Computational Quantum Physics

Philippe de Forcrand (forcrand@phys.ethz.ch)
Philipp Werner (werner@phys.ethz.ch)

ETH Zürich, FS 2009

Chapters 1-4 and parts of chapter 8 are taken from the lecture notes of Prof. M. Troyer.

Contents

Chapter 1

Introduction

1.1 General

For **physics students** the computational quantum physics courses is a recommended prerequisite for any computationally oriented semester thesis, proseminar, diploma thesis or doctoral thesis.

For **computational science and engineering (RW) students** the computational quantum physics courses is part of the "Vertiefung" in theoretical physics.

1.1.1 Lecture Notes

All the lecture notes, source codes, applets and supplementary material can be found on our web page `http://www.itp.phys.ethz.ch/education/lectures_fs09/cqp`.

1.1.2 Exercises

Programming Languages

Except when a specific programming language or tool is explicitly requested you are free to choose any programming language you like. Solutions will often be given either as C++ programs or Mathematica Notebooks.

Computer Access

The lecture rooms offer both Linux workstations, for which accounts can be requested with the computer support group of the physics department in the HPR building, as well as connections for your notebook computers.

1.1.3 Prerequisites

As a prerequisite for this course we expect knowledge of the following topics. Please contact us if you have any doubts or questions.

Computing

- Basic knowledge of UNIX
- At least one procedural programming language such as C, C++, Pascal, Java or FORTRAN. C++ knowledge is preferred.
- Knowledge of a symbolic mathematics program such as Mathematica or Maple.
- Ability to produce graphical plots.

Numerical Analysis

- Numerical integration and differentiation
- Linear solvers and eigensolvers
- Root solvers and optimization
- Statistical analysis

Quantum Mechanics

Basic knowledge of quantum mechanics, at the level of the quantum mechanics taught to computational scientists, should be sufficient to follow the course. If you feel lost at any point, please ask the lecturer to explain whatever you do not understand. We want you to be able to follow this course without taking an advanced quantum mechanics class.

1.1.4 References

1. J.M. Thijssen, *Computational Physics*, Cambridge University Press (1999) ISBN 0521575885
2. Nicholas J. Giordano, *Computational Physics*, Pearson Education (1996) ISBN 0133677230.
3. Harvey Gould and Jan Tobochnik, *An Introduction to Computer Simulation Methods*, 2nd edition, Addison Wesley (1996), ISBN 00201506041
4. Tao Pang, *An Introduction to Computational Physics*, Cambridge University Press (1997) ISBN 0521485924

1.2 Overview

In this class we will learn how to simulate quantum systems, starting from the simple one-dimensional Schrödinger equation to simulations of interacting quantum many body problems in condensed matter physics and in quantum field theories. In particular we will study

- The one-body Schrödinger equation and its numerical solution
- The many-body Schrödinger equation and second quantization
- Approximate solutions to the many body Schrödinger equation
- Path integrals and quantum Monte Carlo simulations
- Numerically exact solutions to (some) many body quantum problems
- Some simple quantum field theories

Chapter 2

Quantum mechanics in one hour

2.1 Introduction

The purpose of this chapter is to refresh your knowledge of quantum mechanics and to establish notation. Depending on your background you might not be familiar with all the material presented here. If that is the case, please ask the lecturers and we will expand the introduction. Those students who are familiar with advanced quantum mechanics are asked to glance over some omissions and are encouraged to help us improve this quick introduction.

2.2 Basis of quantum mechanics

2.2.1 Wave functions and Hilbert spaces

Quantum mechanics is nothing but simple linear algebra, albeit in huge Hilbert spaces, which makes the problem hard. The foundations are pretty simple though.

A pure state of a quantum system is described by a "wave function" $|\Psi\rangle$, which is an element of a Hilbert space $\mathcal{H}$:

$$|\Psi\rangle \in \mathcal{H} \tag{2.1}$$

Usually the wave functions are normalized:

$$|| \, |\Psi\rangle \, || = \sqrt{\langle\Psi|\Psi\rangle} = 1. \tag{2.2}$$

Here the "bra-ket" notation

$$\langle\Phi|\Psi\rangle \tag{2.3}$$

denotes the scalar product of the two wave functions $|\Phi\rangle$ and $|\Psi\rangle$.

The simplest example is the spin-1/2 system, describing e.g. the two spin states of an electron. Classically the spin $\vec{S}$ of the electron (which can be visualized as an internal angular momentum), can point in any direction. In quantum mechanics it is described by a two-dimensional complex Hilbert space $\mathcal{H} = \mathbb{C}^2$. A common choice of

basis vectors are the "up" and "down" spin states

$$|\uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{2.4}$$

$$|\downarrow\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{2.5}$$

This is similar to the classical Ising model, but in contrast to a classical Ising spin that can point only either up or down, the quantum spin can exist in any complex superposition

$$|\Psi\rangle = \alpha|\uparrow\rangle + \beta|\downarrow\rangle \tag{2.6}$$

of the basis states, where the normalization condition (2.2) requires that $|\alpha|^2+|\beta|^2 = 1$.

For example, as we will see below the state

$$|\rightarrow\rangle = \frac{1}{\sqrt{2}}\left(|\uparrow\rangle + |\downarrow\rangle\right) \tag{2.7}$$

is a superposition that describes the spin pointing in the positive x-direction.

2.2.2 Mixed states and density matrices

Unless specifically prepared in a pure state in an experiment, quantum systems in Nature rarely exist as pure states but instead as probabilistic superpositions. The most general state of a quantum system is then described as a density matrix ρ, with unit trace

$$\mathrm{Tr}\rho = 1. \tag{2.8}$$

The density matrix of a pure state is just the projector onto that state

$$\rho_{\mathrm{pure}} = |\Psi\rangle\langle\Psi|. \tag{2.9}$$

For example, the density matrix of a spin pointing in the positive x-direction is

$$\rho_{\rightarrow} = |\rightarrow\rangle\langle\rightarrow| = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}. \tag{2.10}$$

Instead of being in a coherent superposition of up and down, the system could also be in a probabilistic mixed state, with a 50% probability of pointing up and a 50% probability of pointing down, which would be described by the density matrix

$$\rho_{\mathrm{mixed}} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}. \tag{2.11}$$

2.2.3 Observables

Any physical observable is represented by a self-adjoint linear operator acting on the Hilbert space, which in a final dimensional Hilbert space can be represented by a Hermitian matrix. For our spin-1/2 system, using the basis introduced above, the components

S^x, S^y and S^z of the spin in the x-, y-, and z-directions are represented by the Pauli matrices

$$S^x = \frac{\hbar}{2}\sigma_x = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{2.12}$$

$$S^y = \frac{\hbar}{2}\sigma_y = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{2.13}$$

$$S^z = \frac{\hbar}{2}\sigma_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.14}$$

The spin component along an arbitrary unit vector $\hat{e}$ is the linear superposition of the components, i.e.

$$\hat{e}\cdot\vec{S} = e^x S^x + e^y S^y + e^z S^z = \frac{\hbar}{2}\begin{pmatrix} e^z & e^x - ie^y \\ e^x + ie^y & -e^z \end{pmatrix} \tag{2.15}$$

The fact that these observables do not commute but instead satisfy the non-trivial commutation relations

$$[S^x, S^y] = S^x S^y - S^y S^x = i\hbar S^z, \tag{2.16}$$

$$[S^y, S^z] = i\hbar S^x, \tag{2.17}$$

$$[S^z, S^x] = i\hbar S^y, \tag{2.18}$$

will have important consequences.

2.2.4 The measurement process

The outcome of a measurement in a quantum system is usually intrusive and not deterministic. After measuring an observable A, the new wave function of the system will be an eigenvector of A and the outcome of the measurement the corresponding eigenvalue. The state of the system is thus changed by the measurement process!

For example, if we start with a spin pointing up with wave function

$$|\Psi\rangle = |\uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{2.19}$$

or alternatively density matrix

$$\rho_\uparrow = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \tag{2.20}$$

and we measure the x-component of the spin S^x, the resulting measurement will be either $+\hbar/2$ or $-\hbar/2$, depending on whether the spin after the measurement points in the $+$ or $-$ x-direction, and the wave function after the measurement will be either of

$$|\rightarrow\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle) = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \tag{2.21}$$

$$|\leftarrow\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle - |\downarrow\rangle) = \begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix} \tag{2.22}$$

Either of these states will be picked with a probability given by the overlap of the initial wave function by the individual eigenstates:

$$p_{\rightarrow} = ||\langle \rightarrow |\Psi\rangle||^2 = 1/2 \tag{2.23}$$
$$p_{\leftarrow} = ||\langle \leftarrow |\Psi\rangle||^2 = 1/2 \tag{2.24}$$

The final state is a probabilistic superposition of these two outcomes, described by the density matrix

$$\rho = p_{\rightarrow}|\rightarrow\rangle\langle\rightarrow| + p_{\leftarrow}|\leftarrow\rangle\langle\leftarrow| = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}. \tag{2.25}$$

which differs from the initial density matrix $\rho_{\uparrow}$.

If we are not interested in the result of a particular outcome, but just in the average, the expectation value of the measurement can easily be calculated from a wave function $|\Psi\rangle$ as

$$\langle A\rangle = \langle\Psi|A|\Psi\rangle \tag{2.26}$$

or from a density matrix ρ as

$$\langle A\rangle = \mathrm{Tr}(\rho A). \tag{2.27}$$

It is easy to convince yourself that for pure states the two formulations are identical.

2.2.5 The uncertainty relation

If two observables A and B do not commute $[A, B] \neq 0$, they cannot be measured simultaneously. If A is measured first, the wave function is changed to an eigenstate of A, which changes the result of a subsequent measurement of B. As a consequence the values of A and B in a state Ψ cannot be simultaneously known, which is quantified by the famous Heisenberg uncertainty relation which states that if two observables A and B do not commute but satisfy

$$[A, B] = i\hbar \tag{2.28}$$

then the product of the root-mean-square deviations ΔA and ΔB of simultaneous measurements of A and B has to be larger than

$$\Delta A \Delta B \geq \hbar/2 \tag{2.29}$$

For more details about the uncertainty relation, the measurement process or the interpretation of quantum mechanics we refer interested students to an advanced quantum mechanics class or text book.

2.2.6 The Schrödinger equation

The time-dependent Schrödinger equation

After so much introduction the Schrödinger equation is very easy to present. The wave function $|\Psi\rangle$ of a quantum system evolves according to

$$i\hbar\frac{\partial}{\partial t}|\Psi(t)\rangle = H|\Psi(t)\rangle, \tag{2.30}$$

where H is the Hamilton operator. This is just a first order linear differential equation.

The time-independent Schrödinger equation

For a stationary time-independent problem the Schrödinger equation can be simplified. Using the ansatz

$$|\Psi(t)\rangle = \exp(-iEt/\hbar)|\Psi\rangle, \tag{2.31}$$

where E is the energy of the system, the Schrödinger equation simplifies to a linear eigenvalue problem

$$H|\Psi\rangle = E|\Psi\rangle. \tag{2.32}$$

The rest of the semester will be spent solving just this simple eigenvalue problem!

The Schrödinger equation for the density matrix

The time evolution of a density matrix $\rho(t)$ can be derived from the time evolution of pure states, and can be written as

$$i\hbar\frac{\partial}{\partial t}\rho(t) = [H, \rho(t)] \tag{2.33}$$

The proof is left as a simple exercise.

2.2.7 The thermal density matrix

Finally we want to describe a physical system not in the ground state but in thermal equilibrium at a given inverse temperature $\beta = 1/k_B T$. In a classical system each microstate i of energy E_i is occupied with a probability given by the Boltzman distribution

$$p_i = \frac{1}{Z}\exp(-\beta E_i), \tag{2.34}$$

where the partition function

$$Z = \sum_i \exp(-\beta E_i) \tag{2.35}$$

normalizes the probabilities.

In a quantum system, if we use a basis of eigenstates $|i\rangle$ with energy E_i, the density matrix can be written analogously as

$$\rho_\beta = \frac{1}{Z}\sum_i \exp(-\beta E_i)|i\rangle\langle i| \tag{2.36}$$

For a general basis, which is not necessarily an eigenbasis of the Hamiltonian H, the density matrix can be obtained by diagonalizing the Hamiltonian, using above equation, and transforming back to the original basis. The resulting density matrix is

$$\rho_\beta = \frac{1}{Z}\exp(-\beta H) \tag{2.37}$$

where the partition function now is

$$Z = \mathrm{Tr}\exp(-\beta H) \tag{2.38}$$

Calculating the thermal average of an observable A in a quantum system is hence very easy:

$$\langle A\rangle = \mathrm{Tr}(A\rho_\beta) = \frac{\mathrm{Tr}A\exp(-\beta H)}{\mathrm{Tr}\exp(-\beta H)}. \tag{2.39}$$

2.3 The spin-S problem

Before discussing solutions of the Schrödinger equation we will review two very simple systems: a localized particle with general spin S and a free quantum particle.

In section 2.2.1 we have already seen the Hilbert space and the spin operators for the most common case of a spin-1/2 particle. The algebra of the spin operators given by the commutation relations (2.12)-(2.12) allows not only the two-dimensional representation shown there, but a series of $2S+1$-dimensional representations in the Hilbert space $\mathbb{C}^{2S+1}$ for all integer and half-integer values $S = 0, \frac{1}{2}, 1, \frac{3}{2}, 2, \ldots$. The basis states $\{|s\rangle\}$ are usually chosen as eigenstates of the S^z operator

$$S^z|s\rangle = \hbar s|s\rangle, \tag{2.40}$$

where s can take any value in the range $-S, -S+1, -S+2, \ldots, S-1, S$. In this basis the S_z operator is diagonal, and the S^x and S^y operators can be constructed from the "ladder operators"

$$S^+|s\rangle = \sqrt{S(S+1) - s(s+1)}|s+1\rangle \tag{2.41}$$

$$S^-|s\rangle = \sqrt{S(S+1) - s(s-1)}|s-1\rangle \tag{2.42}$$

which increment or decrement the S^z value by 1 through

$$S^x = \frac{1}{2}\left(S^+ + S^-\right) \tag{2.43}$$

$$S^y = \frac{1}{2i}\left(S^+ - S^-\right). \tag{2.44}$$

The Hamiltonian of the spin coupled to a magnetic field $\vec{h}$ is

$$H = -g\mu_B \vec{h} \cdot \vec{S}, \tag{2.45}$$

which introduces nontrivial dynamics since the components of $\vec{S}$ do not commute. As a consequence the spin precesses around the magnetic field direction.

Exercise: Derive the differential equation governing the rotation of a spin starting along the $+x$-direction rotating under a field in the $+z$-direction

2.4 A quantum particle in free space

Our second example is a single quantum particle in an n-dimensional free space. Its Hilbert space is given by all twice-continuously differentiable complex functions over the real space $\mathbb{R}^n$. The wave functions $|\Psi\rangle$ are complex-valued functions $\Psi(\vec{x})$ in n-dimensional space. In this representation the operator $\hat{x}$, measuring the position of the particle is simple and diagonal

$$\hat{x} = \vec{x}, \tag{2.46}$$

while the momentum operator $\hat{p}$ becomes a differential operator

$$\hat{p} = -i\hbar\nabla. \tag{2.47}$$

These two operators do not commute but their commutator is

$$[\hat{x}, \hat{p}] = i\hbar. \tag{2.48}$$

The Schrödinger equation of a quantum particle in an external potential $V(\vec{x})$ can be obtained from the classical Hamilton function by replacing the momentum and position variables by the operators above. Instead of the classical Hamilton function

$$H(\vec{x}, \vec{p}) = \frac{\vec{p}^2}{2m} + V(\vec{x}) \tag{2.49}$$

we use the quantum mechanical Hamiltonian operator

$$H = \frac{\hat{p}^2}{2m} + V(\hat{x}) = -\frac{\hbar^2}{2m}\nabla^2 + V(\vec{x}), \tag{2.50}$$

which gives the famous form

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi + V(\vec{x})\psi \tag{2.51}$$

of the one-body Schrödinger equation.

2.4.1 The harmonic oscillator

As a special exactly solvable case let us consider the one-dimensional quantum harmonic oscillator with mass m and potential $\frac{K}{2}x^2$. Defining momentum $\hat{p}$ and position operators $\hat{q}$ in units where $m = \hbar = K = 1$, the time-independent Schrödinger equation is given by

$$H|n\rangle = \frac{1}{2}(\hat{p}^2 + \hat{q}^2)|n\rangle = E_n|n\rangle \tag{2.52}$$

Inserting the definition of $\hat{p}$ we obtain an eigenvalue problem of an ordinary differential equation

$$-\frac{1}{2}\phi_n''(q) + \frac{1}{2}q^2\phi_n(q) = E_n\phi_n(q) \tag{2.53}$$

whose eigenvalues $E_n = (n + 1/2)$ and eigenfunctions

$$\phi_n(q) = \frac{1}{\sqrt{2^n n!\sqrt{\pi}}} \exp\left(-\frac{1}{2}q^2\right) H_n(q), \tag{2.54}$$

are known analytically. Here the H_n are the Hermite polynomials and $n = 0, 1, \ldots$.

Using these eigenstates as a basis sets we need to find the representation of $\hat{q}$ and $\hat{p}$. Performing the integrals

$$\langle m|\hat{q}|n\rangle \qquad \text{and} \qquad \langle m|\hat{p}|n\rangle \tag{2.55}$$

it turns out that they are nonzero only for $m = n \pm 1$ and they can be written in terms of "ladder operators" a and $a^\dagger$:

$$\hat{q} = \frac{1}{\sqrt{2}}(a^\dagger + a) \tag{2.56}$$

$$\hat{p} = \frac{1}{i\sqrt{2}}(a^\dagger - a) \tag{2.57}$$

$$\tag{2.58}$$

where the raising and lowering operators $a^\dagger$ and a only have the following nonzero matrix elements:

$$\langle n+1|a^\dagger|n\rangle = \langle n|a|n+1\rangle = \sqrt{n+1}. \tag{2.59}$$

and commutation relations

$$[a,a] = [a^\dagger, a^\dagger] = 0 \tag{2.60}$$

$$[a, a^\dagger] = 1. \tag{2.61}$$

It will also be useful to introduce the number operator $\hat{n} = a^\dagger a$ which is diagonal with eigenvalue n: elements

$$\hat{n}|n\rangle = a^\dagger a|n\rangle = \sqrt{n}a^\dagger|n-1\rangle = n||n\rangle. \tag{2.62}$$

To check this representation let us plug the definitions back into the Hamiltonian to obtain

$$\begin{aligned} H &= \frac{1}{2}(\hat{p}^2 + \hat{q}^2) \\ &= \frac{1}{4}\left[-(a^\dagger - a)^2 + (a^\dagger + a)^2\right] \\ &= \frac{1}{2}(a^\dagger a + aa^\dagger) \\ &= \frac{1}{2}(2a^\dagger a + 1) = \hat{n} + \frac{1}{2}, \end{aligned} \tag{2.63}$$

which has the correct spectrum. In deriving the last lines we have used the commutation relation (2.61).

Chapter 3

The quantum one-body problem

3.1 The time-independent 1D Schrödinger equation

We start the numerical solution of quantum problems with the time-indepent one-dimensional Schrödinger equation for a particle with mass m in a Potential $V(x)$. In one dimension the Schrödinger equation is just an ordinary differential equation

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + V(x)\psi(x) = E\psi(x). \tag{3.1}$$

We start with simple finite-difference schemes and discretize space into intervals of length Δx and denote the space points by

$$x_n = n\Delta x \tag{3.2}$$

and the wave function at these points by

$$\psi_n = \psi(x_n). \tag{3.3}$$

3.1.1 The Numerov algorithm

After rewriting the second order differential equation to a coupled system of two first order differential equations, any ODE solver such as the Runge-Kutta method could be applied, but there exist better methods. For the special form

$$\psi''(x) + k(x)\psi(x) = 0, \tag{3.4}$$

of the Schrödinger equation, with $k(x) = 2m(E - V(x))/\hbar^2$ we can derive the Numerov algorithm by starting from the Taylor expansion of ψ_n:

$$\psi_{n\pm1} = \psi_n \pm \Delta x\psi'_n + \frac{\Delta x^2}{2}\psi''_n \pm \frac{\Delta x^3}{6}\psi^{(3)}_n + \frac{\Delta x^4}{24}\psi^{(4)}_n \pm \frac{\Delta x^5}{120}\psi^{(5)}_n + \mathrm{O}(\Delta x^6) \tag{3.5}$$

Adding ψ_{n+1} and ψ_{n-1} we obtain

$$\psi_{n+1} + \psi_{n-1} = 2\psi_n + (\Delta x)^2\psi''_n + \frac{(\Delta x)^4}{12}\psi^{(4)}_n. \tag{3.6}$$

Replacing the fourth derivatives by a finite difference second derivative of the second derivatives

$$\psi_n^{(4)} = \frac{\psi''_{n+1} + \psi''_{n-1} - 2\psi''_n}{\Delta x^2} \tag{3.7}$$

and substituting $-k(x)\psi(x)$ for $\psi''(x)$ we obtain the Numerov algorithm

$$\begin{aligned}\left(1 + \frac{(\Delta x)^2}{12} k_{n+1}\right)\psi_{n+1} = \quad & 2\left(1 - \frac{5(\Delta x)^2}{12} k_n\right)\psi_n \\ & - \left(1 + \frac{(\Delta x)^2}{12} k_{n-1}\right)\psi_{n-1} + \mathrm{O}(\Delta x^6),\end{aligned} \tag{3.8}$$

which is locally of sixth order!

Initial values

To start the Numerov algorithm we need the wave function not just at one but at two initial values and will now present several ways to obtain these.

For potentials $V(x)$ with reflection symmetry $V(x) = V(-x)$ the wave functions need to be either even $\psi(x) = \psi(-x)$ or odd $\psi(x) = -\psi(-x)$ under reflection, which can be used to find initial values:

- For the even solution we use a half-integer mesh with mesh points $x_{n+1/2} = (n + 1/2)\Delta x$ and pick initial values $\psi_(x_{-1/2}) = \psi(x_{1/2}) = 1$.
- For the odd solution we know that $\psi(0) = -\psi(0)$ and hence $\psi(0) = 0$, specifying the first starting value. Using an integer mesh with mesh points $x_n = n\Delta x$ we pick $\psi(x_1) = 1$ as the second starting value.

In general potentials we need to use other approaches. If the potentials vanishes for large distances: $V(x) = 0$ for $|x| \geq a$ we can use the exact solution of the Schrödinger equation at large distances to define starting points, e.g.

$$\psi(-a) = 1 \tag{3.9}$$

$$\psi(-a - \Delta x) = \exp(-\Delta x\sqrt{2mE/\hbar}). \tag{3.10}$$

Finally, if the potential never vanishes we need to begin with a single starting value $\psi(x_0)$ and obtain the second starting value $\psi(x_1)$ by performing an integration over the first time step $\Delta\tau$ with an Euler or Runge-Kutta algorithm.

3.1.2 The one-dimensional scattering problem

The scattering problem is the numerically easiest quantum problem since solutions exist for all energies $E > 0$, if the potential vanishes at large distances ($V(x) \to 0$ for $|x| \to \infty$). The solution becomes particularly simple if the potential is nonzero only on a finite interval $[0, a]$. For a particle approaching the potential barrier from the left ($x < 0$) we can make the following ansatz for the free propagation when $x < 0$:

$$\psi_L(x) = A\exp(-iqx) + B\exp(iqx) \tag{3.11}$$

where A is the amplitude of the incoming wave and B the amplitude of the reflected wave. On the right hand side, once the particle has left the region of finite potential ($x > a$), we can again make a free propagation ansatz,

$$\psi_R(x) = C \exp(-iqx) \tag{3.12}$$

The coefficients A, B and C have to be determined self-consistently by matching to a numerical solution of the Schrödinger equation in the interval $[0, a]$. This is best done in the following way:

- Set $C = 1$ and use the two points a and $a + \Delta x$ as starting points for a Numerov integration.
- Integrate the Schrödinger equation numerically – backwards in space, from a to 0 – using the Numerov algorithm.
- Match the numerical solution of the Schrödinger equation for $x < 0$ to the free propagation ansatz (3.11) to determine A and B.

Once A and B have been determined the reflection and transmission probabilities R and T are given by

$$R = |B|^2/|A|^2 \tag{3.13}$$
$$T = 1/|A|^2 \tag{3.14}$$

3.1.3 Bound states and solution of the eigenvalue problem

While there exist scattering states for all energies $E > 0$, bound states solutions of the Schrödinger equation with $E < 0$ exist only for discrete energy eigenvalues. Integrating the Schrödinger equation from $-\infty$ to $+\infty$ the solution will diverge to $\pm\infty$ as $x \to \infty$ for almost all values. These functions cannot be normalized and thus do not constitute solutions to the Schrödinger equation. Only for some special eigenvalues E, will the solution go to zero as $x \to \infty$.

A simple eigensolver can be implemented using the following shooting method, where we again will assume that the potential is zero outside an interval $[0, a]$:

- Start with an initial guess E
- Integrate the Schrödinger equation for $\psi_E(x)$ from $x = 0$ to $x_f \gg a$ and determine the value $\psi_E(x_f)$
- use a root solver, such as a bisection method (see appendix A.1), to look for an energy E with $\psi_E(x_f) \approx 0$

This algorithm is not ideal since the divergence of the wave function for $x \pm \infty$ will cause roundoff error to proliferate.

