbard Hamiltonian

$$\hat{H}(t) \quad = \quad -J(t)\sum_{l=1}^{L-1}(a_l^\dagger a_{l+1} + a_{l+1}^\dagger a_l) + \frac{U(t)}{2}\sum_{l=1}^{L}\hat{n}_l(\hat{n}_l - 1) \quad .$$

For our subsequent analysis we initialize the system in the ground (pure) state $|\psi(t=0)\rangle$ of $\hat{H}(t=0)$ that we find with the help of the Lanczos algorithm. We then propagate that state with the above time-dependent Bose-Hubbard Hamiltonian according to the Schrödinger equation

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

with the fourth-order Runge-Kutta scheme with adaptive stepsize, explained, e.g., in [182]. This algorithm guarantees an error ϵ per time evolution step by adjusting the stepsize. All shown results were obtained with $\epsilon = 10^{-12}$, and their correctness was confirmed by comparing results obtained with $\epsilon = 10^{-6}$, $\epsilon = 10^{-9}$, and $\epsilon = 10^{-12}$. In other words, our numerical results are accurate to a precision much better than resolvable in our observables shown in the figures below. To be specific, we consider the experimental parameters $V = 10\,E_\mathrm{R}$ and $V_\perp = 30\,E_\mathrm{R}$ as lattice depths, $\mathrm{d}V = 0.2$ as lattice depth modulation, $\lambda = 842\,\mathrm{nm}$ as the wave length of the laser, $a_\mathrm{s} = 5.45\,\mathrm{nm}$ as the scattering length and $m = 86.909\,\mathrm{u}$ as the mass of rubidium-87 [183–185]. For these parameters the system is Mott-insulating ($U/J \approx 27.8$), and the initial ground state has essentially vanishing entropy of entanglement, as can be seen in figure 5.2a). As driving frequency, we chose $\omega = U/\hbar = 12862\,\mathrm{Hz}$, what corresponds to resonant driving in the Mott-insulating regime.

Figure 5.3 shows the dynamical enhancement of entanglement caused by the coherent driving. We start with the ground state of the static system. After 100 ms the driving is switched off and after $t = 150\,\mathrm{ms}$ the lattice depth is increased to $V_0 = 30\,E_\mathrm{R}$ in order to freeze the entanglement dynamics completely. As we conclude from figure 5.3, resonant driving enhances the initial ground state entanglement from a vanishingly small value at time $t = 0$ to a value almost of a maximally entangled state. By comparison with figure 5.2, we conclude an enhancement of two orders of magnitude.

Apparently, entanglement grows rather quickly once the driving is switched on. After about 10 to 20 ms the increase slows down a bit, until entanglement saturates after $t \approx 50\,\mathrm{ms}$. From that time on, entanglement fluctuates around an average value due to the finite interactions in the system. Figure 5.3 shows how these remaining fluctuations smooth out with growing system size. Once entanglement is saturated, the coherent driving can be switched off as it is done at $t = 100\,\mathrm{ms}$ in our specific case. In the subsequent interval $100\,\mathrm{ms} < t < 150\,\mathrm{ms}$ the system evolves under the static Hamiltonian. The entanglement dynamics reduced to small fluctuations that, in particular, decrease with growing system size as it can be seen in figure 5.3. This gives once more evidence that the static interactions on their own have virtually negligible impact on entanglement. For the purpose of verifying entanglement it is nevertheless desirable to freeze the entanglement dynamics completely. This can be achieved by ramping up the lattice depth as done at $t = 150\,\mathrm{ms}$ in figure 5.3. Finally, by increasing the potential barrier as depicted in figure 5.1, the two separated, individually addressable atomic ensembles are obtained.

In the Mott-insulating regime the eigenstates of the static Hamiltonian are essentially completely separable. Modulating the lattice depth drives the system into a coherent superposition over several eigenstates of the static system, what is accompanied by an increase of entanglement. Accordingly, one would expect the saturation time T to depend crucially on the lattice depth modulation, and, indeed, it scales approximately like $T \propto \mathrm{d}V^{-1}$. The saturation time also depends

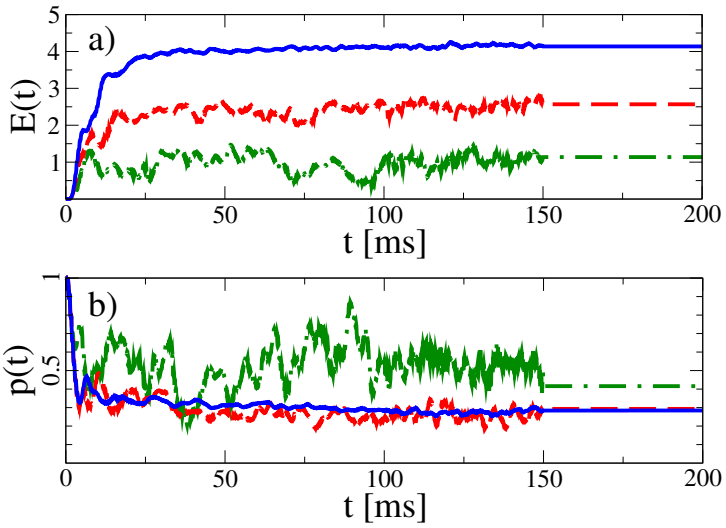


Figure 5.3: Entropy of entanglement E and probability of successful postselection for the driven Bose-Hubbard Hamiltonian (5.4) with $L = 8 = N$ (solid blue), $L = 6 = N$ (dashed red) and $L = 4 = N$ (dash-dotted green). At $t = 100\,\mathrm{ms}$ the coherent driving is stopped and at $t = 150\,\mathrm{ms}$ the lattice depth is increased in order to freeze the entanglement dynamics.

on the system size, as can be seen in figure 5.3, where the $L = 4 = N$ system reaches its largest value before the $L = 6 = N$ system, which saturates before the $L = 8 = N$ system. But this dependence is much weaker than the one on the lattice depth modulation, such that a large system can be entangled as fast as a small system by choosing a larger value of $\mathrm{d}V$. Since the maximally attainable entanglement, however, is limited by the number of separable states that can be coherently superposed, it depends on the system size (as can be seen in figure 5.3a)) but is independent of the driving strength as we explicitly confirmed. Thus, a larger modulation of the lattice depth can accelerate the entanglement generation; however, even very weak coherent driving can yield the same enhancement of entanglement as strong driving. A properly chosen frequency of the driving, on the other hand, is essential for coherently superposing many product states, and, in agreement with [166, 183], we find generation of strongly entangled states only for driving frequencies $\hbar\omega = xU$ with $x = 1/2$, $x = 1$ and $x = 2$.[1]

With increasing number of atoms, also the number of possible distributions of atoms between the two subsystems is growing, so that the probability to find an even distribution might decrease. This decrease is apparent in figure 5.3b), where the probability of an even distribution drops from $p = 0.42$ for 4 atoms to $p = 0.29$ for 6 atoms. For larger atom numbers, however, this decrease essentially stops and finding 8 atoms evenly distributed is essentially as likely as for 6 atoms. This reflects that the probability for even splitting becomes largely independent of the particle number for large N. Thus, one can expect to find evenly distributed particle numbers with substantial probability also in an experiment with significantly more bosons than a numerical simulation can handle.

When the system gets excited, higher bands might get populated, so that it is

[1] We checked this relation on a lattice with $L = 4$ sites by increasing the boson number from $N = 2$ to $N = 12$. The peaks for $x = 1/2$ and $x = 1$ are always there independent of the filling. For $x = 2$ a resonance is found only for incommensurate filling starting at $N = 5$.

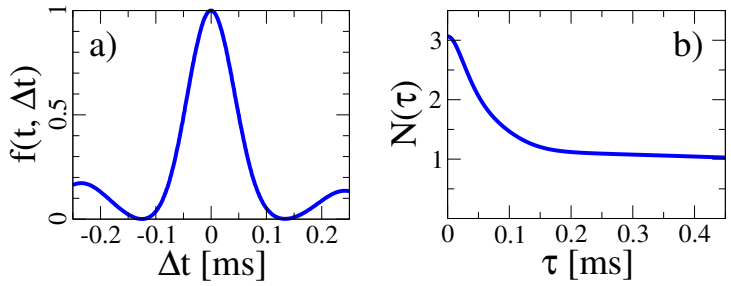


Figure 5.4: a) Fidelity $f(t, \Delta t)$ at $t = 100\,\text{ms}$ for state preparation with imperfect timing. b) Negativity N of the mixed state (5.5) resulting from the measurement of the driven Bose-Hubbard Hamiltonian with $L = 6 = N$ at $t_0 = 100\,\text{ms}$.

not immediately evident that the one-band description remains valid. The gap between the lowest and the first band can be estimated as $\Delta E = \hbar\omega/E_\text{R} = 2\sqrt{V_0/E_\text{R}}$ [31, 171], which, in our case, becomes $\Delta E \approx 10U$. On the other hand, we numerically find the highest energies $E_\text{max} \approx 4U$ (in the largest system with $L = 8 = N$ and for a driving up to 100 ms), what clearly demonstrates that states associated with higher bands are not being populated.

5.5 Robustness

In an experiment, certainly also the timing will be crucial. As figure 5.3 shows, the system evolves rapidly, and fluctuations in the durations of coherent driving or ramping up the barrier that are comparable to system time scales will result in the generation of a mixed state which typically has reduced entanglement. The relevant time scale can be obtained from the fidelity $f(t, \Delta t) = |\langle\psi(t)|\psi(t + \Delta t)\rangle|^2$, where $|\psi(t)\rangle$ is the postselected system state after driving of duration t. The fidelity f is depicted in figure 5.4 for the exemplary case of $N = 6$ particles in a $L = 6$ site lattice. The width of the central peak (full width at half maximum) that determines the minimal required experimental precision reads in this case $\Delta t_\text{m} = 0.1\,\text{ms}$. In a similar fashion, we can also estimate the required precision for all other experimental parameters, such as the potential V ($\Delta V_\text{m} = 0.08\,E_\text{R}$), the perpendicular confinement $V_\perp$ ($\Delta V_{\perp,\text{m}} = 0.12\,E_\text{R}$), the amplitude of the driving $\text{d}V$ ($\Delta\text{d}V_\text{m} = 0.016$) and the driving frequency ω ($\Delta\omega_\text{m} = 14.5\,\text{Hz}$).

To estimate the impact of fluctuations of these parameters on the attainable entanglement, we have to consider the mixed state that is obtained with many repetitions of the experiment with the fluctuating parameter taking different values at each repetition. To be specific, we focus on the inaccuracy in the duration of driving, and we assume that these durations are distributed according to a Gaussian centered around $t_0 = 100\,\text{ms}$ with a standard deviation τ. This gives rise to the mixed state

$$\rho(\tau) \quad = \quad \frac{1}{\tau\sqrt{2\pi}} \int_{-\infty}^{\infty} \text{d}t\, e^{-\frac{(t-t_0)^2}{2\tau^2}} |\psi(t)\rangle\langle\psi(t)| \quad . \tag{5.5}$$

The negativity of this mixed state is depicted in figure 5.4b) as function of the inaccuracy τ of the duration of driving. For $\tau = 0\,\text{ms}$, the situation reduces to the case of pure state entanglement as discussed above; but for finite τ, entanglement is reduced significantly, what is a very generic feature of mixed states. Besides this expected behavior, there are two features that should be stressed.

1. At $\tau = 0\,\mathrm{ms}$ the first derivative of $N(\tau)$ vanishes, so that entanglement turns out to be insensitive to small timing errors. The second-order Taylor expansion reads $N(\tau) \approx 3.064 - 900(\tau/\mathrm{ms})^2$. That is, timing errors below $0.01\,\mathrm{ms}$ imply a change in negativity of less than $3\,\%$.

2. Even in the presence of significantly larger timing errors there is still rather strong entanglement with $N(\tau) \simeq 1$, c.f. the high saturation value in figure 5.4b) for $\tau > 0.15\,\mathrm{ms}$.

In particular, this astonishing robustness against experimental fluctuations underpins that potential that coherent driving offers as means to create entanglement as compared to engineered interactions.

5.6 Conclusions

As recent investigations on driven spin systems suggest [162, 163], the feature of dynamical enhancement of entanglement is not particular to the Bose-Hubbard system, but is a rather generic feature, which is largely independent of detailed system properties. An advantage of the present bosonic system compared to many spin systems is that particle numbers can easily be varied in an experiment, which provides the means to study the generation of entanglement in the entire regime from rather small systems, through the mesoscopic domain, up to the semiclassical regime. In particular, observing the rise and decay of entanglement with increasing particle number will provide us with valuable insight into the emergence of classical behavior in large quantum systems.

Whereas we would like to stress that creating highly entangled states should not be taken as synonymous to attempting to process quantum information, it should also be noted that states in spatial degrees of freedom as considered here have an equally justified footing in quantum information science as spin states, either for direct encoding of quantum information [30, 186, 187] or as a means of communication between distant spins [188].

Part III

Other Classical Simulations

Chapter 6

Detection of Avoided Crossings by Fidelity

Some of the problems considered in the previous chapters can feature quantum chaos that manifests itself in the presence of avoided crossings in their energy spectra. In this chapter, we propose the fidelity, defined as overlap of eigenstates of two slightly different Hamiltonians, as an efficient detector of avoided crossings in the energy spectrum. This new application of fidelity is motivated for model systems, and its value for analyzing complex quantum spectra is underlined by applying it to a random matrix model and a tilted Bose-Hubbard system. This chapter is published in reference [56].

6.1 Introduction

As already mentioned in the previous chapters, the progress in cooling and manipulating ultracold atomic gases in recent years has opened new perspectives on interacting many-body models from condensed matter physics [30, 31, 189]. It led to questions and opportunities beyond conventional solid-state physics, e.g., the direct experimental study of quantum phase transitions [30, 31], the role and engineering of genuine quantum correlations [30, 31, 54, 190], and the phenomenon of quantum chaos in systems that consist of indistinguishable particles [191–197]. In this context, it is possible to detect a quantum phase transition by the change of fidelity (modulus of the overlap between eigenstates of slightly different Hamiltonians) [55, 198, 199], since the ground state of a quantum system changes dramatically at a critical parameter [200].

Up to now, the temporal change of fidelity – as the overlap of the same initial states evolved by different Hamiltonians [201, 202] – has been measured experimentally in wave billiards [203, 204], but also in systems of cold atoms subject to optical potentials [205–209]. Similar techniques may be applied to measure the evolving overlap of two eigenstates where time is substituted by the change of some tunable control parameter. Often a quantum phase transition may be viewed, for finite-size realizations of a system, as an avoided crossing (AC) in parameter space which closes in the thermodynamic limit [200]. A scenario of *many* ACs with a broad distribution of widths [210–212], as a manifestation of a strong coupling of many energy levels, is naturally found in quantum chaotic systems [210]. The dynamical evolution of these systems is determined by the number and distribution

of ACs present in the spectrum. The question then arises whether the applicability of fidelity can be lifted from pure ground-state analysis [213] to detect and characterize ACs in the entire spectrum of a complex quantum system. In this chapter we propose to use the fidelity as a new tool to detect and characterize ACs in quantum spectra [167, 214]. This is corroborated by analytical and numerical results for exemplary quantum systems.

6.1.1 Reader's Guide

This chapter has the following structure. In section 6.2, the considered fidelity measure is defined and illustrated for simplified two- and three-level models. Section 6.3 goes beyond these idealized models and applies our fidelity to more realistic quantum chaotic systems, namely, a random matrix theory model as well as a Bose-Hubbard Hamiltonian. Brief conclusions are given in section 6.4.

6.2 The Fidelity Measure

Given some parameter depending Hamiltonian $\hat{H}(\lambda) = \hat{H}_1 + \lambda\hat{H}_2$, the fidelity [201, 202] between the n-th eigenstates, denoted by $|n\rangle$, of two slightly different Hamiltonians $\hat{H}(\lambda)$ and $\hat{H}(\lambda + \delta\lambda)$ is defined as $f_n(\lambda, \delta\lambda) \equiv |\langle n(\lambda)|n(\lambda + \delta\lambda)\rangle|$. In complex quantum systems with many degrees of freedom, many of the levels of the system are coupled to each other leading to ACs in the spectrum of the Hamiltonian when the parameter λ is changed [210]. To simplify the discussion, we assume a finite size Hilbert space $\mathcal{H}$, where all energy levels are never exactly degenerate. To detect and characterize an AC for a given quantum level n we study the fidelity change [55, 198, 199]

$$S_n(\lambda, \delta\lambda) \equiv \frac{1 - f_n(\lambda, \delta\lambda)}{(\delta\lambda)^2} \tag{6.1}$$

which measures the change of the state $|n\rangle$. For $\delta\lambda \ll 1$, it is independent of $\delta\lambda$, i.e. $S_n(\lambda, \delta\lambda) \approx S_n(\lambda)$, and vanishingly small everywhere except in the vicinity of an AC. The independence of $\delta\lambda$ arises from the fact that the first non-vanishing contribution to f_n in the expansion of the changed state $|n(\lambda + \delta\lambda)\rangle$ is of second order in $\delta\lambda$ [214, 215]. The fidelity measure (6.1) also has the advantage of being applicable locally in the spectrum, where one follows a certain state $|n(\lambda)\rangle$ and its neighbors over a range of parameter values λ to study the ACs they encounter. In addition, it is well-suited for numerical computations, since λ is the only relevant parameter as long as $\delta\lambda$ is sufficiently small. The different limit of large $\delta\lambda$ and hence the coupling over a broad energy band was the focus of a recent work using another generalized fidelity [216]. In contrast, our interest here is the detection and characterization of ACs as local couplings in energy space.