A better solution is to integrate the Schrödinger equation from both sides towards the center:

- We search for a point b with $V(b) = E$

- Starting from $x = 0$ we integrate the left hand side solution $\psi_L(x)$ to a chosen point b and obtain $\psi_L(b)$ and a numerical estimate for $\psi'_L(b) = (\psi_L(b) - \psi_L(b - \Delta x))/\Delta x$.
- Starting from $x = a$ we integrate the right hand solution $\psi_R(x)$ down to the same point b and obtain $\psi_R(b)$ and a numerical estimate for $\psi'_R(b) = (\psi_R(b + \Delta x) - \psi_R(b))/\Delta x$.
- At the point b the wave functions and their first two derivatives have to match, since solutions to the Schrödinger equation have to be twice continuously differentiable. Keeping in mind that we can multiply the wave functions by an arbitrary factor we obtain the conditions

$$\begin{aligned} \psi_L(b) &= \alpha\psi_R(b) && (3.15)\\ \psi'_L(b) &= \alpha\psi'_R(b) && (3.16)\\ \psi''_L(b) &= \alpha\psi''_R(b) && (3.17) \end{aligned}$$

  The last condition is automatically fulfilled since by the choice $V(b) = E$ the Schrödinger equation at b reduces to $\psi''(b) = 0$. The first two conditions can be combined to the condition that the logarithmic derivatives vanish:

$$\frac{d\log\psi_L}{dx}|_{x=b} = \frac{\psi'_L(b)}{\psi_L(b)} = \frac{\psi'_R(b)}{\psi_R(b)} = \frac{d\log\psi_R}{dx}|_{x=b} \qquad (3.18)$$

- This last equation has to be solved for in a shooting method, e.g. using a bisection algorithm

Finally, at the end of the calculation, normalize the wave function.

3.2 The time-independent Schrödinger equation in higher dimensions

The time independent Schrödinger equation in more than one dimension is a partial differential equation and cannot, in general, be solved by a simple ODE solver such as the Numerov algorithm. Before employing a PDE solver we should thus always first try to reduce the problem to a one-dimensional problem. This can be done if the problem factorizes.

3.2.1 Factorization along coordinate axis

A first example is a three-dimensional Schrödinger equation in a cubic box with potential $V(\vec{r}) = V(x)V(y)V(z)$ with $\vec{r} = (x, y, z)$. Using the product ansatz

$$\psi(\vec{r}) = \psi_x(x)\psi_y(y)\psi_z(z) \qquad (3.19)$$

the PDE factorizes into three ODEs which can be solved as above.

3.2.2 Potential with spherical symmetry

Another famous trick is possible for spherically symmetric potentials with $V(\vec{r}) = V(|\vec{r}|)$ where an ansatz using spherical harmonics

$$\psi_{l,m}(\vec{r}) = \psi_{l,m}(r, \theta, \phi) = \frac{u(r)}{r} Y_{lm}(\theta, \phi) \tag{3.20}$$

can be used to reduce the three-dimensional Schrödinger equation to a one-dimensional one for the radial wave function $u(r)$:

$$\left[-\frac{\hbar^2}{2m} \frac{d^2}{dr^2} + \frac{\hbar^2 l(l+1)}{2mr^2} + V(r) \right] u(r) = E u(r) \tag{3.21}$$

in the interval $[0, \infty[$. Given the singular character of the potential for $r \to 0$, a numerical integration should start at large distances r and integrate towards $r = 0$, so that the largest errors are accumulated only at the last steps of the integration.

3.2.3 Finite difference methods

The simplest solvers for partial differential equations, the finite difference solvers can also be used for the Schrödinger equation. Replacing differentials by differences we convert the Schrödinger equation to a system of coupled inear equations. Starting from the three-dimensional Schrödinger equation (we set $\hbar = 1$ from now on)

$$\nabla^2 \psi(\vec{x}) + 2m(V - E(\vec{x}))\psi(\vec{x}) = 0, \tag{3.22}$$

we discretize space and obtain the system of linear equations

$$\begin{aligned} \frac{1}{\Delta x^2} & [\psi(x_{n+1}, y_n, z_n) + \psi(x_{n-1}, y_n, z_n) \\ & + \psi(x_n, y_{n+1}, z_n) + \psi(x_n, y_{n-1}, z_n) \\ & + \psi(x_n, y_n, z_{n+1}) + \psi(x_n, y_n, z_{n-1})] \\ + & \left[2m(V(\vec{x}) - E) - \frac{6}{\Delta x^2} \right] \psi(x_n, y_n, z_n) \quad = \quad 0. \end{aligned} \tag{3.23}$$

For the scattering problem a linear equation solver can now be used to solve the system of equations. For small linear problems Mathematica can be used, or the `dsysv` function of the LAPACK library. For larger problems it is essential to realize that the matrices produced by the discretization of the Schrödinger equation are usually very sparse, meaning that only $O(N)$ of the N^2 matrix elements are nonzero. For these sparse systems of equations, optimized iterative numerical algorithms exist[1] and are implemented in numerical libraries such as in the ITL library.[2]

[1] R. Barret, M. Berry, T.F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods* (SIAM, 1993)

[2] J.G. Siek, A. Lumsdaine and Lie-Quan Lee, *Generic Programming for High Performance Numerical Linear Algebra* in *Proceedings of the SIAM Workshop on Object Oriented Methods for Inter-operable Scientific and Engineering Computing (OO'98)* (SIAM, 1998); the library is availavle on the web at: `http://www.osl.iu.edu/research/itl/`

To calculate bound states, an eigenvalue problem has to be solved. For small problems, where the full matrix can be stored in memory, Mathematica or the `dsyev` eigensolver in the LAPACK library can be used. For bigger systems, sparse solvers such as the Lanczos algorithm (see appendix A.2) are best. Again there exist efficient implementations[3] of iterative algorithms for sparse matrices.[4]

3.2.4 Variational solutions using a finite basis set

In the case of general potentials, or for more than two particles, it will not be possible to reduce the Schrödinger equation to a one-dimensional problem and we need to employ a PDE solver. One approach will again be to discretize the Schrödinger equation on a discrete mesh using a finite difference approximation. A better solution is to expand the wave functions in terms of a finite set of basis functions

$$|\phi\rangle = \sum_{i=1}^{N} a_i |u_i\rangle. \tag{3.24}$$

To estimate the ground state energy we want to minimize the energy of the variational wave function

$$E^* = \frac{\langle\phi|H|\phi\rangle}{\langle\phi|\phi\rangle}. \tag{3.25}$$

Keep in mind that, since we only chose a finite basis set $\{|u_i\rangle\}$ the variational estimate E^* will always be larger than the true ground state energy E_0, but will converge towards E_0 as the size of the basis set is increased, e.g. by reducing the mesh size in a finite element basis.

To perform the minimization we denote by

$$H_{ij} = \langle u_i|H|u_j\rangle = \int d\vec{r} u_i(\vec{r})^* \left(-\frac{\hbar^2}{2m}\nabla^2 + V\right) u_j(\vec{r}) \tag{3.26}$$

the matrix elements of the Hamilton operator H and by

$$S_{ij} = \langle u_i|u_j\rangle = \int d\vec{r} u_i(\vec{r})^* u_j(\vec{r}) \tag{3.27}$$

the overlap matrix. Note that for an orthogonal basis set, S_{ij} is the identity matrix δ_{ij}. Minimizing equation (3.25) we obtain a generalized eigenvalue problem

$$\sum_j H_{ij} a_j = E \sum_k S_{ik} a_k. \tag{3.28}$$

or in a compact notation with $\vec{a} = (a_1, \ldots, a_N)$

$$H\vec{a} = ES\vec{a}. \tag{3.29}$$

[3] http://www.comp-phys.org/software/ietl/

[4] Z. Bai, J. Demmel and J. Dongarra (Eds.), *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide* (SIAM, 2000).

If the basis set is orthogonal this reduces to an ordinary eigenvalue problem and we can use the Lanczos algorithm.

In the general case we have to find orthogonal matrices U such that $U^T S U$ is the identity matrix. Introducing a new vector $\vec{b} = U^{-1}\vec{a}$. we can then rearrange the problem into

$$\begin{aligned} H\vec{a} &= ES\vec{a} \\ HU\vec{b} &= ESU\vec{b} \\ U^T HU\vec{b} &= EU^T SU\vec{b} = E\vec{b} \end{aligned} \tag{3.30}$$

and we end up with a standard eigenvalue problem for $U^T HU$. Mathematica and LAPACK both contain eigensolvers for such generalized eigenvalue problems.

Example: the anharmonic oscillator

The final issue is the choice of basis functions. It is advantageous to make use of known solutions to a similar problem as we will illustrate in the case of an anharmonic oscillator with Hamilton operator

$$\begin{aligned} H &= H_0 + \lambda q^4 \\ H_0 &= \frac{1}{2}(p^2 + q^2), \end{aligned} \tag{3.31}$$

where the harmonic oscillator H_0 was already discussed in section 2.4.1. It makes sense to use the N lowest harmonic oscillator eigenvectors $|n\rangle$ as basis states of a finite basis and write the Hamiltonian as

$$H = \frac{1}{2} + \hat{n} + \lambda \hat{q}^4 = \frac{1}{2} + \hat{n} + \frac{\lambda}{4}(a^\dagger + a)^4 \tag{3.32}$$

Since the operators a and $a^\dagger$ are nonzero only in the first sub or superdiagonal, the resulting matrix is a banded matrix of bandwidth 9. A sparse eigensolver such as the Lanczos algorithm can again be used to calculate the spectrum. Note that since we use the orthonormal eigenstates of H_0 as basis elements, the overlap matrix S here is the identity matrix and we have to deal only with a standard eigenvalue problem. A solution to this problem is provided in a Mathematica notebook on the web page.

The finite element method

In cases where we have irregular geometries or want higher precision than the lowest order finite difference method, and do not know a suitable set of basis function, the finite element method (FEM) should be chosen over the finite difference method. Since explaining the FEM can take a full semester in itself, we refer interested students to classes on solving partial differential equations.

3.3 The time-dependent Schrödinger equation

Finally we will reintroduce the time dependence to study dynamics in non-stationary quantum systems.

3.3.1 Spectral methods

By introducing a basis and solving for the complete spectrum of energy eigenstates we can directly solve the time-dependent problem in the case of a stationary Hamiltonian. This is a consequence of the linearity of the Schrödinger equation.

To calculate the time evolution of a state $|\psi(t_0)\rangle$ from time t_0 to t we first solve the stationary eigenvalue problem $H|\phi\rangle = E|\phi\rangle$ and calculate the eigenvectors $|\phi_n\rangle$ and eigenvalues ϵ_n. Next we represent the initial wave function $|\psi\rangle$ by a spectral decomposition

$$|\psi(t_0)\rangle = \sum_n c_n |\phi_n\rangle. \tag{3.33}$$

Since each of the $|\phi_n\rangle$ is an eigenvector of H, the time evolution $e^{-i\hbar H(t-t_0)}$ is trivial and we obtain at time t:

$$|\psi(t)\rangle = \sum_n c_n e^{-i\hbar\epsilon_n(t-t_0)} |\phi_n\rangle. \tag{3.34}$$

3.3.2 Direct numerical integration

If the number of basis states is too large to perform a complete diagonalization of the Hamiltonian, or if the Hamiltonian changes over time we need to perform a direct integration of the Schrödinger equation. Like other initial value problems of partial differential equations the Schrödinger equation can be solved by the method of lines. After choosing a set of basis functions or discretizing the spatial derivatives we obtain a set of coupled ordinary differential equations which can be evolved for each point along the time line (hence the name) by standard ODE solvers.

In the remainder of this chapter we use the symbol H to refer the representation of the Hamiltonian in the chosen finite basis set. A forward Euler scheme

$$|\psi(t_{n+1})\rangle = |\psi(t_n)\rangle - i\hbar\Delta_t H|\psi(t_n)\rangle \tag{3.35}$$

is not only numerically unstable. It also violates the conservation of the norm of the wave function $\langle\psi|\psi\rangle = 1$. Since the exact quantum evolution

$$\psi(x, t+\Delta_t) = e^{-i\hbar H\Delta_t}\psi(x,t). \tag{3.36}$$

is unitary and thus conserves the norm, we want to look for a unitary approximant as integrator. Instead of using the forward Euler method (3.35) which is just a first order Taylor expansion of the exact time evolution

$$e^{-i\hbar H\Delta_t} = 1 - i\hbar H\Delta_t + \mathrm{O}(\Delta_t^2), \tag{3.37}$$

we reformulate the time evolution operator as

$$e^{-i\hbar H\Delta_t} = \left(e^{i\hbar H\Delta_t/2}\right)^{-1} e^{-i\hbar H\Delta_t/2} = \left(1 + i\hbar H\frac{\Delta_t}{2}\right)^{-1}\left(1 - i\hbar H\frac{\Delta_t}{2}\right) + \mathrm{O}(\Delta_t^3), \tag{3.38}$$

which is unitary!

This gives the simplest stable and unitary integrator algorithm

$$\psi(x,t+\Delta_t) = \left(1 + i\hbar H\frac{\Delta_t}{2}\right)^{-1}\left(1 - i\hbar H\frac{\Delta_t}{2}\right)\psi(x,t) \tag{3.39}$$

or equivalently

$$\left(1 + i\hbar H\frac{\Delta_t}{2}\right)\psi(x,t+\Delta_t) = \left(1 - i\hbar H\frac{\Delta_t}{2}\right)\psi(x,t). \tag{3.40}$$

Unfortunately this is an implicit integrator. At each time step, after evaluating the right hand side a linear system of equations needs to be solved. For one-dimensional problems the matrix representation of H is often tridiagonal and a tridiagonal solver can be used. In higher dimensions the matrix H will no longer be simply tridiagonal but still very sparse and we can use iterative algorithms, similar to the Lanczos algorithm for the eigenvalue problem. For details about these algorithms we refer to the nice summary at `http://mathworld.wolfram.com/topics/Templates.html` and especially the biconjugate gradient (BiCG) algorithm. Implementations of this algorithm are available, e.g. in the Iterative Template Library (ITL).

3.3.3 The split operator method

A simpler and explicit method is possible for a quantum particle in the real space picture with the "standard" Schrödinger equation (2.51). Writing the Hamilton operator as

$$H = \hat{T} + \hat{V} \tag{3.41}$$

with

$$\hat{T} = \frac{1}{2m}\hat{p}^2 \tag{3.42}$$

$$\hat{V} = V(\vec{x}) \tag{3.43}$$

it is easy to see that $\hat{V}$ is diagonal in position space while $\hat{T}$ is diagonal in momentum space. If we split the time evolution as

$$e^{-i\hbar\Delta_t H} = e^{-i\hbar\Delta_t\hat{V}/2}e^{-i\hbar\Delta_t\hat{T}}e^{-i\hbar\Delta_t\hat{V}/2} + \mathrm{O}(\Delta_t^3) \tag{3.44}$$

we can perform the individual time evolutions $e^{-i\hbar\Delta_t\hat{V}/2}$ and $e^{-i\hbar\Delta_t\hat{T}}$ exactly:

$$\left[e^{-i\hbar\Delta_t\hat{V}/2}|\psi\rangle\right](\vec{x}) = e^{-i\hbar\Delta_t V(\vec{x})/2}\psi(\vec{x}) \tag{3.45}$$

$$\left[e^{-i\hbar\Delta_t\hat{T}/2}|\psi\rangle\right](\vec{k}) = e^{-i\hbar\Delta_t||\vec{k}||^2/2m}\psi(\vec{k}) \tag{3.46}$$

in real space for the first term and momentum space for the second term. This requires a basis change from real to momentum space, which is efficiently performed using a Fast Fourier Transform (FFT) algorithm. Propagating for a time $t = N\Delta_t$, two consecutive

applications of $e^{-i\hbar\Delta_t\hat{V}/2}$ can easily be combined into a propagation by a full time step $e^{-i\hbar\Delta_t\hat{V}}$, resulting in the propagation:

$$\begin{aligned} e^{-i\hbar\Delta_t H} &= \left(e^{-i\hbar\Delta_t\hat{V}/2}e^{-i\hbar\Delta_t\hat{T}}e^{-i\hbar\Delta_t\hat{V}/2}\right)^N + \mathrm{O}(\Delta_t^2) \\ &= e^{-i\hbar\Delta_t\hat{V}/2}\left[e^{-i\hbar\Delta_t\hat{T}}e^{-i\hbar\Delta_t\hat{V}}\right]^{N-1} e^{-i\hbar\Delta_t\hat{T}}e^{-i\hbar\Delta_t\hat{V}/2} \end{aligned} \tag{3.47}$$

and the discretized algorithm starts as

$$\psi_1(\vec{x}) = e^{-i\hbar\Delta_t V(\vec{x})/2}\psi_0(\vec{x}) \tag{3.48}$$

$$\psi_1(\vec{k}) = \mathcal{F}\psi_1(\vec{x}) \tag{3.49}$$

where $\mathcal{F}$ denotes the Fourier transform and $\mathcal{F}^{-1}$ will denote the inverse Fourier transform. Next we propagate in time using full time steps:

$$\psi_{2n}(\vec{k}) = e^{-i\hbar\Delta_t||\vec{k}||^2/2m}\psi_{2n-1}(\vec{k}) \tag{3.50}$$

$$\psi_{2n}(\vec{x}) = \mathcal{F}^{-1}\psi_{2n}(\vec{k}) \tag{3.51}$$

$$\psi_{2n+1}(\vec{x}) = e^{-i\hbar\Delta_t V(\vec{x})}\psi_{2n}(\vec{x}) \tag{3.52}$$

$$\psi_{2n+1}(\vec{k}) = \mathcal{F}\psi_{2n+1}(\vec{x}) \tag{3.53}$$

except that in the last step we finish with another half time step in real space:

$$\psi_{2N+1}(\vec{x}) = e^{-i\hbar\Delta_t V(\vec{x})/2}\psi_{2N}(\vec{x}) \tag{3.54}$$

This is a fast and unitary integrator for the Schrödinger equation in real space. It could be improved by replacing the locally third order splitting (3.44) by a fifth-order version involving five instead of three terms.

Chapter 4

Introduction to many-body quantum mechanics

4.1 The complexity of the quantum many-body problem

After learning how to solve the 1-body Schrödinger equation, let us next generalize to more particles. If a single body quantum problem is described by a Hilbert space $\mathcal{H}$ of dimension $\dim\mathcal{H} = d$ then N *distinguishable* quantum particles are described by the tensor product of N Hilbert spaces

$$\mathcal{H}^{(N)} \equiv \mathcal{H}^{\otimes N} \equiv \bigotimes_{i=1}^{N} \mathcal{H} \tag{4.1}$$

with dimension d^N.

As a first example, a single spin-1/2 has a Hilbert space $\mathcal{H} = \mathbb{C}^2$ of dimension 2, but N spin-1/2 have a Hilbert space $\mathcal{H}^{(N)} = \mathbb{C}^{2^N}$ of dimension 2^N. Similarly, a single particle in three dimensional space is described by a complex-valued wave function $\psi(\vec{x})$ of the position $\vec{x}$ of the particle, while N distinguishable particles are described by a complex-valued wave function $\psi(\vec{x}_1, \ldots, \vec{x}_N)$ of the positions $\vec{x}_1, \ldots, \vec{x}_N$ of the particles. Approximating the Hilbert space $\mathcal{H}$ of the single particle by a finite basis set with d basis functions, the N-particle basis approximated by the same finite basis set for single particles needs d^N basis functions.

This exponential scaling of the Hilbert space dimension with the number of particles is a big challenge. Even in the simplest case – a spin-1/2 with $d = 2$, the basis for $N = 30$ spins is already of of size $2^{30} \approx 10^9$. A single complex vector needs 16 GByte of memory and will not fit into the memory of your PC anymore.

This challenge will be to addressed later in this course by learning about

1. approximative methods, reducing the many-particle problem to a single-particle problem
2. quantum Monte Carlo methods for bosonic and magnetic systems
3. brute-force methods solving the exact problem in a huge Hilbert space for modest numbers of particles

4.2 Indistinguishable particles

4.2.1 Bosons and fermions

In quantum mechanics we assume that elementary particles, such as the electron or photon, are indistinguishable: there is no serial number painted on the electrons that would allow us to distinguish two electrons. Hence, if we exchange two particles the system is still the same as before. For a two-body wave function $\psi(\vec{q}_1, \vec{q}_2)$ this means that

$$\psi(\vec{q}_2, \vec{q}_1) = e^{i\phi}\psi(\vec{q}_1, \vec{q}_2), \tag{4.2}$$

since upon exchanging the two particles the wave function needs to be identical, up to a phase factor $e^{i\phi}$. In three dimensions the first homotopy group is trivial and after doing two exchanges we need to be back at the original wave function[1]

$$\psi(\vec{q}_1, \vec{q}_2) = e^{i\phi}\psi(\vec{q}_2, \vec{q}_1) = e^{2i\phi}\psi(\vec{q}_1, \vec{q}_2), \tag{4.3}$$

and hence $e^{2i\phi} = \pm 1$:

$$\psi(\vec{q}_2, \vec{q}_1) = \pm\psi(\vec{q}_1, \vec{q}_2) \tag{4.4}$$

The many-body Hilbert space can thus be split into orthogonal subspaces, one in which particles pick up a $-$ sign and are called fermions, and the other where particles pick up a $+$ sign and are called bosons.

Bosons

For bosons the general many-body wave function thus needs to be symmetric under permutations. Instead of an arbitrary wave function $\psi(\vec{q}_1, \ldots, \vec{q}_N)$ of N particles we use the symmetrized wave function

$$\Psi^{(S)} = \mathcal{S}_+\psi(\vec{q}_1, \ldots, \vec{q}_N) \equiv \mathcal{N}_S \sum_p \psi(\vec{q}_{p(1)}, \ldots, \vec{q}_{p(N)}), \tag{4.5}$$

where the sum goes over all permutations p of N particles, and $\mathcal{N}_S$ is a normalization factor.

[1] As a side remark we want to mention that in two dimensions the first homotopy group is $\mathbb{Z}$ and not trivial: it matters whether we move the particles clock-wise or anti-clock wise when exchanging them, and two clock-wise exchanges are not the identity anymore. Then more general, anyonic, statistics are possible.

Fermions

For fermions the wave function has to be antisymmetric under exchange of any two fermions, and we use the anti-symmetrized wave function

$$\Psi^{(A)}\mathcal{S}_-\psi(\vec{q}_1,\ldots,\vec{q}_N) \equiv \mathcal{N}_A \sum_p \mathrm{sgn}(p)\psi(\vec{q}_{p(1)},\ldots,\vec{q}_{p(N)}), \tag{4.6}$$

where $\mathrm{sgn}(p) = \pm 1$ is the sign of the permutation and $\mathcal{N}_A$ again a normalization factor.

A consequence of the antisymmetrization is that no two fermions can be in the same state as a wave function

$$\psi(\vec{q}_1,\vec{q}_2) = \phi(\vec{q}_1)\phi(\vec{q}_2) \tag{4.7}$$

since this vanishes under antisymmetrization:

$$\Psi(\vec{q}_1,\vec{q}_2) = \psi(\vec{q}_1,\vec{q}_2) - \psi(\vec{q}_2,\vec{q}_1) = \phi(\vec{q}_1)\phi(\vec{q}_2) - \phi(\vec{q}_2)\phi(\vec{q}_1) = 0 \tag{4.8}$$

Spinful fermions

Fermions, such as electrons, usually have a spin-1/2 degree of freedom in addition to their orbital wave function. The full wave function as a function of a generalized coordinate $\vec{x} = (\vec{q},\sigma)$ including both position $\vec{q}$ and spin σ.

4.2.2 The Fock space

The Hilbert space describing a quantum many-body system with $N = 0, 1, \ldots, \infty$ particles is called the Fock space. It is the direct sum of the appropriately symmetrized single-particle Hilbert spaces $\mathcal{H}$:

$$\bigoplus_{N=0}^{\infty} S_\pm \mathcal{H}^{\otimes n} \tag{4.9}$$

where S_+ is the symmetrization operator used for bosons and S_- is the anti-symmetrization operator used for fermions.

The occupation number basis

Given a basis $\{|\phi_1\rangle,\ldots,|\phi_L\rangle\}$ of the single-particle Hilbert space $\mathcal{H}$, a basis for the Fock space is constructed by specifying the number of particles n_i occupying the single-particle wave function $|f_1\rangle$. The wave function of the state $|n_1,\ldots,n_L\rangle$ is given by the appropriately symmetrized and normalized product of the single particle wave functions. For example, the basis state $|1,1\rangle$ has wave function

$$\frac{1}{\sqrt{2}}\left[\phi_1(\vec{x}_1)\phi_2(\vec{x}_2) \pm \phi_1(\vec{x}_2)\phi_2(\vec{x}_1)\right] \tag{4.10}$$

where the $+$ sign is for bosons and the $-$ sign for fermions.

For bosons the occupation numbers n_i can go from 0 to ∞, but for fermions they are restricted to $n_i = 0$ or 1 since no two fermions can occupy the same state.

The Slater determinant

The antisymmetrized and normalized product of N single-particle wave functions ϕ_i can be written as a determinant, called the Slater determinant

$$\mathcal{S}_- \prod_{i_1}^{N} \phi_i(\vec{x}_i) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\vec{x}_1) & \cdots & \phi_N(\vec{x}_1) \\ \vdots & & \vdots \\ \phi_1(\vec{x}_N) & \cdots & \phi_N(\vec{x}_N) \end{vmatrix}. \tag{4.11}$$

Note that while the set of Slater determinants of single particle basis functions forms a basis of the fermionic Fock space, the general fermionic many body wave function is a linear superposition of many Slater determinants and cannot be written as a single Slater determinant. The Hartee Fock method, discussed below, will simplify the quantum many body problem to a one body problem by making the approximation that the ground state wave function can be described by a single Slater determinant.

4.2.3 Creation and annihilation operators

Since it is very cumbersome to work with appropriately symmetrized many body wave functions, we will mainly use the formalism of second quantization and work with creation and annihilation operators.

The annihilation operator $a_{i,\sigma}$ associated with a basis function $|\phi_i\rangle$ is defined as the result of the inner product of a many body wave function $|\Psi\rangle$ with this basis function $|\phi_i\rangle$. Given an N-particle wave function $|\Psi^{(N)}\rangle$ the result of applying the annihilation operator is an $N-1$-particle wave function $|\tilde{\Psi}^{(N)}\rangle = a_i|\Psi^{(N)}\rangle$. It is given by the appropriately symmetrized inner product

$$\tilde{\Psi}(\vec{x}_1, \ldots, \vec{x}_{N-1}) = \mathcal{S}_\pm \int d\vec{x}_N f_i^\dagger(\vec{x}_N)\Psi(\vec{x}_1, \ldots, \vec{x}_N). \tag{4.12}$$

Applied to a single-particle basis state $|\phi_j\rangle$ the result is

$$a_i|\phi_j\rangle = \delta_{ij}|0\rangle \tag{4.13}$$

where $|0\rangle$ is the "vacuum" state with no particles.