6.2.1 Two-State Model

Let us first discuss an isolated AC which can locally be described in nearly-degenerate perturbation theory as an effective two-level system. It is then represented by a Hamiltonian $\hat{H}(\lambda) = \lambda\sigma^Z + g\sigma^X$, with a real coupling g between the levels (σ^X and σ^Z denote Pauli matrices), showing an AC at $\lambda = 0$ of width $c = 2g$. The eigenstates are easily found [210, 217] and from them we calculate the fidelity

for the two-level system:

$$f_{\pm}(\lambda,\delta\lambda) = \frac{g^2+\lambda(\bar{\lambda}-\lambda)+\lambda^2+\bar{\lambda}\sqrt{g^2+\lambda^2}+\lambda\sqrt{g^2+\bar{\lambda}^2}+\sqrt{[g^2+\lambda^2][g^2+\bar{\lambda}^2]}}{2\sqrt{\left[g^2+\lambda\left(\lambda\pm\sqrt{g^2+\lambda^2}\right)\right]\left[g^2+\bar{\lambda}\left(\bar{\lambda}\pm\sqrt{g^2+\bar{\lambda}^2}\right)\right]}} \quad ,$$

where we used the shorthand notation $\bar{\lambda} \equiv \lambda + \delta\lambda$. To obtain the fidelity change in the limit $\delta\lambda \ll 1$, we need to expand the expression for the fidelity in a power series for $\delta\lambda$ and keep only the leading term proportional to $(\delta\lambda)^2$. The final expression is the same for both eigenstates (indexed by $\pm$) and has the simple form:

$$S_{\pm}(\lambda) = \frac{1}{8}\left(\frac{g}{g^2+\lambda^2}\right)^2 \quad . \tag{6.2}$$

This is the square of a Lorentzian and differs significantly from zero only near the AC at $\lambda = 0$. This formula already allows us a good understanding of isolated ACs, as, for example, the peak width is easily computed as $\sigma^{\mathrm{FWHM}} = 2g\sqrt{\sqrt{2}-1}$. On the other hand an AC can be characterized by the ratio between the local energy level curvature and the distance between the two repelling energy levels. We call the absolute value of this ratio *renormalized curvature* $C_n(\lambda)$ and find

$$C_{\pm}(\lambda) \equiv \left|\frac{1}{\Delta(\lambda)}\frac{\partial^2 E_{\pm}(\lambda)}{\partial\lambda^2}\right| = 4S_{\pm}(\lambda) \tag{6.3}$$

for the two-level system. For higher-dimensional systems we expand the wave function $|n(\lambda+\delta\lambda)\rangle$ in second order in $\delta\lambda$ and find

$$S_n(\lambda) = \frac{1}{2}\sum_{m\neq n}\frac{|\langle m(\lambda)|\hat{H}_2|n(\lambda)\rangle|^2}{[E_n-E_m]^2} \approx \frac{|\langle n'(\lambda)|\hat{H}_2|n(\lambda)\rangle|^2}{2\,[E_n-E_{n'}]^2} \quad ,$$

where we reduced the sum near an isolated AC to the nearest neighboring level n'. Similarly, one obtains for the renormalized curvature [218]

$$C_n(\lambda) = \left|\frac{2}{\Delta(\lambda)}\sum_{m\neq n}\frac{|\langle m(\lambda)|\hat{H}_2|n(\lambda)\rangle|^2}{E_n-E_m}\right| \approx 2\frac{|\langle n'(\lambda)|\hat{H}_2|n(\lambda)\rangle|^2}{[E_n-E_{n'}]^2} = 4S_n(\lambda) \quad . \tag{6.4}$$

The relation $C_n \approx 4S_n$ thus holds as long as the effect of other levels can be neglected close to a single AC.

6.2.2 Beyond the Two-Level Approximation

ACs in higher dimensional systems are not totally isolated, but other levels can contribute to the evolution of a quantum state as the parameter λ is varied. Consider two energy levels approaching each other as $\lambda \to 0$, and a third level being well separated by a distance ϵ in energy and weakly coupled to the first two levels. A Hamiltonian model for such a situation reads

$$\hat{H}(\lambda) = \begin{pmatrix} -\lambda & g & g_{13} \\ g & \lambda & g_{23} \\ g_{13} & g_{23} & \epsilon \end{pmatrix}, \quad g_{ij}, \epsilon \in \mathbb{R} \quad , \tag{6.5}$$

where we limited ourselves to real couplings. Since the first two levels become nearly degenerate and are well-separated from the third one, we can write this in degenerate perturbation theory[1] close to the crossing as

$$\hat{H}_{\mathrm{PT}}(\lambda) = \begin{pmatrix} -\lambda + \frac{g_{13}^2}{\epsilon} & g + \frac{g_{13}g_{23}}{\epsilon} \\ g + \frac{g_{13}g_{23}}{\epsilon} & \lambda + \frac{g_{23}^2}{\epsilon} \end{pmatrix} + \mathcal{O}(\epsilon^{-2}) \quad . \tag{6.6}$$

This reduces the three-level system to an effective two-level system taking the effect of the distant level perturbatively into account. The same procedure can be applied, in principal, to higher dimensional systems. The minimal distance c between the two levels of equation (6.6) is thus changed by the influence of the distant third level in first order to

$$c_{\mathrm{PT}} = 2|g|\sqrt{\left(1 + \tfrac{g_{13}g_{23}}{2g\epsilon}\right)^2 + \left(\tfrac{g_{23}^2 - g_{13}^2}{2g\epsilon}\right)^2} \approx 2|g|\left(1 + \frac{g_{13}g_{23}}{2g\epsilon}\right) \quad , \tag{6.7}$$

where we kept only the leading order behavior. The minimal distance in an isolated AC is accordingly only slightly changed, provided that the coupling to the third level is not much larger than between the two encountering levels and that the third level is well-separated from them. We need to compute the eigenstates $|E_\pm(\lambda + \delta\lambda, \epsilon)\rangle$ of equation (6.6) and then take their overlap for slightly different parameter values to obtain the fidelity, i.e., $f_\pm(\lambda, \delta\lambda, \epsilon) = |\langle E_\pm(\lambda, \epsilon)|E_\pm(\lambda + \delta\lambda, \epsilon)\rangle|$. The fidelity change can be computed by taking the second derivative of the fidelity at $\delta\lambda = 0$. The full expression is very long and difficult to grasp. Expanding it in inverse powers of ϵ and including just the first order correction to the simple two-level system, the fidelity change under the influence of a third not too close level is then given by

$$S_\pm^{\mathrm{PT}}(\lambda, \epsilon) = \frac{1}{8}\frac{g^2}{(g^2 + \lambda^2)^2}\left[1 - \frac{2}{\epsilon}\frac{(gg_{13} + \lambda g_{23})(gg_{23} - \lambda g_{13})}{g(g^2 + \lambda^2)} + \mathcal{O}\left(\epsilon^{-2}\right)\right] \quad .$$

The correction due to the third level is also λ-dependent and changes the peak height at $\lambda = 0$. Let us also include the second order correction to the fidelity change at $\lambda = 0$ here

$$S_\pm^{\mathrm{PT}}(\lambda = 0, \epsilon) = \frac{1}{8g^2}\left[1 - \frac{2}{\epsilon}\frac{g_{13}g_{23}}{g} - \frac{1}{2\epsilon^2}\frac{g_{13}^4 - 8g_{13}^2g_{23}^2 + g_{23}^4}{g^2} + \mathcal{O}(\epsilon^{-3})\right] \quad .$$

If all off-diagonal matrix elements are of similar magnitude, the effect of the third level is characterized by its inverse distance to the AC. This underlines our claim that the effect of a third level on an AC is not too strong, provided that the level is not very close. But the latter does not take place when three levels undergo a joint AC, i.e., if there were no off-diagonal matrix elements coupling the levels they would all cross in one point. Such a situation cannot be reduced to an effective two-level system. We will in the following also study numerically the behavior of the fidelity change in exactly this case, where the third level cannot be considered a simple perturbation to the two-level system, i.e., when the approximation of an *isolated* AC breaks down.

Three crossing levels can be generated, e.g., by the following real symmetric Hamiltonian

$$\hat{H}(\lambda) = \begin{pmatrix} -\lambda & a & b \\ a & 0 & c \\ b & c & \lambda \end{pmatrix} \quad , \tag{6.8}$$

[1] An excellent overview on this topic gives, e.g., reference [219].

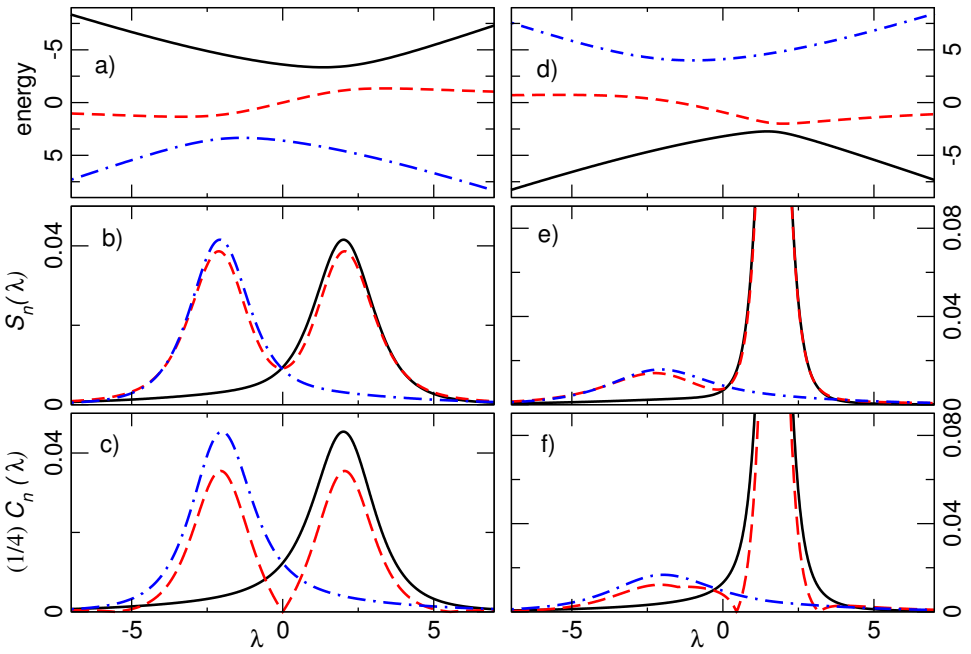


Figure 6.1: (a) Energy spectrum of equation (6.8) for $a = 0, b = 2, c = 3$. All levels are coupled and the spectrum shows two close ACs; (b) fidelity change $S_n(\lambda)$ and (c) renormalized curvature $C_n(\lambda)$ for the energy levels of (a). (d) Energy spectrum for $a = 1, b = 2, c = 3$. All three levels are now directly coupled and the spectrum shows two close ACs. (e) $S_n(\lambda)$ and (f) $C_n(\lambda)$ for the energy levels of (d).

which generalizes the above 2×2-model. Figure 6.1 shows that the fidelity change, defined in equation (6.1), is able to detect and to distinguish two nearby ACs in this system. Furthermore it reflects specific features of an AC in the shape of its peak, i.e., depending on the coupling g, $S_n(\lambda)$ shows a narrow peak of height $S(\lambda = 0) = 1/(8g^2)$.

We see already in this simple example that the renormalized curvature captures the form of the fidelity change $S_n(\lambda)$ close to an AC, with deviations arising from the admixture of a further level, which first and foremost affects the local curvature, i.e., the numerator in equation (6.3). But it also demonstrates that the fidelity change $S(\lambda)$ itself is still effective in detecting and characterizing the ACs.

6.3 Application to Complex Systems

6.3.1 Quantum Chaos Model

A highly dense spectrum with many and possibly overlapping ACs is encountered in quantum chaotic systems as described by Random Matrix Theory (RMT) [210]. A prime example having such a dense complex spectrum is the combination of two random matrices drawn from the Gaussian orthogonal ensemble (GOE) [210]

$$H(\lambda) = \cos(\lambda)H_1 + \sin(\lambda)H_2 \quad , \quad H_1, H_2 \in \mathrm{GOE} \quad . \tag{6.9}$$

The distribution of minimal distances c at the ACs (normalized to unit mean) is then given by a Gaussian distribution $P(c) = (2/\pi)\exp[-c^2/\pi]$ [211]. Using our fidelity measure, we can directly detect the ACs in this system (by a numerical

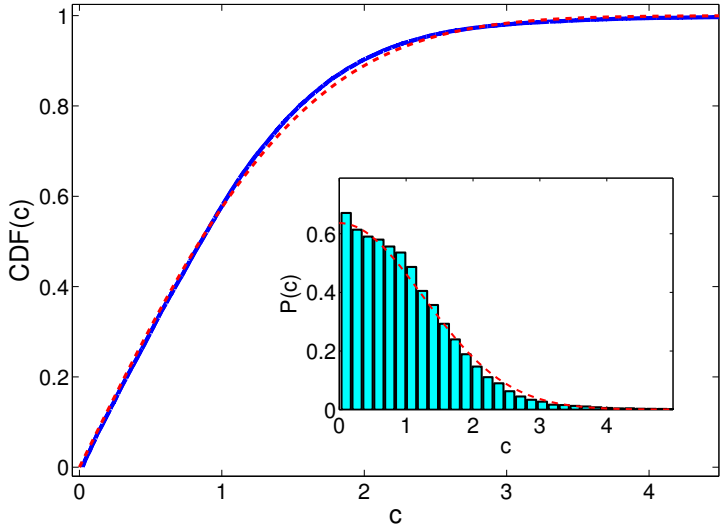


Figure 6.2: Cumulative distribution of ACs determined from the fidelity change maxima for the RMT model of equation (6.9) with dim $H_{1,2} = 1024$ and $\lambda \in [0, \pi[$ showing ca. 30,000 ACs: the numerical distribution (solid line) in excellent agreement with the RMT prediction $\mathrm{CDF}(c) = \mathrm{erf}(c/\sqrt{\pi})$ (dashed line). *Inset:* Distribution of widths of the ACs $P(c)$ (histogram) and the RMT prediction (dashed line).

search for maxima of the S-function) and estimate also their widths. In the vicinity of a local maximum, the S-function has a Lorentzian shape as in equation (6.2) even in very dense quantum chaotic spectra. Under this assumption, we can thus extract the width of the AC as $c = 2g = 1/\sqrt{2S_{\max}}$, c.f. equation (6.2), from the local maximum $S_{\max}$. Averaging over many ACs, the fidelity allows the verification of the RMT prediction with high accuracy. This is demonstrated in figure 6.2 for large random matrices.

6.3.2 Bose-Hubbard System

To further exemplify the value of our fidelity measure, we apply it to a one-dimensional Bose-Hubbard Hamiltonian with additional Stark force [193, 196, 197, 220]. This example of a many-body Wannier-Stark system can be realized with ultracold atoms in optical lattices and the relevant parameters may be changed using well-known experimental techniques [30, 31]. This model describes N particles on L lattice sites, with hopping between adjacent sites and a local on-site interaction. As exemplified in [193, 196, 197], a gauge transformation into the force accelerated frame of reference turns a constant Stark force into a time-dependent phase $\exp(\pm iFt)$ with periodicity $T_B = 2\pi/F$ (the Bloch period). The corresponding Hamiltonian reads

$$\hat{H}(t) = -\frac{J}{2}\sum_{l=1}^{L}(\mathrm{e}^{iFt}a_{l+1}^{\dagger}a_l + \mathrm{H.c.}) + \frac{U}{2}\sum_{l=1}^{L}\hat{n}_l(\hat{n}_l - 1) \quad , \tag{6.10}$$

where $a_l^{\dagger}$ (a_l) creates (annihilates) a boson at site l and $\hat{n}_l = a_l^{\dagger}a_l$ is the number of bosons at site l. The parameter J is the hopping matrix element, U the interaction energy for two atoms occupying the same site, and F the Stark force. Periodic boundary conditions are imposed for $\hat{H}(t)$, such that the Hamiltonian and the one-

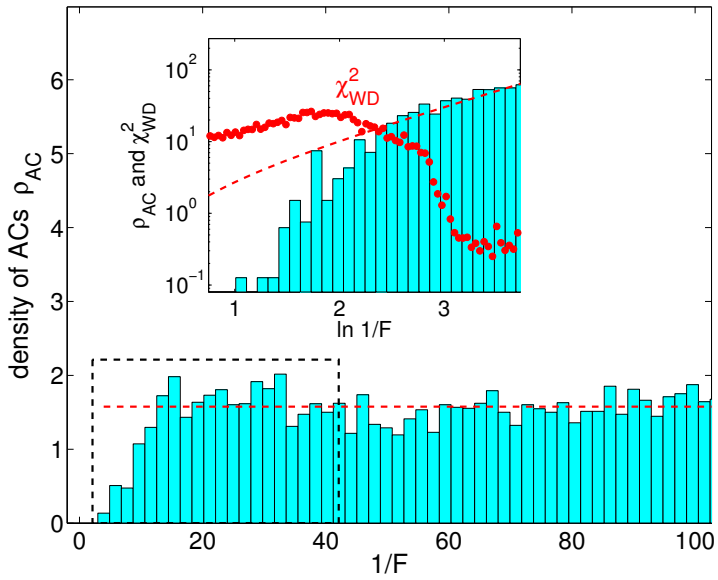


Figure 6.3: Density of ACs ρ_{AC} in the quasienergy spectrum of the Floquet operator of our Bose-Hubbard model for varying F and fixed $J = 0.038$, $U = 0.032$, $N = L = 6$. The number of ACs as detected by the fidelity change increases with $1/F$ and saturates around $1/F \approx 20$ to an average value which is shown by the dashed line. *Inset:* Magnification of the region marked by the box on logarithmic scale with a comparison to a χ^2 test (with small values for good Wigner-Dyson statistics [196, 197]).

period Floquet operator $\hat{U}_F(T_B) = \mathcal{T}\exp\left(-i\int_0^{T_B}\hat{H}(t)\mathrm{d}t\right)$ (where $\mathcal{T}$ denotes time-ordering) decompose into a sum of operators for specific quasimomenta κ [193]. In the following, we use F as a control parameter. For $J \approx U \ll F$ the quasienergy spectrum (eigenphases of $\hat{U}_F(T_B)$) is dominated by the force F and the system is regular. Decreasing the force to $J \approx U \gtrsim F$ the quasienergy spectrum reorders and the coupling between the levels becomes more important. For fillings of order unity, e.g. $N/L \approx 1$, the system is quantum chaotic in this regime and the spectrum obeys Wigner-Dyson statistics [193, 196, 197]. As F is varied one observes an increasing number of ACs as the spectrum is changing and additionally many broad ACs once the quantum chaotic region is reached.

To illustrate the crossover between regions with few and many ACs, we study the density of ACs as detected by the fidelity change S_n, when changing the system parameter λ. In a histogram, the density $\rho_{\mathrm{AC}}(\lambda)$ is defined via $\rho_{\mathrm{AC}}(\lambda)\cdot\mathrm{d}\lambda \equiv N_{\mathrm{AC}}(\lambda)/\dim\mathcal{H}$, comparing the number of ACs $N_{\mathrm{AC}}(\lambda)$ in the interval $[\lambda, \lambda+\mathrm{d}\lambda]$ to the total number of energy levels $\dim\mathcal{H}$. This is shown in the main part of figure 6.3 where we observe no ACs at large F, i.e., small values of $1/F$, and an increasing number of ACs for larger values of $1/F$ that saturates around $1/F \approx 20$.

The mentioned transition between regular and chaotic spectral properties for $J \approx U \gtrsim F$ and approximately integer filling in the tilted system can be visualized by comparing the actual level spacing distribution to a Wigner-Dyson distribution using a standard statistical χ^2 test [196, 197]. This is displayed in the inset of figure 6.3 along with the density of ACs in figure 6.3. The fidelity change $S(1/F)$ detects ACs and shows the same qualitative behavior as the spectral statistics along the crossover from regular to chaotic dynamics: in regions of good Wigner-Dyson

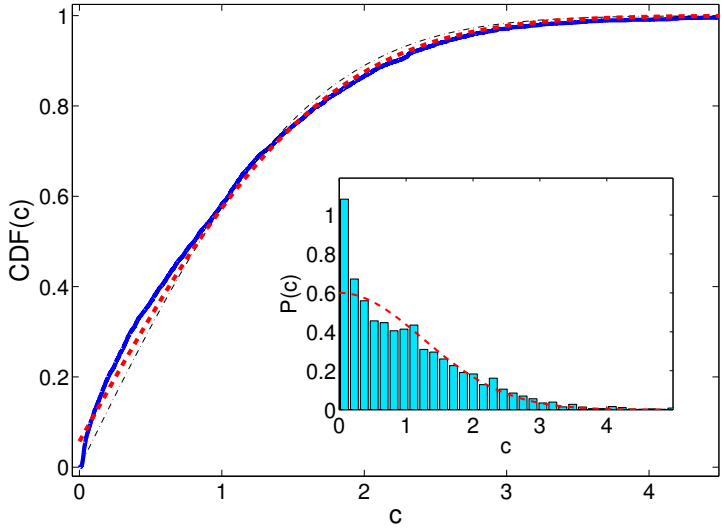


Figure 6.4: Cumulative distribution of ACs determined from the fidelity change maxima for the system of equation (6.10). Shown are the numerical distribution (solid line), the best fit for a mixed RMT spectrum (thick dashed line, chaotic part $\gamma \approx 0.94$), and the RMT prediction for a purely chaotic spectrum (thin dashed-dotted line). Parameters: $N = 6, L = 7, J = 0.038, U = 0.032, F = 1/39 \ldots 1/35$. *Inset:* Distribution of widths of ACs for the same model (histogram) and equation (6.11) with $\gamma \approx 0.94$ (dashed line). The enhancement close to $c = 0$ arises from regular "solitonic" states [221] in the spectrum.

statistics we find a high density of ACs compared to a smaller number of ACs in the regular regime. The crossover beginning for $\log(1/F) \approx 2$, where the density of ACs rises above unity, i.e., on average each energy level undergoes more than one AC in the unit interval. The transition is complete for $\log(1/F) \approx 3$ where the χ^2 test saturates around a low value. However, the density of ACs alone is not able to distinguish regular from chaotic dynamics. Instead the ACs need to have a broad distribution of widths which is reflected in the distribution $P(c)$ introduced above.