The creation operator $a_i^\dagger$ is defined as the adjoint of the annihilation operator a_i. Applying it to the vacuum "creates" a particle with wave function ϕ_i:

$$|\phi_i\rangle = a_i^\dagger|0\rangle \tag{4.14}$$

For sake of simplicity and concreteness we will now assume that the L basis functions $|\phi_i\rangle$ of the single particle Hilbert space factor into $L/(2S+1)$ orbital wave functions $f_i(\vec{q})$ and $2S+1$ spin wave functions $|\sigma\rangle$, where $\sigma = -S, -S+1, ..., S$. We will write creation and annihilation operators $a_{i,\sigma}^\dagger$ and $a_{i,\sigma}$ where i is the orbital index and σ the spin index. The most common cases will be spinless bosons with $S=0$, where the spin index can be dropped and spin-1/2 fermions, where the spin can be up $(+1/2)$ or down $(-1/2)$.

Commutation relations

The creation and annihilation operators fulfill certain canonical commutation relations, which we will first discuss for an orthogonal set of basis functions. We will later generalize them to non-orthogonal basis sets.

For bosons, the commutation relations are the same as that of the ladder operators discussed for the harmonic oscillator (2.61):

$$[a_i, a_j] = [a_i^\dagger, a_j^\dagger] = 0 \tag{4.15}$$

$$[a_i, a_j^\dagger] = \delta_{ij}. \tag{4.16}$$

For fermions, on the other hand, the operators anticommute

$$\begin{aligned} \{a_{j\sigma'}^\dagger, a_{i\sigma}\} = \{a_{i\sigma}^\dagger, a_{j\sigma'}\} &= \delta_{\sigma\sigma'}\delta_{ij} \\ \{a_{i\sigma}, a_{j\sigma'}\} = \{a_{i\sigma}^\dagger, a_{j\sigma'}^\dagger\} &= 0. \end{aligned} \tag{4.17}$$

The anti-commutation implies that

$$(a_i^\dagger)^2 = a_i^\dagger a_i^\dagger = -a_i^\dagger a_i^\dagger \tag{4.18}$$

and that thus

$$(a_i^\dagger)^2 = 0, \tag{4.19}$$

as expected since no two fermions can exist in the same state.

Fock basis in second quantization and normal ordering

The basis state $|n_1, \ldots, n_L\rangle$ in the occupation number basis can easily be expressed in terms of creation operators:

$$|n_1, \ldots, n_L\rangle = \prod_{i=1}^{L} (a_i^\dagger)^{n_i} |0\rangle = (a_1^\dagger)^{n_1} (a_2^\dagger)^{n_2} \cdots (a_L^\dagger)^{n_L} |0\rangle \tag{4.20}$$

For bosons the ordering of the creation operators does not matter, since the operators commute. For fermions, however, the ordering matters since the fermionic creation operators anticommute: and $a_1^\dagger a_2^\dagger |0\rangle = -a_1^\dagger a_2^\dagger |0\rangle$. We thus need to agree on a specific ordering of the creation operators to define what we mean by the state $|n_1, \ldots, n_L\rangle$. The choice of ordering does not matter but we have to stay consistent and use e.g. the convention in equation (4.20).

Once the normal ordering is defined, we can derive the expressions for the matrix elements of the creation and annihilation operators in that basis. Using above normal ordering the matrix elements are

$$a_i |n_1, \ldots, n_i, \ldots, n_L\rangle = \delta_{n_i,1} (-1)^{\sum_{j=1}^{i-1} n_i} |n_1, \ldots, n_i - 1, \ldots, n_L\rangle \tag{4.21}$$

$$a_i^\dagger |n_1, \ldots, n_i, \ldots, n_L\rangle = \delta_{n_i,0} (-1)^{\sum_{j=1}^{i-1} n_i} |n_1, \ldots, n_i + 1, \ldots, n_L\rangle \tag{4.22}$$

where the minus signs come from commuting the annihilation and creation operator to the correct position in the normal ordered product.

4.2.4 Nonorthogonal basis sets

In simulating the electronic properties of atoms and molecules below we will see that the natural choice of single particle basis functions centered around atoms will necessarily give a non-orthogonal set of basis functions. This is no problem, as long as the definition of the annihilation and creation operators is carefully generalized. For this generalization it will be useful to introduce the fermion field operators $\psi^\dagger_\sigma(\vec{r})$ and $\psi_\sigma(\vec{r})$, creating and annihilating a fermion localized at a single point $\vec{r}$ in space. Their commutation relations are simply

$$\begin{aligned} \{\psi^\dagger_{\sigma'}(\vec{r}), \psi_\sigma(\vec{r'})\} = \{\psi^\dagger_\sigma(\vec{r}), \psi_{\sigma'}(\vec{r'})\} &= \delta_{\sigma\sigma'}\delta(\vec{r}-\vec{r'}) \\ \{\psi_\sigma(\vec{r}), \psi_{\sigma'}(\vec{r'})\} = \{\psi^\dagger_\sigma(\vec{r}), \psi^\dagger_{\sigma'}(\vec{r'})\} &= 0. \end{aligned} \tag{4.23}$$

The scalar products of the basis functions define a matrix

$$S_{ij} = \int d^3\vec{r} f^*_i(\vec{r}) f_j(\vec{r}), \tag{4.24}$$

which is in general *not* the identity matrix. The associated annihilation operators $a_{i\sigma}$ are again defined as scalar products

$$a_{i\sigma} = \sum_j (S^{-1})_{ij} \int d^3\vec{r} f^*_j(\vec{r}) \psi_\sigma(\vec{r}). \tag{4.25}$$

The non-orthogonality causes the commutation relations of these operators to differ from those of normal fermion creation- and annihilation operators:

$$\begin{aligned} \{a^\dagger_{i\sigma}, a_{j\sigma'}\} &= \delta_{\sigma\sigma'}(S^{-1})_{ij} \\ \{a_{i\sigma}, a_{j\sigma'}\} = \{a^\dagger_{i\sigma}, a^\dagger_{j\sigma'}\} &= 0. \end{aligned} \tag{4.26}$$

Due to the non-orthogonality the adjoint $a^\dagger_{i\sigma}$ does *not* create a state with wave function f_i. This is done by the operator $\hat{a}^\dagger_{i\sigma}$, defined through:

$$\hat{a}^\dagger_{i\sigma} = \sum_j S_{ji} a^\dagger_{i\sigma}, \tag{4.27}$$

which has the following simple commutation relation with $a_{j\sigma}$:

$$\{\hat{a}^\dagger_{i\sigma}, a_{j\sigma}\} = \delta_{ij}. \tag{4.28}$$

The commutation relations of the $\hat{a}^\dagger_{i\sigma}$ and the $\hat{a}_{j\sigma'}$ are:

$$\begin{aligned} \{\hat{a}^\dagger_{i\sigma}\hat{a}_{j\sigma'}\} &= \delta_{\sigma\sigma'}S_{ij} \\ \{\hat{a}_{i\sigma}, \hat{a}_{j\sigma'}\} = \{\hat{a}^\dagger_{i\sigma}, \hat{a}^\dagger_{j\sigma'}\} &= 0. \end{aligned} \tag{4.29}$$

We will need to keep the distinction between a and $\hat{a}$ in mind when dealing with non-orthogonal basis sets.

Chapter 5

Density Matrix Renormalization Group

5.1 Introduction

The density matrix renormalization group (DMRG) method is a variational method which allows to obtain accurate ground state wave functions of one-dimensional quantum systems. The adaptation of this technique to higher dimensions is a very active area of research. Following White and Noack[1] we will present the DMRG method as it was originally invented. To understand the insights which led to the development of DMRG, it is useful to briefly discuss the failure of "real-space RG" for the simple problem of a particle in a box.

5.2 Real-space RG for the "particle in a box"

We would like to compute the ground state energy and wave function of a 1-dimensional electron in an infinitely deep, flat potential well (particle in a box). After discretization of space, and in appropriate units the Hamiltonian of this system (Laplacian) becomes

$$H = \begin{pmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & \dots \\ & & \dots & \dots \end{pmatrix}. \tag{5.1}$$

One might be tempted to try the following real-space RG procedure:

1. Isolate a block of length $L \rightarrow$ matrices H_L and T_L such that

$$H = \begin{pmatrix} H_L & T_L & & \\ T_L & H_L & T_L & \\ & T_L & H_L & \dots \\ & & \dots & \dots \end{pmatrix}. \tag{5.2}$$

[1] R. Noack and S. White, Lecture notes in physics, Vol. 538 (Springer, 1999), Chapter 2.

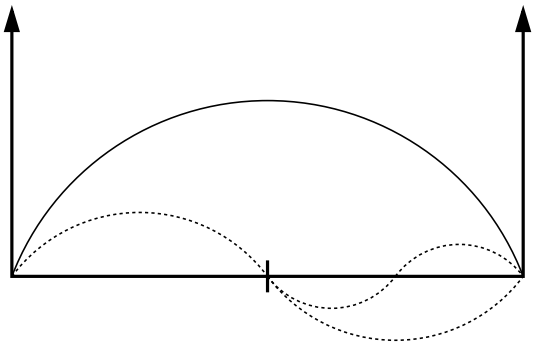

Figure 5.1: Sketch of the ground state wave function (solid line) of the "particle in a box". This ground state wave function cannot be adequately reproduced by joining low-energy wave functions for a box of half the size (dashed lines).

2. Diagonalize $H_L \rightarrow$ basis transformation matrix O which contains as columns the eigenvectors v_1, v_2, ... corresponding to eigenenergies $E_1 < E_2 < \ldots$. Define the $(\dim H_L) \times m$ matrix O_L by keeping only the m eigenvectors with lowest eigenenergy.

3. Compute the $m \times m$ matrices

$$\bar{H}_L = O_L^\dagger H_L O_L, \tag{5.3}$$
$$\bar{T}_L = O_L^\dagger T_L O_L. \tag{5.4}$$

4. Double the size of the system by combining two blocks of size $\bar{L}$ to form a block of size $2L$:

$$H_{2L} = \begin{pmatrix} \bar{H}_L & \bar{T}_L \\ \bar{T}_L & \bar{H}_L \end{pmatrix}, \quad T_{2L} = \begin{pmatrix} 0 & 0 \\ \bar{T}_L & 0 \end{pmatrix}. \tag{5.5}$$

Repeat step 2 with $2L \rightarrow L$.

This procedure gives very poor results for the ground state energy. The reason for this is that the wave function of the block of length $2L$ is constructed from low-energy wave functions for a block of length L. In particular, these wave functions will have a "node" in the middle, unlike the true ground state wave function which is node-less (see Fig. 5.1).

To fix this problem of wrong boundary conditions, one could

- keep low energy wave functions for different combinations of boundary conditions (fixed-fixed, fixed-free, free-fixed, free-free),
- embed the block into a larger "environment" which can provide appropriate boundary conditions.

The latter idea led to the development of DMRG, an RG type algorithm which is based on the truncation of a *reduced density matrix*.

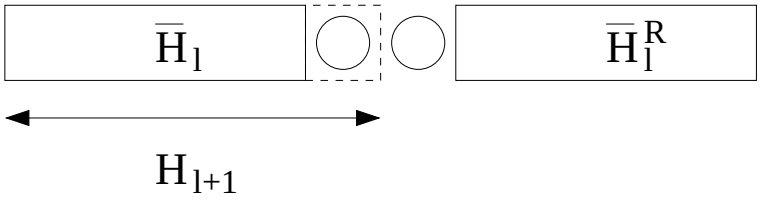

Figure 5.2: Linear growth procedure of the infinite-size DMRG algorithm. $\bar{H}_l$ denotes the system block in the reduced basis and $\bar{H}_l^R$ the environment block. H_{l+1} is the new system block obtained by adding one site.

5.3 Density matrix renormalization group

Suppose we have a "superblock" consisting of a "system" with states $|i\rangle$ and an "environment" with states $|j\rangle$. Let the ground state of the superblock be

$$|\psi\rangle = \sum_{ij} \psi_{ij} |i\rangle |j\rangle. \tag{5.6}$$

The corresponding density matrix is $\rho_{\text{superblock}} = |\psi\rangle\langle\psi|$, and the reduced density matrix of the system is

$$\rho_{\text{system}}^{\text{red}} = \text{Tr}_{\text{environment}}[\rho_{\text{superblock}}] = \sum_{i,i'} \Big(\sum_j \psi_{ij} \psi_{i'j}^* \Big) |i\rangle\langle i'|. \tag{5.7}$$

That is, the elements of the reduced density matrix are given by $(\rho_{\text{system}}^{\text{red}})_{i,i'} = \sum_j \psi_{ij} \psi_{i'j}^*$.

If A is an operator which acts only on the system block, then the expectation value of this operator is given by

$$\langle A \rangle = \text{Tr}_{\text{system}} \Big(\rho_{\text{system}}^{\text{red}} A \Big). \tag{5.8}$$

Let $\{|u^\alpha\rangle\}$ be the eigenbasis of $\rho_{\text{system}}^{\text{red}}$ corresponding to eigenvalues $\omega_\alpha \geq 0$, $\sum_\alpha w_\alpha = 1$. Since

$$\langle A \rangle = \sum_\alpha w_\alpha \langle u^\alpha | A | u^\alpha \rangle \tag{5.9}$$

one can expect that the effect of neglecting a state with weight $w_\alpha \approx 0$ will be small. White therefore suggested that a good description of the system can be obtained by keeping the m eigenstates of $\rho_{\text{system}}^{\text{red}}$ with the largest eigenvalues w_α.

5.3.1 Infinite-size DMRG

The infinite-size algorithm is illustrated in Fig. 5.2. $\bar{H}_l$ denotes the system block in the reduced basis and $\bar{H}_l^R$ the environment block obtained by mirroring the system block. The algorithm is based on a linear growth procedure and proceeds by iterating the following steps:

1. Form a superblock of size $L = 2l + 2$ by combining $\bar{H}_l$, two single sites and $\bar{H}_l^R$.

2. Diagonalize the superblock Hamiltonian $H_L^{\text{superblock}} \rightarrow$ ground state $|\psi\rangle \rightarrow$ density matrix $\rho = |\psi\rangle\langle\psi|$.

3. Compute the reduced density matrix for the new system block of length $l+1$: $\rho^{\text{red}}_{\text{system}(l+1)} = \text{Tr}_{\text{environment}(l+1)}\rho$. Obtain the m eigenvectors corresponding to the largest eigenvalues and from these the transformation matrix O_L to the reduced basis.

4. Transform H_{l+1} and other operators of the new system block to the reduced density matrix eigenbasis: $\bar{H}_{l+1} = O_L^\dagger H_{l+1} O_L$, $\bar{A}_{l+1} = O_L^\dagger A_{l+1} O_L$.

Repeat with $l \rightarrow l+1$.

The main differences to the real-space RG are (i) that the diagonalization of the superblock and the reduction of the basis in the system block take place in systems of different size, and (ii) that the new basis states are determined from a diagonalization of the reduced density matrix, not the Hamiltonian.

5.3.2 Finite-size DMRG

To study finite systems, one runs the infinite-size algorithm until the system size is reached. From that point on the environment block is chosen such that the size of the superblock remains constant. If the system block is grown by one site, the environment block is reduced by one site. Once the minimum size for the environment block is reached, the "sweeping direction" is reversed (system block shrinks, environment block grows). In this algorithm, at each step, we have to store the operators for the block which has grown. An illustration is shown in Fig. 5.3.

5.4 Example: $S = 1/2$ Heisenberg chain

The Hamiltonian of the Heisenberg spin chain is given by

$$H = -J \sum_{\langle i,j \rangle} \vec{S}_i \cdot \vec{S}_j = -J \sum_{\langle i,j \rangle} \left\{ S_i^z S_{i+1}^z + \frac{1}{2}(S_i^+ S_{i+1}^- + S_i^- S_{i+1}^+) \right\}, \tag{5.10}$$

and for spin-1/2 the operators in the "S_z" basis are

$$S^z = \frac{1}{2} \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}, \quad S^+ = \begin{pmatrix} & 1 \\ 0 & \end{pmatrix}, \quad S^- = \begin{pmatrix} & 0 \\ 1 & \end{pmatrix}. \tag{5.11}$$

To treat this Hamiltonian, one has to store $m \times m$ representations of S_l^z and S_l^+ ($S_l^- = (S_l^+)^\dagger$) for l the right or left end site of the block. Two blocks B_1 and B_2 are then joined together as follows:

$$[H_{B_1 B_2}]_{ij,i'j'} = [H_{B_1}]_{ii'}\delta_{jj'} + [H_{B_2}]\delta_{ii'} + [S_l^z]_{ii'}[S_{l+1}^z]_{jj'} + \frac{1}{2}[S_l^+]_{ii'}[S_{l+1}^-]_{jj'} + \frac{1}{2}[S_l^-]_{ii'}[S_{l+1}^+]_{jj'}. \tag{5.12}$$

Here, B_1 can be the system block $\bar{H}_l$ and B_2 the additional site, or B_1 can represent the system block (with additional site) and B_2 the environment block (also with additional site).

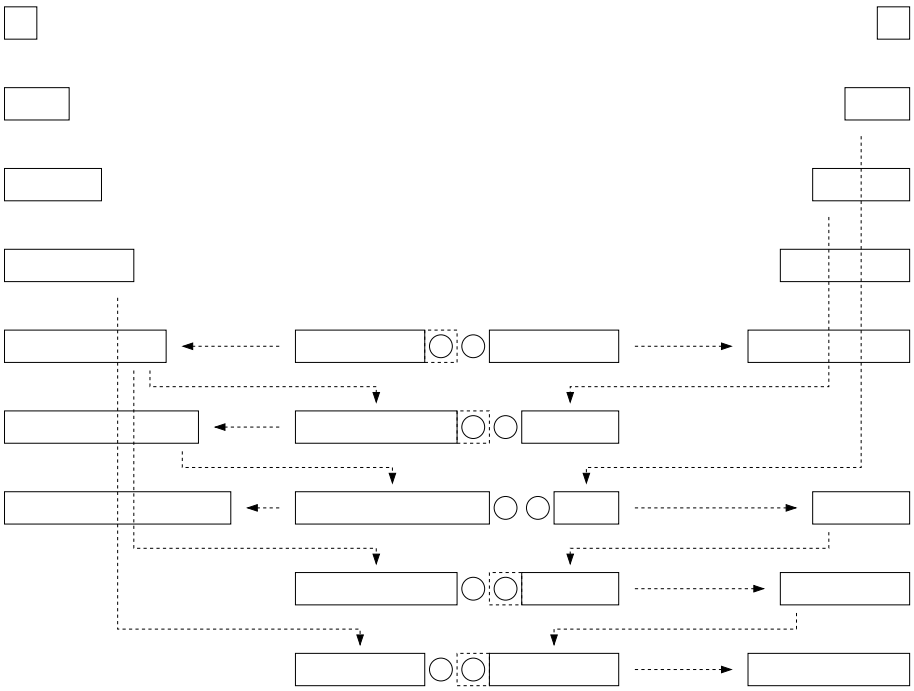

Figure 5.3: Illustration of a finite-system DMRG "sweep". The rectangles on the left and right represent stored configurations for the system and environment, respectively.

5.4.1 Measurements

Measurements are made using the superblock ground state $|\psi\rangle$. If A is defined on the system block and we have updated $[A]_{ii'}$ at every step ($A \leftarrow O_L^\dagger A O_L$), then the ground state expectation value of A is approximated by

$$\langle\psi|A|\psi\rangle = \sum_{ii'j} \psi^*_{ij}\psi_{i'j}[A]_{ii'}. \tag{5.13}$$

5.5 DMRG for the "particle in a box"

We started this chapter by analyzing the failure of real-space RG for the simple problem of a particle in a box. DMRG yields very accurate ground state energies for this problem. However, the application of the DMRG procedure to the particle in a box is not entirely straightforward. In this simple problem, the dimension of the Hilbert space grows linearly, rather than exponentially with the size L of the system (number of discretization steps).

Let us introduce the zero particle state $|0\rangle$ and the 1 particle states $\{|i\rangle\}_{i=1,\ldots,L}$, where $|i\rangle$ represents a particle localized at position i. Dividing the box into a system ($k = 1, \ldots, l$) and an environment ($k = l+1, \ldots, L$) we can write the (ground state)

wave function $|\psi\rangle = \sum_{k=1}^{L} \psi(k)|k\rangle$ in the following product form:

$$|\psi\rangle = \sum_{i,j=0}^{l} \psi_{ij}|i\rangle|j\rangle, \tag{5.14}$$

$$\psi_{ij} = \begin{pmatrix} 0 & \psi(l+1) & \dots & \psi(L) \\ \psi(1) & & & \\ \dots & & 0 & \\ \psi(l) & & & \end{pmatrix}. \tag{5.15}$$

Here, the upper left block corresponds to the zero particle state and the lower right block to two particle states. The reduced density matrix $(\rho^{\rm red}_{\rm system})_{ii'} = \sum_j \psi_{ij}\psi^*_{ji'}$ for the system $(i, i' = 1, \dots, l)$ becomes

$$\rho^{\rm red}_{\rm system} = \begin{pmatrix} \psi_1\psi_1^* & \psi_1\psi_2^* & \dots & \psi_1\psi_l^* \\ \psi_2\psi_1^* & \psi_2\psi_2^* & \dots & \\ \dots & \dots & & \\ \psi_l\psi_1^* & & & \psi_l\psi_l^* \end{pmatrix}. \tag{5.16}$$

The eigenvector of this matrix corresponding to the (only) non-zero eigenvalue is

$$|u\rangle = \frac{1}{\sqrt{w_l}} \begin{pmatrix} \psi_1 \\ \psi_2 \\ \dots \\ \psi_l \end{pmatrix}, \tag{5.17}$$

with $w_l = \sum_{k=1}^{l} |\psi_k|^2$, and thus simply corresponds to the projection of the ground state wave function onto the left block. Since we keep only this one state for the system, we don't need to compute a density matrix at all.

In the "site" basis, we denote the one (normalized) state representing the system by $\{L_j\}_{j=1,\dots,l}$, and the one (normalized) state representing the environment by $\{R_j\}_{j=l+3,\dots,L}$, while the two explicitly treated sites between the left and right blocks have basis states $|l\rangle$ and $|l+1\rangle$. A wave function in the reduced basis can be encoded by four coefficients a_1, a_2, a_3 and a_4:

$$\psi = \{\psi_j\}_{j=1,\dots,L} = \{\{a_1 L_j\}_{j=1,\dots,l}, a_2, a_3, \{a_4 R_j\}_{j=l+3,\dots,L}\}. \tag{5.18}$$

The Hamiltonian matrix element between two states ψ and ψ' becomes

$$\langle\psi|H|\psi'\rangle = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \end{pmatrix} \begin{pmatrix} H_{11} & T_{12} & & \\ T_{12} & 2 & -1 & \\ & -1 & 2 & T_{34} \\ & & T_{34} & H_{44} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}, \tag{5.19}$$

with $H_{11} = \langle L|H|L\rangle$, $H_{44} = \langle R|H|R\rangle$, $T_{12} = \langle L|H|l+1\rangle = -L_l$ and $T_{34} = \langle l+2|H|R\rangle = -R_{l+3}$.

The procedure for growing the left block (computing an improved estimate for $L^{(l+1)}$ given $L^{(l)}$ and $R^{(l+3)}$) requires the following steps:

1. Diagonalize the 4×4 superblock Hamiltonian defined in Eq. (5.19) $\rightarrow$ ground state wave function $\langle\psi| = (a_1, a_2, a_3, a_4)$.

2. The new basis state $L^{(l+1)'}$ is given by the properly normalized projection onto $\{|L^{(l)}\rangle, |l+1\rangle\}$:
$$L^{(l+1)'} = \frac{1}{\sqrt{a_1^2 + a_2^2}} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \equiv \begin{pmatrix} a_1' \\ a_2' \end{pmatrix}. \tag{5.20}$$

3. Construct the new 4×4 Hamiltonian
$$H'_{11} = \langle L^{(l+1)'}|H|L^{(l+1)'}\rangle = {a_1'}^2 H_{11} + 2{a_2'}^2 + 2a_1'a_2'T_{12}, \tag{5.21}$$
$$T'_{12} = -L^{(l+1)'}. \tag{5.22}$$

Then repeat with $l \rightarrow l+1$.

Note that a_2' is the value of the updated wave function at site $l+1$, and thus can be used to plot ψ while sweeping back and forth.

Chapter 6

Path integrals and quantum Monte Carlo

6.1 Introduction

In this chapter, we continue the numerical study of a quantum many-body system. The system is described by its wave-function $\psi(\vec{r}_1, \vec{r}_2, .., \vec{r}_N)$, which is a mapping $\mathbf{R}^N \to \mathbf{C}$. Finding the exact wave-function is not practically feasible unless N is very small. The previous chapter used a *spectral* approach, i.e. decomposed the wave-function onto a carefully chosen set of basis functions, and found the best approximate solution of the Schrödinger equation in the corresponding vector space. In this chapter, we use the technique of Monte Carlo to obtain a *noisy* but *unbiased* estimator of the wave-function. This strategy goes by the general name of Quantum Monte Carlo, which covers many variants. I will describe the two main ones, called Diffusion Monte Carlo (or Green's function Monte Carlo) and Path integral Monte Carlo. Diffusion Monte Carlo has developed into a precise tool to compute the groundstate properties of systems with $\mathcal{O}(10^3)$ particles. Path integral Monte Carlo is technically similar, but gives finite-temperature properties. The path integral formalism carries over to the relativistic case, and forms an essential basis for the study of quantum field theory (to be reviewed in Chapter 8). Although some familiarity with Monte Carlo algorithms is desirable, the necessary basic facts about Monte Carlo are summarized in the first section.