By using the fidelity change in order to detect and characterize ACs, we can resolve further remarkable details in the full spectrum. With this method we are, e.g., able to detect a small number of regular states [221] traversing the chaotic sea of energy levels in the chaotic regime of the tilted Bose-Hubbard model. In this case the distribution of widths of ACs is a mixture of regular and quantum chaotic distributions:

$$P(c) = (1-\gamma)\delta(c) + \frac{2\gamma^2}{\pi\bar{c}} \exp\left[-\frac{\gamma^2 c^2}{\pi\bar{c}^2}\right] \quad , \tag{6.11}$$

with a chaotic part of weight $0 \leq \gamma \leq 1$ [222]. A finite regular component makes itself visible as a strong enhancement of $P(c)$ close to zero, c.f., the inset of figure 6.4.

We are able to estimate the size of this component by analyzing the cumulative distribution function $\mathrm{CDF}(c) = 1 - \gamma + \gamma \ \mathrm{erf}\left(\frac{\gamma c}{\sqrt{\pi}}\right)$. The result is shown in the main part of figure 6.4, where we plot the numerically obtained distribution and the best χ^2-fit including a finite regular component. We obtain a chaotic part of $\gamma \approx 0.94$, corresponding to ca. 6% of regular levels, in good agreement with counting 7 regular levels out of 132 by direct inspection of the spectrum. Except for the identification of single regular levels [221], this has so far not been detected in the

tilted Bose-Hubbard model by other statistical measures. The reported results are obtained for periodic boundary conditions applied to the Hamiltonian of equation (6.10), but we found a qualitatively similar picture for hard-wall boundary conditions, as used in reference [221]. Our results underline the value of fidelity as a measure for detecting ACs with high resolution in energy spectra.

6.4 Conclusions

We showed that quantum fidelity is perfectly suited to detect and characterize ACs in the energy spectrum. It therefore connects information about the wave function of a system with its spectrum, without direct reference to the energy levels by using only the overlap of wave functions.[2] This has been exemplified for simple models and for complex quantum systems showing many ACs. The fidelity, therefore, proves very useful to study many-body systems, also beyond their ground-state properties [55, 198, 199].

We expect a clear advantage of the fidelity change compared to spectral statistics in the sense that it can be applied, in principle, also just *locally* in the spectrum. This means that, if one is interested only in local spectral properties of a system, it is sufficient to follow a small number of levels to characterize the behavior of a system. For larger systems, computing the entire spectrum and all eigenstates is in general difficult, but the fidelity allows an analysis of parts of the spectrum providing local spectral information. To make use of this advantage, one may resort to numerical algorithms optimized to access just a subset of eigenstates, e.g., the Lanczos algorithm [224, 225].

[2] A similar connection, yet in the deep semiclassical regime, between spectral properties and fidelity is identified in reference [223].

Appendix A

Generalized Householder Transformations for the Complex Symmetric Eigenvalue Problem

In this appendix, we present results which are not directly related to this thesis. They were nevertheless obtained during the Ph.D. and we include them here for completeness.

Efficient diagonalization routines are of fundamental importance for the analysis of quantum systems. In some cases, one encounters the necessity to diagonalize non–Hermitian matrices for which algorithms provided by standard libraries are not optimal. Here, we develop an intuitive and scalable algorithm for the diagonalization of complex symmetric matrices, which can arise from the projection of pseudo–Hermitian and complex scaled Hamiltonians onto a suitable basis set of "trial" states. The algorithm is based on generalized Householder transformations and relies on iterative similarity transformations $T \to T' = Q^{\mathrm{t}}\, T\, Q$, where Q is a complex and orthogonal, but not unitary, matrix, i.e, $Q^{\mathrm{t}} = Q^{-1}$ but $Q^{\dagger} \neq Q^{-1}$. We present numerical reference data to support the scalability of the algorithm. We construct the generalized Householder transformations from the notion that the conserved scalar product of eigenstates Ψ_n and Ψ_m of a pseudo–Hermitian quantum mechanical Hamiltonian can be reformulated in terms of the generalized indefinite inner product $\int \mathrm{d}x\, \Psi_n(x,t)\, \Psi_m(x,t)$, where the integrand is locally defined, and complex conjugation is avoided. A few example calculations are described which illustrate the physical origin of the ideas used in the construction of the algorithm. This chapter is published in reference [57].

A.1 Introduction

Complex symmetric matrices $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{t}}$ arise naturally from the projection of a complex scaled ("resonance-generating") Hamiltonian onto a basis of quantum mechanical "trial" states. For suitably chosen parameters, the diagonalization of a matrix of this type leads to (accurate) approximations for the resonance energies,

and the resonance eigenstates, of the complex scaled Hamiltonian [226–229]. The resonance energies are manifestly complex; the width of the quantum state enters the complex energy as $E = \mathrm{Re}\, E - \frac{i}{2}\Gamma$, where Γ is the width (we use natural units with $\hbar = c = \epsilon_0 = 1$ throughout this chapter). A paradigmatic example [226] is the complex-scaled cubic anharmonic oscillator Hamiltonian $h_3 = -\frac{1}{2}\exp(-2i\theta)\,\partial_x^2 + \frac{1}{2}\exp(2i\theta)\,x^2 + G\exp(3i\theta)\,x^3$, where G is a coupling parameter and $0 < \theta < \pi/5$ is a complex rotation angle.

However, complex scaled Hamiltonians are not the only source of complex symmetric matrices in theoretical physics. For example, if one projects the pseudo–Hermitian ($\mathcal{PT}$-symmetric) anharmonic oscillator [230, 231] Hamiltonian $H_3 = -\frac{1}{2}\,\partial_x^2 + \frac{1}{2}\,x^2 + i\,G\,x^3$ with an imaginary cubic perturbation ($G > 0$) onto a basis of harmonic oscillator eigenstates, then one obtains a complex symmetric (but not Hermitian) matrix, the eigenvalues of which are real.

As shown in references [227, 232], the complex resonance energies of the "real" cubic anharmonic oscillator h_3 are connected with the real eigenenergies of the imaginary cubic perturbation H_3 via a dispersion relation. The same holds true for all anharmonic oscillators of odd degree. Some of the numerical calculations were instrumental in providing additional evidence for the generalization [227] of the so-called Bender–Wu formulas [233, 234], which describe the large-order asymptotic growth of the perturbative coefficients of an arbitrary energy eigenvalue of an even anharmonic oscillator, to odd anharmonic oscillators. Indeed, the conjectures on nonperturbative quantization conditions, described in references [227, 228] had been checked against high-precision numerical data before the results were presented.

The purpose of this chapter is threefold: First, to illustrate the numerical procedures underlying the numerical verification of the conjectured generalized quantization conditions, second, to describe an intuitive and scalable (in terms of the numerical precision) matrix diagonalization algorithm which seems to be particularly suited for the treatment of complex symmetric matrices. Our algorithm has a certain "twist" in the sense that it is based on generalized Householder transformations. The generalized Householder matrices $\boldsymbol{H}_v$ have manifestly complex entries but are not Hermitian unlike the familiar formalism (see p. 225 of reference [235]). Instead, they are orthogonal matrices with the property $\boldsymbol{H}_v\,\boldsymbol{H}_v^{\mathrm{t}} = \mathbb{1}$. To the best of our knowledge, these generalized Householder reflections have not appeared in the standard literature [92, 236–239] on matrix diagonalization procedures before.

Finally, the third purpose of the chapter is to illustrate a few properties of the eigenenergies of pseudo–Hermitian Hamiltonians, based on calculations done with the algorithm presented here. Let us anticipate one of the observations made in the course of the calculations. At face value, the conserved scalar product [240] (under the time evolution governed by a pseudo–Hermitian anharmonic oscillator) is given as the integral $\int \mathrm{d}x\,\Psi_n^*(-x,t)\,\Psi_m(x,t)$; this expression involves a nonlocal integrand with function evaluations at x and $-x$. Typically, eigenstates of complex symmetric matrices are orthogonal with regard to a conceptually much simpler scalar product, namely, the indefinite inner product $\int \mathrm{d}x\,\Psi_n(x,t)\,\Psi_m(x,t)$, where the integrand is locally defined, and complex conjugation is avoided. However, the two scalar products are related, as described in this chapter, and this observation has motivated the construction of the matrix diagonalization algorithm presented here.

A.1.1 Reader's Guide

We thus proceed by describing the algorithm in section A.2 and the physical motivation for its development in section A.3, together with a few example calculations. Conclusions are reserved for section A.4.[1]

A.2 Complex Symmetric Eigenvalues and Eigenvectors

The quantum mechanics considered here is formulated in an infinite-dimensional vector space (Hilbert space) of functions. However, once a numerical evaluation of eigenvalues of a particular Hamiltonian is pursued, the quantum mechanical Hamiltonian needs to be projected onto a suitable basis set of wave functions, leading to a finite-dimensional matrix.

When a pseudo–Hermitian Hamiltonian such as the imaginary cubic oscillator is projected onto a basis set consisting of harmonic-oscillator eigenfunctions, one obtains a complex symmetric (not Hermitian!) matrix. Typically, the generalized indefinite inner product [241–243] naturally emerges as a tool in the analysis of pseudo–Hermitian quantum mechanics, and one would thus naturally assume that the indefinite inner product might be useful in the development of a suitable matrix diagonalization algorithm. Indeed, the approximate calculation of eigenvalues of quantum mechanical Hamiltonians naturally emerges as a task in the analysis of the quantum dynamics induced by the pseudo–Hermitian time evolution.

Such an algorithm will be described in the following; it essentially relies on two steps: In the first, the complex symmetric input matrix is transformed to tridiagonal form, and in the second step, the tridiagonal matrix is diagonalized to machine accuracy. The first step uses the concept of the complex inner product in an absolutely essential manner; it is based on generalized Householder reflection matrices. The complex symmetric input matrix is transformed to tridiagonal form in a single computation whose computational cost is of order n^2 (here, "tridiagonal form" refers to a form where the diagonal, as well as the sub- and superdiagonal entries of the matrix are nonzero). For the second step, one has a number of methods available; we shall briefly outline a method based on iterative QL decompositions with implicit (Wilkinson) shifts.

A.2.1 Tridiagonalization

We first define the indefinite inner product for finite-dimensional n-vectors as follows,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_* = \sum_{i=1}^{n} x_i \, y_i = \boldsymbol{x}^{\mathrm{t}} \cdot \boldsymbol{y} \,, \tag{A.1}$$

where $\boldsymbol{x}^{\mathrm{t}}$ is a row vector, whereas $\boldsymbol{y}$ is a column vector. Note that the entries of $\boldsymbol{x}$ and $\boldsymbol{y}$ may be complex numbers, but complex conjugation of either $\boldsymbol{x}$ or $\boldsymbol{y}$ is avoided. We use generalized Householder reflection matrices $\boldsymbol{H}_v$, which have the

[1] Notice that matrices and vectors are bold here for better readability.

properties,

$$\boldsymbol{H}_v = \mathbb{1} - \frac{2}{\langle v, v\rangle_*}\, v \otimes v^{\mathrm{t}}\,,\ \boldsymbol{H}_v\ \boldsymbol{x} = \boldsymbol{x} - 2\boldsymbol{u}\,\langle \boldsymbol{u}, \boldsymbol{x}\rangle_*\ ,\ \boldsymbol{u} = \frac{\boldsymbol{v}}{|\boldsymbol{v}|_*}\,,\ |\boldsymbol{v}|_* = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v}\rangle_*}\,. \tag{A.2}$$

Here, the dyadic product of a column and a row vector is denoted by the symbol $\otimes$, and the branch cut of the square root function in the calculation of $|v|_*$ is along the negative real axis.

The generalized Householder reflection matrix $\boldsymbol{H}_v$ is symmetric, i.e., $\boldsymbol{H}_v = \boldsymbol{H}_v^{\mathrm{t}}$. Furthermore, the Householder reflections are square roots of the unit matrix,

$$\boldsymbol{H}_v^2 = \boldsymbol{H}_v\ \boldsymbol{H}_v^{\mathrm{t}} = \boldsymbol{H}_v^{\mathrm{t}}\ \boldsymbol{H}_v = \mathbb{1} - 4\boldsymbol{u} \otimes \boldsymbol{u}^t + 4\,\boldsymbol{u} \otimes \boldsymbol{u}^t\ \langle \boldsymbol{u}, \boldsymbol{u}\rangle_* = \mathbb{1}. \tag{A.3}$$

A characteristic property of the Householder reflections is that the parameter vector v can be adjusted so that the input vector $\boldsymbol{x}$ is projected onto a particular axis, upon the calculation of $\boldsymbol{H}_v\ \boldsymbol{y}$. We set

$$\boldsymbol{v} = \boldsymbol{y} + |\boldsymbol{y}|_*\ \hat{\boldsymbol{e}}_n\,, \tag{A.4}$$

where $\hat{\boldsymbol{e}}_n$ is the "last" unit vector in the n-dimensional space, and we verify that

$$\begin{aligned}\boldsymbol{H}_v\ \boldsymbol{y} &= \boldsymbol{y} - \frac{2}{\langle \boldsymbol{v}, \boldsymbol{v}\rangle_*}\ \langle \boldsymbol{v}, \boldsymbol{y}\rangle_*\ \boldsymbol{v} = \boldsymbol{y} - \frac{2\left(\boldsymbol{y}^{\mathrm{t}} + |\boldsymbol{y}|_*\ \hat{\boldsymbol{e}}_n^{\mathrm{t}}\right)\cdot \boldsymbol{y}}{\left(\boldsymbol{y}^{\mathrm{t}} + |\boldsymbol{y}|_*\ \hat{\boldsymbol{e}}_n^{t}\right)\cdot\left(\boldsymbol{y} + |\boldsymbol{y}|_* \hat{\boldsymbol{e}}_n\right)}\boldsymbol{v} \\ &= \boldsymbol{y} - \frac{2\left(|\boldsymbol{y}|_*^2 + |\boldsymbol{y}|_* y_n\right)}{2|\boldsymbol{y}|_*^2 + 2|\boldsymbol{y}|_* y_n}\boldsymbol{v} = \boldsymbol{y} - \boldsymbol{v} = -|\boldsymbol{y}|_* \hat{\boldsymbol{e}}_n\,.\end{aligned} \tag{A.5}$$

Furthermore,

$$\boldsymbol{y}^{\mathrm{t}}\boldsymbol{H}_v = (\boldsymbol{H}_v \boldsymbol{y})^{\mathrm{t}} = -|\boldsymbol{y}|_* \hat{\boldsymbol{e}}_n^{\mathrm{t}}\,. \tag{A.6}$$

The results from equations (A.5) and (A.6) are useful in the tridiagonalization procedure. Let $\boldsymbol{A}$ be the matrix we want to tridiagonalize. In the first step, we choose the column vector $\boldsymbol{y}_{n-1}$ to consist of the first $n-1$ elements of the last column of $\boldsymbol{A}$,

$$\boldsymbol{y}_{n-1} = \begin{pmatrix} A_{1\,n} \\ A_{2\,n} \\ \vdots \\ A_{n-1\,n} \end{pmatrix}. \tag{A.7}$$

By defining $\boldsymbol{B}_{n-1}$ as an $(n-1)\times(n-1)$ matrix where $B_{ij} = A_{ij}$, for $i, j = 1, \ldots, n-1$, we can write $\boldsymbol{A}$ as

$$\boldsymbol{A} = \left(\begin{array}{c|c} \boldsymbol{B}_{n-1} & \boldsymbol{y}_{n-1} \\ \hline \boldsymbol{y}_{n-1}^{\mathrm{t}} & A_{nn} \end{array}\right). \tag{A.8}$$

In the spirit of the Householder reflection, we set

$$\boldsymbol{v}_{n-1} = \boldsymbol{y}_{n-1} + |\boldsymbol{y}_{n-1}|_*\ \hat{\boldsymbol{e}}_{n-1}\,. \tag{A.9}$$

We can then construct $\boldsymbol{H}_{v_{n-1}}$, which will be a Householder matrix of rank $n-1$. The complex $n \times n$ matrix $\boldsymbol{H}_{n-1}$ is defined as

$$\boldsymbol{H}_{n-1} = \left(\begin{array}{c|c} \boldsymbol{H}_{v_{n-1}} & \boldsymbol{0} \\ \hline \boldsymbol{0}^{\mathrm{t}} & 1 \end{array}\right). \tag{A.10}$$

Then,

$$\boldsymbol{A}' = \boldsymbol{H}_{n-1}\, \boldsymbol{A}\, \boldsymbol{H}_{n-1} = \left(\begin{array}{c|c} \boldsymbol{H}_{\boldsymbol{v}_{n-1}} \boldsymbol{B}_{n-1} \boldsymbol{H}_{\boldsymbol{v}_{n-1}} & \boldsymbol{H}_{\boldsymbol{v}_{n-1}} \boldsymbol{y}_{n-1} \\ \hline \boldsymbol{y}_{n-1}^{\mathrm{t}} \boldsymbol{H}_{\boldsymbol{v}_{n-1}} & A_{nn} \end{array}\right). \tag{A.11}$$

Using equation (A.5) and (A.6), this can be reduced to

$$\boldsymbol{A}' = \left(\begin{array}{ccc|c} & & & 0 \\ & \boldsymbol{B}'_{n-1} & & \vdots \\ & & & -|\boldsymbol{y}_{n-1}|_* \\ \hline 0 & \cdots & -|\boldsymbol{y}_{n-1}|_* & A_{nn} \end{array}\right), \tag{A.12}$$

where

$$\boldsymbol{B}'_{n-1} = \boldsymbol{H}_{\boldsymbol{v}_{n-1}} \boldsymbol{B}_{n-1} \boldsymbol{H}_{\boldsymbol{v}_{n-1}}. \tag{A.13}$$

For the second step we choose $\boldsymbol{y}_{n-2}$ to be the first $n-2$ elements of the second to last column of $\boldsymbol{A}'$,

$$\boldsymbol{y}_{n-2} = \begin{pmatrix} A'_{1\,n-1} \\ A'_{2\,n-1} \\ \vdots \\ A'_{n-2\,n-1} \end{pmatrix}, \quad \boldsymbol{v}_{n-2} = \boldsymbol{y}_{n-2} + |\boldsymbol{y}_{n-2}|_* \, \hat{\boldsymbol{e}}_{n-2}. \tag{A.14}$$

The Householder matrix $\boldsymbol{H}_{\boldsymbol{v}_{n-2}}$ is of rank $n-2$, and $\boldsymbol{H}_{n-2}$ is defined as

$$\boldsymbol{H}_{n-2} = \left(\begin{array}{c|c} \boldsymbol{H}_{\boldsymbol{v}_{n-2}} & \boldsymbol{0} \\ \hline \boldsymbol{0}^{\mathrm{t}} & \mathbb{1}_{2\times 2} \end{array}\right). \tag{A.15}$$

We write $\boldsymbol{A}'$ as follows,

$$\boldsymbol{A}' = \left(\begin{array}{c|c|c} \boldsymbol{B}_{n-2} & \boldsymbol{y}_{n-2} & \boldsymbol{0} \\ \hline \boldsymbol{y}_{n-2}^{\mathrm{t}} & A'_{n-1\,n-1} & -|\boldsymbol{y}_{n-1}|_* \\ \hline \boldsymbol{0}^{\mathrm{t}} & -|\boldsymbol{y}_{n-1}|_* & A'_{nn} \end{array}\right), \tag{A.16}$$

where we observe that $A'_{nn} = A_{nn}$. We then calculate $\boldsymbol{A}''$ using the similarity transformation

$$\boldsymbol{A}'' = \boldsymbol{H}_{n-2}\, \boldsymbol{A}'\, \boldsymbol{H}_{n-2}. \tag{A.17}$$

The result is

$$\boldsymbol{A}'' = \left(\begin{array}{ccc|c|c} & & & 0 & \\ & \boldsymbol{B}'_{n-2} & & \vdots & \boldsymbol{0} \\ & & & -|\boldsymbol{y}_{n-2}|_* & \\ \hline 0 & \cdots & -|\boldsymbol{y}_{n-2}|_* & a'_{n-1\,n-1} & -|\boldsymbol{y}_{n-1}|_* \\ \hline & \boldsymbol{0}^{\mathrm{t}} & & -|\boldsymbol{y}_{n-1}|_* & A'_{nn} \end{array}\right), \tag{A.18}$$

where

$$\boldsymbol{B}'_{n-2} = \boldsymbol{H}_{v_{n-2}}\, \boldsymbol{B}_{n-2}\, \boldsymbol{H}_{v_{n-2}}\,. \tag{A.19}$$

A total of $n-2$ iterations of this process leads to a tridiagonal matrix $\boldsymbol{T}$, where

$$\boldsymbol{T} = \boldsymbol{Z}^{-1}\, \boldsymbol{A}\, \boldsymbol{Z}\,, \tag{A.20}$$

with

$$\boldsymbol{Z} = \boldsymbol{H}_{n-1}\, \boldsymbol{H}_{n-2} \ldots \boldsymbol{H}_2\,, \qquad \boldsymbol{Z}^{-1} = \boldsymbol{Z}^{\mathrm{t}} = \boldsymbol{H}_2\, \boldsymbol{H}_3 \ldots \boldsymbol{H}_{n-1}\,. \tag{A.21a}$$

In order to write a computationally efficient algorithm, it is helpful to observe that the explicit calculation of the $\boldsymbol{Z}$ matrix actually is unnecessary. Obviously, the only computationally nontrivial step in the iteration of the Householder transformations consists in the calculation of the matrix

$$\begin{aligned}\boldsymbol{B}' =& \boldsymbol{H}_v\, \boldsymbol{B}\, \boldsymbol{H}_v = \left(\mathbb{1} - 2\,\frac{\boldsymbol{v}\otimes\boldsymbol{v}^{\mathrm{t}}}{|\boldsymbol{v}|_*^2}\right)\, \boldsymbol{B}\, \left(\mathbb{1} - 2\,\frac{\boldsymbol{v}\otimes\boldsymbol{v}^{\mathrm{t}}}{|\boldsymbol{v}|_*^2}\right) \\ =& \boldsymbol{B} - \boldsymbol{v}\otimes\boldsymbol{u}^{\mathrm{t}} - \boldsymbol{u}\otimes\boldsymbol{v}^{\mathrm{t}} + 2q\,\boldsymbol{v}\otimes\boldsymbol{v}^{\mathrm{t}} = \boldsymbol{B} - \boldsymbol{v}\otimes\boldsymbol{w}^{\mathrm{t}} - \boldsymbol{w}\otimes\boldsymbol{v}^{\mathrm{t}}\,,\end{aligned} \tag{A.22}$$

where we skip a few algebraic steps in the derivation and use the definitions

$$p = \frac{1}{2}|\boldsymbol{v}|_*^2\,, \quad \boldsymbol{u} = \frac{\boldsymbol{B}\,\boldsymbol{v}}{p}\,, \quad q = \frac{\boldsymbol{v}^{\mathrm{t}}\cdot\boldsymbol{u}}{2p}\,, \quad \boldsymbol{w} = \boldsymbol{u} - q\boldsymbol{v}\,. \tag{A.23}$$

It is advantageous to calculate, for each iteration, the vector $\boldsymbol{v}$, then p, $\boldsymbol{u}$, q, $\boldsymbol{w}$ and finally $\boldsymbol{B}'$.

A.2.2 Diagonalization

The tridiagonal matrix $\boldsymbol{T}$ obtained in equation (A.20) is sparsely populated; the only nonvanishing entries are on the diagonal, the superdiagonal and the subdiagonal. It can be written in the form

$$\boldsymbol{T} = \begin{pmatrix} D_1 & E_1 & & & \\ E_1 & D_2 & E_2 & & \\ & E_2 & \ddots & \ddots & \\ & & \ddots & D_{n-1} & E_{n-1} \\ & & & E_{n-1} & D_n \end{pmatrix}. \tag{A.24}$$

In principle, a number of methods are available for the diagonalization of such sparsely populated matrices. One of these is based on QL factorization. In its most basic version [244,245], the QL factorization implements the similarity transformations by first calculating the decomposition of a symmetric triangular input matrix $\boldsymbol{T}$, as given by $\boldsymbol{T} = \boldsymbol{Q}\,\boldsymbol{L}$ where $\boldsymbol{Q}$ is an orthogonal matrix ($\boldsymbol{Q}^{\mathrm{t}} = \boldsymbol{Q}^{-1}$), and $\boldsymbol{L}$ is a lower diagonal matrix. One then implements the similarity transformations by simply calculating $\boldsymbol{T}' = \boldsymbol{L}\,\boldsymbol{Q} = \boldsymbol{Q}^{-1}\,\boldsymbol{T}\,\boldsymbol{Q}$. This corresponds to an iterative similarity transformation $\boldsymbol{T} = \boldsymbol{Q}\,\boldsymbol{T}'\,\boldsymbol{Q}^{-1} = \boldsymbol{Q}\,\boldsymbol{Q}'\,\boldsymbol{T}''\,\boldsymbol{Q}'^{-1}\boldsymbol{Q}^{-1}$ and so on. The plain QL factorization is known to be an efficient algorithm for wide classes of input matrices [92,238]. If the input matrix is triangular, one can show [236,246] that the rate of convergence in the Kth iteration goes as $(\lambda_i/\lambda_{i+1})^K$, for an ordered sequence of eigenvalues $|\lambda_1| < |\lambda_2| < \cdots < |\lambda_n|$ of an $n \times n$ input matrix. When the matrix $\boldsymbol{T}$ is diagonalized to machine accuracy, a fixed point of the similarity transformation is reached. For complex input matrices [239,247], the common form of the QL

decomposition calls for $\boldsymbol{Q}$ to be unitary ($\boldsymbol{Q}^{-1} = \boldsymbol{Q}^{\dagger}$) rather than complex and symmetric ($\boldsymbol{Q}^{-1} = \boldsymbol{Q}^{\mathrm{t}}$). Our $\boldsymbol{Q}$ matrices have the latter property and represent a slight generalization of the commonly accepted version of the QL decomposition.

We use a so-called Wilkinson shift in order to enhance the rate of convergence, as described in section 8.13 of reference [238]. The "implicit shift" involves a guess σ for a specific eigenvalue of $\boldsymbol{T}$, and the ensuing implementation of the similarity transformation $\boldsymbol{T} \rightarrow \boldsymbol{T}'$ is known as "chasing the bulge". One performs the decomposition on a matrix shifted by the guess for the eigenvalue,

$$\boldsymbol{T} - \sigma \, \mathbb{1}_{n\times n} = \boldsymbol{Q}\,\boldsymbol{L}\,, \tag{A.25a}$$

and uses the fact that

$$\boldsymbol{T}' = \boldsymbol{L}\,\boldsymbol{Q} + \sigma \, \mathbb{1}_{n\times n} = \boldsymbol{Q}^{-1}\,\boldsymbol{T}\,\boldsymbol{Q}\,. \tag{A.25b}$$

Indeed, in a computationally efficient algorithm, neither $\boldsymbol{Q}$ nor $\boldsymbol{L}$ are ever explicitly computed. One takes advantage of the fact that the similarity transformation (A.25) is equivalent to a series of Jacobi [248] and Givens [249] rotations, as described in the following.

In the first step, the implicit "Wilkinson" shift σ is calculated [236, 246] from one of the eigenvalues of the 2×2 matrix in the upper left corner of (A.24),

$$\begin{pmatrix} D_1 & E_1 \\ E_1 & D_2 \end{pmatrix}. \tag{A.26}$$

It reads as follows,

$$\sigma = \frac{D_2 + D_1}{2} \pm \sqrt{\left(\frac{D_2 - D_1}{2}\right)^2 + E_1^2}\,. \tag{A.27}$$

The $\pm$ sign is chosen such as to minimize the complex modulus $|\sigma - D_1|$ of the difference of σ and the diagonal entry D_1. The guess for the eigenvalue is calculated for the upper left corner of the input matrix, while, in the QL decomposition, the implicitly shifted Jacobi rotation $\boldsymbol{J}$ is a complex orthogonal (but not unitary) matrix which rotates the lower right corner of the tridiagonal input matrix as follows,

$$\boldsymbol{J} = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & c & s \\ & & & -s & c \end{pmatrix}, \qquad \boldsymbol{J}^{\mathrm{t}}\,\boldsymbol{J} = \mathbb{1}_{n\times n}\,, \tag{A.28}$$

with manifestly complex entries $c^2 + s^2 = 1$ (but in general $|c|^2 + |s|^2 \neq 1$),

$$c = \frac{D_n - \sigma}{\sqrt{(D_n - \sigma)^2 + E_{n-1}^2}}\,, \qquad s = \frac{E_{n-1}}{\sqrt{(D_n - \sigma)^2 + E_{n-1}^2}}\,. \tag{A.29}$$

The transformed matrix $\boldsymbol{T}' = \boldsymbol{J}^t\,\boldsymbol{T}\,\boldsymbol{J}$ has the form

$$\boldsymbol{T}' = \begin{pmatrix} \ddots & \ddots & & & \\ \ddots & D'_{n-3} & E'_{n-3} & & \\ & E'_{n-3} & D'_{n-2} & E'_{n-2} & F' \\ & & E'_{n-2} & D'_{n-1} & E'_{n-1} \\ & & F' & E'_{n-1} & D'_n \end{pmatrix} \tag{A.30}$$

with an obvious "bulge" (entry F) at the elements $T'_{n-2\,n} = T'_{n\,n-2} \neq 0$. The bulge can be "chased upward" using a generalized (complex and symmetric, but not Hermitian) Givens rotation,

$$\boldsymbol{G} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c & s & & \\ & & & -s & c & & \\ & & & & & 1 & \end{pmatrix}, \qquad \boldsymbol{G}^{\mathrm{t}}\,\boldsymbol{G} = \mathbb{1}_{n\times n}\,, \tag{A.31}$$

where again $c^2 + s^2 = 1$, and

$$c = \frac{E'_{n-1}}{\sqrt{E'^2_{n-1} + T'^2_{n\,n-2}}}\,, \qquad s = \frac{T'_{nn-2}}{\sqrt{E'^2_{n-1} + T'^2_{n\,n-2}}}\,. \tag{A.32}$$

The second transformation leads to $\boldsymbol{T}'' = \boldsymbol{G}^{\mathrm{t}}\,\boldsymbol{T}'\,\boldsymbol{G}$ with

$$\boldsymbol{T}'' = \begin{pmatrix} \ddots & \ddots & & & \\ \ddots & D''_{n-3} & E''_{n-3} & F'' & \\ & E''_{n-3} & D''_{n-2} & E''_{n-2} & \\ & F'' & E''_{n-2} & D''_{n-1} & E''_{n-1} \\ & & & E''_{n-1} & D''_{n} \end{pmatrix}. \tag{A.33}$$

Upon a Givens rotation, one updates the entries on the diagonal and sub-(super-)diagonal of the tridiagonal matrix $\boldsymbol{T} \to \boldsymbol{T}' \to \boldsymbol{T}''$. The additional element of the "bulge" can be stored as a single variable. After $n-2$ (Givens) $\boldsymbol{G}_j$ rotations with $j = n-2, \ldots 1$, starting from $\boldsymbol{G}_{n-2} \equiv \boldsymbol{G}$ and continuing to $\boldsymbol{G}_1$, the bulge has disappeared, and $\boldsymbol{T}$ again assumes a tridiagonal form. The orthogonal transformation $\boldsymbol{Q}$ from equation (A.25b) is identified as

$$\boldsymbol{Q} = \boldsymbol{J}\;\boldsymbol{G}_{n-2}\;\boldsymbol{G}_{n-3}\cdots\boldsymbol{G}_1\,. \tag{A.34}$$

In general, the convergence toward the eigenvalues in the Kth iteration is improved [236,246] to $[(\lambda_i - \sigma)/(\lambda_{i+1} - \sigma)]^K$, again for an ordered sequence of eigenvalues $|\lambda_1| < |\lambda_2| < \cdots < |\lambda_n|$. The similarity transformations are iterated until the off-diagonal element E_1 is zeroed to machine accuracy. One then repeats the process for the lower right $(n-1) \times (n-1)$ submatrix of $\boldsymbol{T}$, then, for the $(n-2) \times (n-2)$ submatrix of $\boldsymbol{T}$, each time zeroing the first off-diagonal element, until $\boldsymbol{T}$ is diagonalized to machine accuracy.

A.2.3 Numerical Reference Data

The algorithmic procedure described above leads to a matrix diagonalization algorithm for complex symmetric matrices, which can find both the eigenvalues and eigenvectors of the original input matrix $\boldsymbol{A}$. We have checked numerical results obtained for wide classes of $\mathcal{PT}$-symmetric anharmonic oscillators against numerous published data. An example of an interesting alternative procedure for the calculation of the eigenenergies is given by the moment method [250–252], which relies on a Fourier transformation of the Schrödinger equation and is based on a recursive calculation of the moments which define the series expansion of

the wave function in momentum space. Applied to the positive-definite measure $S(x) = |\psi(x)|^2$, the method has been shown to generate numerical approximations for the discrete states of the $-\mathrm{i}\,x^3$ non–Hermitian potential [253], as well as a potential proportional to $i\,x^3 + i\,\alpha\,x$, which induces $\mathcal{PT}$-symmetry breaking, manifestly complex eigenvalues [254,255]. The wave functions of eigenstates have also been studied, including Stokes and Anti-Stokes lines, for both the imaginary cubic oscillator [256] as well as generalized $(i\,x)^N$-potentials [257]. Furthermore, we have used the algorithm for the calculation of resonance and anti-resonance energies of the "real" cubic perturbation (potential proportional to x^3) and other odd anharmonic oscillators. We note that the eigenvalues of the $\mathcal{PT}$-symmetric imaginary cubic perturbation and the Hermitian, but not essentially self-adjoint real cubic oscillator are related by a dispersion relation [227,232].

For reference, let us consider the two Hamiltonians

$$h_3 = -\frac{1}{2}\,\mathrm{e}^{-2i\theta}\,\partial_x^2 + \frac{1}{2}\,\mathrm{e}^{2i\theta}\,x^2 + \mathrm{e}^{3i\theta}\,x^3\,, \qquad 0 < \theta < \frac{\pi}{5}\,, \tag{A.35a}$$

$$H_3 = -\frac{1}{2}\,\partial_x^2 + \frac{1}{2}\,x^2 + i\,x^3\,. \tag{A.35b}$$

The first of these involves a complex scaling transformation, which gives rise to manifestly complex resonance energy eigenvalues. The complex scaling transformation is "dual" to the resummation of the perturbation series to the complex resonance energies, which has been discussed in references [258–260]. Using a multi-precision arithmetic implementation [261] of the algorithm described in section A.2, we easily obtain the first two resonance energy eigenvalues of h_3 as follows,

$$\begin{aligned}\epsilon_0 = {} & 0.61288\,84333\,07754\,62425\,88175\,01988\,65141\,37333\,39788\,30718\,29420\,66181 \\ & -0.40859\,26669\,32267\,28315\,94988\,68767\,16051\,62709\,74834\,43840\,39990\,97532\,i\,,\end{aligned} \tag{A.36a}$$

$$\begin{aligned}\epsilon_1 = {} & 2.18041\,38375\,36348\,77123\,01619\,63541\,74113\,12471\,72136\,83505\,89744\,59041 \\ & -1.52620\,76556\,93032\,51000\,68539\,46967\,49562\,44459\,06099\,84880\,44103\,55220\,i\,.\end{aligned} \tag{A.36b}$$

All of the given decimals are significant; the 60-figure precision is obtained in a basis of roughly 1000 harmonic oscillator eigenstates and can be enhanced if desired. The ground-state energy of the Hamiltonian H_3 and its first-excited-state energy read as follows,

$$E_0 = 0.79734\,26075\,08906\,18903\,90809\,60791\,01316\,30972\,44534\,48033\,11575\,78578\,, \tag{A.37a}$$

$$E_1 = 2.77352\,49851\,95379\,71540\,58170\,00015\,53014\,23108\,48902\,82968\,52057\,22959\,. \tag{A.37b}$$

We again reemphasize that the precision of these results can be easily enhanced, as it was necessary to test some of the conjectured generalized quantization conditions for anharmonic oscillators presented in references [226–229]. For typical double-precision (16 decimals) and quadruple-precision (32 decimals) calculations, we observe that the timings for matrix diagonalization using our algorithm are comparable to those using the routine `ZGEEVX` built into the `LAPACK`

library [239]. Using a dedicated, concise implementation of the algorithm discussed in sections A.2.1 and A.2.2, we were even able to obtain timings which exceed the speed of LAPACK's by up to 50% for typical applications (matrices of rank 500×500); however, the speed-up may be compiler-specific (we were using gfortran version 4.7.2).[2]

A.3 Pseudo–Hermitian Quantum Mechanics

Let us try to explore the physical motivation for the construction of the matrix diagonalization algorithm presented in section A.2, on the basis of pseudo–Hermitian ($\mathcal{PT}$-symmetric) quantum mechanics. A $\mathcal{PT}$-symmetric Hamilton operator H fulfills the relation

$$H = \mathcal{P}\,\mathcal{T}\,H\,\mathcal{T}\,\mathcal{P} = \mathcal{P}\,H^\dagger\,\mathcal{P}\,, \tag{A.38}$$

where $H^\dagger$ is obtained [240] from H by the replacement $i \to -i$, which is the same as the Hermitian adjoint if all other terms in the Hamiltonian are explicitly real (rather than complex). The parity and time reversal are denoted as $\mathcal{P}$ and $\mathcal{T}$, respectively. If a Hamiltonian H fulfills a relation of the type $H = \eta^{-1}\,H^\dagger\,\eta$, then H is said to be pseudo–Hermitian [262]. $\mathcal{PT}$-symmetry can thus be interpreted as a special case of pseudo-Hermiticity ($\eta = \mathcal{P}$), even if there is a certain "clash" with the original definition from reference [262], where it was assumed that η is a positive-definite operator. By contrast, $\mathcal{P}$ may have the negative eigenvalue -1.