6.2 Recall: Monte Carlo essentials

• Monte Carlo is a powerful stochastic technique to estimate ratios of high-dimensional integrals.
- Simple example: how much is π ? See Fig.6.1.

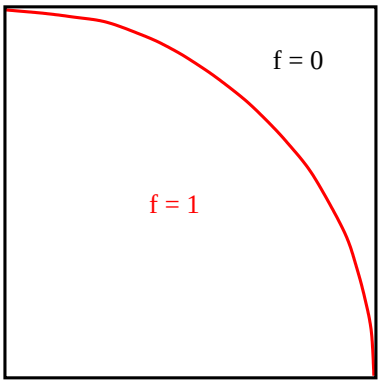

Figure 6.1: Elementary Monte Carlo experiment: draw two random numbers x, y uniformly distributed in $[0, 1]$. If $x^2 + y^2 < 1$, then $f = 1$, otherwise $f = 0$. $\langle f \rangle = \pi/4$.

- Complicated example: $\langle W \rangle = \sum_{\text{states } i} W_i \exp(-\beta E_i) / \sum_{\text{states } i} \exp(-\beta E_i)$

In this latter case, the sum is over all states, but most states have a high energy and contribute negligibly. Therefore, it is efficient to use *importance sampling*, namely to *sample* all states with a probability $\propto \exp(-\beta E(\text{state}))$. The expectation value above becomes $\langle W \rangle = \sum_{\text{sampled states } i} W_i / \sum_{\text{sampled states} i} 1$, ie. $\langle W \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} W_i$. Statistical errors around the exact value shrink as $N^{-1/2}$ thanks to the

• Central limit theorem:

$$\frac{1}{N} \sum_{i=1}^{N} x_i = \bar{x} + \delta x \quad \text{(noisy)} \tag{6.1}$$

$$\langle \delta x \rangle = 0 \quad \text{(unbiased)} \tag{6.2}$$

$$\delta x \sim \mathcal{O}(\frac{1}{\sqrt{N}}) \tag{6.3}$$

More precisely, $\langle (\delta x)^2 \rangle = \frac{\langle (x-\bar{x})^2 \rangle}{N}$. Variance reduction techniques reduce the prefactor in the $\frac{1}{\sqrt{N}}$ error.

• A Monte Carlo simulation is a Markov chain, ie. a Markov process in discrete time.

At each step, the system may change from state i to state j with probability P_{ij} fixed. The matrix elements of the Markov matrix P satisfy:

$$P_{ij} \geq 0 \tag{6.4}$$

$$\sum_j P_{ij} = 1 \tag{6.5}$$

At each step, the probability distribution V of the possible states is multiplied by P^T: $V^{k+1} = P^T V^k = (P^k)^T V^0$. Eq.(6.5) shows that P has a right eigenvector ($V_k = 1 \ \forall k$) with eigenvalue 1. Under assumption of ergodicity ($\forall i, j \ \exists k$ s.t. $(P^k)_{ij} > 0$) (and regularity: $\exists k$ s.t. $\forall i, j \ (P^k)_{ij} > 0$), a Markov matrix has a *unique* left eigenvector with eigenvalue 1, i.e. a *stationary distribution* V^*. All other eigenvalues are < 1, ensuring exponential convergence to V^*.

• How to design a Markov matrix with stationary distribution V^* ?
A sufficient condition is *detailed balance*:

$$\frac{P_{ij}}{P_{ji}} = \frac{V_j^*}{V_i^*} \tag{6.6}$$

• A simple algorithm which satisfies detailed balance is the Metropolis algorithm:
Step 1. Starting from state $S_{\rm old}$, choose a candidate state $S_{\rm cand}$ by taking a random step: $S_{\rm cand} = R \circ S_{\rm old}$, drawn from an even distribution: ${\rm Prob}(R^{-1}) = {\rm Prob}(R)$.
Step 2. Accept the candidate state $S_{\rm cand}$ as the next state $S_{\rm new}$ in the Markov chain with probability

$${\rm Prob}(S_{\rm new} = S_{\rm cand}) = \min(1, V^*(S_{\rm cand})/V^*(S_{\rm old})) \tag{6.7}$$

If $S_{\rm cand}$ is rejected, set $S_{\rm new} = S_{\rm old}$ in the Markov chain.

6.3 Notation and general idea

For simplicity of notation, I consider a single particle in one dimension. The wavefunction $\psi(x,t)$ is a mapping $\mathbf{R} \times \mathbf{R} \to \mathbf{C}$. The ket $|\psi(t)\rangle = \int dx \psi(x,t)|x\rangle$ is a state vector in the Hilbert space. $|x\rangle$ is an eigenstate of the position operator $\mathbf{X}$: $\mathbf{X}|x\rangle = x|x\rangle$, with the completeness relation $\int dx |x\rangle\langle x| = \mathbf{1}$.

The time-dependent Schrödinger equation $i\hbar \frac{d}{dt}|\psi\rangle = H|\psi\rangle$ has for solution

$$|\psi(t)\rangle = \exp(-\frac{i}{\hbar}Ht)|\psi(0)\rangle \tag{6.8}$$

Now change the time to pure imaginary, $\tau = it$ (also called performing a Wick rotation to Euclidean time), so that

$$|\psi(\tau)\rangle = \exp(-\frac{\tau}{\hbar}H)|\psi(0)\rangle \tag{6.9}$$

and expand in eigenstates of H: $H|\psi_k\rangle = E_k|\psi_k\rangle, k = 0, 1, .., E_0 \leq E_1 \leq ...$

$$|\psi(\tau)\rangle = \sum_k \exp(-\frac{\tau}{\hbar}E_k)\langle\psi_k|\psi(0)\rangle|\psi_k\rangle \tag{6.10}$$

It is apparent that $\frac{\tau}{\hbar}$ acts as an *inverse temperature*, and that $|\psi(\tau)\rangle$ will become proportional to the groundstate $|\psi_0\rangle$ as $\tau \to +\infty$, provided $|\psi(0)\rangle$ is not orthogonal to it. Diffusion and path integral Monte Carlo both simulate an imaginary time evolution to obtain information on low-energy states.

Moreover, in both approaches, the time evolution is performed as a sum over histories, each history having a different probability. Note that the same description is used for financial predictions, so you might learn something really useful here...

6.4 Expression for the Green's function

As a starting point, notice that the Schrödinger equation is linear, so that its solution $|\psi(t)\rangle = \exp(-\frac{i}{\hbar}Ht)|\psi(0)\rangle$ is equivalent to

$$\psi(x,t) = \int dx_0 G(x,t;x_0,0)\psi(x_0,0) \tag{6.11}$$

where

$$G(x,t;x_0,0) \equiv \langle x|\exp(-\frac{i}{\hbar}Ht)|x_0\rangle \tag{6.12}$$

is the Green's function, i.e. the solution of the Schrödinger equation for $\psi(x,0) = \delta(x-x_0)$ (also called transition amplitude, or matrix element). Check:

$$\begin{aligned}
|\psi(t)\rangle &= \int dx\psi(x,t)|x\rangle \\
&= \int dx|x\rangle \int dx_0 G(x,t;x_0,0)\psi(x_0,0) \\
&= \int dx \int dx_0 |x\rangle\langle x|\exp(-\frac{i}{\hbar}Ht)|x_0\rangle\psi(x_0,0) \\
&= \exp(-\frac{i}{\hbar}Ht)|\psi(0)\rangle
\end{aligned}$$

- Important property of Green's function:

$$\int dx_1 G(x,t;x_1,t_1)G(x_1,t_1;x_0,0) = G(x,t;x_0,0) \tag{6.13}$$

On its way from x_0 at $t=0$ to x at time t, the particle passes *somewhere* at time t_1. Check:

$$\int dx_1 \langle x|\exp(-\frac{i}{\hbar}H(t-t_1))|x_1\rangle\langle x_1|\exp(-\frac{i}{\hbar}Ht_1)|x_0\rangle = \langle x|\exp(-\frac{i}{\hbar}Ht)|x_0\rangle$$

- Divide t into N intervals $\delta t = t/N$; take $N \to \infty$ at the end.

$$\langle x|\exp(-\frac{i}{\hbar}Ht)|x_0\rangle = \int dx_1 dx_2..dx_{N-1} \prod_{k=1}^{N} \langle x_k|\exp(-\frac{i\delta t}{\hbar}H)|x_{k-1}\rangle \tag{6.14}$$

with $x_N \equiv x$. The task is to evaluate an elementary matrix element $\langle x_k|\exp(-\frac{i\delta t}{\hbar}H)|x_{k-1}\rangle$.

- Evaluation of $\langle x_k|\exp(-\frac{i\delta t}{\hbar}H)|x_{k-1}\rangle$:

Problem: the Hamiltonian $H = \frac{p^2}{2m} + V(x)$ is an operator made of two pieces which do not commute. The potential energy operator $V(x)$ is diagonal in position space $|x\rangle$. The kinetic energy operator $\frac{p^2}{2m}$ is diagonal in momentum space $|p\rangle$. The change of basis position $\leftrightarrow$ momentum is encoded in the matrix elements $\langle x|p\rangle = \exp(\frac{i}{\hbar}px)$. Normally, $e^A e^B = e^{A+B+\frac{1}{2}[A,B]+\cdots}$ (Baker-Campbell-Hausdorff). Here, we neglect commutator terms which are $\mathcal{O}(\delta t^2)$ since we consider $\delta t \to 0$, and write

$$\exp(-\frac{i\delta t}{\hbar}H) \approx \exp(-\frac{i\delta t}{\hbar}\frac{V}{2})\exp(-\frac{i\delta t}{\hbar}\frac{p^2}{2m})\exp(-\frac{i\delta t}{\hbar}\frac{V}{2})$$

then insert a complete set of momentum states $\int \frac{dp}{2\pi}|p\rangle\langle p| = \mathbf{1}$ between successive factors. $\langle x_k|\exp(-\frac{i\delta t}{\hbar}H)|x_{k-1}\rangle$ becomes

$$\exp\left[-\frac{i\delta t}{\hbar}\frac{V(x_k)+V(x_{k-1})}{2}\right]\int\frac{dp_k}{2\pi}\int\frac{dp_{k-1}}{2\pi}\langle x_k|p_k\rangle\langle p_k|\exp(-\frac{i\delta t}{\hbar}\frac{p^2}{2m})|p_{k-1}\rangle \quad \langle p_{k-1}|x_{k-1}\rangle$$
$$= \cdots\exp(\frac{i}{\hbar}p_k x_k)\delta(p_k-p_{k-1})\exp(-\frac{i\delta t}{\hbar}\frac{p_k^2}{2m})\exp(-\frac{i}{\hbar}p_{k-1}x_{k-1})$$
$$= \cdots\int\frac{dp_k}{2\pi}\exp\left(\frac{i}{\hbar}p_k(x_k-x_{k-1})\right)\exp\left(-\frac{i\delta t}{\hbar}\frac{p_k^2}{2m}\right)$$

The last expression is a Gaussian integral. It can be evaluated by completing the square:

$$\int\frac{dp_k}{2\pi}\exp\left[-i\left(p_k\sqrt{\frac{\delta t}{2m\hbar}}-\frac{x_k-x_{k-1}}{2\hbar\sqrt{\frac{\delta t}{2m\hbar}}}\right)^2\right]\exp\left[+i\frac{(x_k-x_{k-1})^2}{4\hbar^2(\frac{\delta t}{2m\hbar})}\right]$$
$$= \text{constant } C\times\exp\left[\frac{i\delta t}{\hbar}\frac{1}{2}m(\frac{x_k-x_{k-1}}{\delta t})^2\right]$$

Putting everything together, one obtains

$$\langle x_k|\exp(-\frac{i\delta t}{\hbar}H)|x_{k-1}\rangle \approx C\exp\left[-\frac{i\delta t}{\hbar}\left(-\frac{1}{2}m(\frac{x_k-x_{k-1}}{\delta t})^2+\frac{V(x_k)+V(x_{k-1})}{2}\right)\right] \tag{6.15}$$

Finally, rotate to imaginary time $\tau = it$:

$$\boxed{\langle x_k|\exp(-\frac{\delta\tau}{\hbar}H)|x_{k-1}\rangle \approx C\exp\left[-\frac{\delta\tau}{\hbar}\left(+\frac{1}{2}m(\frac{x_k-x_{k-1}}{\delta\tau})^2+\frac{V(x_k)+V(x_{k-1})}{2}\right)\right]} \tag{6.16}$$

The exponent is just what one would expect for the Hamiltonian: $(\frac{1}{2}mv^2+V)$. The constant C is independent of x_k, x_{k-1} and drops out of all observables. Eq.(6.16) is at the core of the simulation algorithms. It becomes exact as $\delta\tau \to 0$.

6.5 Diffusion Monte Carlo

The idea of DMC is to evolve in imaginary time a large number m of "walkers" (aka "replicas", "particles"), each described by its position $x^j, j = 1,..,m$ at time $\tau = k\delta\tau$. The groundstate wave-function $\psi_0(x)$ is represented by the average density of walkers at large time: $\psi_0(x) = \lim_{k\to\infty}\langle\delta(x^j - x)\rangle$.

For this to be possible, the groundstate wave-function must be real positive everywhere. One always has the freedom to choose a global (indep. of x) phase. It turns out that, for a bosonic system, the groundstate wave-function can be chosen real positive. Excited states wave-functions (e.g. with angular momentum) have nodes; fermionic groundstate wave-functions also have nodes. These difficulties are discussed in 6.5.2.

The simplest form of DMC assigns to each walker j, at timestep k, a position x_k^j and a weight w_k^j. A convenient starting configuration is $\psi(x,\tau=0) = \delta(x-x_0)$, so

that all the walkers are in the same state: $x_0^j = x_0, w_0^j = 1\ \forall j$. The time evolution of the wave-function $\psi(x,\tau)$, described by Eq.(6.11), simplifies to

$$\psi(x,\tau) = G(x,\tau;x_0,0) \tag{6.17}$$

and $G(x,\tau;x_0,0)$ is a product of elementary factors Eq.(6.16). Each elementary factor is factorized into its kinetic part and its potential part, which are applied in succession to each walker. Namely, given a walker j at position x_{k-1} with weight w_{k-1}, the new position x_k and weight w_k are obtained as follows:
• **Step 1**. The kinetic part gives for x_k a Gaussian distribution centered around x_{k-1}. This distribution can be *sampled* stochastically, by drawing $(x_k - x_{k-1})$ from a Gaussian distribution with variance $\frac{\hbar\delta\tau}{m}$. This step corresponds to *diffusion* around x_{k-1}, and gives its name to the algorithm. The formal reason is that the time-dependent Schrödinger equation for a free particle, in imaginary time, is identical to the heat equation: $\frac{d\psi}{d\tau} = \frac{\hbar}{2m}\nabla^2\psi$.
• **Step 2**. The potential part modifies the weight: $w_k = w_{k-1}\exp(-\frac{\delta\tau}{\hbar}\frac{V(x_k)+V(x_{k-1})}{2})$. Both factors together allow for a stochastic representation of $\psi(x,k\delta\tau)$:

$$\psi(x,k\delta\tau) = \langle w_k \delta(x_k - x)\rangle \tag{6.18}$$

where $\langle..\rangle$ means averaging over the m walkers.

One problem with this algorithm is that the weights w^j will vary considerably from one walker to another, so that the contribution of many walkers to the average Eq.(6.18) will be negligible, and the computer effort to simulate them will be wasted. All walkers should maintain identical weights for best efficiency if possible. This can be achieved by "cloning" the important walkers and "killing" the negligible ones, again stochastically, by Step 3:
• **Step 3**. Compute the nominal weight $w^* \equiv \frac{1}{m}\sum_j w^j$. Replace each walker j by a number of clones (all with weight w^*) equal to $\mathrm{int}(\frac{w^j}{w^*} + r)$, where r is a random number distributed uniformly in $[0,1[$. You can check that the average over r of this expression is $\frac{w^j}{w^*}$, so that each walker is replaced, on average, by its appropriate number of equal-weight clones. Note that the total number m of walkers will fluctuate by $\mathcal{O}(\sqrt{m})$.

With these 3 very simple steps, one can already obtain interesting results. Two technical modifications are customary:
- One limits the maximum number of clones at step 3. As can be seen from Eq.(6.16), the number of clones increases when the walker reaches a region of small potential. Since DMC is often used for Coulombic systems where the potential is unbounded from below, this seems like a wise precaution. In any case, if this maximum number of clones is reached, it indicates that the variation in the weight over a single step is large, and thus that the stepsize $\delta\tau$ is too large.
- As formulated, the nominal weight w^* varies as $\exp(-\frac{E_0}{\hbar}\tau)$ for large τ. To avoid overflow or underflow, one introduces a trial energy E_T, and multiplies all the weights w^j by $\exp(+\frac{E_T}{\hbar}\delta\tau)$ after each step. Stability of the weights is achieved when $E_T = E_0$, which gives a simple means to compute the groundstate energy.

Let me now describe two powerful modifications of this simple algorithm.

6.5.1 Importance sampling

Fluctuations in the estimated wave-function come mostly from Step 2. The variation of the weight of a single walker signals a waste of computer effort: a weight which tends to 0 indicates that the walker approaches a forbidden region (e.g. for a particle in a box, at the edge of the box); a weight which diverges occurs at a potential singularity (e.g. at the location of the charge in a Coulomb potential). In both cases, it would be advantageous to incorporate prior knowledge of the wave-function in the diffusion step, so that walkers would be discouraged or encouraged to enter the respective regions, and the weights would vary less.

This can be accomplished by choosing a "guiding" or "trial" wave-function $\psi_T(x)$, and evolving in imaginary time the product $\Phi(x,\tau) \equiv \psi_T(x)\psi(x,\tau)$. From the Schrödinger equation obeyed by $\psi(x,\tau)$:

$$-\hbar\frac{d}{d\tau}\psi = (-\frac{\hbar^2}{2m}\nabla^2 + V)\psi \tag{6.19}$$

one obtains the equation obeyed by $\Phi(x,\tau)$:

$$\begin{aligned} -\hbar\frac{d}{d\tau}\Phi &= \psi_T(-\frac{\hbar^2}{2m}\nabla^2 + V)\psi \quad (\psi_T \text{ is time} - \text{independent}) && (6.20)\\ &= (-\frac{\hbar^2}{2m}\nabla^2 + V)\Phi + \frac{\hbar^2}{2m}(2\vec{\nabla}\psi_T\cdot\vec{\nabla}\psi + \psi\nabla^2\psi_T) && (6.21)\\ &= \left[-\frac{\hbar^2}{2m}\nabla^2 + \frac{\hbar^2}{m}(\frac{\vec{\nabla}\psi_T}{\psi_T})\cdot\vec{\nabla} + \left(V - \frac{\hbar^2}{2m}\frac{\nabla^2\psi_T}{\psi_T}\right)\right]\ \Phi && (6.22) \end{aligned}$$

The first two terms define a new diffusion, with a drift term which can be simply added to the diffusion in Step 1. The last two terms define a new potential, to be used in the update of the weights in Step 2.

Let us check the effect of a good trial wave-function on the walkers' weight, by choosing for ψ_T the groundstate ψ_0 (assuming ψ_0 is known). In that case, the last term in Eq.(6.22) is equal to $(V-(V-E_0))\Phi$. The new potential is simply E_0, which is a constant independent of x. In Step 2, the weights of all walkers are multiplied by the same factor. As a result, in Step 3, no walker is either cloned or killed.

6.5.2 Fermionic systems

A major application of Diffusion Monte Carlo is to compute the energy of electrons in a molecule or a crystal, given some positions for the atomic nuclei. However, the electrons are indistinguishable fermions, and the wave-function should be anti-symmetric under interchange of any two of them. Thus, if it is positive when electrons 1 and 2 are at positions $(\vec{r}_1,\vec{r}_2)$, it must be negative when they are at $(\vec{r}_2,\vec{r}_1)$. Configuration space (which has dimension $3N$ for N electrons) has nodal surfaces separating positive and negative regions. If the location of these nodal surfaces was known, then one could perform distinct simulations (as many as there are disconnected regions), in each region of definite sign: If ψ_0 is positive, then apply the above algorithm, with a potential barrier preventing the walkers to cross the nodal surface. If ψ_0 is negative, apply the

same algorithm with the substitution $\psi \leftarrow -\psi$. Usually, the location of the nodal surfaces is not known, and some ansatz is made. This strategy is called the *fixed node approximation*.

It is clear that fixing the nodal surface away from its groundstate location can only increase the energy, so that the fixed node approximation gives a variational upper bound to the groundstate energy. One can try to relax the constraint, and move the nodal surface in the direction which most lowers the energy. The gradient of the wavefunction, which must be continuous across the nodal surface, can help in this relaxation strategy. It is unclear to me how well this approach works.

This difficulty is one avatar of the infamous "sign problem" usually associated with fermions.

6.5.3 Useful references

- *Introduction to the Diffusion Monte Carlo Method*, by I. Kosztin, B. Faber and K. Schulten, arXiv:physics/9702023.
- *Monte Carlo methods*, by M. H. Kalos and P. A. Whitlock, Wiley pub., 1986. See Chapter 8.
- A Java demonstration of DMC by I. Terrell can be found at *http://camelot.physics.wm.edu/ ian/dmcview/dmcview.php*

6.6 Path integral Monte Carlo

6.6.1 Main idea

To go from Diffusion Monte Carlo to Path integral Monte Carlo, all that is necessary is a simple change of viewpoint, associated in practice with a different usage of computer memory, as illustrated Fig. 6.2.

• Diffusion Monte Carlo considers *many* walkers $\{x^j\}$ at *one* imaginary time $\tau = k\delta\tau$.

• Path integral Monte Carlo considers *one* walker at *many* times, i.e. one *path* $x(\tau)$.

One practical advantage is that the weights of the walkers, with their undesirable fluctuations, can now be eliminated. Importance sampling can be implemented exactly, using e.g. the Metropolis algorithm. To see this, rewrite the transition amplitude Eq.(6.14) in imaginary time $\tau = it$:

$$\begin{aligned}
\langle x|\exp(-\frac{\tau}{\hbar}H)|x_0\rangle &= \int dx_1 dx_2..dx_{N-1}\prod_{k=1}^{N}\langle x_k|\exp(-\frac{\delta\tau}{\hbar}H)|x_{k-1}\rangle \\
&\approx C^N\int dx_1 dx_2..dx_{N-1}\prod_{k=1}^{N}\exp\left[-\frac{\delta\tau}{\hbar}\left(+\frac{1}{2}m(\frac{x_k-x_{k-1}}{\delta\tau})^2+\frac{V(x_k)+V(x_{k-1})}{2}\right)\right] \\
&= C^N\int dx_1 dx_2..dx_{N-1}\exp\left[-\frac{\delta\tau}{\hbar}\left(+\frac{1}{2}m\sum_k(\frac{x_k-x_{k-1}}{\delta\tau})^2+\sum_k\frac{V(x_k)+V(x_{k-1})}{2}\right)\right]
\end{aligned} \tag{6.23}$$

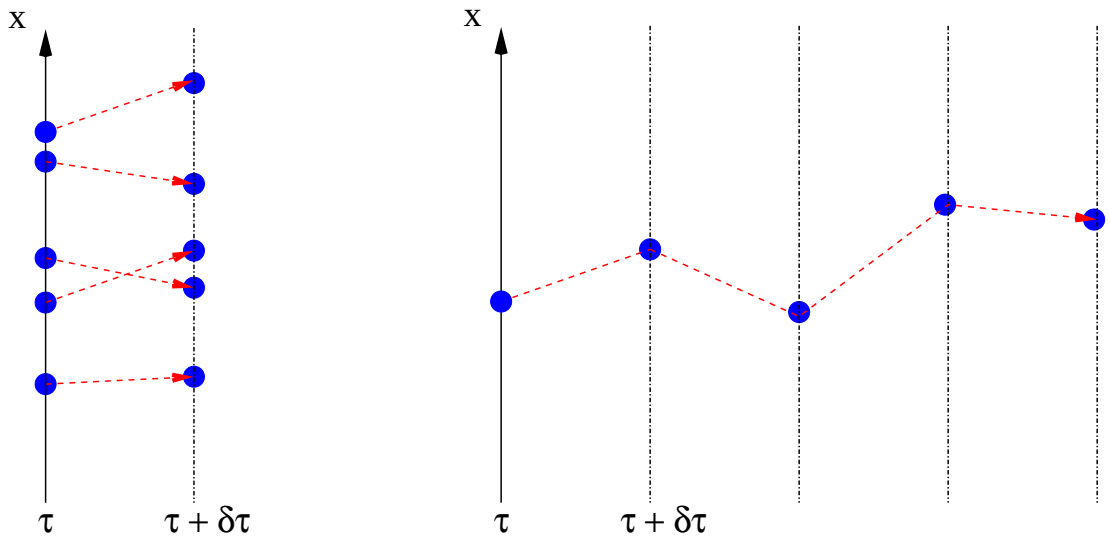

Figure 6.2: Comparison between Diffusion Monte Carlo (*left*) and Path integral Monte Carlo (*right*): many paths or many time-steps.

This represents the desired *stationary distribution* of the Markov chain corresponding to our Monte Carlo process. To converge to this distribution, starting from an arbitrary configuration (e.g. $x_k = 0\ \forall k$), perform many *sweeps*, where one sweep consists of updating once each x_{k_0}, keeping all $x_k, k \neq k_0$ fixed. The induced probability distribution for x_{k_0} is, up to an irrelevant proportionality constant:

$$\text{Prob}(x_{k_0}) \propto \exp\left[-\frac{\delta\tau}{\hbar}\left(+\frac{1}{2}m\left(\left(\frac{x_{k_0+1}-x_{k_0}}{\delta\tau}\right)^2+\left(\frac{x_{k_0}-x_{k_0-1}}{\delta\tau}\right)^2\right)+V(x_{k_0})\right)\right] \tag{6.24}$$

Therefore, a Metropolis update of x_{k_0} is implemented as:
Step 1. Propose a candidate $x^{\text{new}} = x_{k_0} + \delta x$, where δx is drawn from an even distribution (e.g. uniform in $[-\Delta, +\Delta]$).
Step 2. Accept x^{new} as the new value of x_{k_0} with the Metropolis probability

$$\min\left(1, \text{Prob}(x^{\text{new}})/\text{Prob}(x_{k_0})\right)$$

and $\text{Prob}(x)$ given by Eq.(6.24). Note that the irrelevant proportionality constant in Eq.(6.24) cancels in the ratio. As usual, the optimal step-size, governed by Δ, is the result of a trade-off: a large step causes large variations in Eq.(6.24), and thereby a small average acceptance probability; a small step causes a slow evolution in Monte Carlo time, resulting in a large *auto-correlation* between successive path configurations $\{x_k\}$. A rough rule of thumb is to adjust the step size so that the average acceptance is around 1/2.