For reference, we continue our analysis with the well-known imaginary cubic perturbation [226–231], which we had already employed in section A.2.3. It is described by the Hamiltonian

$$H_3 = -\frac{1}{2}\,\partial_x^2 + \frac{1}{2}\,x^2 + i\,G\,x^3\,, \qquad G > 0\,, \tag{A.39}$$

which for $G = 1$ reduces to equation (A.35b). The eigenfunctions of H_3 are manifestly complex, in contrast to those of the quartic anharmonic oscillator,

$$H_4 = -\frac{1}{2}\,\partial_x^2 + \frac{1}{2}\,x^2 + g\,x^4\,, \qquad g > 0\,, \tag{A.40}$$

where the eigenstate wave functions can be chosen as purely real. For a $\mathcal{PT}$-symmetric system, the scalar product

$$\langle\psi(t)|\phi(t)\rangle_{\mathcal{PT}} \equiv \int \mathrm{d}x\,\psi^*(x,t)\,\mathcal{P}\phi(x,t) \tag{A.41}$$

is conserved under time evolution if both ψ and ϕ fulfill the time-dependent Schrödinger equation $i\partial_t\psi(t) = H_3\psi(t)$, and $i\partial_t\phi(t) = H_3\phi(t)$. However, the integrand in equation (A.41) is manifestly "nonlocal" because $\psi^*(x,t)\mathcal{P}\phi(x,t) = \psi^*(x,t)\phi(-x,t)$; it depends on function evaluations at x and $-x$. This is in contrast to the ordinary scalar product

$$\langle\psi(t)|\phi(t)\rangle \equiv \int \mathrm{d}x\,\psi^*(x,t)\,\phi(x,t)\,, \tag{A.42}$$

where the first argument is complex conjugated, and the (generalized) indefinite inner product

$$\langle\psi(t)|\phi(t)\rangle_* \equiv \int \mathrm{d}x\,\psi(x,t)\,\phi(x,t)\,, \tag{A.43}$$

[2] The gfortran compiler is available at http://hpc.sourceforge.net.

where none of the arguments are complex conjugated. The Hamiltonian H_3 given in equation (A.39) involves a manifestly complex potential, which we denote as $W(x)$,

$$W(x) = \frac{1}{2}\, x^2 + i\, G\, x^3 = V(x)\, \mathrm{e}^{i\, \arg(V(x))}\,, \tag{A.44a}$$

$$V(x) = |W(x)| = \sqrt{\tfrac{1}{4}\, x^4 + G^2\, x^6}\,. \tag{A.44b}$$

The modulus $V(x) = |W(x)|$ tends to infinity as $x \to \pm\infty$. For purely real potentials like the "confining" quartic oscillator given in equation (A.40), intuition suggests that the "bulk" of the probability density of the eigenstate wave function should be concentrated in the "classically allowed" region, i.e., in the region where the eigenenergy E is greater than the potential, i.e., $E > V(x)$ [where $V(x) \in \mathbb{R}$]. For a manifestly complex potential, the condition $E > W(x)$ [with $W(x) \in \mathbb{C}$] cannot be applied because the complex numbers are not ordered. In the following, we consider a few example calculations of energy eigenvalues of the imaginary cubic perturbation (A.39), which illustrate these observations. All of these have been accomplished using the algorithm presented in section A.2.

A.3.1 Example Calculations

We can formally split the Hamiltonian H_3 into a "real part" and an "imaginary part" as follows,

$$\mathrm{Re}\, H_3 = -\tfrac{1}{2}\, \partial_x^2 + \tfrac{1}{2} x^2\,, \qquad \mathrm{Im}\, H_3 = i\, G\, x^3\,. \tag{A.45}$$

Likewise, we can also split the eigenstate wave function $\psi_n(x)$ into real and imaginary parts,

$$\psi_n(x) = \mathrm{Re}\, \psi_n(x) + i\, \mathrm{Im}\, \psi_n(x)\,. \tag{A.46}$$

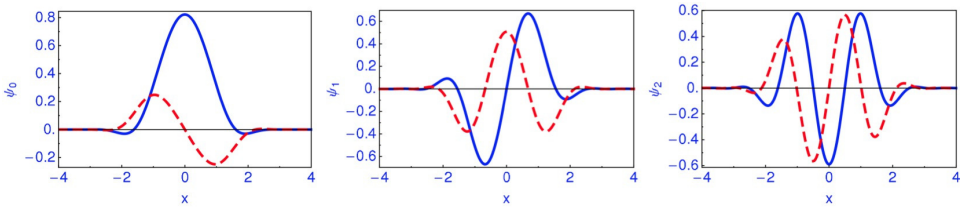
Based on the decomposition (A.45), one can show that if $\mathrm{Re}\, \psi_n(x)$ is even under parity and $\psi_n(x)$ is an eigenstate of H_3 with real eigenvalue of ϵ_n, then $\mathrm{Im}\, \psi_n(x)$ has to be parity-odd, and vice versa. Also, if $\mathrm{Re}\, \psi_n(x)$ is odd under parity, then $\mathrm{Im}\, \psi_n(x)$ has to be even under parity. Because the parity operator $\mathcal{P}$ does not commute with the Hamiltonian H_3, the eigenstates of H_3 are not eigenstates of parity. Furthermore, because the potential is manifestly complex, so are the wave functions. Yet, numerical evidence drawn from figure A.1 suggests that if the global phase of the wave function is appropriately chosen, both real as well as imaginary parts of the eigenstates are eigenstates of parity, individually. These eigenstates are naturally obtained if one diagonalizes an approximation to the cubic Hamiltonian obtained by projection onto a suitably large basis set of harmonic oscillator eigenstates.

The antilinear $\mathcal{PT}$ operator commutes with the Hamiltonian, and the eigenfunctions of H_3 are also $\mathcal{PT}$ eigenstates [240]. The precise eigenvalue of $\mathcal{PT}$ may, however, depend on the phase assigned to $\psi_n(x)$ because $\mathcal{T}$ is an antilinear operator. Let us first investigate the phase conventions used in figure A.1, where the real and imaginary parts of the wave function are, alternatingly, even and odd under parity as we proceed to higher excited states. The appropriate eigenvalues are thus

$$\mathcal{PT}\, \psi_n(x) = \psi_n^*(-x) = (-1)^n\, \psi(x)\,. \tag{A.47}$$

However, the eigenvalue of the wave functions $\Psi_n(x) = (-i)^n \psi_n(x)$ with respect to the $\mathcal{PT}$ operator is unity,

$$\mathcal{PT}[(-i)^n\, \psi_n(x)] = i^n\, \psi_n^*(-x) = [(-i)^n\, \psi(x)]\,. \tag{A.48}$$



(a) (b) (c)

Figure A.1: Figure (a) displays the ground-state wave function of the imaginary cubic Hamiltonian (A.39) for $G = 1.0$. This wave function is manifestly complex. The real part $\mathrm{Re}\,\psi_0(x)$ (even under parity) is plotted using solid lines, and the dashed curve corresponds to the parity-odd imaginary part $\mathrm{Im}\,\psi_0(x)$. For the first excited state (still, $G = 1.0$), the real part is odd, while the imaginary part is even under parity [see figure (b)]. The second excited state [figure (c)] has an even real part, while its imaginary part is odd. The global complex phase of the wave function is chosen so that the real part $\mathrm{Re}\,\psi_{n=0,1,2}(x)$ of the wave functions has the same qualitative behavior as the eigenstate wave function of the quartic oscillator displayed in figure A.2(a).

For two eigenstates of the $\mathcal{PT}$-symmetric system, the conserved $\mathcal{PT}$-symmetric scalar product simply is the indefinite inner product, as defined in equation (2.4.2) of reference [242],

$$\begin{aligned}\langle \Psi_n | \Psi_m \rangle_{\mathcal{PT}} &= \int \mathrm{d}x\, \Psi_n^*(x) \mathcal{P}\Psi_m(x) = \int \mathrm{d}x\, \left(\mathcal{PT}\Psi_n\right)(x)\, \Psi_m(x) \\ &= \int \mathrm{d}x\, \Psi_n(x)\, \Psi_m(x) = \langle \Psi_n | \Psi_m \rangle_* \,. \end{aligned} \tag{A.49}$$

The integrand in equation (A.49) is "local" in the sense that it depends only on wave functions at x, not $-x$. The non-local character of the integrand in equation (A.41) has otherwise been called into question and has given rise to rather sophisticated attempts at finding an alternative, appropriate interpretation [263]. The natural normalization condition for eigenstates of complex, symmetric matrices (including infinite-dimensional matrices) is given by [242]

$$\langle \psi_n | \psi_m \rangle_* = \delta_{nm}\,, \qquad \langle \Psi_n | \Psi_m \rangle_* = (-1)^n\, \delta_{nm}\,, \tag{A.50}$$

and involves the indefinite inner product defined in equation (A.43). The equivalence shown in equation (A.49) and the orthogonality properties (A.50) provide the main motivation for the construction of the generalized Householder transformation (A.2); indeed, the generalized inner product (A.49) reduces to (A.1) for finite-dimensional vector spaces. "Half" of the Ψ_n eigenstates acquire a negative $\mathcal{PT}$-symmetric norm, as described by the prefactor $(-1)^n$. For two time-dependent states of the $\mathcal{PT}$-symmetric system, given as

$$\chi(t) = \sum_n a_n(t)\, |\Psi_n\rangle\,, \quad \rho(t) = \sum_m b_m(t)\, |\Psi_m\rangle\,, \tag{A.51}$$

with $a_n = (-1)^n \langle \Psi_n | \chi \rangle$, and $b_n = (-1)^n \langle \Psi_n | \rho \rangle$, the $\mathcal{PT}$-symmetric scalar product is calculated as follows,

$$\begin{aligned}\langle \chi(t) | \rho(t) \rangle_{\mathcal{PT}} &= \sum_{nm} a_n^*(t)\, b_m(t) \int \mathrm{d}x\, \Psi_n^*(x) \mathcal{P}\Psi_m(x) = \sum_{nm} a_n^*(t)\, b_m(t) \int \mathrm{d}x\, \Psi_n(x) \Psi_m(x) \\ &= \sum_n (-1)^n\, a_n^*(t)\, b_n(t) = \langle \vec{a}(t) | \vec{b}(t) \rangle_{\mathrm{even}} - \langle \vec{a}(t) | \vec{b}(t) \rangle_{\mathrm{odd}}\,. \end{aligned} \tag{A.52}$$

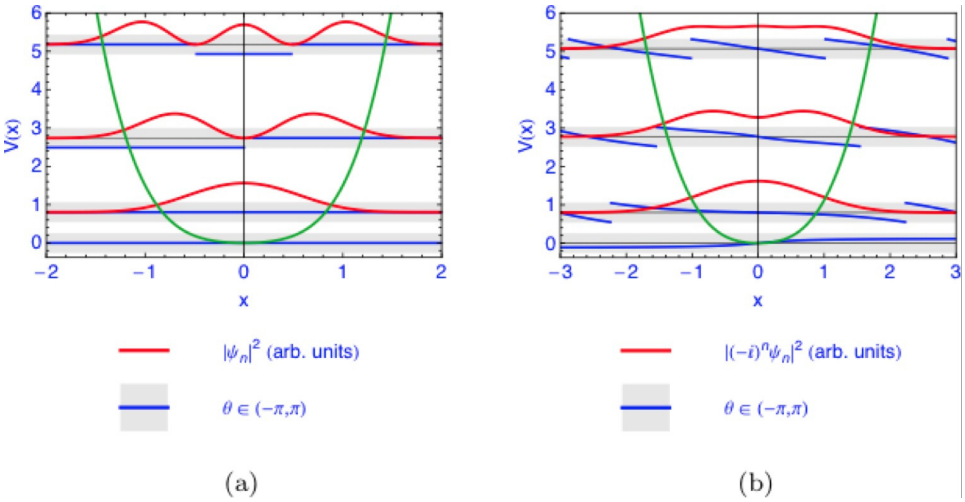


Figure A.2: In figure (a), we plot the probability density $\rho = |\psi(x)|^2$ of the quartic oscillator's ground state and first two excited states, in a potential $V(x) = \frac{1}{2}x^2 + g\,x^4$ with $g = 1.0$. Although the wave functions of the quartic potential are purely real, we use a modulus-phase plot for the real wave functions in figure (a). A sign change then corresponds to a jump in the complex phase from zero (for a positive real number) to $-\pi$ (for a negative real number). In figure (b), we give a modulus-phase plot of the eigenstate wave functions of the imaginary cubic perturbation, where the complex phase of the wave function is displayed in the shaded region. The complex phase $\theta = \theta(x)$ in the decomposition $(-i)^n\,\psi_n(x) = |\psi_n(x)|\,\exp[i\,\theta(x)]$ covers the interval $[-\pi, \pi)$. Here, the ψ_n are the wave functions of figure A.1, multiplied by a phase factor $(-i)^n$ [see also equation (A.48)].

with an obvious identification of $\langle\vec{a}(t)|\vec{b}(t)\rangle_{\text{even}}$ and $\langle\vec{a}(t)|\vec{b}(t)\rangle_{\text{odd}}$. Note that one cannot suppress the factor $(-1)^n$ in the second-to-last line of equation (A.52) by a change in the global phase factor of the wave functions. The factor either occurs because of the $\mathcal{PT}$-symmetric eigenvalue of the ψ_n, or because of the alternating sign of the norm of the Ψ_n. The $\mathcal{PT}$-symmetric time evolution is separately unitary in the space of the coefficients $a_n(t)$, $b_n(t)$ with (i) even n and (ii) odd n, as denoted by an appropriate subscript in equation (A.52). In some sense, the $\mathcal{PT}$-symmetric time evolution leads to a natural "splitting" of the Hilbert space into two subspaces, those of the function with negative $\mathcal{PT}$-symmetric norm and those with positive norm, according to the second equality in equation (A.50). The same pattern has recently been observed in a field-theoretical context [264, 265]: Half of the states of the generalized Dirac equation with a pseudo-scalar mass term acquire a negative Fock-space norm.

From ordinary, Hermitian, quantum mechanics, it is known that the $L^2(\mathbb{R})$ eigenfunctions of a Hermitian operator are in some sense confined to spatial regions where the eigenenergy E_n is larger than the local value of the potential, $E > V(x)$. An intuitive understanding can be obtained if we interpret the potential in terms of a modulus and a phase, according to equation (A.44a). We use the rationale of accompanying a plot of the modulus of a function by a grey band to convey complex phase information in figures A.2 and A.3, where the complex eigenstate wave functions shown previously in figure A.1 are plotted in terms of

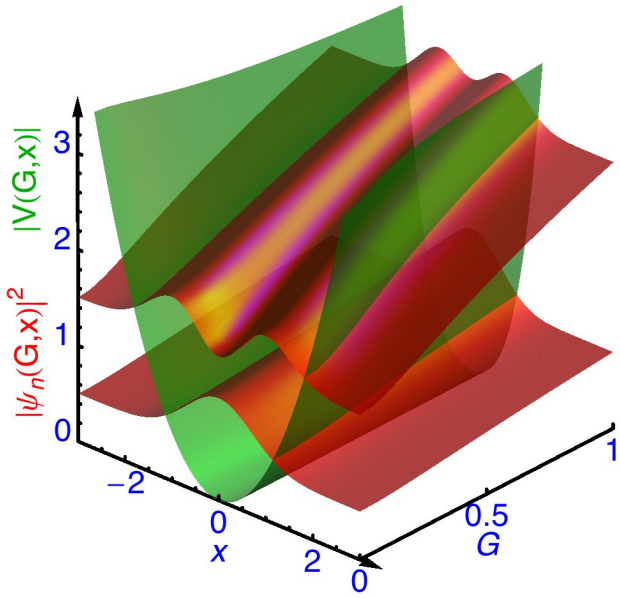


Figure A.3: Illustration of the confinement mechanism for the imaginary cubic potential described by the Hamiltonian (A.39), for the ground and the first excited state. The bulk of the modulus square of the wave function is centered in the "allowed" region where the (real rather than complex) energy $E > V(x) = V(G,x) = |W(G,x)|$. The potential is plotted in green (the "trough"-like structure), whereas the moduli of the wave functions are plotted in red (the "wave-like structures"). The squares of the moduli of the wave functions either have a single maximum (ground state), or two maxima (first excited state). The ground state wave function has a modulus square $|\psi_0(x)|^2 = |\psi_0(G,x)|^2$ as a function of G and x. As G increases, the bound-state energy (which is equal to the base line of the wave function curve at any given value of G) increases, and the modulus of the potential forms a more narrow trough to which the ground-state wave function is confined. The same is true for the first excited state. The central minimum of the modulus square of the first-excited state wave function is clearly visible.

the decomposition

$$\psi_n(x) = |\psi_n(x)|\, \mathrm{e}^{i\, \arg(\psi_n(x))}\,. \tag{A.53}$$

The curves in figures A.2 and A.3 show that indeed, the wave functions of the $\mathcal{PT}$-symmetric oscillator are concentrated [in the sense of a large absolute value of the integrand in equation (A.49)] to a region where $E_n > V(x) = |W(x)|$, where $W(x)$ is the complex-valued $\mathcal{PT}$-symmetric potential (A.44a). In figure A.3, we illustrate that the "confinement" mechanism holds for all values in the range $0 < G < 1$. The interlacing of zeros in complex Sturm–Liouville problems for $\mathcal{PT}$-eigenfunctions, has been discussed in reference [266]. In agreement with the conclusions of reference [266], we find that both the real as well as imaginary parts of the complex eigenfunctions have an infinite number of zeros, individually, when the argument x of the wave function covers the real numbers. Furthermore, this statement even holds for an infinitesimally small, but nonvanishing coupling G in the imaginary cubic perturbation $i\,G\,x^3$.

A.4 Conclusions

In section A.2, we have presented an efficient algorithm for the calculation of eigenvalues of complex symmetric (not Hermitian) matrices. The algorithm is scalable in terms of the desired numerical accuracy and relies on generalized Householder reflections which "depopulate" the input matrix by projecting the entries onto the sub- and super-diagonals, using the generalized inner product which avoids the complex conjugation of the first argument. We find that a subsequent diagonalization of the tridiagonal matrix obtained from the Householder transformations, using generalized Jacobi and Givens rotation matrices (which are again complex and symmetric but not unitary) leads to an efficient eigenvalue solver. Numerical reference data are provided in section A.2.3, and we reemphasize that many of the previously reported numerical tests of generalized quantization conditions for anharmonic oscillators [226–229] rely on the numerical methods described in this chapter. The indefinite inner product can indeed be useful in numerical algorithms; in reference [243], the indefinite inner product had been used previously in a reformulation of the Rayleigh quotient, within an adaptation of the Jacobi–Davidson method for complex symmetric matrices (which has nothing to do with the Jacobi rotations used in our algorithm). We might add that in contrast to the Jacobi–Davidson method, our algorithm does not require a Gram–Schmidt orthogonalization step and is based on a generalized inner product which draws its inspiration from physics.