6.6.2 Finite temperature

There is no need to keep the initial and final states $|x_0\rangle$ and $|x\rangle$ fixed. They prevent the system from approaching the groundstate except in the middle $\sim \tau/2$. A simple

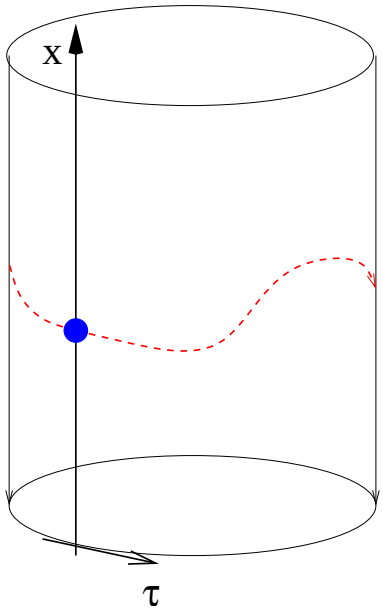

Figure 6.3: After compactification of the Euclidean time direction, the paths become closed loops, and the path integral becomes identical to the partition function of the quantum mechanical system, where the inverse temperature is (proportional to) the Euclidean time-extent.

and far-reaching modification improves the situation.

Impose that the initial and final states be identical, and let them be any state $|x\rangle$. Then the amplitude $\langle x| \exp(-\frac{\tau}{\hbar}H)|x_0\rangle$ becomes

$$\sum_x \langle x| \exp(-\frac{\tau}{\hbar}H)|x\rangle = \text{Tr } \exp(-\frac{\tau}{\hbar}H) \tag{6.25}$$

This expression looks like a statistical mechanics partition function, at temperature $k_B T = \hbar/\tau$. Indeed, it can be expanded in an eigenbasis of H:

$$Z \equiv \text{Tr } \exp(-\frac{\tau}{\hbar}H) = \sum_k \langle \psi_k| \exp(-\frac{\tau}{\hbar}H)|\psi_k\rangle = \sum_k \exp(\frac{\tau}{\hbar}E_k) \tag{6.26}$$

which is identical to the partition function for our 1-d quantum-mechanical system at temperature $k_B T = \hbar/\tau$.

Geometrically, the identity of $|x\rangle$ and $|x_0\rangle$ can be ensured by making the imaginary time direction *compact*, as in Fig. 6.3. The paths are loops drawn on a cylinder. The perimeter of the cylinder is (up to a factor $\hbar/k_B$) the inverse temperature. The $T \to 0$ limit is obtained when the imaginary time extent is made infinite.

Again, the partition function of our 1-d quantum mechanical system at temperature T is *identical* to the partition function of a classical gas of loops $x(\tau')$, $\tau' \in [0, \tau = \frac{\hbar}{k_B T}]$, distributed according to the probability

$$\frac{1}{Z} \exp\left[-\frac{1}{\hbar}\int_0^\tau d\tau' \left(\frac{1}{2}m(\frac{dx}{d\tau'})^2 + V(x)\right)\right] \tag{6.27}$$

which is the $\delta\tau \to 0$ limit of Eq.(6.23) for periodic paths, and

$$Z = \int_{x(\tau)=x(0)} \mathcal{D}x(\tau') \ \exp\left[-\frac{1}{\hbar}\int_0^\tau d\tau' \left(\frac{1}{2}m(\frac{dx}{d\tau'})^2 + V(x)\right)\right] \tag{6.28}$$

Now, the expression "path integral" is clear: Z is an integral over paths $x(\tau')$. Since the integral is over functions of τ', Z is called a functional integral. The exponent $\int d\tau' \left(\frac{1}{2}m(\frac{dx}{d\tau'})^2 + V(x)\right)$ is called the *Euclidean action*.

Furthermore, notice the role of $\hbar$ in the exponent of Eq.(6.27): it governs the magnitude of the fluctuations, and therefore plays a role similar to a *temperature*. In the limit $\hbar \to 0$, all fluctuations are suppressed and the only path surviving in the path integral is that which minimizes the exponent: this is the classical limit, where the path from $(x_0, 0)$ to (x, τ) is deterministic, and we recover the *minimal action principle*. When $\hbar \neq 0$, fluctuations are allowed around this special path: these fluctuations are *quantum fluctuations*.

We will consider the path integral more formally again when discussing field theories in Chapter 7.

6.6.3 Observables

Since one can identify the Euclidean time-extent τ of the system and its inverse temperature, let us use the standard notation $\beta = 1/(k_B T) = \tau/\hbar$. The expectation value of an observable W is given by

$$\langle W \rangle = \frac{1}{Z}\mathrm{Tr}\left(W \exp(-\beta H)\right) \tag{6.29}$$

which gives, when expanded in an eigenbasis of H

$$\langle W \rangle = \frac{\sum_k \langle\psi_k|W|\psi_k\rangle \exp(-\beta E_k)}{\sum_k \langle\psi_k|\psi_k\rangle \exp(-\beta E_k)} \tag{6.30}$$

$$\underset{\beta\to\infty}{=} \frac{\langle\psi_0|W|\psi_0\rangle}{\langle\psi_0|\psi_0\rangle} \tag{6.31}$$

so that groundstate properties can be obtained for large Euclidean time extent.

A simple example is the position projector $W = |x_0\rangle\langle x_0| = \delta(x_0)$. Substituting above gives

$$\langle \delta(x_0) \rangle \underset{\beta\to\infty}{=} |\langle\psi_0|x_0\rangle|^2 = |\psi_0(x_0)|^2 \tag{6.32}$$

so that the number of paths going through x_0 (at any Euclidean time) is proportional to the square of the wave-function.

Let us compare with Diffusion Monte Carlo: there, one had $\langle\delta(x_0)\rangle \propto \psi_0(x_0)$, or $\propto \psi_T(x_0)\psi_0(x_0)$ when using a guiding wave-function ψ_T. For the optimal choice $\psi_T = \psi_0$ where all walkers have weight 1, one recovers the Path integral result.

One can also measure unequal-time correlations, like the two-point correlator $\delta(x,\Delta\beta)\delta(x,0)$:

$$\begin{aligned}\langle\delta(x,\Delta\beta)\delta(x,0)\rangle &= \frac{\sum_k\langle\psi_k|\exp(-(\beta-\Delta\beta)H)\ \delta(x)\ \exp(-\Delta\beta H)\ \delta(x)|\psi_k\rangle}{\sum_k\langle\psi_k|\exp(-\beta H)|\psi_k\rangle} \\ &= \frac{\sum_{k,l}|\langle\psi_k|\delta(x)|\psi_l\rangle|^2\exp(-(\beta-\Delta\beta)E_k)\exp(-\Delta\beta E_l)}{\sum_k\exp(-\beta E_k)} \\ &\underset{\Delta\beta\ll\beta\to\infty}{\approx} |\langle\psi_0|\delta(x)|\psi_0\rangle|^2+|\langle\psi_1|\delta(x)|\psi_0\rangle|^2\exp(-\Delta\beta(E_1-E_0))+\cdots\end{aligned} \tag{6.33}$$

where a complete set of states $\sum_l|\psi_l\rangle\langle\psi_l|$ was inserted to go from the first to the second line. The final expression shows that the *connected* correlator $\langle\delta(x,\Delta\beta)\delta(x,0)\rangle-\langle\delta(x)\rangle^2$ approaches 0 exponentially fast, and that the rate of decay (in Euclidean time $\Delta\beta$) gives the *energy gap* (E_1-E_0) between the groundstate and the first excited state.

The groundstate energy itself is more delicate to obtain, for the following reason. Consider the kinetic energy part of Eq.(6.23). It defines a Gaussian distribution for (x_k-x_{k-1}), with variance $\langle(x_k-x_{k-1})^2\rangle\sim\mathcal{O}(\delta\tau)$, as appropriate for Brownian motion: the mean-square distance grows linearly with time. However, this implies that the mean-sqare velocity $\langle(\frac{x_k-x_{k-1}}{\delta\tau})^2\rangle$, i.e. the kinetic energy, grows as $1/\delta\tau$. This is, again, expected from Brownian motion: a Brownian path is continuous, but not differentiable (it is full of kinks). As a result, the reference groundstate energy is infinite when $\delta\tau\to 0$.

A way around this problem is to adopt a different definition of the kinetic energy, replacing $(x_k-x_{k-1})^2$ by a "split-point" alternative $(x_{k+1}-x_k)(x_k-x_{k-1})$. Alternatively, one can make use of the virial theorem which gives for the kinetic energy $\langle\frac{p^2}{2m}\rangle=\frac{1}{2}\langle x\frac{dV}{dx}\rangle$. See exercises.

6.6.4 Monte Carlo simulation

The Monte Carlo simulation samples paths $x(\tau)$, discretized at N points. Thus, the configuration of the system at any step of the Markov chain is defined by the positions $\{x_k,k=1,..,N\}$. To manipulate dimensionless numbers, set $\hbar=k_B=1$ and use the conventional notation $a\equiv\delta\tau$. a, which is called the *lattice spacing* has the dimension of time (equivalently, of length). Dimensionless variables are

$$\begin{aligned}\hat{m} &= am \\ \hat{x} &= x/a \\ \hat{V}(\hat{x}) &= aV(x) \\ \hat{T} &= aT = 1/N\end{aligned}$$

The probability of a path $\{\hat{x}_k,k=1,..,N\}$ is given by Eq.(6.23):

$$\frac{1}{Z}\exp\left[-\left(\frac{1}{2}\hat{m}\sum_k(\hat{x}_k-\hat{x}_{k-1})^2+\sum_k\hat{V}(\hat{x}_k)\right)\right] \tag{6.34}$$

and can be sampled with the Metropolis algorithm, as already outlined. Two requirements must be kept in mind:

• For the groundstate to dominate the partition function, the temperature must be small enough that the excited states have decayed away, namely $\beta(E_1 - E_0) \gg 1$. In dimensionless units, this translates into

$$N \gg (\hat{E}_1 - \hat{E}_0)^{-1} \tag{6.35}$$

In other words, a clear exponential decay of the connected correlation Eq.(6.33) must be observed. Violation of this requirement can be interpreted as a finite-size effect.
• For discretization errors $\mathcal{O}(a^2)$ to be small, all correlation lengths must be large compared to a. This translates into

$$(\hat{E}_1 - \hat{E}_0)^{-1} \gg 1 \tag{6.36}$$

For the harmonic oscillator $\hat{V}(\hat{x}) = \frac{1}{2}\omega^2\hat{x}^2$, these two requirements become $1 \ll \omega^{-1} \ll N$. The idea is to choose ω small enough that a continuum extrapolation $a \to 0$ can be performed, while keeping a constant, low enough temperature. While decreasing a, the computer requirements increase for two reasons: N increases, and the number of Monte Carlo sweeps required to obtain a statistically uncorrelated path increases too. The latter phenomenon is called *critical slowing down*. It can be largely alleviated, in the quantum-mechanical case, by cluster or loop algorithms.

6.6.5 Useful references

- *A statistical approach to Quantum Mechanics*, by M. Creutz and B. Freedman, Annals of Physics 132 (1981) 427.
- A Java demonstration of Path integral Monte Carlo by A. Santamaria can be found at *http://fisteo12.ific.uv.es/∼santamar/qapplet/metro.html*. Note that the parameters of the quartic potential can be adjusted interactively.

Chapter 7

An Introduction to Quantum Field Theory

7.1 Introduction

This chapter is a generalization of Sec. 6.6 on Path Integral Monte Carlo. Instead of considering one (or N) quantum-mechanical particles as was done there, the idea is now to consider a quantum *field*, which contains infinitely many degrees of freedom. However, in practice, we are going to simulate a large but finite number of degrees of freedom, and extrapolate at the end. So there really is not much difference with Sec. 6.6.

The formal basis for Quantum Field Theory is an important subject, which will not be covered here. The goal of this chapter is to convey some understanding of *simulations* of quantum field theories, by appealing to intuition rather than rigor.

7.2 Path integrals: from classical mechanics to field theory

Consider first the case of a single, classical particle with Hamiltonian $H = \frac{p^2}{2m} + V$. Hamilton's equations describe the time-evolution of this particle:

$$\frac{dq}{dt} = +\frac{\partial H}{\partial p} \quad \longrightarrow \quad \dot{q} = \frac{p}{m} \tag{7.1}$$

$$\frac{dp}{dt} = -\frac{\partial H}{\partial q} \quad \longrightarrow \quad \dot{p} = -\nabla V \tag{7.2}$$

The usual point of view is to start from initial conditions $(q, \dot{q})$ at time $t = 0$, and evolve q and $\dot{q}$ according to the coupled ODEs above. Note, however, that the boundary conditions can instead be split between the beginning and the end of the evolution. In particular, one can specify the beginning and ending coordinates $(q(0), q(t))$. There is a unique path $q(t'), t' \in [0, t]$, which satisfies the above equations, and specifies the initial and final velocities. To find this path, it is convenient to change viewpoint and consider the action $S = \int_0^t dt' \mathcal{L}(q, \dot{q})$, where $\mathcal{L}$ is the Lagrangian $\frac{1}{2}m\dot{q}^2 - V(q)$. One

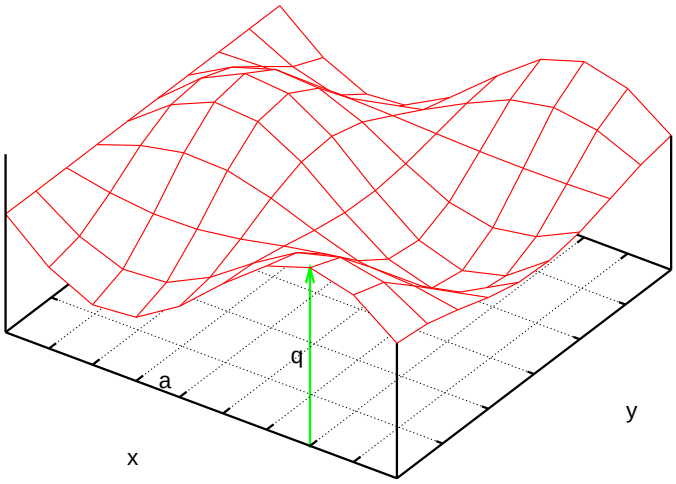

Figure 7.1: A field configuration $q(x, y)$, discretized on a square grid of spacing a.

then invokes the principle of stationary (or least) action, from which can be derived the Euler-Lagrange equations

$$\frac{\partial \mathcal{L}}{\partial q_i} = \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu q_i)} \quad . \tag{7.3}$$

Note that the notion of action is more general than the notion of Hamiltonian: some systems have an action, but no Hamiltonian. This was in fact the original motivation for Feynman to develop the path integral formalism in his Ph.D. thesis: he was interested in systems having non-local interactions in time (with an interaction term $q(t)q(t')$).

Consider now many particles interacting with each other, with Lagrangian $\mathcal{L} = \sum_i \frac{1}{2} m \dot{q}_i^2 - V(\{q_i\})$ and take q_i to represent the z-coordinate of particle i, whose x and y coordinates are fixed on a square grid of spacing a. Furthermore, take the interaction between particles to be of the form $\sum_{\langle ij \rangle}(q_i - q_j)^2$, where $\langle ij \rangle$ stands for nearest-neighbours on the grid, as if springs were connecting i and j. Low-energy configurations will then have almost the same q-value at neighbouring grid points, so that the configuration $\{q_i\}$ will be smooth and look like a "mattress" as in Fig. 7.1.

When the grid spacing a is reduced to zero, the configuration $\{q_i\}$ becomes a *classical field* $q(\vec{x})$ ($\vec{x} \in \mathcal{R}^2$ in this example), with infinitely many degrees of freedom. The action of this field is specified by its Lagrangian density $\mathcal{L} = \frac{1}{2}\partial_\mu q \partial^\mu q - \frac{1}{2} m_0^2 q^2 - V(q)$ where the first term is the continuum version of $(q_i - q_j)^2$ (with $\partial_\mu q \partial^\mu q = \dot{q}^2 - |\vec{\nabla} q|^2$), the second one is a harmonic term corresponding to a mass, and the last term describes the local (anharmonic) interaction, e.g. $\propto q^4$ [1]. The action is then $S = \int_0^t dt' dx dy \mathcal{L}(q(x, y, t'))$. Note that the Lagrangian density $\mathcal{L}$ satisfies several symmetries: it is invariant under translations and rotations in the (x, y) plane, and under the sign flip $q(\vec{x}) \to -q(\vec{x})\ \forall \vec{x}$, at least for an interaction $\propto q^4$. Each continuous symmetry leads to a conserved quantity: energy-momentum for translations, angular momentum for rotations. We will see the importance of the discrete symmetry $q \leftrightarrow -q$ later.

[1] One could think of additional interaction terms, constructed from *higher derivatives* of the field. They are not considered here because they lead to non-renormalizable theories.

Now we can consider quantum effects on the above system. As a result of quantum fluctuations, the path from $q(t=0)$ to $q(t)$ is no longer unique. As in quantum mechanics, all paths contribute with an amplitude $\propto \exp(+\frac{i}{\hbar}\int_0^t dt'\mathcal{L})$, from which it becomes clear that the magnitude of relevant fluctuations in the action is $\hbar$. One can then follow the strategy of Sec. 6.6 and make time purely imaginary, by introducing $\tau = it \in \mathcal{R}$. The immediate result is that idt' above becomes $d\tau'$, so that the amplitude becomes real. The other change is $\dot{q}^2 \to -(\partial_\tau q)^2$, so that an overall minus sign can be taken out, leading to the amplitude

$$\exp(-\frac{1}{\hbar}S_E) \tag{7.4}$$

where $S_E = \int d\tau' d\vec{x}\mathcal{L}_E$ is the Euclidean action, and

$$\mathcal{L}_E = \frac{1}{2}(\partial_\mu\phi)^2 + \frac{1}{2}m_0^2\phi^2 + V(\phi) \tag{7.5}$$

is the Euclidean Lagrangian density, and the field q is now denoted by ϕ as is customary. The first term $(\partial_\mu\phi)^2 = (\partial_\tau\phi)^2 + |\vec{\nabla}\phi|^2$ is now symmetric between space and time, so that the metric is *Euclidean* in $(d+1)$ dimensions (d spatial dimensions, plus Euclidean time).

It is worth summarizing the sign flips which occurred in the kinetic energy T and the potential energy U during the successive steps we have just taken. We started with the Hamiltonian $H = T + U$, then considered the Lagrangian $L = T - U$. Going to imaginary time changes the sign of T. Finally, we take out an overall minus sign in the definition of L_E, so that paths with the smallest action are the most likely. This leads to the Euclidean Lagrangian density $L_E = T + U$, which is identical to the Hamiltonian we started from, except that the momentum p is replaced by the derivative $\partial_0\phi$.

It is also useful to perform some elementary dimensional analysis. Since it appears in the exponent of the amplitude Eq.(7.4), the Euclidean action S_E is *dimensionless* (we set $\hbar = 1$). Hence the Euclidean Lagrangian density has mass dimension $(d+1)$, and therefore the field ϕ has mass dimension $\frac{d-1}{2}$. This is interesting, because if we take the "normal" number of spatial dimensions $d = 3$ and the interaction term $V(\phi) = \frac{g_0}{4!}\phi^4$, then g_0 is a dimensionless number. It makes sense then to perform a Taylor expansion of this theory in powers of g_0 about the free case $g_0 = 0$: this is the scope of perturbation theory. Here, we will try to obtain non-perturbative results, by directly simulating the theory at some large value of g_0.

We have so far considered a field $\phi(\vec{x},\tau)$ which takes values in $\mathcal{R}$. It is easy to generalize the Lagrangian density Eq.(7.5) to cases when ϕ takes values in $\mathcal{C}$, or has several components forming a vector $\vec{\phi} \equiv \begin{pmatrix} \phi_1 \\ .. \\ \phi_N \end{pmatrix}$, perhaps with a constraint $\sum_N \phi_k^2 = 1$, depending on the desired symmetries. Typically, the Euclidean Lagrangian density is the starting, defining point of a quantum field theory.

Finally, we can introduce a finite temperature T, exactly as we did in the quantum-mechanical case: we make the Euclidean time direction *compact*: $\tau \in [0, \beta = \frac{1}{T}]$, and impose periodic boundary conditions on the field ϕ: $\phi(\vec{x},\beta) = \phi(\vec{x},0)\ \forall\vec{x}$. This works

for the same reason as in quantum mechanics: the partition function

$$Z = \int_{periodic} \mathcal{D}\phi \exp(-\int_0^\beta d\tau' d\vec{x} \mathcal{L}_E(\phi)) \tag{7.6}$$

is a weighted sum of eigenstates of the Hamiltonian: $Z = \sum_i \exp(-\beta E_i)$. We will be concerned here with the $T = 0$ situation. In that case, the two-point correlator provides a means to measure the mass gap $(E_1 - E_0)$:

$$\langle \phi(\vec{x}, \tau)\phi(\vec{x}, 0)\rangle - \langle\phi\rangle^2 \underset{\tau\to\infty}{=} c^2 \exp(-(E_1 - E_0)\tau) \tag{7.7}$$

or equivalently the correlation length $\xi = (E_1 - E_0)^{-1}$. The lowest energy state, with energy E_0, is the vacuum, which contains particle-antiparticle pairs because of quantum fluctuations, but whose net particle number is zero. The first excited state, with energy E_1, contains one particle at rest. Call its mass $m_R = E_1 - E_0$. Then this mass can be obtained from the decay of the two-point correlator, as $m_R = 1/\xi$. This is the "true", measurable mass of the theory, and it is not equal to the mass m_0 used in the Lagrangian density. m_R is called the *renormalized* mass, while m_0 is the *bare* mass. Similarly, the "true" strength g_R of the interaction can be measured from 4-correlators of ϕ, and it is not equal to the coupling g_0 used in the Lagrangian density: g_0 is the bare coupling, g_R the renormalized coupling.

7.3 Numerical study of ϕ^4 theory

Here, we show that very important results in Quantum Field Theory can be extracted from simulations of the $4d$ Ising model. Our starting point is the continuum Euclidean action:

$$S_E = \int d\tau d^3x \left[\frac{1}{2}(\partial_\mu\phi_0)^2 + \frac{1}{2}m_0^2\phi_0^2 + \frac{g_0}{4!}\phi_0^4\right] \tag{7.8}$$

where the subscript 0 is to emphasize that we are dealing with bare quantities (field, mass and coupling), and the coupling normalization $1/4!$ is conventional. We discretize the theory on a hypercubic $(4d)$ lattice with spacing a [2]. After the usual replacements $\int d\tau d^3x \;\to\; a^4 \sum_{\text{sites } x}$ and $\partial_\mu\phi_0 \;\to\; \frac{\phi_0(x+\hat{\mu})-\phi_0(x)}{a}$, we end up with the lattice action

$$S_L = \sum_x \left[-2\kappa \sum_\mu \phi(x)\phi(x+\hat{\mu}) + \phi(x)^2 + \lambda(\phi(x)^2 - 1)^2 - \lambda\right] \tag{7.9}$$

where we use the new variables ϕ, κ and λ defined by

$$a\phi_0 = \sqrt{2\kappa}\;\phi \tag{7.10}$$

$$a^2 m_0^2 = \frac{1-2\lambda}{\kappa} - 8 \tag{7.11}$$

$$g_0 = \frac{6\lambda}{\kappa^2} \tag{7.12}$$

[2] The lattice spacing is taken to be the same in space and in time for simplicity; one could consider different values a_s and a_τ.

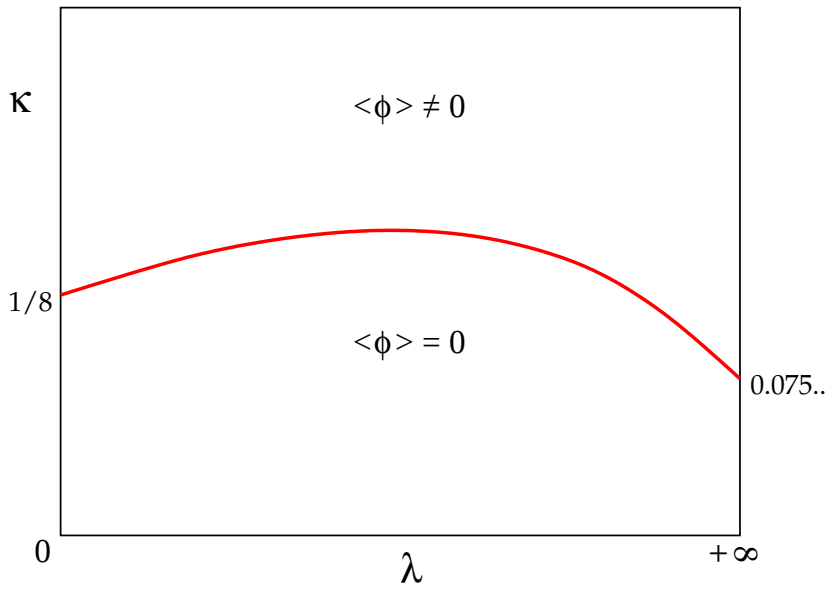

Figure 7.2: Phase diagram of the lattice theory defined by Eq.(7.9). The two phases are separated by a line of second-order phase transitions.