We would like to illustrate and comment on the algorithm by pointing out a few possible modifications and intricacies of the methods used. The above version of the algorithm described in section A.2 is based on the QL rather than QR decomposition. In the (shifted) iterated QL decomposition, one starts the calculation of the eigenvalue guess from the upper left corner of the input matrix but calculates the Givens and Jacobi rotations from the lower right, i.e., one "chases the bulge upward". In that case, the "uppermost" eigenvalue of the input matrix converges first; the guess σ approximates the true eigenvalue to machine accuracy. If the complex symmetric Hamiltonian is obtained from a basis set of "trial" quantum states the first of which approximates the state of lowest energy of the perturbed system, then the QL decomposition as opposed to the QR decomposition ensures that the "ground-state energy converges first". Still, it is an instructive exercise to modify the algorithm so that the "eigenvalue guess" σ is first calculated for the "lower

rightmost" eigenvalue. One then calculates the Jacobi and Givens rotations which "chase the bulge" from the upper left to the lower right of the tridiagonal matrix. In that case, the "lower rightmost" eigenvalue converges first. This may be useful in particular cases where the "highest" eigenvalue is of particular interest.

One has a few options for controlling the convergence of the algorithm: For example, in many applications, the element D_i may be declared to have "converged" if $D_i + E_i$ equals D_i to machine accuracy, but this criterion may be too restrictive in some cases, especially, when multi-precision arithmetic is being used. In that case, it should be replaced by a criterion which states that $|E_i / D_i|$ is less than a specific, predefined accuracy, say 10^{-64} (for so-called "octuple precision" arithmetic, with 64 decimals).

Our QL factorization involves manifestly complex symmetric $\boldsymbol{Q}$ matrices. Typically, routines built into modern computer algebra system use a unitary matrix $\boldsymbol{Q}$ for such decompositions. These routines use manifestly different matrices than those employed in the approach described above and therefore cannot be used, say, in a meaningful comparison to an implementation of the above algorithm.

In the iterated, shifted QL decomposition of the tridiagonal matrix $\boldsymbol{T}$, the most common pitfall consists in a "premature zero", i.e., in an entry on the sub- or superdiagonal E_j which becomes zero to machine accuracy before the "target element" $E_{i<j}$ for which the current guess σ is calculated has converged to the desired accuracy. In that case, the tridiagonal matrix naturally divides into two matrices (two "irreducible representations") which have to be considered separately. Typically, this phenomenon occurs when the entries in the original input matrix $\boldsymbol{A}$ have a somewhat irregular pattern (e.g., random matrices). The necessity to partition the tridiagonal matrix upon the occurrence of premature zeroes is described rather scarcely in the literature; some lecture notes on the matter can be found in section 11.4 of [267] and near the end of section 3.6.2 of reference [268]. The division into two matrices is called "partitioning" in section 4.7 of [269]. For matrices obtained from regularly distributed, trial basis states, which typically occur in theoretical physics, we have not observed this phenomenon.

As a final remark, we would like to mention that a plain iterated QL or QR decomposition leads to a rather efficient, but not optimized, convergence of the eigenvalue problem, especially for regular input entries in the matrix $\boldsymbol{A}$. The QL and QR decompositions of the input matrix $\boldsymbol{A}$ can be calculated using generalized Householder reflections: For QL, one starts from the rightmost column vector of $\boldsymbol{A}$ and projects it onto its last element; for QR, on starts from the leftmost column vector of $\boldsymbol{A}$ and projects it onto its first element; the subsequent Householder reflections are constructed from the "deflated" $(n-1) \times (n-1)$ submatrices, in either direction [92, 238]. Skipping the tridiagonalization step, one can thus construct generalized QL and QR factorization-based matrix diagonalization routines where $\boldsymbol{Q}$ is manifestly complex and symmetric, but not unitary. Otherwise, the implicit shift (A.25) leads to improved convergence to numerical approximations of the eigenvalues.

The usefulness of the generalized Householder reflections used in the algorithm has a connection to the underlying physics. Indeed, we find that the most natural interpretation of the conserved scalar product in the $\mathcal{PT}$-symmetric time evolution is in terms of the generalized indefinite inner product defined for finite-dimensional vectors in equation (A.1) and for Hilbert space vectors in equation (A.43). This inner product is linear in both arguments and avoids complex conjugation. Attempts at finding an alternative, appropriate interpretation [263] seem too complicated to take precedence over the immediate identification of the

$\mathcal{PT}$-symmetric scalar product of eigenvectors in terms of the indefinite inner product. The integrand immediately becomes "local" and one avoids integrations over eigenfunctions evaluated at the (possibly very distant) points x and $-x$. We reemphasize that the indefinite inner product is crucial in the analysis of resonance eigenvectors [242] for the "real" cubic perturbation, and that the eigenvalues of the "real" and "imaginary" cubic perturbation are related by a dispersion relation [227,232]. So, in some sense, the importance of the indefinite inner product for the analysis of the imaginary cubic Hamiltonian had to be expected, and the emergence of an efficient matrix diagonalization algorithm for such matrices is only natural (see section A.3).

It has been stressed in the literature that the scalar product $\langle\psi(t)|\phi(t)\rangle_{\mathcal{PT}}$ as defined in equation (A.41) is not positive definite. This has been used as an argument against the viability of $\mathcal{PT}$-symmetric Hamiltonians for the description of natural phenomena. However, one may counter argue that the same problem persists with regard to the relativistic Klein-Gordon equation where the time-like component of the conserved Noether current can become negative (see chapter 2 of reference [270]). The Klein-Gordon equation is assumed to describe a charged scalar field like the charged component of the Higgs (doublet) field [271] (the latter is usually assumed to vanish under a gauge transformation, and the remaining neutral component of the Higgs doublet is expanded about its vacuum expectation value). Strictly speaking, one has to reinterpret the timelike component of the conserved Noether current as a charge density, not a probability density. Analogously, the timelike component of the conserved Noether current of the pseudo–Hermitian, generalized Dirac Hamiltonian (with a pseudo-scalar mass term) may naturally be interpreted as a non-positive definite "weak-interaction density" (see reference [265]). Equation (A.52) suggests that the Hilbert space, under the $\mathcal{PT}$-symmetric time evolution, is split into two "halves", one of which entails negative $\mathcal{PT}$-symmetric norm (analogous to the "right-handed neutrinos"), and the other has positive $\mathcal{PT}$-symmetric norm (analogous to the "left-handed neutrinos" within the model proposed in reference [265]). We recall that "half" of the Ψ_n eigenstates acquire a negative $\mathcal{PT}$-symmetric norm under a very natural choice of the global complex phase [see equation (A.51)]. The $\mathcal{PT}$-symmetric norm, in turn, can be formulated in terms of the generalized indefinite inner product, on which this chapter is based.

Bibliography

[1] D. Ceperley, *Ground state of the fermion one-component plasma: A Monte Carlo study in two and three dimensions*, Phys. Rev. B **18**, 3126 (1978).

[2] D. M. Ceperley and B. J. Alder, *Ground State of the Electron Gas by a Stochastic Method*, Phys. Rev. Lett. **45**, 566 (1980).

[3] P. Hohenberg and W. Kohn, *Inhomogeneous Electron Gas*, Phys. Rev. **136**, B864 (1964).

[4] W. Kohn and L. J. Sham, *Self-Consistent Equations Including Exchange and Correlation Effects*, Phys. Rev. **140**, A1133 (1965).

[5] S. R. White, *Density Matrix Formulation for Quantum Renormalization Groups*, Phys. Rev. Lett. **69**, 2863 (1992).

[6] T. Nishino, *Density Matrix Renormalization Group Method for 2D Classical Models*, J. Phys. Soc. Japan **64**, 3598 (1995).

[7] G. Vidal, *Efficient Classical Simulation of Slightly Entangled Quantum Computations*, Phys. Rev. Lett. **91**, 147902 (2003).

[8] G. Vidal, *Efficient Simulation of One-Dimensional Quantum Many-Body Systems*, Phys. Rev. Lett. **93**, 040502 (2004).

[9] S. R. White and A. E. Feiguin, *Real-Time Evolution Using the Density Matrix Renormalization Group*, Phys. Rev. Lett. **93**, 076401 (2004).

[10] A. J. Daley, C. Kollath, U. Schollwöck, and G. Vidal, *Time-dependent density-matrix renormalization-group using adaptive effective Hilbert spaces*, J. Stat. Mech.: Theor. Exp., P04005 (2004).

[11] F. Verstraete and J. I. Cirac, *Renormalization algorithms for Quantum Many-Body Systems in two and higher dimensions*, arXiv:cond-mat/0407066v1 (2004).

[12] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, *Matrix Product Density Operators: Simulation of Finite-Temperature and Dissipative Systems*, Phys. Rev. Lett. **93**, 207204 (2004).

[13] F. Verstraete, D. Porras, and J. I. Cirac, *Density Matrix Renormalization Group and Periodic Boundary Conditions: A Quantum Information Perspective*, Phys. Rev. Lett. **93**, 227205 (2004).

[14] G. Vidal, *Classical Simulation of Infinite-Size Quantum Lattice Systems in One Spatial Dimension*, Phys. Rev. Lett. **98**, 070201 (2007).

[15] G. Vidal, *Entanglement Renormalization*, Phys. Rev. Lett. **99**, 220405 (2007).

[16] G. Vidal, *Class of Quantum Many-Body States That Can Be Efficiently Simulated*, Phys. Rev. Lett. **101**, 110501 (2008).

[17] J. Jordan, R. Orús, G. Vidal, F. Verstraete, and J. I. Cirac, *Classical Simulation of Infinite-Size Quantum Lattice Systems in Two Spatial Dimensions*, Phys. Rev. Lett. **101**, 250602 (2008).

[18] C. V. Kraus, N. Schuch, F. Verstraete, and J. I. Cirac, *Fermionic projected entangled pair states*, Phys. Rev. A **81**, 052338 (2010).

[19] I. Pižorn and F. Verstraete, *Fermionic implementation of projected entangled pair states algorithm*, Phys. Rev. B **81**, 245110 (2010).

[20] T. Barthel, C. Pineda, and J. Eisert, *Contraction of fermionic operator circuits and the simulation of strongly correlated fermions*, Phys. Rev. A **80**, 042333 (2009).

[21] C. Pineda, T. Barthel, and J. Eisert, *Unitary circuits for strongly correlated fermions*, Phys. Rev. A **81**, 050303(R) (2010).

[22] P. Corboz, R. Orús, B. Bauer, and G. Vidal, *Simulation of strongly correlated fermions in two spatial dimensions with fermionic projected entangled-pair states*, Phys. Rev. B **81**, 165104 (2010).

[23] P. Corboz, J. Jordan, and G. Vidal, *Simulation of fermionic lattice models in two dimensions with projected entangled-pair states: Next-nearest neighbor Hamiltonians*, Phys. Rev. B **82**, 245119 (2010).

[24] Z. Y. Xie, J. Chen, J. F. Yu, X. Kong, B. Normand, and T. Xiang, *Tensor Renormalization of Quantum Many-Body Systems Using Projected Entangled Simplex States*, Phys. Rev. X **4**, 011025 (2014).

[25] P. Corboz, S. R. White, G. Vidal, and M. Troyer, *Stripes in the two-dimensional $t - J$ model with infinite projected entangled-pair states*, Phys. Rev. B **84**, 041108(R) (2011).

[26] P. Corboz, T. M. Rice, and M. Troyer, *Competing States in the t-J Model: Uniform d-Wave State versus Stripe State*, Phys. Rev. Lett. **113**, 046402 (2014).

[27] R. P. Feynman, *Simulating physics with computers*, Intern. J. Theor. Phys. **21**, 467 (1982).

[28] I. Buluta and F. Nori, *Quantum Simulators*, Science **326**, 108 (2009).

[29] J. I. Cirac and P. Zoller, *Goals and opportunities in quantum simulation*, Nature Phys. **8**, 264 (2012).

[30] I. Bloch, *Quantum coherence and entanglement with ultracold atoms in optical lattices*, Nature **453**, 1016 (2008).

[31] I. Bloch, J. Dalibard, and W. Zwerger, *Many-body physics with ultracold gases*, Rev. Mod. Phys. **80**, 885 (2008).

[32] T. Esslinger, *Fermi-Hubbard Physics with Atoms in an Optical Lattice*, Annu. Rev. Condens. Matter Phys. **1**, 129 (2010).

[33] I. Bloch, J. Dalibard, and S. Nascimbène, *Quantum simulations with ultracold quantum gases*, Nature Phys. **8**, 267 (2012).

[34] D. Jaksch, C. Bruder, J. I. Cirac, C. W. Gardiner, and P. Zoller, *Cold Bosonic Atoms in Optical Lattices*, Phys. Rev. Lett. **81**, 3108 (1998).

[35] M. Greiner, O. Mandel, T. Esslinger, T. W. Hänsch, and I. Bloch, *Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms*, Nature **415**, 39 (2002).

[36] R. Jördens, N. Strohmaier, K. Günter, H. Moritz, and T. Esslinger, *A Mott insulator of fermionic atoms in an optical lattice*, Nature **455**, 204 (2008).

[37] U. Schneider, L. Hackermüller, S. Will, Th. Best, I. Bloch, T. A. Costi, R. W. Helmes, D. Rasch, and A. Rosch, *Metallic and Insulating Phases of Repulsively Interacting Fermions in a 3D Optical Lattice*, Science **322**, 1520 (2008).

[38] W. S. Bakr, J. I. Gillen, A. Peng, S. Fölling, and M. Greiner, *A quantum gas microscope for detecting single atoms in a Hubbard-regime optical lattice*, Nature **462**, 74 (2009).

[39] J. F. Sherson, C. Weitenberg, M. Endres, M. Cheneau, I. Bloch, and S. Kuhr, *Single-atom-resolved fluorescence imaging of an atomic Mott insulator*, Nature **467**, 68 (2010).

[40] S. Trotzky, Y.-A. Chen, A. Flesch, I. P. McCulloch, U. Schollwöck, J. Eisert, and I. Bloch, *Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional Bose gas*, Nature Phys. **8**, 325 (2012).

[41] F. Verstraete, V. Murg, and J. I. Cirac, *Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems*, Adv. Phys. **57**, 143 (2008).

[42] H. C. Jiang, Z. Y. Weng, and T. Xiang, *Accurate Determination of Tensor Network State of Quantum Lattice Models in Two Dimensions*, Phys. Rev. Lett. **101**, 090603 (2008).

[43] I. Pižorn, L. Wang, and F. Verstraete, *Time evolution of projected entangled pair states in the single-layer picture*, Phys. Rev. A **83**, 052321 (2011).

[44] M. Lubasch, J. I. Cirac, and M.-C. Bañuls, *Unifying projected entangled pair state contractions*, New J. Phys. **16**, 033014 (2014).

[45] M. Lubasch, J. I. Cirac, and M.-C. Bañuls, *Algorithms for finite projected entangled pair states*, Phys. Rev. B **90**, 064425 (2014).

[46] A. D. Becke, *Density-functional exchange-energy approximation with correct asymptotic behavior*, Phys. Rev. A **38**, 3098 (1988).

[47] A. D. Becke, *Density-functional thermochemistry. iii. the role of exact exchange*, J. Chem. Phys. **98**, 5648 (1993).

[48] J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple*, Phys. Rev. Lett. **77**, 3865 (1996).

[49] J. P. Perdew, K. Burke, and M. Ernzerhof, *Errata: Generalized Gradient Approximation Made Simple*, Phys. Rev. Lett. **78**, 1396 (1997).

[50] R. Peverati and D. G. Truhlar, *The quest for a universal density functional: The accuracy of density functionals across a broad spectrum of databases in chemistry and physics*, arXiv:1212.0944v4 (2013).

[51] E. Dagotto, *Correlated electrons in high-temperature superconductors*, Rev. Mod. Phys. **66**, 763 (1994).

[52] M. Lubasch, V. Murg, U. Schneider, J. I. Cirac, and M.-C. Bañuls, *Adiabatic Preparation of a Heisenberg Antiferromagnet Using an Optical Superlattice*, Phys. Rev. Lett. **107**, 165301 (2011).

[53] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge (2007).

[54] M. Lubasch, F. Mintert, and S. Wimberger, *Dynamical enhancement of spatial entanglement in massive particles*, Phys. Rev. A **84**, 063615 (2011).

[55] P. Zanardi and N. Paunković, *Ground state overlap and quantum phase transitions*, Phys. Rev. E **74**, 031123 (2006).

[56] P. Plötz, M. Lubasch, and S. Wimberger, *Detection of avoided crossings by fidelity*, Physica A **390**, 1363 (2011).

[57] J. H. Noble, M. Lubasch, and U. D. Jentschura, *Generalized Householder transformations for the complex symmetric eigenvalue problem*, Eur. Phys. J. Plus **128**, 93 (2013).

[58] M. Zwolak and G. Vidal, *Mixed-State Dynamics in One-Dimensional Quantum Lattice Systems: A Time-Dependent Superoperator Renormalization Algorithm*, Phys. Rev. Lett. **93**, 207205 (2004).

[59] Y.-Y. Shi, L.-M. Duan, and G. Vidal, *Classical simulation of quantum many-body systems with a tree tensor network*, Phys. Rev. A **74**, 022320 (2006).

[60] V. Murg, F. Verstraete, and J. I. Cirac, *Efficient Evaluation of Partition Functions of Inhomogeneous Many-Body Spin Systems*, Phys. Rev. Lett. **95**, 057206 (2005).

[61] V. Murg, F. Verstraete, and J. I. Cirac, *Variational study of hard-core bosons in a two-dimensional optical lattice using projected entangled pair states*, Phys. Rev. A **75**, 033605 (2007).

[62] V. Murg, F. Verstraete, and J. I. Cirac, *Exploring frustrated spin systems using projected entangled pair states*, Phys. Rev. B **79**, 195119 (2009).

[63] M.-C. Bañuls, D. Pérez-García, M. M. Wolf, F. Verstraete, and J. I. Cirac, *Sequentially generated states for the study of two-dimensional systems*, Phys. Rev. A **77**, 052306 (2008).

[64] A. W. Sandvik and G. Vidal, *Variational Quantum Monte Carlo Simulations with Tensor-Network States*, Phys. Rev. Lett. **99**, 220602 (2007).

[65] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, *Simulation of Quantum Many-Body Systems with Strings of Operators and Monte Carlo Tensor Contractions*, Phys. Rev. Lett. **100**, 040501 (2008).

[66] R. Orús and G. Vidal, *Simulation of two-dimensional quantum systems on an infinite lattice revisited: Corner transfer matrix for tensor contraction*, Phys. Rev. B **80**, 094403 (2009).

[67] M. Levin and C. P. Nave, *Tensor Renormalization Group Approach to Two-Dimensional Classical Lattice Models*, Phys. Rev. Lett. **99**, 120601 (2007).

[68] Z. Y. Xie, H. C. Jiang, Q. N. Chen, Z. Y. Weng, and T. Xiang, *Second Renormalization of Tensor-Network States*, Phys. Rev. Lett. **103**, 160601 (2009).