Note in particular the multiplication of ϕ_0 by a to form a dimensionless variable, since ϕ_0 has mass dimension 1. The original formulation had two bare parameters, m_0 and g_0. They have been mapped into two bare parameters, κ and λ. This discretized theory can be simulated by standard Monte Carlo algorithms like Metropolis, on a hypercubic lattice of L sites in each direction. To minimize finite-size effects, periodic boundary conditions are usually imposed in each direction.

The behaviour of our system is easy to understand qualitatively in the two limits $\lambda = 0$ and $\lambda = +\infty$.

• When $\lambda = 0$, the interaction is turned off. This is the free theory, which has two phases depending on the value of κ: a disordered or symmetric phase $\langle\phi\rangle = 0$ when κ is small, and an ordered phase $\langle\phi\rangle \neq 0$ when κ is large. Thus, the symmetry $\phi \leftrightarrow -\phi$ is spontaneously broken when $\kappa > \kappa_c = \frac{1}{8}$, which corresponds to the vanishing of the mass m_0.

• When $\lambda = +\infty$, fluctuations of ϕ away from the values ± 1 cost an infinite amount of energy. Thus, ϕ is restricted to ± 1, and our theory reduces to the $4d$ Ising model with coupling 2κ. As in lower dimensions, the Ising model undergoes a *second-order phase transition* corresponding to the spontaneous breaking of the symmetry $\phi \leftrightarrow -\phi$, for a critical value $\kappa_c \approx 0.075$.

For intermediate values of λ, again a second-order transition takes place, leading to the phase diagram depicted Fig. 7.2.

The existence of a second-order phase transition is crucial: it allows us to define a *continuum limit* of our lattice theory. Remember that the "true", renormalized mass m_R can be extracted from the exponential decay of the 2-point correlator

$$\langle\phi(x)\phi(y)\rangle - \langle\phi\rangle^2 \underset{|x-y|\to\infty}{\propto} \exp(-m_R|x-y|) = \exp(-\frac{|x-y|}{\xi}) \qquad (7.13)$$

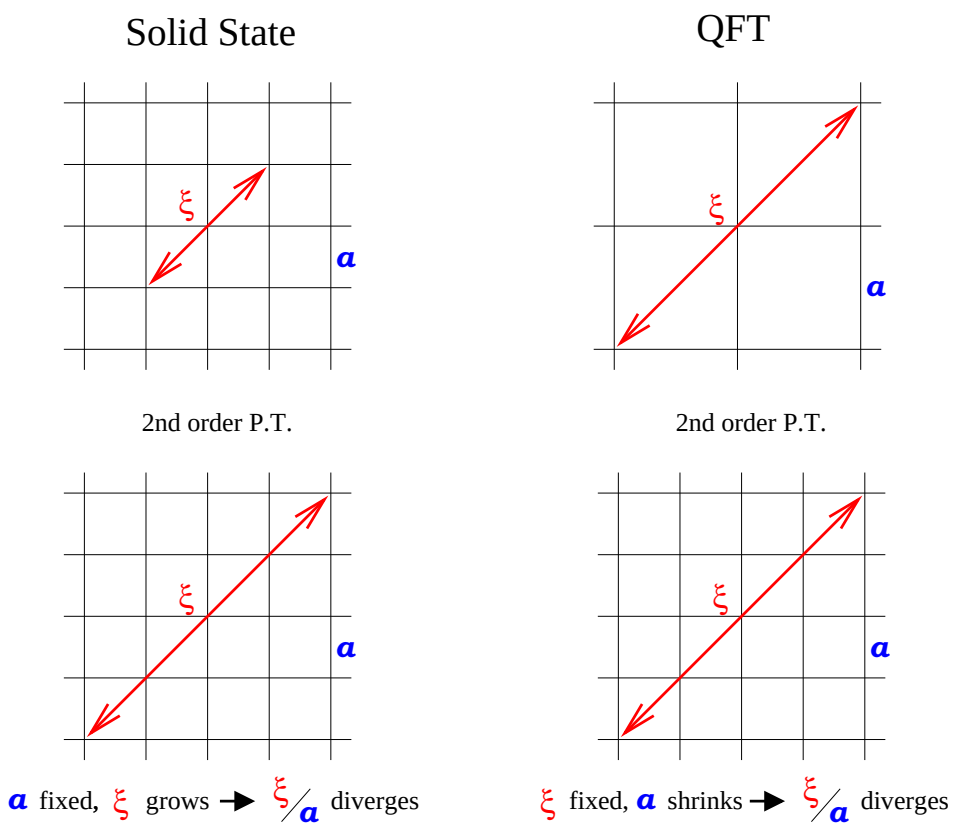

Figure 7.3: Two different viewpoints on a second-order phase transition: in solid-state physics (*left*), the crystal is "real" and the physical correlation length diverges; in quantum field theory (*right*), the correlation length is "real", and the lattice spacing shrinks to zero.

(see Eq.(7.7)). On the lattice, we can only measure the dimensionless combination $am_R = \frac{1}{\xi/a}$, and the separation $|x-y|$ can only be measured in lattice units, i.e. $\frac{|x-y|}{a}$. Taking the continuum limit $a \to 0$ (while keeping m_R fixed) forces the correlation length *measured in lattice units*, i.e. ξ/a, to diverge. This only occurs when the lattice theory has a second-order phase transition (or higher order).

Therefore, the interpretation of a second-order phase transition is different between solid-state physics and lattice field theory. In the first case, the lattice spacing has a physical meaning, like the distance between two ions in a crystal: the lattice is "for real", and the correlation length really diverges at a second-order critical point. In the lattice field theory, the correlation length is "for real", and the lattice spacing a shrinks to zero at the critical point. This is illustrated in Fig. 7.3.

In this latter case, one must be careful that the physical box size (La) also shrinks as $a \to 0$. In order to obtain a controlled continuum limit at constant physical volume, one must increase the number of lattice points L in each direction keeping (La) fixed.

Going back to our ϕ^4 theory, one sees that a continuum limit can be defined for any value of the bare coupling $\lambda \in [0, +\infty]$, by tuning κ to its critical value $\kappa_c(\lambda)$. An interesting question is: what is the value of the "true", renormalized coupling as a

function of λ ? The answer is clear when $\lambda = 0$: the theory is free, and the coupling is zero, whether bare or renormalized. To obtain a non-zero answer, a reasonable strategy is to maximize the bare coupling, and thus to consider the Ising limit $\lambda = +\infty$. The renormalized coupling is extracted from the strength of the 4-spin correlation, normalized as explained in the Exercises. The rather surprising answer is that the renormalized coupling is zero, just like for $\lambda = 0$. In fact, the renormalized coupling is always zero for any choice of λ. In other words, the renormalized ϕ^4 theory is *free*, no matter the value of the bare coupling! The formal statement is that the ϕ^4 theory is "trivial". Note that this is only true in $(3 + 1)$ dimensions. In lower dimensions, the renormalized coupling is non-zero.

Now, why is this finding important? The Standard Model of particle interactions contains a Higgs particle, which gives a mass to all other particles by coupling to them. The field theory describing the Higgs particle is very much like the ϕ^4 theory we have just studied, except that the field ϕ is now a complex doublet $\begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}$. The bare parameters are chosen so that the system is in the broken-symmetry phase, where ϕ has a non-zero expectation value. The masses of all particles are proportional to $\langle\phi\rangle$, therefore it is crucial that $\langle\phi\rangle \neq 0$. In turn, this symmetry breaking is only possible if the coupling λ is non-zero. But, as we have seen, the "true", renormalized value of λ is zero. Therefore, we have a logical inconsistency. The consequence is that the Standard Model cannot be the final, exact description of particle interactions. Some new physics beyond the Standard Model must exist, and from the numerical study of the lattice theory near $\kappa_c(\lambda)$, one can set the energy scale for this new physics to become visible at around 600-800 GeV. This argument is so powerful that it has been used in the design of the Large Hadron Collider (LHC), to be turned on imminently at CERN, and which will study collisions up to about 1000 GeV only.

7.4 Monte Carlo algorithms for the Ising model

It is most efficient to perform numerical simulations of the ϕ^4 theory in its Ising limit. The partition function to sample is then simply

$$Z = \sum_{\sigma_x = \pm 1} \exp(+J \sum_{\langle xy \rangle} \sigma_x \sigma_y) \tag{7.14}$$

with a homogeneous ferromagnetic interaction $J = 2\kappa > 0$ between pairs of spins σ_x, σ_y at nearest-neighbour sites $\langle xy \rangle$. This partition function can be sampled by the general Metropolis algorithm reviewed in Chapter 6. However, the autocorrelation (Monte Carlo) time necessary to produce statistically independent configurations grows with the correlation length ξ as ξ^z, where the dynamic critical exponent z is about 2. Simulations near the second-order phase transition, where ξ diverges, become very demanding in computer resources. This phenomenon is called critical slowing down.

Two alternatives to the Metropolis algorithm have been discovered, the cluster algorithm (Swendsen & Wang, 1987) and the worm algorithm (Prokof'ev & Svistunov, 2001), which give $z \sim 0$ to 0.5 (z varies with the dimensionality of the system) for the Ising model. They have revolutionized the numerical study of the systems where they

work efficiently. We now describe them briefly. In both cases, new degrees of freedom b_{xy} are introduced, associated with *bonds*, i.e. nearest-neighbour pairs of lattice sites xy.

7.4.1 Cluster algorithm

The idea is to rewrite the Boltzmann weight for each single spin pair as

$$\begin{aligned} \exp(+J\sigma_x\sigma_y) &= e^{-J} + (e^{+J} - e^{-J})\delta_{\sigma_x,\sigma_y} && (7.15) \\ &= \sum_{b_{xy}=0,1} \left[e^{-J}\delta_{b_{xy},0} + (e^{+J} - e^{-J})\delta_{\sigma_x,\sigma_y}\delta_{b_{xy},1}\right] && (7.16) \end{aligned}$$

The combined system of spins σ_x and bond variables b_{xy} can be updated by two alternating kinds of updates:
1. Heatbath on the bonds, keeping the spins frozen:
The probabilities of assigning 0 or 1 to b_{xy} can be read off eq.(7.16). They depend whether the two spins are aligned or not.
- If $\sigma_x \neq \sigma_y$, only the first term in eq.(7.16) survives, and the only option is to set b_{xy} to zero.
- If $\sigma_x = \sigma_y$, then the normalized probabilities of assigning values 0 and 1 are $\frac{e^{-J}}{e^{-J}+(e^{+J}-e^{-J})} = e^{-2J}$ and $\frac{e^{+J}-e^{-J}}{e^{-J}+(e^{+J}-e^{-J})} = 1 - e^{-2J}$ respectively.
2. Heatbath on the spins, keeping the bonds frozen:
Again from eq.(7.16), the value $b_{xy} = 0$ does not constrain the spins σ_x and σ_y. On the contrary, the value $b_{xy} = 1$ comes with $\delta_{\sigma_x,\sigma_y}$, so that the two spins must have the same orientation. This applies to all spins connected by $b = 1$ bonds, which form *clusters*. Thus, all spins in a cluster must flip together. And spins in different clusters do not interact. Therefore, the new orientation $\pm$ of each cluster can be chosen independently of the others with probability 1/2.[3]

As a bonus, the correlation of two arbitrarily distant spins $\sigma_{x_1}, \sigma_{x_2}$ can easily be estimated: it is zero if they belong to different clusters ($C(x_1) \neq C(x_2)$), one if they belong to the same cluster ($C(x_1) = C(x_2)$). Hence, an improved estimator can be formed for the spin-spin correlation and for the magnetic susceptibility:

$$\begin{aligned} \langle\sigma_{x_1}\sigma_{x_2}\rangle &= \langle\delta_{C(x_1),C(x_2)}\rangle && (7.17) \\ \chi \equiv \langle\sum_x \sigma_0\sigma_x\rangle &= \frac{1}{V}\langle\sum_{x_1}\sigma_{x_1}\sum_{x_2}\sigma_{x_2}\rangle && (7.18) \\ &= \frac{1}{V}\langle\sum_C |C|\sigma_{x_C}\sum_{C'}|C'|\sigma_{x_{C'}}\rangle && (7.19) \\ &= \frac{1}{V}\langle\sum_C |C|^2\rangle && (7.20) \end{aligned}$$

where $|C|$ is the size of cluster C. The tremendous variance reduction offered by these improved estimators is clear: a small expectation value for the spin-spin correlation

[3] A single-cluster variant (Wolff, 1989) is even more efficient.

results from measurements giving mostly 0 or rarely 1, instead of resulting from the cancellation of $+1$ and -1 measurements. Also, the connection between the susceptibility and the cluster size guarantees that the cluster size will diverge at the phase transition, very much like real magnetization domains.

It is this latter connection which is delicate to enforce when formulating a cluster algorithm for a more complicated statistical model. Ising-type variables can often be *embedded* as partial reflections in more complicated spin variables. A valid cluster algorithm can be implemented, which samples the correct partition function. But its efficiency is not necessarily as hoped for. Such limitations appear to apply, in particular, to lattice *gauge* theories.

7.4.2 Worm algorithm

The idea is to rewrite the Boltzmann weight for each single spin pair as

$$\begin{aligned} \exp(+J\sigma_x\sigma_y) &= \cosh J + \sinh J\ (\sigma_x\sigma_y) && (7.21)\\ &= \sum_{b_{xy}=0,1} \left[\cosh J\ \delta_{b_{xy},0} + \sinh J\ (\sigma_x\sigma_y)\delta_{b_{xy},1}\right] && (7.22) \end{aligned}$$

Note that a more general strategy consists of Taylor-expanding the exponential into $\sum_{l=0}^{+\infty} \frac{(J\sigma_x\sigma_y)^l}{l!}$. The simpler expansion eq.(7.22) is appropriate for the Ising model. This expansion can be substituted for every term in the partition function, after which one carries the summation over spins σ_x:

$$\begin{aligned} Z = \sum_{\{\sigma_x\}} \prod_{\langle xy\rangle} e^{+J\sigma_x\sigma_y} &= \sum_{\{b_{xy}\}} \sum_{\{\sigma_x\}} \prod_{\langle xy\rangle} \left[\cosh J\ \delta_{b_{xy},0} + \sinh J\ (\sigma_x\sigma_y)\delta_{b_{xy},1}\right] && (7.23)\\ &= (\cosh J)^{Vd} \sum_{\{b_{xy}=0,1\}}{}' (\tanh J)^{\sum b_{xy}} && (7.24) \end{aligned}$$

where the constant factor $(\cosh J)^{Vd}$, with Vd the total number of bonds, is irrelevant, and where $\sum'$ includes only the bond configurations which satisfy the *constraint* $\sum_y b_{xy} = 0 \mod 2\ \forall x$. This constraint results from the $\pm$ summation over each spin σ_x: only even powers of σ_x survive, all contributing the same factor[4]. Thus, in the final expression, one has traded spin for bond variables which can take value 0 or 1 with probability $\frac{1}{1+\tanh J}$ and $\frac{\tanh J}{1+\tanh J}$ respectively. The difficulty is to update them while respecting the "tight-packing" constraint, which forces the non-zero bonds to form *closed loops*. The beautiful idea of the worm algorithm consists of enlarging the space of configurations sampled by Monte Carlo, by cleverly relaxing the constraint (i.e. "cutting the loops").

First, notice that a similar expression to eq.(7.24) can be obtained for the modified

[4]It is most instructive to derive the equivalent algorithm when a magnetic field h is turned on, contributing $h\sum_x \sigma_x$ to the energy.

partition function with two sources σ_{x_1} and σ_{x_2}:

$$Z_{x_1,x_2} \equiv \sum_{\{\sigma_x\}} \sigma_{x_1}\sigma_{x_2} \prod_{\langle xy\rangle} e^{+J\sigma_x\sigma_y} \tag{7.25}$$

$$= (\cosh J)^{Vd} \sum_{\{b_{xy}=0,1\}}{}'' (\tanh J)^{\sum b_{xy}} \tag{7.26}$$

where now the constraint $\sum''$ means that $\sum_y b_{xy} = 1 \mod 2$ for $x = x_1$ or x_2, 0 otherwise. Instead of a gas of closed loops, we now have 2 "dangling ends". Furthermore, when x_1 and x_2 coincide, $Z_{x_1,x_2} = Z$. Thus, it makes sense to sample Z_{x_1,x_2}: a bond configuration contributing to Z will be produced every time $x_1 = x_2$. In-between configurations are not wasted, since they sample Z_{x_1,x_2}, which is (up to a normalization factor $1/Z$) the spin-spin correlation $\langle\sigma_{x_1}\sigma_{x_2}\rangle$.

A simple algorithm sampling $\sum_{x_1,x_2} Z_{x_1,x_2}$ works as follows, starting from a closed-loop configuration contributing to Z (e.g. all bond variables set to zero):
1. Select at random a site x_1, and set $x_2 = x_1$.
2. Select at random one of the $2d$ links touching x_2, and call its other end x_c.
3. Propose a change of b_{x_2,x_c}, from 0 to 1 or from 1 to 0, and accept it with the Metropolis probability $\min(1, \tanh J) = \tanh J$ or $\min(1, 1/\tanh J) = 1$ respectively. If the change is accepted, set $x_2 = x_c$.
4. Increment Z_{x_1,x_2} by 1.
5. If $x_2 = x_1$ increment Z by 1 and go back to (1); else go back to (2).
This algorithm moves around the "head" x_2 and the "tail" x_1 of a "worm", hence the name. Two successive contributions to Z differ by a global change in the loop configuration, obtained by a sequence of local steps. The spin-spin correlation and the susceptibility can be extracted as

$$\langle\sigma_{x_1}\sigma_{x_2}\rangle = \frac{1}{Z} Z_{x_1,x_2} \tag{7.27}$$

$$\chi = \frac{1}{V}\frac{1}{Z}\sum_{x_1,x_2} Z_{x_1,x_2} \tag{7.28}$$

The efficiency of the worm algorithm is comparable to that of the best, single-cluster variant of the cluster algorithm. Moreover, the same algorithm applies straightforwardly to the ϕ^4 theory away from its Ising limit, or to a complex $|\phi|^4$ theory. Its limitations are not completely known yet, and its further application, e.g. to fermionic systems, is the subject of current research.

7.5 Gauge theories

Of the four forces known in Nature, at least three (the strong, the weak and the electromagnetic forces) are described by gauge theories. In addition to the usual "matter" fields (electrons, quarks), these theories contain "gauge" fields (photons, W and Z bosons, gluons) which "mediate" the interaction: the interaction between, say, two electrons is caused by the exchange of photons between them. This is analogous to the exchange

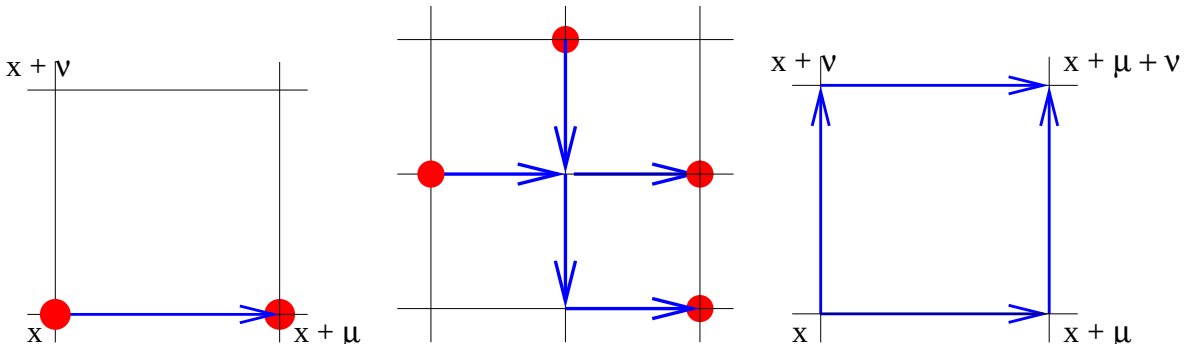

Figure 7.4: Graphical representation (*left*) of the gauge-invariant nearest-neighbour interaction: $\phi^*(x)\phi(x+\hat{\mu})$ becomes $\phi^*(x)U_\mu(x)\phi(x+\hat{\mu})$; (*middle*) an example of a gauge-invariant 4-point correlation; (*right*) the smallest closed loop is the plaquette, with associated matrix $U_\mu(x)U_\nu(x+\hat{\mu})U_\mu^\dagger(x+\hat{\nu})U_\nu^\dagger(x)$.

of momentum which occurs when one person throws a ball at another, and the other catches it. In this way, two particles interact when they are far apart, even though the Lagrangian contains only local interactions. Moreover, gauge theories are invariant under a larger class of symmetries, namely *local* (x-dependent) symmetries.

7.5.1 QED

As an example, let us consider here a variant of Quantum ElectroDynamics (QED), called scalar QED, where electrons are turned into bosons. A simple modification of the previous ϕ^4 theory is required: in order to represent charged bosons, the field ϕ, instead of being real, is made complex, $\phi(x) \in \mathcal{C}$. The continuum Euclidean action becomes

$$S_E = \int d\tau d^3x \left[|\partial_\mu \phi_0|^2 + m_0^2 |\phi_0|^2 + \frac{g_0}{4!} |\phi_0|^4 \right] \tag{7.29}$$

and, after discretization on the lattice:

$$S_L = \sum_x \left[-\kappa \sum_\mu (\phi^*(x)\phi(x+\hat{\mu}) + \text{h.c.}) + |\phi(x)|^2 + \lambda(|\phi(x)|^2 - 1)^2 - \lambda \right] \tag{7.30}$$

S_L is invariant under the *global* (x-independent) rotation $\phi(x) \to \exp(i\alpha)\phi(x)\ \forall x$. The idea is now to promote this symmetry to a *local* one, where α may depend on x. It is clear that the derivative term $\phi^*(x)\phi(x+\hat{\mu})$ is not invariant under this transformation. Invariance is achieved by introducing new degrees of freedom, namely complex phases (elements of $U(1)$) which live on the links between nearest-neighbours. Call $U_\mu(x) = \exp(i\theta_\mu(x))$ the link variable starting at site x in direction μ, and give it the initial value 1, i.e. $\theta_\mu(x) = 0\ \forall x, \mu$. Modify the derivative term as follows:

$$\phi^*(x)\phi(x+\hat{\mu}) \ \to \ \phi^*(x)U_\mu(x)\phi(x+\hat{\mu}) \tag{7.31}$$

This term is now invariant under a *local* transformation $\phi(x) \to \exp(i\alpha(x))\phi(x)$, with $\alpha(x) \neq \alpha(x+\hat{\mu})$, provided that $U_\mu(x)$ also transforms:

$$\begin{aligned} \phi(x) &\to \exp(i\alpha x)\phi(x) && (7.32) \\ U_\mu(x) &\to \exp(i\alpha(x))U_\mu(x)\exp(-i\alpha(x+\hat{\mu})) && (7.33) \end{aligned}$$

The significance of the new variables $U_\mu(x)$ and of the new expression for the discretized derivative can be elucidated by expressing $\theta_\mu(x) = eaA_\mu(x)$, and considering the continuum limit $a \to 0$. To lowest order in a, the derivative $\partial\mu$ becomes the *covariant derivative* $D_\mu \equiv \partial_\mu + ieA_\mu$, and the transformation eq.(7.33) is a *gauge transformation* for A_μ: $A_\mu(x) \to A_\mu(x) - e\partial_\mu\alpha(x)$. Thus, our link variables $U_\mu(x)$ represent the gauge potential $A_\mu(x)$ associated with the electromagnetic field, and eq.(7.31) describes the interaction of our bosonic electrons with the photon. To complete the lattice discretization of QED, what is missing is the energy of the electromagnetic field, namely $\frac{1}{2}\int d\vec{x}d\tau(|\vec{E}|^2(x) + |\vec{B}|^2(x))$. We identify its lattice version below.

It becomes simple to construct n-point correlators of ϕ which are invariant under the local transformation eq.(7.33): all the fields ϕ need to be connected by "strings" of gauge fields, made of products of gauge links U_μ as in Fig. 7.4. Under a local gauge transformation, the phase changes $\alpha(x)$ will always cancel out between ϕ and the attached U, or between two successive U's.

There also exists another type of gauge-invariant object. Consider the product of links $\prod_{x\to x} U$ around a closed loop, starting and ending at x. It transforms as

$$\prod_{x\to x} U \;\to\; \exp(i\alpha(x)) \prod_{x\to x} U \;\exp(-i\alpha(x)) \qquad (7.34)$$

which is invariant since all the U's are complex phase factors which commute with each other. Thus, another valid term to add to the [real] lattice action is the real part of any closed loop, summed over translations and rotations to preserve the other desired symmetries. The simplest version of such a term is to take elementary square loops of size a, made of 4 links going around a *plaquette*: $P_{\mu\nu}(x) \equiv U_\mu(x)U_\nu(x+\hat{\mu})U_\mu^\dagger(x+\hat{\nu})U_\nu^\dagger(x)$. Thus, the complete action of our scalar QED theory is

$$\sum_x |\phi(x)|^2 - \kappa \sum_x \sum_\mu (\phi^*(x)U_\mu(x)\phi(x+\hat{\mu}) + \text{h.c.}) - \beta \sum_x \sum_{\mu\neq\nu} \text{Re}(P_{\mu\nu}(x)) \qquad (7.35)$$

The plaquette term looks geometrically like a curl. Indeed, substituting $U_\mu(x) = \exp(ieaA_\mu(x))$ and expanding to leading-order in a yields $\text{Re}(P_{\mu\nu}(x)) \approx \exp(iea(A_\mu(x) + A_\nu(x+\hat{\mu}) - A_\mu(x+\hat{\nu}) - A_\nu(x)))$, so that

$$\text{Re}(P_{\mu\nu}(x)) \approx 1 - \frac{1}{2}e^2a^4(\partial_\mu A_\nu - \partial_\nu A_\mu)^2 \qquad (7.36)$$

so that the last term in eq.(7.35) becomes, up to an irrelevant constant, $-\beta e^2 \frac{1}{2}\int d\vec{x}d\tau(|\vec{E}|^2 + |\vec{B}|^2)$, where one has expressed the electric and magnetic fields $\vec{E}$ and $\vec{B}$ in terms of the gauge potential A_μ. It suffices then to set $\beta = 1/e^2$ to recover the energy of an electro-magnetic field.