[69] S. Östlund and S. Rommer, *Thermodynamic Limit of Density Matrix Renormalization*, Phys. Rev. Lett. **75**, 3537 (1995).

[70] L. Wang, I. Pižorn, and F. Verstraete, *Monte Carlo simulation with tensor network states*, Phys. Rev. B **83**, 134421 (2011).

[71] L. Wang and F. Verstraete, *Cluster update for tensor network states*, arXiv:1110.4362v1 (2011).

[72] F. Verstraete and J. I. Cirac, *Matrix product states represent ground states faithfully*, Phys. Rev. B **73**, 094423 (2006).

[73] M. B. Hastings, *An area law for one-dimensional quantum systems*, J. Stat. Mech.: Theor. Exp., P08024 (2007).

[74] U. Schollwöck, *The density-matrix renormalization group*, Rev. Mod. Phys. **77**, 259 (2005).

[75] U. Schollwöck, *The density-matrix renormalization group in the age of matrix product states*, Ann. Phys. **326**, 96 (2011).

[76] R. Orús, *A practical introduction to tensor networks: Matrix product states and projected entangled pair states*, Ann. Phys. **349**, 117 (2014).

[77] M. B. Hastings, *Solving gapped Hamiltonians locally*, Phys. Rev. B **73**, 085115 (2006).

[78] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, *Computational Complexity of Projected Entangled Pair States*, Phys. Rev. Lett. **98**, 140506 (2007).

[79] B. Pirvu, V. Murg, J. I. Cirac, and F. Verstraete, *Matrix product operator representations*, New J. Phys. **12**, 025012 (2010).

[80] H. 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*, Phys. Rev. Lett. **101**, 010504 (2008).

[81] J. I. Cirac, D. Poilblanc, N. Schuch, and F. Verstraete, *Entanglement spectrum and boundary theories with projected entangled-pair states*, Phys. Rev. B **83**, 245134 (2011).

[82] G. Evenbly and R. N. C. Pfeifer, *Improving the efficiency of variational tensor network algorithms*, Phys. Rev. B **89**, 245118 (2014).

[83] G. De las Cuevas, N. Schuch, D. Pérez-García, and J. I. Cirac, *Purifications of multipartite states: limitations and constructive methods*, New J. Phys. **15**, 123021 (2013).

[84] S. Todo and K. Kato, *Cluster Algorithms for General-S Quantum Spin Systems*, Phys. Rev. Lett. **87**, 047203 (2001).

[85] A. F. Albuquerque *et al.* (ALPS collaboration), *The ALPS project release 1.3: Open-source software for strongly correlated systems*, J. Magn. Magn. Mater. **310**, 1187 (2007).

[86] B. Bauer *et al.* (ALPS collaboration), *The ALPS project release 2.0: open source software for strongly correlated systems*, J. Stat. Mech.: Theor. Exp., P05001 (2011).

[87] Z. Y. Xie, J. Chen, M. P. Qin, J. W. Zhu, L. P. Yang, and T. Xiang, *Coarse-graining renormalization by higher-order singular value decomposition*, Phys. Rev. B **86**, 045139 (2012).

[88] A. Hosseinkhani, B. G. Dezfouli, F. Ghasemipour, A. T. Rezakhani, and H. Saberi, *Uncontrolled disorder effects in fabricating photonic quantum simulators on a kagome geometry: A projected-entangled-pair-state versus exact-diagonalization analysis*, Phys. Rev. A **89**, 062324 (2014).

[89] M. Kliesch, D. Gross, and J. Eisert, *Matrix product operators and states - NP-hardness and undecidability*, arXiv:1404.4466v1 (2014).

[90] J. B. Keller, *Closest Unitary, Orthogonal and Hermitian Operators to a Given Operator*, Math. Mag. **48**, 192 (1975).

[91] H. Kalis, D. Klagges, R. Orús, and K. P. Schmidt, *Fate of the cluster state on the square lattice in a magnetic field*, Phys. Rev. A **86**, 022317 (2012).

[92] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore and London, 3rd edition (1996).

[93] A. W. Sandvik, *Finite-size scaling of the ground-state parameters of the two-dimensional Heisenberg model*, Phys. Rev. B **56**, 11678 (1997).

[94] B. Bauer, G. Vidal, and M. Troyer, *Assessing the accuracy of projected entangled-pair states on infinite lattices*, J. Stat. Mech.: Theor. Exp., P09006 (2009).

[95] F. Mezzacapo, N. Schuch, M. Boninsegni, and J. I. Cirac, *Ground-state properties of quantum many-body systems: entangled-plaquette states and variational Monte Carlo*, New J. Phys. **11**, 083026 (2009).

[96] A. Sfondrini, J. Cerrillo, N. Schuch, and J. I. Cirac, *Simulating two- and three-dimensional frustrated quantum systems with string-bond states*, Phys. Rev. B **81**, 214426 (2010).

[97] E. M. Stoudenmire and S. R. White, *Studying Two-Dimensional Systems with the Density Matrix Renormalization Group*, Annu. Rev. Condens. Matter Phys. **3**, 111 (2012).

[98] S. R. White and A. L. Chernyshev, *Neél Order in Square and Triangular Lattice Heisenberg Models*, Phys. Rev. Lett. **99**, 127004 (2007).

[99] P. Corboz and F. Mila, *Crystals of Bound States in the Magnetization Plateaus of the Shastry-Sutherland Model*, Phys. Rev. Lett. **112**, 147203 (2014).

[100] S. Singh, R. N. C. Pfeifer, and G. Vidal, *Tensor network states and algorithms in the presence of a global $U(1)$ symmetry*, Phys. Rev. B **83**, 115125 (2011).

[101] B. Bauer, P. Corboz, R. Orús, and M. Troyer, *Implementing global Abelian symmetries in projected entangled-pair state algorithms*, Phys. Rev. B **83**, 125106 (2011).

[102] H. W. J. Blöte and Y. Deng, *Cluster Monte Carlo simulation of the transverse Ising model*, Phys. Rev. E **66**, 066110 (2002).

[103] S. Yan, D. A. Huse, and S. R. White, *Spin-Liquid Ground State of the $S = 1/2$ Kagome Heisenberg Antiferromagnet*, Science **332**, 1173 (2011).

[104] S. Depenbrock, I. P. McCulloch, and U. Schollwöck, *Nature of the Spin-Liquid Ground State of the $S = 1/2$ Heisenberg Model on the Kagome Lattice*, Phys. Rev. Lett. **109**, 067201 (2012).

[105] The Numerical Algorithms Group (NAG), *The NAG Library*, http://www.nag.com, Oxford, United Kingdom (2013).

[106] M. Lubasch, J. I. Fuks, H. Appel, A. Rubio, J. I. Cirac, and M.-C. Bañuls, *Density functional theory beyond the local density approximation*, in preparation.

[107] R. M. Dreizler and E. K. U. Gross, *Density Functional Theory*, Springer (1996).

[108] W. Kohn, *Nobel Lecture: Electronic structure of matter-wave functions and density functionals*, Rev. Mod. Phys. **71**, 1253 (1999).

[109] Nobelprize.org, *"The Nobel Prize in Chemistry 1998"*, Nobel Media AB (2013).

[110] S. Redner, *Citation Statistics from 110 Years of Physical Review*, Physics Today **58**, 49 (2005).

[111] N. Schuch and F. Verstraete, *Computational complexity of interacting electrons and fundamental limitations of density functional theory*, Nature Phys. **5**, 732 (2009).

[112] C. Lee, W. Yang, and R. G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density*, Phys. Rev. B **37**, 785 (1988).

[113] G. F. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid*, Cambridge University Press, Cambridge (2005).

[114] F. C. Alcaraz and K. Capelle, *Density functional formulations for quantum chains*, Phys. Rev. B **76**, 035109 (2007).

[115] M. Levy, *Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem*, Proc. Natl. Acad. Sci. U.S.A. **76**, 6062 (1979).

[116] C. R. Jacob and M. Reiher, *Spin in Density-Functional Theory*, Int. J. Quantum Chem. **2012**, 3661 (2012).

[117] K. Peirs, D. Van Neck, and M. Waroquier, *Algorithm to derive exact exchange-correlation potentials from correlated densities in atoms*, Phys. Rev. A **67**, 012505 (2003).

[118] E. M. Stoudenmire, L. O. Wagner, S. R. White, and K. Burke, *Exact density functional theory with the density matrix renormalization group*, arXiv:1107.2394v1 (2011).

[119] I. V. Oseledets, *Compact matrix form of the d-dimensional tensor decomposition*, INM RAS (2009).

[120] I. V. Oseledets and E. E. Tyrtyshnikov, *TT-cross approximation for multidimensional arrays*, Lin. Alg. Appl. **432**, 70 (2010).

[121] https://github.com/oseledets/TT-Toolbox, *"TT-Toolbox Version 2.2"*, I. V. Oseledets (2014).

[122] S. McKinley and M. Levine, *Lecture notes: "Cubic Spline Interpolation"*, http://online.redwoods.edu/instruct/darnold/laproj/fall98/skymeg/proj.pdf (1998).

[123] H. Appel, *Private communication*.

[124] D. G. Tempel, T. J. Martínez, and N. T. Maitra, *Revisiting Molecular Dissociation in Density Functional Theory: A Simple Model*, J. Chem. Theory Comput. **5**, no. 4, 770-780 (2009).

[125] N. Helbig, J. I. Fuks, M. Casula, M. J. Verstraete, M. A. L. Marques, I. V. Tokatly, and A. Rubio, *Density functional theory beyond the linear regime: Validating an adiabatic local density approximation*, Phys. Rev. A **83**, 032503 (2011).

[126] L. O. Wagner, E. M. Stoudenmire, K. Burke, and S. R. White, *Guaranteed Convergence of the Kohn-Sham Equations*, Phys. Rev. Lett. **111**, 093003 (2013).

[127] C. G. Broyden, *A Class of Methods for Solving Nonlinear Simultaneous Equations*, Math. Comput. **19**, no. 92, 577-593 (1965).

[128] G. P. Kerker, *Efficient iteration scheme for self-consistent pseudopotential calculations*, Phys. Rev. B **23**, 3082 (1981).

[129] D. D. Johnson, *Modified Broyden's method for accelerating convergence in self-consistent calculations*, Phys. Rev. B **38**, 12807 (1988).

[130] D. R. Bowler and M. J. Gillan, *An efficient and robust technique for achieving self consistency in electronic structure calculations*, Chem. Phys. Lett. **325**, 473 (2000).

[131] F. Werner, O. Parcollet, A. Georges, and S. R. Hassan, *Interaction-Induced Adiabatic Cooling and Antiferromagnetism of Cold Fermions in Optical Lattices*, Phys. Rev. Lett. **95**, 056401 (2005).

[132] R. Jördens, L. Tarruell, D. Greif, T. Uehlinger, N. Strohmaier, H. Moritz, T. Esslinger, L. De Leo, C. Kollath, A. Georges, V. Scarola, L. Pollet, E. Burovski, E. Kozik, and M. Troyer, *Quantitative Determination of Temperature in the Approach to Magnetic Order of Ultracold Fermions in an Optical Lattice*, Phys. Rev. Lett. **104**, 180401 (2010).

[133] D. C. McKay and B. DeMarco, *Cooling in strongly correlated optical lattices: prospects and challenges*, Rep. Prog. Phys. **74**, 054401 (2011).

[134] T.-L. Ho, *The Intrinsic Difficulties of Constructing Strongly Correlated States of Lattice Quantum Gases by Connecting Up Pre-engineered Isolated Atomic Clusters*, arXiv:0808.2677v1 (2008).

[135] A. S. Sørensen, E. Altman, M. Gullans, J. V. Porto, M. D. Lukin, and E. Demler, *Adiabatic preparation of many-body states in optical lattices*, Phys. Rev. A **81**, 061603(R) (2010).

[136] U. Schneider, L. Hackermüller, S. Will, Th. Best, I. Bloch, T. A. Costi, R. W. Helmes, D. Rasch, and A. Rosch, *Supporting Online Material for "Metallic and Insulating Phases of Repulsively Interacting Fermions in a 3D Optical Lattice"*, Science **322**, 1520 (2008).

[137] U. Schneider, L. Hackermüller, J. P. Ronzheimer, S. Will, S. Braun, Th. Best, I. Bloch, E. Demler, S. Mandt, D. Rasch, and A. Rosch, *Breakdown of diffusion: From collisional hydrodynamics to a continuous quantum walk in a homogeneous Hubbard model*, arXiv:1005.3545v1 (2010).

[138] U. Schneider, L. Hackermüller, J. P. Ronzheimer, S. Will, S. Braun, Th. Best, I. Bloch, E. Demler, S. Mandt, D. Rasch, and A. Rosch, *Fermionic transport and out-of-equilibrium dynamics in a homogeneous Hubbard model with ultracold atoms*, Nature Physics **8**, 213 (2012).

[139] J. Sebby-Strabley, M. Anderlini, P. S. Jessen, and J. V. Porto, *Lattice of double wells for manipulating pairs of cold atoms*, Phys. Rev. A **73**, 033605 (2006).

[140] S. Trotzky, P. Cheinet, S. Fölling, M. Feld, U. Schnorrberger, A. M. Rey, A. Polkovnikov, E. A. Demler, M. D. Lukin, and I. Bloch, *Time-Resolved Observation and Control of Superexchange Interactions with Ultracold Atoms in Optical Lattices*, Science **319**, 295 (2008).

[141] S. Trotzky, Y.-A. Chen, U. Schnorrberger, P. Cheinet, and I. Bloch, *Controlling and Detecting Spin Correlations of Ultracold Atoms in Optical Lattices*, Phys. Rev. Lett. **105**, 265303 (2010).

[142] C. Zener, *Non-Adiabatic Crossing of Energy Levels*, Proc. R. Soc. A **137**, 696 (1932).

[143] M. Matsumoto, Ch. Yasuda, S. Todo, and H. Takayama, *Ground-state phase diagram of quantum Heisenberg antiferromagnets on the anisotropic dimerized square lattice*, Phys. Rev. B **65**, 014407 (2001).

[144] G. M. Bruun, O. F. Syljuåsen, K. G. L. Pedersen, B. M. Andersen, E. Demler, and A. S. Sørensen, *Antiferromagnetic noise correlations in optical lattices*, Phys. Rev. A **80**, 033622 (2009).

[145] L. Hulthén, *Ph.D. thesis "Über das Austauschproblem eines Kristalles"*, Ark. Mat. Astron. Fys. **26A**, 1 (1938).

[146] A. W. Sandvik and H. G. Evertz, *Loop updates for variational and projector quantum Monte Carlo simulations in the valence-bond basis*, Phys. Rev. B **82**, 024407 (2010).

[147] I. Affleck, D. Gepner, H. J. Schulz, and T. Ziman, *Critical behaviour of spin-s Heisenberg antiferromagnetic chains: analytic and numerical results*, J. Phys. A: Math. Gen. **22**, 511 (1989).

[148] M. Boninsegni and N. V. Prokof'ev, *Phase diagram of an anisotropic bosonic t-J model*, Phys. Rev. B **77**, 092502 (2008).

[149] J. J. García-Ripoll, M. A. Martin-Delgado, and J. I. Cirac, *Implementation of Spin Hamiltonians in Optical Lattices*, Phys. Rev. Lett. **93**, 250405 (2004).

[150] J. Estève, C. Gross, A. Weller, S. Giovanazzi, and M. K. Oberthaler, *Squeezing and entanglement in a Bose-Einstein condensate*, Nature **455**, 1216 (2008).

[151] M. F. Riedel, P. Böhi, Y. Li, T. W. Hänsch, A. Sinatra, and P. Treutlein, *Atom-chip-based generation of entanglement for quantum metrology*, Nature **464**, 1170 (2010).

[152] G. K. Brennen, C. M. Caves, P. S. Jessen, and I. H. Deutsch, *Quantum Logic Gates in Optical Lattices*, Phys. Rev. Lett. **82**, 1060 (1999).

[153] D. Jaksch, H. J. Briegel, J. I. Cirac, C. W. Gardiner, and P. Zoller, *Entanglement of Atoms via Cold Controlled Collisions*, Phys. Rev. Lett. **82**, 1975 (1999).

[154] T. Calarco, E. A. Hinds, D. Jaksch, J. Schmiedmayer, J. I. Cirac, and P. Zoller, *Quantum gates with neutral atoms: Controlling collisional interactions in time-dependent traps*, Phys. Rev. A **61**, 022304 (2000).

[155] R. Ionicioiu and P. Zanardi, *Quantum-information processing in bosonic lattices*, Phys. Rev. A **66**, 050301(R) (2002).

[156] J. Eisert, M. B. Plenio, S. Bose, and J. Hartley, *Towards Quantum Entanglement in Nanoelectromechanical Devices*, Phys. Rev. Lett. **93**, 190402 (2004).

[157] S. Bose, *Entanglement from the Dynamics of an Ideal Bose Gas in a Lattice*, arXiv:cond-mat/0610024v1 (2006).

[158] C. E. Creffield, *Quantum Control and Entanglement using Periodic Driving Fields*, Phys. Rev. Lett. **99**, 110501 (2007).

[159] F. Galve, L. A. Pachón, and D. Zueco, *Bringing Entanglement to the High Temperature Limit*, Phys. Rev. Lett. **105**, 180501 (2010).

[160] J. Cai, G. G. Guerreschi, and H. J. Briegel, *Quantum Control and Entanglement in a Chemical Compass*, Phys. Rev. Lett. **104**, 220502 (2010).

[161] J. Cai, S. Popescu, and H. J. Briegel, *Dynamic entanglement in oscillating molecules and potential biological implications*, Phys. Rev. E **82**, 021921 (2010).

[162] F. Galve, D. Zueco, S. Kohler, E. Lutz, and P. Hänggi, *Entanglement resonance in driven spin chains*, Phys. Rev. A **79**, 032332 (2009).

[163] J. Li and G. S. Paraoanu, *Generation and propagation of entanglement in driven coupled-qubit systems*, New J. Phys. **11**, 113020 (2009).

[164] S. Ghosh, T. F. Rosenbaum, G. Aeppli, and S. N. Coppersmith, *Entangled quantum state of magnetic dipoles*, Nature **425**, 48 (2003).

[165] H. Häffner, C. F. Roos, and R. Blatt, *Quantum computing with trapped ions*, Phys. Rep. **469**, 155 (2008).

[166] T. Stöferle, H. Moritz, C. Schori, M. Köhl, and T. Esslinger, *Transition from a Strongly Interacting 1D Superfluid to a Mott Insulator*, Phys. Rev. Lett. **92**, 130403 (2004).

[167] M. Lubasch, *Diploma thesis "Quantum chaos and entanglement in the Bose-Hubbard model"*, University of Heidelberg (2009).

[168] J. Simon, W. S. Bakr, R. Ma, M. E. Tai, P. M. Preiss, and M. Greiner, *Quantum simulation of antiferromagnetic spin chains in an optical lattice*, Nature **472**, 307 (2011).

[169] M. Endres, M. Cheneau, T. Fukuhara, C. Weitenberg, P. Schauß, C. Gross, L. Mazza, M.-C. Bañuls, L. Pollet, I. Bloch, and S. Kuhr, *Observation of Correlated Particle-Hole Pairs and String Order in Low-Dimensional Mott Insulators*, Science **334**, 200 (2011).