Note that it is our demand to preserve invariance under the local transformation eq.(7.33) which has led us to the general form of the action eq.(7.35). We could have considered larger loops instead of plaquettes. But in the continuum limit $a \to 0$, these loops would yield the same continuum action. So the form of the QED action is essentially dictated by the local gauge symmetry.

One can now study the scalar QED theory defined by eq.(7.35) using Monte Carlo simulations, for any value of the bare couplings (κ, β). Contrary to continuum perturbation theory, one is not limited to small values of e (i.e. large β).

7.5.2 QCD

Other gauge theories have basically the *same* discretized action eq.(7.35). What changes is the group to which the link variables $U_\mu(x)$ belong. For QCD, these variables represent the gluon field, which mediates the interaction between quarks. Quarks come in 3 different colors and are thus represented by a 3-component vector at each site[5]. Hence, the link variables are 3×3 matrices. The gauge-invariant piece associated with a closed loop is the *trace* of the corresponding matrix, thanks to cyclic invariance of the trace in eq.(7.34). No other changes are required to turn lattice QED into lattice QCD!

As emphasized in Sec. 7.3, the Euclidean Lagrangian density defines the lattice theory. The continuum limit can be obtained by approaching a critical point. For QCD, the critical point is $\beta \to +\infty$, i.e. $g_0 = 0$ since $\beta \propto 1/g_0^2$ as in QED. As we have seen, the vanishing of the bare coupling does not imply much about the true, renormalized coupling.

7.6 Overview

The formulation of lattice QCD is due to K. Wilson (1974). First Monte Carlo simulations were performed by M. Creutz in 1980, on a 4^4 lattice. A goal of early simulations was to check whether quarks were confined. This can be demonstrated by considering the potential $V(r)$ between a quark and an anti-quark separated by distance r. Contrary to the case of QED where the potential $\propto 1/r$ saturates as $r \to \infty$, in QCD the potential keeps rising linearly, $V(r) \sim \sigma r$, so that it takes an infinite amount of energy to completely separate the quark and anti-quark. Equivalently, the force between the two objects goes to a *constant* σ. The energy of the quark-antiquark pair grows as if it was all concentrated in a tube of finite diameter. In fact, describing the quark-antiquark pair as an infinitely thin vibrating string is a very good approximation, as shown in the state-of-the-art Monte Carlo data Fig. 7.5, now performed on 64^4 lattices. To control discretization errors, the lattice spacing must be about $1/10^{\text{th}}$ of the correlation length or less. To control finite-volume effects, the lattice size must be about 3 times the correlation length or more. This implies lattices of minimum size 30^4, which have only become within reach of a reasonable computer budget in recent years.

The above simulations considered only the effect of gluons: since gluons carry a color charge (in contrast to the photon which is electrically neutral), they can lead to

[5]This would be the full story if quarks were bosons. Because they are fermions, each color component is in fact a 4-component complex vector, called Dirac *spinor*.

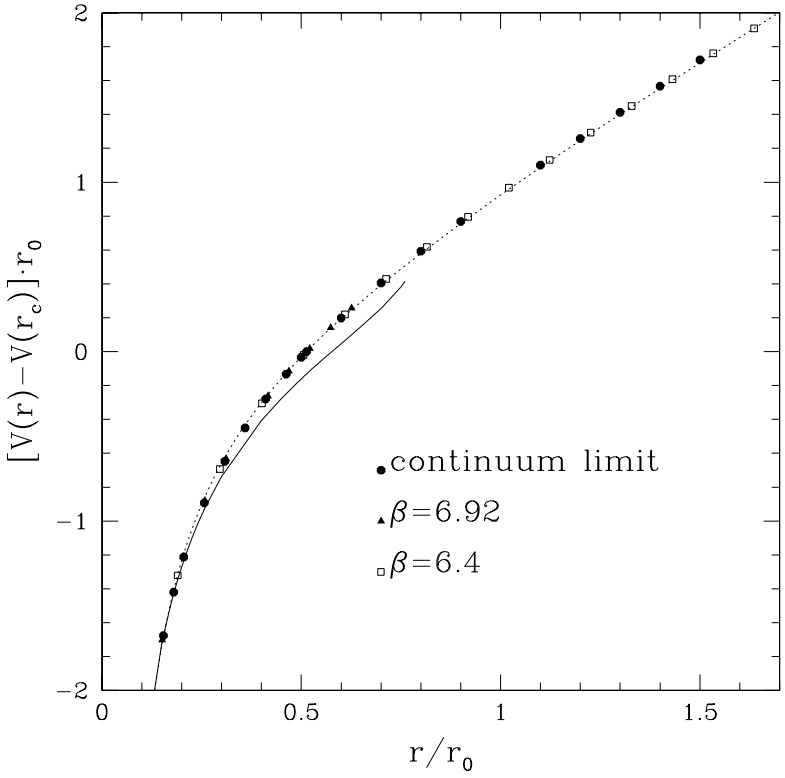

Figure 7.5: Potential $V(r)$ between a static quark and antiquark, as a function of their separation r. Data obtained at 2 values of the lattice spacing (finite values of β) are extrapolated to the continuum limit ($\beta \to \infty$). At short distance, the potential is Coulomb-like because the interaction becomes weak (the solid line shows the prediction of perturbation theory). At large distance, the potential rises linearly, which shows that it takes an infinite energy to separate the two objects: quarks are confined. A simple model of a vibrating string (dotted line) gives a good description, down to remarkably short distances.

complex effects like the confinement of charges introduced in the gluon system. But to study QCD proper, quarks must be simulated also. This is more difficult because quarks are fermions, i.e. non-commuting variables. The strategy is to integrate them out analytically. This integration induces a more complicated interaction among the remaining gluonic link variables. Actually, this interaction is *non-local*, which increases the algorithmic complexity of the Monte Carlo simulation. An efficient, exact simulation algorithm, called Hybrid Monte Carlo, was only discovered in 1987 (see bibliography). Even so, the simulation of so-called "full QCD", on lattices of size 30^4 or larger, requires a computer effort $\mathcal{O}(1)$ Teraflop$\times$ year, which has forced the community to evolve into large collaborations using dedicated computers.

Using these resources, one is now able to reproduce the masses of quark and anti-quark bound states, i.e. mesons and baryons, to a few percent accuracy. The impact of neglecting the effect of quarks or including them is nicely illustrated in Fig. 7.6. Some predictions have also been made for the properties of mesons made of charm or bottom quarks, currently being studied in particle accelerators.

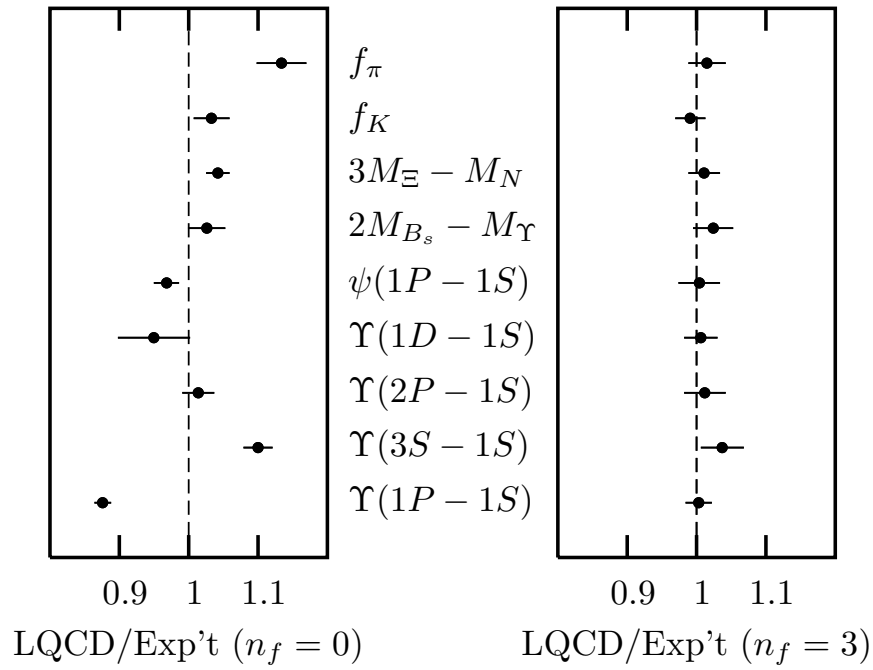

Figure 7.6: Comparison of lattice and experimental measurements of various quantities. The left figure shows the ratios lattice/experiment, for a lattice model which neglects quark effects (the number N_f of quark flavors is set to zero). The right figure shows the same ratios, for a lattice model including 3 quark flavors, all with equal masses.

Another essential purpose of QCD simulations is to quantify QCD effects in experimental tests of the electroweak Standard Model. By checking whether experimental results are consistent with the Standard Model, one can expose inconsistencies which would be the signature of new, beyond-the-standard-model physics. To reveal such inconsistencies, one must first determine precisely the predictions of the Standard Model. This entails the determination of QCD effects, which can only be obtained from lattice QCD simulations.

Finally, another direction where QCD simulations have been playing a major role is that of high temperature. The confinement of quarks, which is an experimental fact at normal temperatures, is believed to disappear at very high temperatures $\mathcal{O}(100)$ MeV $\sim \mathcal{O}(10^{12})$ K. This new state of matter, where quarks and gluons form a plasma, is being probed by accelerator experiments which smash heavy ions against each other. Lattice simulations provide an essential tool to predict the properties of this plasma.

7.7 Useful references

- *Computational Physics*, by J.M. Thijssen, Cambridge Univ. Press (2007), second edition. See Chapter 15.
- *Quantum Field Theory in a nutshell*, by A. Zee, Princeton Univ. Press (2003). This book reads like a thriller.

- *Hybrid Monte Carlo*, by S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, Phys. Lett. B **195** (1987) 216.
- *Nonuniversal critical dynamics in Monte Carlo simulations*, by R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. **58** (1987) 86.
- *Collective Monte Carlo Updating for Spin Systems*, by U. Wolff, Phys. Rev. Lett. **62** (1989) 361.
- *Worm Algorithms for Classical Statistical Models*, by N. Prokof'ev and B. Svistunov, Phys. Rev. Lett. **87** (2001) 160601.

Chapter 8

Electronic structure of molecules and solids

8.1 Introduction

In this chapter we will discuss the arguably most important quantum many body problem – the electronic structure problem – relevant for almost all properties of matter relevant in our daily life. With $O(10^{23})$ atoms in a typical piece of matter, the exponential scaling of the Hilbert space dimension with the number of particles is a nightmare. In this chapter we will discuss approximations used in quantum chemistry that reduce the problem to a polynomial one, typically scaling like $O(N^4)$. These methods map the problem to an effective single-particle problem and work only as long as correlations between electrons are weak.

8.2 The electronic structure problem

For many atoms (with the notable exception of Hydrogen and Helium which are so light that quantum effects are important), the nuclei of atoms are so much heavier than the electrons that we can view them as classical particles and can consider them as stationary for the purpose of calculating the properties of the electrons. Using this Born-Oppenheimer approximation the Hamiltonian operator for the electrons becomes

$$H = \sum_{i=1}^{N} \left(-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r}_i) \right) + \sum_{i<j} \frac{e^2}{|\vec{r}_i - \vec{r}_j|} \tag{8.1}$$

where the potential of the M atomic nuclei with charges Z_i at the locations $\vec{R}_i$ is given by

$$V(\vec{r}) = -e^2 \sum_{i=1}^{M} \frac{Z_i}{|\vec{R}_i - \vec{r}|}. \tag{8.2}$$

Using a basis set of L orbital wave functions $\{f_i\}$, the matrix elements of the one-

body and two-body parts of the Hamilton operator (8.1) are

$$t_{ij} = \int d^3\vec{r} f_i^*(\vec{r}) \left(\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r}) \right) f_j(\vec{r}) \tag{8.3}$$

$$V_{ijkl} = e^2 \int d^3\vec{r} \int d^3\vec{r}' f_i^*(\vec{r}) f_j^*(\vec{r'}) \frac{1}{|\vec{r} - \vec{r}'|} f_k(\vec{r}) f_l(\vec{r}') \tag{8.4}$$

and the Hamilton operator can be written in second quantized notation as ($a_{i\sigma}^\dagger$ creates an electron with spin σ in orbital f_i)

$$H = \sum_{ij\sigma} t_{ij} a_{i\sigma}^\dagger a_{j\sigma} + \frac{1}{2} \sum_{ijkl\sigma\sigma'} V_{ijkl} a_{i\sigma}^\dagger a_{j\sigma'}^\dagger a_{l\sigma'} a_{k\sigma}. \tag{8.5}$$

8.3 Basis functions

Before attempting to solve the many body problem we will discuss basis sets for single particle wave functions.

8.3.1 The electron gas

For the free electron gas with Hamilton operator

$$H = -\sum_{i=1}^{N} \frac{\hbar^2}{2m}\nabla^2 + e^2 \sum_{i<j} v_{ee}(\vec{r}_i, \vec{r}_j) \tag{8.6}$$

$$v_{ee}(\vec{r}, \vec{r}') = \frac{1}{|\vec{r} - \vec{r}'|} \tag{8.7}$$

the ideal choice for basis functions are plane waves

$$\psi_{\vec{k}}(\vec{r}) = \exp(-i\vec{k}\vec{r}). \tag{8.8}$$

Such plane wave basis functions are also commonly used for band structure calculations of periodic crystals.

At low density the electron gas forms a Wigner crystal. Then a better choice of basis functions are eigenfunctions of harmonic oscillators centered around the classical equilibrium positions.

8.3.2 Atoms and molecules

Which functions should be used as basis functions for atoms and molecules? We can let ourselves be guided by the exact solution of the Hydrogen atom and use the so-called *Slater-Type-Orbitals* (STO):

$$f_{inlm}(r, \theta, \phi) \propto r^{n-1} e^{-\zeta_i r} Y_{lm}(\theta, \phi). \tag{8.9}$$

These wave functions have the correct asymptotic radial dependence and the correct angular dependence. The values ζ_i are optimized so that the eigenstates of isolated atoms are reproduced as accurately as possible.

The main disadvantage of the STOs becomes apparent when trying to evaluate the matrix elements in equation (8.4) for basis functions centered around two different nuclei at position $\vec{R}_A$ and $\vec{R}_B$. There we have to evaluate integrals containing terms like

$$\frac{1}{|\vec{r}-\vec{r}'|}e^{-\zeta_i|\vec{r}-\vec{R}_A|}e^{-\zeta_j|\vec{r}-\vec{R}_B|} \tag{8.10}$$

which cannot be solved in any closed form.

The *Gauss-Type-Orbitals* (GTO)

$$f_{ilmn}(\vec{r}) \propto x^l y^m z^n e^{-\zeta_i r^2} \tag{8.11}$$

simplify the evaluation of matrix elements, as Gaussian functions can be integrated easily and the product of Gaussian functions centered at two different nuclei is again a single Gaussian function:

$$e^{-\zeta_i|\vec{r}-\vec{R}_A|^2}e^{-\zeta_j|\vec{r}-\vec{R}_B|^2} = Ke^{-\zeta|\vec{r}-\vec{R}|^2} \tag{8.12}$$

with

$$K = e^{-\frac{\zeta_i\zeta_j}{\zeta_i+\zeta_j}|\vec{R}_A-\vec{R}_B|^2}, \tag{8.13}$$

$$\zeta = \zeta_i+\zeta_j, \tag{8.14}$$

$$\vec{R} = \frac{\zeta_i\vec{R}_A+\zeta_j\vec{R}_B}{\zeta_i+\zeta_j}. \tag{8.15}$$

Also the term $\frac{1}{|\vec{r}-\vec{r}'|}$ can be rewritten as an integral over a Gaussian function

$$\frac{1}{|\vec{r}-\vec{r}'|} = \frac{2}{\sqrt{\pi}}\int_0^\infty dt e^{-t^2(\vec{r}-\vec{r}')^2} \tag{8.16}$$

and thus all the integrals (8.4) reduce to purely Gaussian integrals which can be performed analytically. The resulting speed-up more than outweighs the larger number of GTO orbitals needed (compared to STO).

As there are $O(L^4)$ integrals of the type (8.4), quantum chemistry calculations typically scale as $O(N^4)$. Modern algorithms reduce the effort to approximately $O(N)$, since the overlap of basis functions at large distances becomes negligibly small.

Independent of whether one chooses STOs or GTOs, extra care must be taken to account for the non-orthogonality of these basis functions.

8.3.3 Pseudo-potentials

The electrons in inner, fully occupied shells do not contribute to the chemical bindings. To simplify the calculations they can be replaced by pseudo-potentials, modeling the inner shells. Only the outer shells (including the valence shells) are then modeled using basis functions. The pseudo-potentials are chosen such that calculations for isolated atoms are as accurate as possible.

8.4 The Hartree Fock method

8.4.1 The Hartree-Fock approximation

The Hartree-Fock approximation, developed in the early 1930s by D. R. Hartree, J. C. Slater and V. A. Fock, is based on the assumption of independent electrons. It starts from an ansatz for the N-particle wave function as a Slater determinant of N single-particle wave functions:

$$\Phi(\vec{r}_1,\sigma_1;\ldots;\vec{r}_N,\sigma_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\vec{r}_1,\sigma_1) & \cdots & \phi_N(\vec{r}_1,\sigma_1) \\ \vdots & & \vdots \\ \phi_1(\vec{r}_N,\sigma_N) & \cdots & \phi_N(\vec{r}_N,\sigma_N) \end{vmatrix}. \tag{8.17}$$

The orthogonal single particle wave functions ϕ_μ are chosen so that the energy is minimized.

For numerical calculations a finite basis has to be introduced, as discussed in the previous section. Quantum chemists distinguish between the self-consistent-field (SCF) approximation in a finite basis set and the Hartree-Fock (HF) limit, working in a complete basis. In physics both are known as Hartree-Fock approximation.

8.4.2 The Hartree-Fock equations in nonorthogonal basis sets

It will be easiest to perform the derivation of the Hartree-Fock equations in a second quantized notation. To simplify the discussion we assume closed-shell conditions, where each orbital is occupied by both an electron with spin up and one with spin down. We start by writing the Hartree Fock wave function (8.17) in second quantized form:

$$|\Phi\rangle = \prod_{\mu,\sigma} c^\dagger_{\mu\sigma}|0\rangle, \tag{8.18}$$

where $c^\dagger_{\mu\sigma}$ creates an electron in the orbital $\phi_\mu(\mathbf{r},\sigma)$. As these wave functions are orthogonal the $c^\dagger_{\mu\sigma}$ satsify the usual fermion anticommutation relations. Next we expand the $c^\dagger_{\mu\sigma}$ in terms of the creation operators $\hat{a}^\dagger_{n\sigma}$ of our finite basis set $\{f_i\}$ (Greek subscripts refer to the Hartree-Fock single particle orbitals and Roman subscripts to the single particle basis functions):

$$c^\dagger_{\mu\sigma} = \sum_{n=1}^{L} d_{\mu n}\hat{a}^\dagger_{n\sigma} \tag{8.19}$$

and find that (up to permutation signs)

$$a_{j\sigma}|\Phi\rangle = a_{j\sigma}\prod_{\mu,\sigma'} c^\dagger_{\mu\sigma'}|0\rangle = \sum_\nu d_{\nu j} \prod_{\mu\sigma'\neq\nu\sigma} c^\dagger_{\mu\sigma'}|0\rangle. \tag{8.20}$$

In order to evaluate the expectation value $\langle\Phi|H|\Phi\rangle$, which we want to minimize, we introduce the bond-order matrix

$$P_{ij} = \sum_\sigma \langle\Phi|a^\dagger_{i\sigma}a_{j\sigma}|\Phi\rangle = 2\sum_\nu d^*_{\nu i}d_{\nu j}, \tag{8.21}$$

where we have made use of the closed-shell conditions to sum over the spin degrees of freedom. The one-body term of H is now simply $\sum_{ij} P_{ij} t_{ij}$. Next we rewrite the interaction part $\langle \Phi | a^\dagger_{i\sigma} a^\dagger_{k\sigma'} a_{l\sigma'} a_{j\sigma} | \Phi \rangle$ in terms of the P_{ij}. We find that if $\sigma = \sigma'$

$$\langle \Phi | a^\dagger_{i\sigma} a^\dagger_{k\sigma} a_{l\sigma} a_{j\sigma} | \Phi \rangle = \langle \Phi | a^\dagger_{i\sigma} a_{j\sigma} | \Phi \rangle \langle \Phi | a^\dagger_{k\sigma} a_{l\sigma} | \Phi \rangle - \langle \Phi | a^\dagger_{i\sigma} a_{l\sigma} | \Phi \rangle \langle \Phi | a^\dagger_{k\sigma} a_{j\sigma} | \Phi \rangle \tag{8.22}$$

and if $\sigma \neq \sigma'$:

$$\langle \Phi | a^\dagger_{i\sigma} a^\dagger_{k\sigma'} a_{l\sigma'} a_{j\sigma} | \Phi \rangle = \langle \Phi | a^\dagger_{i\sigma} a_{j\sigma} | \Phi \rangle \langle \Phi | a^\dagger_{k\sigma'} a_{l\sigma'} | \Phi \rangle \tag{8.23}$$

Then the energy is (again summing over the spin degrees of freedom):

$$E_0 = \sum_{ij} t_{ij} P_{ij} + \frac{1}{2} \sum_{ijkl} \left(V_{ijkl} - \frac{1}{2} V_{ilkj} \right) P_{ij} P_{kl}. \tag{8.24}$$

We now need to minimize the energy E_0 under the condition that the $|\phi_{\mu\sigma}\rangle$ are normalized:

$$2 = \sum_\sigma \langle \phi_{\mu\sigma} | \phi_{\mu\sigma} \rangle = 2 \sum_{i,j} d^*_{\mu i} d_{\mu j} S_{ij}. \tag{8.25}$$

Using Lagrange multipliers ϵ_μ to enforce this constraint we have to minimize

$$\sum_{ij} t_{ij} P_{ij} + \frac{1}{2} \sum_{ijkl} \left(V_{ijkl} - \frac{1}{2} V_{ilkj} \right) P_{ij} P_{kl} - \sum_\mu \epsilon_\mu 2 \sum_{i,j} d^*_{\mu i} d_{\mu j} S_{ij}. \tag{8.26}$$

Setting the derivative with respect to $d^*_{\mu i}$ to zero we end up with the Hartree-Fock equations for a finite basis set:

$$\sum_{j=1}^{L} (\xi_{ij} - \epsilon_\mu S_{ij}) d_{\mu j} = 0, \tag{8.27}$$

where

$$\xi_{ij} = t_{ij} + \sum_{kl} \left(V_{ijkl} - \frac{1}{2} V_{ilkj} \right) P_{kl}. \tag{8.28}$$

This is again a generalized eigenvalue problem of the form $Ax = \lambda Bx$ and looks like a one-particle Schrödinger equation. However, since the ξ_{ij} (via the P_{kl}) depend on the solution it is a nonlinear and not a linear eigenvalue problem. The equation is solved iteratively (always using the new solution to define ξ_{ij} for the next iteration), until convergence is reached.

The eigenvalues ϵ_μ of ξ do not directly correspond to energies of the orbitals, as the Fock operator counts the V-terms twice. Thus we obtain the total ground state energy from the Fock operator eigenvalues by subtracting the double counted part:

$$E_0 = \sum_{\mu=1}^{N} \epsilon_\mu - \frac{1}{2} \sum_{ijkl} \left(V_{ijkl} - \frac{1}{2} V_{ilkj} \right) P_{ij} P_{kl}. \tag{8.29}$$

8.4.3 Configuration-Interaction

The Hartree-Fock calculation is based on a set of single-particle equations $(\hat{T}+\hat{V}[\vec{\psi}])\vec{\psi} = E\vec{\psi}$ in which each electron feels the presence of other electrons only in the "averaged" form of an effective potential. It provides a *mean-field solution*, which does not correctly capture electron-electron correlations. To improve the method, and to allow the calculation of excited states, often the "configuration-interaction" (CI) method is used.

Starting from the Hartree-Fock ground state

$$|\psi_{HF}\rangle = \prod_{\mu=1}^{N} c_\mu^\dagger |0\rangle \tag{8.30}$$

one or two of the $c_\mu^\dagger$ are replaced by other orbitals $c_i^\dagger$:

$$|\psi_0\rangle = \left(1 + \sum_{i,\mu} \alpha_\mu^i c_i^\dagger c_\mu + \sum_{i<j,\mu<\nu} \alpha_{\mu\nu}^{ij} c_i^\dagger c_j^\dagger c_\mu c_\nu\right)|\psi_{HF}\rangle. \tag{8.31}$$

Note that the wave function ψ_0 can no longer be expressed as a single Slater determinant and therefore contains correlations. The energy of the system is then minimized using this variational ansatz. In a problem with N occupied and M empty orbitals this leads to a matrix eigenvalue problem with dimension $1 + NM + N^2M^2$. Using the Lanczos algorithm the low lying eigenstates can then be calculated in $\mathrm{O}((N+M)^2)$ steps. Further improvements are possible by allowing more than only double-substitutions.

8.5 Thomas-Fermi theory

In 1927, shortly after the introduction of the Schrödinger equation, L. H. Thomas and E. Fermi independently proposed a semi-classical theory for atoms or molecules with a large number N of electrons. Their idea was to express the energy as a functional of the electron density. This formulation in terms of a single, physically intuitive variable $n(r)$ avoids the complications (exponentially growing Hilbert space) associated with the solution of the Schrödinger equation for $3N$ degrees of freedom.