[170] M. P. A. Fisher, P. B. Weichman, G. Grinstein, and D. S. Fisher, *Boson localization and the superfluid-insulator transition*, Phys. Rev. B **40**, 546 (1989).

[171] W. Zwerger, *Mott-Hubbard transition of cold atoms in optical lattices*, J. Opt. B: Quantum Semiclass. Opt. **5**, S9 (2003).

[172] H. M. Wiseman and J. A. Vaccaro, *Entanglement of Indistinguishable Particles Shared between Two Parties*, Phys. Rev. Lett. **91**, 097902 (2003).

[173] N. Schuch, F. Verstraete, and J. I. Cirac, *Nonlocal Resources in the Presence of Superselection Rules*, Phys. Rev. Lett. **92**, 087904 (2004).

[174] N. Schuch, F. Verstraete, and J. I. Cirac, *Quantum entanglement theory in the presence of superselection rules*, Phys. Rev. A **70**, 042310 (2004).

[175] M. R. Dowling, A. C. Doherty, and H. M. Wiseman, *Entanglement of indistinguishable particles in condensed-matter physics*, Phys. Rev. A **73**, 052323 (2006).

[176] Z. Y. Ou and L. Mandel, *Violation of Bell's Inequality and Classical Probability in a Two-Photon Correlation Experiment*, Phys. Rev. Lett. **61**, 50 (1988).

[177] Y. H. Shih and C. O. Alley, *New Type of Einstein-Podolski-Rosen-Bohm Experiment Using Pairs of Light Quanta Produced by Optical Parametric Down Conversion*, Phys. Rev. Lett. **61**, 2921 (1988).

[178] J. G. Rarity and P. R. Tapster, *Experimental Violation of Bell's Inequality Based on Phase and Momentum*, Phys. Rev. Lett. **64**, 2495 (1990).

[179] P. G. Kwiat, K. Mattle, H. Weinfurtner, and A. Zeilinger, *New High-Intensity Source of Polarization-Entangled Photon Pairs*, Phys. Rev. Lett. **75**, 4337 (1995).

[180] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, *Concentrating partial entanglement by local operations*, Phys. Rev. A **53**, 2046 (1996).

[181] G. Vidal and R. F. Werner, *Computable measure of entanglement*, Phys. Rev. A **65**, 032314 (2002).

[182] Chapters 16.1 and 16.2, *Numerical Recipes in Fortran 77*, 2nd edition (available at http://www.nr.com/oldverswitcher.html) (1992).

[183] C. Kollath, A. Iucci, T. Giamarchi, W. Hofstetter, and U. Schollwöck, *Spectroscopy of Ultracold Atoms by Periodic Lattice Modulations*, Phys. Rev. Lett. **97**, 050402 (2006).

[184] A. Zenesini, H. Lignier, D. Ciampini, O. Morsch, and E. Arimondo, *Coherent Control of Dressed Matter Waves*, Phys. Rev. Lett. **102**, 100403 (2009).

[185] A. Zenesini, H. Lignier, G. Tayebirad, J. Radogostowicz, D. Ciampini, R. Mannella, S. Wimberger, O. Morsch, and E. Arimondo, *Time-Resolved Measurement of Landau-Zener Tunneling in Periodic Potentials*, Phys. Rev. Lett. **103**, 090403 (2009).

[186] C. Gneiting and K. Hornberger, *Bell Test for the Free Motion of Material Particles*, Phys. Rev. Lett. **101**, 260503 (2008).

[187] J. D. Jost, J. P. Home, J. M. Amini, D. Hanneke, R. Ozeri, C. Langer, J. J. Bollinger, D. Leibfried, and D. J. Wineland, *Entangled mechanical oscillators*, Nature **459**, 683 (2009).

[188] J. I. Cirac and P. Zoller, *Quantum Computations with Cold Trapped Ions*, Phys. Rev. Lett. **74**, 4091 (1995).

[189] E. Arimondo and S. Wimberger, *Tunneling of Ultracold Atoms in Time-Independent Potentials*, chapter 11 in the book "Dynamical Tunneling: Theory and Experiment" edited by S. Keshavamurthy and P. Schlagheck, CRC Press (2011).

[190] D. Witthaut, F. Trimborn, and S. Wimberger, *Dissipation Induced Coherence of a Two-Mode Bose-Einstein Condensate*, Phys. Rev. Lett. **101**, 200402 (2008).

[191] J. Madroñero, A. Ponomarev, A. R. R. Carvalho, S. Wimberger, C. Viviescas, A. Kolovsky, K. Hornberger, P. Schlagheck, A. Krug, and A. Buchleitner, *Quantum Chaos, Transport, and Control - in Quantum Optics*, pages 33-73 in volume 53 of Adv. At. Mol. Opt. Phys. edited by G. Rempe and M. O. Scully, Elsevier (2006).

[192] G. Montambaux, D. Poilblanc, J. Bellissard, and C. Sire, *Quantum Chaos in Spin-Fermion Models*, Phys. Rev. Lett. **70**, 497 (1993).

[193] A. R. Kolovsky and A. Buchleitner, *Floquet-Bloch operator for the Bose-Hubbard model with static field*, Phys. Rev. E **68**, 056213 (2003).

[194] A. R. Kolovsky and A. Buchleitner, *Quantum chaos in the Bose-Hubbard model*, Europhys. Lett. **68**, 632 (2004).

[195] P. Buonsante and S. Wimberger, *Engineering many-body quantum dynamics by disorder*, Phys. Rev. A **77**, 041606(R) (2008).

[196] A. Tomadin, R. Mannella, and S. Wimberger, *Many-Body Interband Tunneling as a Witness of Complex Dynamics in the Bose-Hubbard Model*, Phys. Rev. Lett. **98**, 130402 (2007).

[197] A. Tomadin, R. Mannella, and S. Wimberger, *Many-body Landau-Zener tunneling in the Bose-Hubbard model*, Phys. Rev. A **77**, 013606 (2008).

[198] P. Buonsante and A. Vezzani, *Ground-State Fidelity and Bipartite Entanglement in the Bose-Hubbard Model*, Phys. Rev. Lett. **98**, 110601 (2007).

[199] S.-J. Gu, H.-M. Kwok, W.-Q. Ning, and H.-Q. Lin, *Fidelity susceptibility, scaling, and universality in quantum critical phenomena*, Phys. Rev. B **77**, 245109 (2008).

[200] S. Sachdev, *Quantum Phase Transitions*, Cambridge University Press, Cambridge (2001).

[201] T. Gorin, T. Prosen, T. H. Seligman, and M. Žnidarič, *Dynamics of Loschmidt echoes and fidelity decay*, Phys. Rep. **435**, 33 (2006).

[202] Ph. Jacquod and C. Petitjean, *Decoherence, entanglement and irreversibility in quantum dynamical systems with few degrees of freedom*, Adv. Phys. **58**, 67 (2009).

[203] C. Dembowski, B. Dietz, T. Friedrich, H.-D. Gräf, A. Heine, C. Mejía-Monasterio, M. Miski-Oglu, A. Richter, and T. H. Seligman, *First Experimental Evidence for Quantum Echoes in Scattering Systems*, Phys. Rev. Lett. **93**, 134102 (2004).

[204] R. Höhmann, U. Kuhl, and H.-J. Stöckmann, *Algebraic Fidelity Decay for Local Perturbations*, Phys. Rev. Lett. **100**, 124101 (2008).

[205] S. Schlunk, M. B. d'Arcy, S. A. Gardiner, D. Cassettari, R. M. Godun, and G. S. Summy, *Signatures of Quantum Stability in a Classically Chaotic System*, Phys. Rev. Lett. **90**, 054101 (2003).

[206] M. F. Andersen, A. Kaplan, T. Grünzweig, and N. Davidson, *Decay of Quantum Correlations in Atom Optics Billiards with Chaotic and Mixed Dynamics*, Phys. Rev. Lett. **97**, 104102 (2006).

[207] S. Wu, A. Tonyushkin, and M. G. Prentiss, *Observation of Saturation of Fidelity Decay with an Atom Interferometer*, Phys. Rev. Lett. **103**, 034101 (2009).

[208] S. Wimberger and A. Buchleitner, *Saturation of fidelity in the atom-optics kicked rotor*, J. Phys. B: At. Mol. Opt. Phys. **39**, L145 (2006).

[209] M. Abb, I. Guarneri, and S. Wimberger, *Pseudoclassical theory for fidelity of nearly resonant quantum rotors*, Phys. Rev. E **80**, 035206(R) (2009).

[210] F. Haake, *Quantum Signatures of Chaos*, Springer, Berlin and Heidelberg (1991).

[211] J. Zakrzewski and M. Kuś, *Distributions of Avoided Crossings for Quantum Chaotic Systems*, Phys. Rev. Lett. **67**, 2749 (1991).

[212] S.-J. Wang and Q. Jie, *General features of quantum chaos and its relevance to nuclear physics*, Phys. Rev. C **63**, 014309 (2000).

[213] P. Giorda and P. Zanardi, *Quantum chaos and operator fidelity metric*, Phys. Rev. E **81**, 017203 (2010).

[214] P. Plötz, *Ph.D. thesis: "Complex Dynamics of Ultracold Atoms"*, University of Heidelberg (available at http://archiv.ub.uni-heidelberg.de/volltextserver/11123/) (2010).

[215] W.-L. You, Y.-W. Li, and S.-J. Gu, *Fidelity, dynamic structure factor, and susceptibility in critical phenomena*, Phys. Rev. E **76**, 022101 (2007).

[216] M. Hiller, T. Kottos, and T. Geisel, *Wave-packet dynamics in energy space of a chaotic trimeric Bose-Hubbard system*, Phys. Rev. A **79**, 023621 (2009).

[217] L. D. Landau and E. M. Lifshitz, *Paragraph 79 in "Quantum Mechanics: Non-relativistic Theory"*, Elsevier (1977).

[218] P. Pechukas, *Distribution of Energy Eigenvalues in the Irregular Spectrum*, Phys. Rev. Lett. **51**, 943 (1983).

[219] J. H. Shirley, *Ph.D. thesis: "Interaction of a quantum system with a strong oscillating field"*, California Institute of Technology (available at http://resolver.caltech.edu/CaltechETD:etd-05142008-103758) (1963).

[220] P. Plötz, J. Madroñero, and S. Wimberger, *Collapse and revival in inter-band oscillations of a two-band Bose-Hubbard model*, J. Phys. B: At. Mol. Opt. Phys. **43**, 081001 (2010).

[221] H. Venzl, T. Zech, B. Oleś, M. Hiller, F. Mintert, and A. Buchleitner, *Solitonic eigenstates of the chaotic Bose-Hubbard Hamiltonian*, Appl. Phys. B **98**, 647 (2010).

[222] X. Yang and J. Burgdörfer, *Statistics of avoided crossings for generic quantum systems*, Phys. Rev. A **48**, 83 (1993).

[223] H. Kohler, I. E. Smolyarenko, C. Pineda, T. Guhr, F. Leyvraz, and T. H. Seligman, *Surprising Relations between Parametric Level Correlations and Fidelity Decay*, Phys. Rev. Lett. **100**, 190404 (2008).

[224] C. Lanczos, *An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators*, J. Res. Nat. Bur. Stand. **45**, 255 (1950).

[225] A. Buchleitner, D. Delande, and J.-C. Gay, *Microwave ionization of three-dimensional hydrogen atoms in a realistic numerical experiment*, J. Opt. Soc. Am. B **12**, 505 (1995).

[226] U. D. Jentschura, A. Surzhykov, M. Lubasch, and J. Zinn-Justin, *Structure, time propagation and dissipative terms for resonances*, J. Phys. A: Math. Theor. **41**, 095302 (2008).

[227] U. D. Jentschura, A. Surzhykov, and J. Zinn-Justin, *Unified Treatment of Even and Odd Anharmonic Oscillators of Arbitrary Degree*, Phys. Rev. Lett. **102**, 011601 (2009).

[228] U. D. Jentschura, A. Surzhykov, and J. Zinn-Justin, *Multi-instantons and exact results III: Unification of even and odd anharmonic oscillators*, Ann. Phys. **325**, 1135 (2010).

[229] J. Zinn-Justin and U. D. Jentschura, *Imaginary cubic perturbation: numerical and analytic study*, J. Phys. A: Math. Theor. **43**, 425301 (2010).

[230] C. M. Bender and S. Boettcher, *Real Spectra in Non-Hermitian Hamiltonians Having $\mathcal{PT}$ Symmetry*, Phys. Rev. Lett. **80**, 5243 (1998).

[231] C. M. Bender, S. Boettcher, and P. N. Meisinger, *$\mathcal{PT}$-symmetric quantum mechanics*, J. Math. Phys. **40**, 2201 (1999).

[232] C. M. Bender and G. V. Dunne, *Large-order perturbation theory for a non-Hermitian $\mathcal{PT}$-symmetric Hamiltonian*, J. Math. Phys. **40**, 4616 (1999).

[233] C. M. Bender and T. T. Wu, *Anharmonic Oscillator*, Phys. Rev. **184**, 1231 (1969).

[234] C. M. Bender and T. T. Wu, *Large-Order Behavior of Perturbation Theory*, Phys. Rev. Lett. **27**, 461 (1971).

[235] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Springer, Berlin, 3rd edition (2002).

[236] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, Oxford (1965).

[237] J. H. Wilkinson and C. H. Reinsch, *Linear Algebra. Handbook for Automatic Computation*, Springer, Heidelberg (1971).

[238] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice Hall, Englewood Cliffs (1998).

[239] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *Lapack Users' Guide*, Society for Industrial and Applied Mathematics, Philadelphia, 3rd edition (1999).

[240] C. M. Bender, J. Brod, A. Refig, and M. E. Reuter, *The $\mathcal{C}$ operator in $\mathcal{PT}$-symmetric quantum theories*, J. Phys. A: Math. Gen. **37**, 10139 (2004).

[241] I. Gohberg, P. Lancaster, and L. Rodman, *Matrices and Indefinite Scalar Products*, Birkhäuser Verlag, Basel (1983).

[242] N. Moiseyev, *Quantum theory of resonances: calculating energies, widths and cross-sections by complex scaling*, Phys. Rep. **302**, 211 (1998).

[243] P. Arbenz and M. E. Hochstenbach, *A Jacobi-Davidson Method for Solving Complex Symmetric Eigenvalue Problems*, SIAM J. Sci. Comput. **25**, 1655 (2004).

[244] J. G. F. Francis, *The QR Transformation - Part 1*, Comp. J. **4**, 265 (1961).

[245] J. G. F. Francis, *The QR Transformation - Part 2*, Comp. J. **4**, 332 (1962).

[246] J. H. Wilkinson, *Global Convergence of Tridiagonal QR Algorithm with Origin Shifts*, Lin. Alg. Appl. **1**, 409 (1968).

[247] S. Wolfram, *The Mathematica Book*, Cambridge University Press, Cambridge, 4th edition (1999).

[248] C. G. J. Jacobi, *Über ein leichtes Verfahren die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen*, Journal für die reine und angewandte Mathematik **1846**, 51 (1846).

[249] W. Givens, *Computation of Plain Unitary Rotations Transforming a General Matrix to Triangular Form*, J. Soc. Indust. Appl. Math. **6**, 26 (1958).

[250] C. R. Handy and D. Bessis, *Rapidly Convergent Lower Bounds for the Schrödinger-Equation Ground-State Energy*, Phys. Rev. Lett. **55**, 931 (1985).

[251] C. R. Handy, D. Bessis, and T. D. Morley, *Generating quantum energy bounds by the moment method: A linear-programming approach*, Phys. Rev. A **37**, 4557 (1988).

[252] C. R. Handy, D. Bessis, G. Sigismondi, and T. D. Morley, *Rapidly Converging Bounds for the Ground-State Energy of Hydrogenic Atoms in Superstrong Magnetic Fields*, Phys. Rev. Lett. **60**, 253 (1988).

[253] C. R. Handy, *Generating converging eigenenergy bounds for the discrete states of the $-ix^3$ non-Hermitian potential*, J. Phys. A: Math. Gen. **34**, L271 (2001).

[254] C. R. Handy, D. Khan, X.-Q. Wang, and C. J. Tymczak, *Multiscale reference function analysis of the $\mathcal{PT}$ symmetry breaking solutions for the $P^2 + iX^3 + i\alpha X$ Hamiltonian*, J. Phys. A: Math. Gen. **34**, 5593 (2001).

[255] C. R. Handy, *Generating converging bounds to the (complex) discrete states of the $P^2 + iX^3 + i\alpha X$ Hamiltonian*, J. Phys. A: Math. Gen. **34**, 5065 (2001).

[256] C. R. Handy and X.-Q. Wang, *Extension of a spectral bounding method to complex rotated Hamiltonians, with applications to $p^2 - ix^3$*, J. Phys. A: Math. Gen. **34**, 8297 (2001).

[257] Z. Yan and C. R. Handy, *Extension of a spectral bounding method to the pt-invariant states of the $-(iX)^N$ non-Hermitian potential*, J. Phys. A: Math. Gen. **34**, 9907 (2001).

[258] V. Franceschini, V. Grecchi, and H. J. Silverstone, *Complex energies from real perturbation series for the LoSurdo-Stark effect in hydrogen by Borel-Padé approximants*, Phys. Rev. A **32**, 1338 (1985).

[259] U. D. Jentschura, *Resummation of the divergent perturbation series for a hydrogen atom in an electric field*, Phys. Rev. A **64**, 013403 (2001).

[260] E. Caliceti, *Distributional Borel summability for vacuum polarization by an external electric field*, J. Math. Phys. **44**, 2026 (2003).

[261] D. H. Bailey, *A Fortran 90-based multiprecision system*, ACM Trans. Math. Soft. **21**, 379 (1995).

[262] W. Pauli, *On Dirac's New Method of Field Quantization*, Rev. Mod. Phys. **15**, 175 (1943).

[263] K. Abhinav, A. Jayannavar, and P. K. Panigrahi, *Conserved correlation in $\mathcal{PT}$-symmetric systems: Scattering and bound states*, Ann. Phys. **331**, 110 (2013).

[264] U. D. Jentschura and B. J. Wundt, *Localizability of tachyonic particles and neutrinoless double beta decay*, Eur. Phys. J. C **72**, 1894 (2012).

[265] U. D. Jentschura and B. J. Wundt, *From Generalized Dirac Equations to a Candidate for Dark Energy*, ISRN High Energy Phys. **2013**, 374612 (2013).

[266] C. M. Bender, S. Boettcher, and V. M. Savage, *Conjecture on the interlacing of zeros in complex Sturm-Liouville problems*, J. Math. Phys. **41**, 6381 (2000).

[267] G. Fasshauer, Chapter 11, *Lecture notes: "477/577 Numerical Linear Algebra / Computational Mathematics I"*, available at http://www.math.iit.edu/~fass/477577_Chapter_11.pdf (2006).

[268] P. Arbenz, Chapter 3, *Lecture notes: "Numerical Methods for Solving Large Scale Eigenvalue Problems"*, available at http://people.inf.ethz.ch/arbenz/ewp/Lnotes/chapter3.pdf (2012).

[269] L. E. Henderson, *Ph.D. thesis: "Testing eigenvalue software"*, University of Arizona (available at http://arizona.openrepository.com/arizona/handle/10150/185744) (1991).

[270] C. Itzykson and J. B. Zuber, *Quantum Field Theory*, McGraw-Hill, New York (1980).

[271] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Perseus, Cambridge, Massachusetts (1995).