The Thomas-Fermi energy functional for an electron in an external potential $V_{\text{ext}}(r)$ is ($\hbar = m = e = 1$)

$$E^{TF}[n] = \frac{3}{10}(3\pi^2)^{2/3}\int dr(n(r))^{5/3} - \int dr V_{\text{ext}}(r)n(r) + \frac{1}{2}\int drdr'\frac{n(r)n(r')}{|r-r'|}. \tag{8.32}$$

The first term corresponds to E_{kin}^{TF}, which is approximated as the kinetic energy of a non-interacting electron gas with density $n(r)$: $E_{\text{kin}}^{TF} = \int drn(r)e_{\text{kin}}^{\text{non-int}}(n(r))$, with $e_{\text{kin}}^{\text{non-int}}(n) = \frac{3}{10}k_F(n)^2 = \frac{3}{10}(3\pi^2 n)^{2/3}$ the non-interacting kinetic energy per electron. The second term describes the effect of an external potential and the third term the classical electrostatic (Hartree) energy. The ground state density and energy are found by minimizing $E^{TF}[n]$ subject to the constraint

$$\int drn(r) = N. \tag{8.33}$$

This can be done by introducing a Lagrange multiplier (chemical potential) μ and minimizing the functional

$$\Omega^{TF}[n] = E^{TF}[n] - \mu\Big\{\int dr n(r) - N\Big\}. \tag{8.34}$$

One finds that the density and potential satisfy

$$\frac{1}{2}(3\pi^2)^{2/3} n(r)^{2/3} - V_{\rm ext}(r) + V_{\rm Hartree}(r) - \mu = 0, \tag{8.35}$$

with

$$V_{\rm Hartree}(r) = \int dr \frac{n(r')}{|r - r'|} \tag{8.36}$$

itself depending on the density. Solving this problem is much simpler than solving a many-body Schrödinger equation. However, while the Thomas-Fermi theory correctly reproduces certain trends, it gives poor predictions for most practical applications. In particular, it cannot explain chemical bonding.

8.6 Density functional theory

8.6.1 Hohenberg-Kohn theorem

In 1963, P. Hohenberg and W. Kohn proved that the density $n(r)$ completely characterizes a quantum mechanical system, so that an exact description of the ground state properties in terms of $n(r)$ is, at least in principle, possible. Note that in Thomas-Fermi theory, the density $n(r)$, through Eqs. (8.36) and (8.35) fixes the external potential $V_{\rm ext}$ which defines the system. The general proof is remarkably simple. We present it here for a non-degenerate ground state.[1]

Let $n(r)$ be the ground state density for N electrons in the potential $v_1(r)$, corresponding to the ground state wave-function ψ_1 and ground state energy E_1. Then

$$E_1 = \langle\psi_1|H_1|\psi_1\rangle = \int dr v_1(r) n(r) + \langle\psi_1|T + U|\psi_1\rangle, \tag{8.37}$$

where H_1 is the total Hamiltonian corresponding to v_1, T the kinetic energy operator and U the interaction energy operator. Now assume that there exists a second potential $v_2(r) \neq v_1(r) + \text{const}$ with ground state wave function $\psi_2 \neq e^{i\phi}\psi_1$ corresponding to the same density $n(r)$. Then

$$E_2 = \langle\psi_2|H_2|\psi_2\rangle = \int dr v_2(r) n(r) + \langle\psi_2|T + U|\psi_2\rangle. \tag{8.38}$$

Since ψ_1 is assumed to be non degenerate,

$$E_1 < \langle\psi_2|H_1|\psi_2\rangle = \int dr v_1(r) n(r) + \langle\psi_2|T + U|\psi_2\rangle = E_2 + \int dr (v_1(r) - v_2(r)) n(r). \tag{8.39}$$

[1] This section is based on the 1996 nobel lecture of Walter Kohn, which provides a nice overview of DFT (http://nobelprize.org/nobel_prizes/chemistry/laureates/1998/kohn-lecture.pdf)

Similarly (without assuming that ψ_2 is non-degenerate)

$$E_2 \leq \langle\psi_1|H_2|\psi_1\rangle = \int dr v_2(r)n(r) + \langle\psi_1|T+U|\psi_1\rangle = E_1 + \int dr(v_2(r)-v_1(r))n(r). \tag{8.40}$$

Adding Eqs. (8.39) and (8.40) leads to the contradiction $E_1 + E_2 < E_2 + E_1$ and we must conclude that the assumption of the existence of a second potential $v_2 \neq v_1 + \text{const}$ giving the same $n(r)$ must be wrong.

Since $n(r)$ determines both N and (up to an irrelevant constant) $v(r)$ it provides the Hamiltonian and particle number of the system. Therefore, $n(r)$ implicitly determines all the properties which can be derived from H via solution of the Schrödinger equation.

8.6.2 Hohenberg-Kohn variational principle

The ground state energy can be calculated from the Rayleigh Ritz minimal principle

$$E = \min_{\tilde{\psi}} \langle\tilde{\psi}|H|\tilde{\psi}\rangle, \tag{8.41}$$

where $\tilde{\psi}$ is a normalized trial wave function corresponding to N electrons. Following Levy and Lieb, we can carry out this minimization in two stages. First, we fix a trial density $\tilde{n}(r)$ and denote by $\tilde{\psi}^\alpha_{\tilde{n}}$ the trial functions corresponding to this $\tilde{n}$ (every $\tilde{\psi}^\alpha$ corresponds to a density $\tilde{n}^\alpha$ obtained by integrating $\langle\tilde{\psi}^\alpha|\tilde{\psi}^\alpha\rangle$ over all variables except the first and multiplying by N). The constrained energy minimum, with $\tilde{n}$ fixed is then defined as

$$E^{LL}[\tilde{n}] \equiv \min_\alpha \langle\tilde{\psi}^\alpha_{\tilde{n}}|H|\tilde{\psi}^\alpha_{\tilde{n}}\rangle = \int dr v(r)\tilde{n}(r) + F[\tilde{n}], \tag{8.42}$$

$$F[\tilde{n}] \equiv \min_\alpha \langle\tilde{\psi}^\alpha_{\tilde{n}}|T+U|\tilde{\psi}^\alpha_{\tilde{n}}\rangle. \tag{8.43}$$

In a second step, E^{LL} is minimized over all $\tilde{n}$:

$$E = \min_{\tilde{n}} E^{LL}[\tilde{n}] = \min_{\tilde{n}} \Big\{ \int dr v(r)\tilde{n}(r) + F[\tilde{n}(r)] \Big\}. \tag{8.44}$$

A few comments are in order. First, the formulation of Levy and Lieb avoids the issue of "v-representability" (the question whether a density $n(r)$ which integrates to N is a possible ground state density for some $v(r)$). The functional $F[\tilde{n}]$ requires no explicit knowledge of $v(r)$. Furthermore, Eq. (8.43) gives a definition of the *universal* functional of the density $F[\tilde{n}]$ as the sum of the kinetic and interaction energy associated with $\tilde{n}$. While this definition involving $3N$-dimensional trial wave functions leads back to the unsolvable Schrödinger equation, important formal progress has nevertheless been made. We now have an *exact* formulation of the ground state densities and energies entirely in terms of the density distribution $\tilde{n}(r)$ and of a unique, though not explicitly known functional of the density $F[\tilde{n}]$. The practical relevance of this formulation will depend on whether useful approximate forms of $F[\tilde{n}]$ can be found.

Thomas-Fermi theory corresponds to the approximations

$$T \approx \int dr n(r) \frac{3}{10} k_F(n)^2, \tag{8.45}$$

$$U \approx \frac{1}{2} \int dr dr' \frac{n(r)n(r')}{|r-r'|}. \tag{8.46}$$

The poor performance is largely due to the inadequate representation of the kinetic energy as the integral of the mean kinetic energy per electron of a uniform electron gas with density $n(r)$ ($\frac{3}{10}k_F(n)^2$). The relationship between density and kinetic energy is very subtle. For example, a metal and a (Mott) insulator can have very similar $n(r)$.

Hartree(-Fock) theory, which is based on single-particle equations for orbitals ϕ_j, describes atomic ground states much better than Thomas-Fermi theory. A reformulation of the Hohenberg-Kohn variational principle in terms of single-particle orbitals should therefore provide a better starting point for approximations.

8.6.3 Kohn-Sham equations

In 1965 W. Kohn and his postdoc L. Sham extracted a set of Hartree-like, but formally exact equations from the Hohenberg-Kohn variational principle. The Hartree(-Fock) equations have the form of a Schrödinger equation for non-interacting particles moving in an effective external potential v_{eff}. A simple derivation of the Kohn-Sham equations makes use of the analogy to the non-interacting model to define v_{eff}. In the non-interacting case, the ground state energy and density can be obtained by calculating the eigenfunctions ϕ_j and eigenvalues ϵ_j of the non-interacting single-particle equation

$$\Big(-\frac{1}{2}\nabla^2 + v(r) - \epsilon_j \Big) \phi_j = 0. \tag{8.47}$$

Taking into account the spin of electrons (and assuming a closed shell configuration) we obtain the energy and density

$$E = 2 \sum_{j=1}^{N/2} \epsilon_j, \tag{8.48}$$

$$n(r) = 2 \sum_{j=1}^{N/2} |\phi_j|^2. \tag{8.49}$$

On the other hand, for the non-ineracting system, the Hohenberg-Kohn variational principle reads

$$E \leq E^{LL}[\tilde{n}] = \int dr v(r) \tilde{n}(r) + T^{\text{non-int}}[\tilde{n}], \tag{8.50}$$

with $T^{\text{non-int}}[\tilde{n}]$ denoting the kinetic energy of the ground-state of non-interacting electrons with density $\tilde{n}$. The minimum of E^{LL} under the constraint $\int dr \tilde{n}(r) = N$ can be computed by introducing a Lagrange-multiplier ϵ. The stationary point with respect to variations in $\tilde{n}(r)$ satisfies

$$v(r) + \frac{\delta}{\delta \tilde{n}(r)} T^{\text{non-int}}[\tilde{n}(r)]\Big|_{\tilde{n}=n} - \epsilon = 0. \tag{8.51}$$

We now return to the interacting problem and express the universal functional $F[\tilde{n}]$ in the form

$$F[\tilde{n}(r)] \equiv T^{\text{non-int}}[\tilde{n}(r)] + \frac{1}{2}\int dr dr' \frac{n(r)n(r')}{|r-r'|} + E_{\text{xc}}[\tilde{n}(r)]. \tag{8.52}$$

This equation, in which we have separated the kinetic energy functional for non-interacting electrons and the Hartree energy, *defines* the so-called exchange-correlation energy functional $E_{\text{xc}}[\tilde{n}]$. The Hohenberg-Kohn variational principle for interacting electrons then takes the form

$$E \leq E^{LL}[\tilde{n}] = \int dr v(r)n(r) + T^{\text{non-int}}[\tilde{n}] + \frac{1}{2}\int dr dr' \frac{n(r)n(r')}{|r-r'|} + E_{\text{xc}}[\tilde{n}(r)], \tag{8.53}$$

and the corresponding Euler-Lagrange equation is satisfied for

$$v_{\text{eff}}(r) + \frac{\delta}{\delta\tilde{n}(r)} T^{\text{non-int}}[\tilde{n}(r)]\Big|_{\tilde{n}=n} - \epsilon = 0, \tag{8.54}$$

where

$$v_{\text{eff}}(r) = v(r) + \int dr' \frac{n(r')}{|r-r'|} + v_{xc}(r), \tag{8.55}$$

$$v_{xc}(r) \equiv \frac{\delta}{\delta\tilde{n}(r)} E_{xc}[\tilde{n}(r)]\Big|_{\tilde{n}=n}. \tag{8.56}$$

Since the form of Eq. (8.54) is identical to the non-interacting one (8.51), except that the external potential $v(r)$ is replaced by the effective potential (8.55), we conclude that the ground state density $n(r)$ can be obtained by solving the single-particle equation

$$\Big(-\frac{1}{2}\nabla^2 + v_{\text{eff}} - \epsilon_j\Big)\phi_j = 0, \tag{8.57}$$

$$n(r) = 2\sum_{j=1}^{N/2} |\phi_j(r)|^2, \tag{8.58}$$

$$v_{\text{eff}} = v(r) + \int dr' \frac{n(r')}{|r-r'|} + v_{xc}(r). \tag{8.59}$$

Note that $v_{xc}(r)$, the local exchange-correlation potential defined in Eq. (8.56), depends on the entire density distribution. The self-consistent equations (8.57), (8.58), (8.59) are called the *Kohn-Sham equations*. They build the basis of density functional theory as it is currently practiced.

Neither the wave-functions ϕ_j nor the eigenenergies ϵ_j have any obvious physical interpretation, except that the ϕ_j give the true ground state energy via Eq. (8.58) and that the highest occupied ϵ_j is related to the ionization energy. Nevertheless, in solid-state physics the ϵ_j are used to obtain an approximate band structure.

8.6.4 Local density approximation (LDA)

The real importance of the Kohn-Sham formulation of the many-body problem in terms of an auxiliary independent particle problem lies in the fact that it enables useful approximations. The kinetic energy term (of the non-interacting effective model) is taken care of explicitly, so the approximations can be limited to the exchange-correlation functional $E_{xc}[n]$ or the corresponding exchange-correlation potential v_{xc}. The most important approximations have a quasi-local form

$$E_{xc}[n(r)] = \int dr e_{xc}(r; [n(\tilde{r})]) n(r), \tag{8.60}$$

where $e_{xc}(r; [n(\tilde{r})])$ denotes an exchange-correlation energy per particle at r, which is a functional of the density distribution $\tilde{n}(r)$ in the vicinity of r.

The simplest approximation is the so-called *local density approximation* (LDA) which was introduced in the original paper by Kohn and Sham. Here, one uses the exchange-correlation energy of a uniform electron gas of density n, $e_{xc}^{\text{uniform}}(n(r))$, which is a function (not a functional) of $n(r)$:

$$E_{xc}^{LDA}[n(r)] \equiv \int dr e_{xc}^{\text{uniform}}(n(r)) n(r). \tag{8.61}$$

The exchange energy of the uniform electron gas can be calculated analytically and evaluates to

$$e_x^{\text{uniform}}(n) = -\frac{3}{4\pi} k_F(n) = -\frac{3}{4\pi} \Big(\frac{9\pi}{4}\Big)^{1/3} \frac{1}{r_s} = -\frac{0.458}{r_s}, \tag{8.62}$$

where the density is expressed in terms of the radius r_s of a sphere containing one electron: $(4\pi/3) r_s^3 = 1/n$. The correlation energy $e_c^{\text{uniform}}(n)$ can be obtained by fitting quantum Monte Carlo results for several values of n. A widely used form is due to Perdew and Zunger,

$$e_c^{\text{uniform}}(n) \approx \begin{cases} -0.048 + 0.031 \ln(r_s) - 0.0116 r_s + 0.0002 r_s \ln(r_s), & r_s < 1 \\ -0.142/(1 + 1.953\sqrt{r_s} + 0.333 r_s), & r_s > 1 \end{cases} . \tag{8.63}$$

LDA is exact for the uniform electron gas and one might naively expect that its application might be restricted to systems characterized by a slowly varying density. However, it turned out that this approximation works better than anyone, including the inventors, could have imagined. It gives reasonable results even in atomic systems with large density variations. Today, LDA is the basis of an entire industry spanning solid state physics and quantum chemistry. The large practical impact of his work has earned W. Kohn the 1996 Nobel prize in chemistry.

8.6.5 Beyond LDA

Improvements of the LDA have been an intense field of research in quantum chemistry. The "local spin density approximation" (LSDA) uses separate densities for electrons with spin up and down. The "generalized gradient approximation" (GGA) and its

variants use functions depending not only on the density, but also on its derivatives:

$$E_{xc}^{LDA}[n(r)] = \int dr e_{xc}^{\text{uniform}}(n(r))n(r), \tag{8.64}$$

$$E_{xc}^{GGA}[n(r)] = \int dr f^{(1)}(n(r), |\nabla n(r)|)n(r), \tag{8.65}$$

$$E_{xc}^{\cdots}[n(r)] = \int dr f^{(2)}(n(r), |\nabla n(r)|, \nabla^2 n(r))n(r). \tag{8.66}$$

Note that $f^{(1)}(n(r), |\nabla n(r)|)n(r)$ is a function, not a functional of n and $|\nabla n|$. The GGA typically improves the bonding energies for small molecules by a factor 3-5.

Despite the success of LDA and its generalizations, one must keep in mind that these approximations can fail completely in situations which lack any resemblance to a non-interacting electron gas. This is the case for strongly correlated electrons systems (transition metal oxides and actinide compounds where electrons occupy partially filled, narrow d or f orbitals), which must be treated by other methods. One promising approach for these materials is the combination of an LDA treatment of the weakly-correlated orbitals with an explicit simulation of the strongly correlated d or f electrons within the so-called dynamical mean field (DMFT) framework. This technique will be discussed in Chapter 9.

8.7 Car-Parinello molecular dynamics

In the lecture on "Computational Statistical Physics" you have learned about the molecular dynamics method, in which atoms move on classical trajectories under forces, such as those derived from a Lennard-Jones potential, which have been previously calculated in quantum mechanical simulations. It would be more accurate to use a full quantum mechanical force calculation at every time step instead of using such static forces that have been extracted from previous simulations.

Roberto Car (currently in Princeton) and Michele Parinello (currently at ETH) have combined density functional theory with molecular dynamics to do just that. Their method, Car-Parinello molecular dynamics (CPMD) allows much better simulations of molecular vibration spectra and of chemical reactions.

The atomic nuclei are propagated using classical molecular dynamics, but the electronic forces which move them are estimated using density functional theory:

$$M_n \ddot{R}_n = -\frac{\partial E[n(r,t), R_n(t)]}{\partial R_n}. \tag{8.67}$$

Here M_n and R_n are the masses and locations of the atomic nuclei and $E[n(r,t), R_n]$ the energy functional (for example in the local density approximation).

As the solution of the full electronic problem at every time step is a very time consuming task we do not want to start every time from scratch. Instead CPMD uses the previous values of the noninteracting electron wave functions ϕ_j of the DFT calculation (8.57) and evolves them to the ground state for the current positions of the nuclei via an artificial molecular dynamics. Hence both the nuclei R_n and the wave functions ϕ_j evolve in the same molecular dynamics scheme.

Ignoring the constraint of orthogonality of the ϕ_j and remembering that the wave-functions determine the density $n(r)$ through Eq. (8.58), we can write the Lagrangian for this auxiliary problem as

$$L' = \sum_j m_{\text{aux}} \langle \dot{\phi}_j | \dot{\phi}_j \rangle - \tilde{E}[\phi_j, \phi_j^*, R_n], \tag{8.68}$$

where m_{eff} is a "mass" associated with the Kohn-Sham wave functions. The constraint of orthogonality of the wave functions is $\langle \phi_i | \phi_j \rangle = \delta_{ij}$ and can be enforced by introducing Lagrange multipliers Λ_{ij}:

$$L = L' + \sum_{ij} \Lambda_{ij} \langle \phi_i | \phi_j \rangle. \tag{8.69}$$

The Euler-Lagrange equation then yields

$$m_{\text{eff}} \ddot{\phi}_i = -\frac{\delta \tilde{E}[\phi_j, \phi_j^*, R_n]}{\delta \phi_i^*} + \sum_j \Lambda_{ij} \phi_i. \tag{8.70}$$

The artificial mass m_{aux} needs to be chosen much lighter than the nuclear masses so that the electronic structure adapts quickly to the movements of the nuclei. In actual simulations one evolves the expansion coefficients $d_{\mu n}$ of an expansion in terms of the basis functions as in equation (8.19) instead of evolving the wave functions. This gives the equations of motion

$$m_{\text{eff}} \ddot{d}_{\mu n} = -\frac{\partial \tilde{E}}{\partial d_{\mu n}^*} + \sum_\nu \Lambda_{\mu\nu} \sum_l S_{nl} d_{\nu l}. \tag{8.71}$$

There are various algorithms to determine the $\Lambda_{\mu\nu}$ so that the wave functions stay orthonormal during the time evolution. We refer to text books and special lectures on CPMD for details.

Chapter 9

Dynamical mean field theory

9.1 Introduction

In chapter 8 we discussed density functional theory, which "solves" the electronic structure problem by mapping it to an auxiliary single-particle problem. In the local density approximation, this approach works well for materials with weak electron-electron correlations. There are, however, classes of materials whose properties cannot be understood on the LDA level, and for which a more accurate description of electron-electron interactions is needed. Prominent examples are transition metal, lanthanide and actinide compounds, which contain partially filled $3d$, $4f$ and $5f$ orbitals. Electron-electron correlations in these narrow orbitals are large and the competition between localization (atomic-like behavior) and delocalization (band-like behavior) leads to complicated phase diagrams and remarkable physical properties. For example, superconductivity with high transition temperature occurs in cuprates and in the recently discovered iron based materials. Manganites exhibit a very large change in the resistivity as a function of applied magnetic field, whereas in cerium, an electronically driven phase transition results in large volume changes. Theoretically, these phenomena are still poorly understood. In part, this is due to the difficulty of deriving appropriate effective Hamiltonians, but in particular it reflects the difficulty of solving even the simplest models. The development of powerful numerical tools to treat correlated fermionic lattice models is currently one of the frontiers in theoretical condensed matter physics, and considerable progress has been made over the past decade.

In this chapter, we will first discuss the Hubbard model, a simple fermionic lattice model which captures the competition between localization and delocalization. We will then discuss how this model can be approximately solved using dynamical mean field theory (DMFT) and in the last section, we will show how the combination of band structure calculations and dynamical mean field theory enables an "ab-initio" simulation of strongly correlated materials.

9.2 Hubbard model

The Hamiltonian of the electronic structure problem can be written in second quantized notation as (see chapter 8)

$$H = \sum_{k\sigma} \epsilon_k a^\dagger_{k\sigma} a_{k\sigma} + \frac{1}{2} \sum_{k_1 k_2 k'_1 k'_2 \sigma\sigma'} \langle k_1 k_2 | V | k'_1 k'_2 \rangle a^\dagger_{k_1\sigma} a^\dagger_{k_2\sigma'} a_{k'_2\sigma'} a_{k'_1\sigma} \quad (9.1)$$

$$= \sum_{ij\sigma} t_{ij} a^\dagger_{i\sigma} a_{j\sigma} + \frac{1}{2} \sum_{ii'jj'\sigma\sigma'} \langle ii' | V | jj' \rangle a^\dagger_{i\sigma} a^\dagger_{i'\sigma'} a_{j'\sigma'} a_{j\sigma}. \quad (9.2)$$

If $a^\dagger_{i\sigma}$ creates an electron of spin σ in the Wannier orbital $\phi(r - R_i)$ centered at lattice site R_i, then the hopping parameters and matrix elements of the Coulomb interaction are

$$t_{ij} = \int dr \phi^*(r - R_i) \left(-\frac{\hbar^2}{2m} \nabla^2 + V_{\rm ext}(r) \right) \phi(r - R_j), \quad (9.3)$$

$$\langle ii' | V | jj' \rangle = e^2 \int dr dr' \phi^*(r - R_i) \phi^*(r' - R_{i'}) \frac{1}{|r - r'|} \phi(r - R_j) \phi(r' - R_{j'}). \quad (9.4)$$

If the Wannier orbitals are very localized, as it is the case for narrow d- and f-bands, then one can neglect all the matrix elements (9.4) except those with $i = i' = j = j'$, and all the hoppings in Eq. (9.3) except those between nearest neighbor sites $\langle ij \rangle$. Defining

$$U \equiv e^2 \int dr dr' \frac{|\phi(r)|^2 |\phi(r')|^2}{|r - r'|} \quad (9.5)$$

we thus obtain the Hubbard model

$$H_{\rm Hubbard} = t \sum_{\langle ij \rangle \sigma} a^\dagger_{i\sigma} a_{j\sigma} - \mu \sum_{i\sigma} n_{i\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow} \quad (9.6)$$

in which the interaction U acts only locally on each lattice site and electrons only hop between nearest neighbor sites with amplitude t. Note that the Hubbard interaction is diagonal in real-space, while the hopping is not. In momentum space, the interaction term is complicated, while the kinetic energy term is diagonal. Electrons of spin σ in the Bloch states $\psi_k(r) = \frac{1}{\sqrt{N}} \sum_j e^{ik \cdot R_j} \phi(r - R_j)$ are created by the operators $a^\dagger_{k\sigma} = \frac{1}{\sqrt{N}} \sum_j e^{ik \cdot R_j} a^\dagger_{j\sigma}$. In the Bloch basis, the matrix elements of the Coulomb interaction become

$$\langle k_1 k_2 | V | k'_1 k'_2 \rangle = e^2 \int dr \int dr' \psi^*_{k_1}(r) \psi^*_{k_2}(r') \frac{1}{|r - r'|} \psi_{k'_1}(r) \psi_{k'_2}(r'), \quad (9.7)$$

while the kinetic term is diagonal with dispersion ϵ_k given by the lattice Fourier transform of the hoppings t_{ij}. On a hypercubic lattice with hopping t between nearest neighbor sites only, the dispersion takes the form

$$\epsilon_k^{\rm 1d} = 2t \cos(k), \quad (9.8)$$

$$\epsilon_k^{\rm 2d\ square} = 2t[\cos(k_x) + \cos(k_y)], \quad (9.9)$$

$$\epsilon_k^{\rm 3d\ square} = 2t[\cos(k_k) + \cos(k_y) + \cos(k_z)]. \quad (9.10)$$

The one-band Hubbard model is not only applicable to s-orbitals, but depending on filling and crystal field splittings also to d- and f-electron systems. While the five d-levels are degenerate in a rotationally invariant environment, this degeneracy is lifted in a crystalline solid. A cubic lattice splits the 5 levels into 3+2, while tetragonal and orthorhombic distortions reduce the symmetries even further and lead to the splitting of all five levels. Undoped cuprates, with 9 electrons in the Cu-d orbitals, may thus be described by the half-filled 1-band Hubbard model. Hence the big interest in understanding the properties of this simple model Hamiltonian and in particular in the question whether or not the repulsive Hubbard model becomes superconducting upon doping. For other materials, such as the recently discovered iron-based superconductors (with 6 electrons in the Fe-d bands and apparently small crystal field splittings) multi-band versions of the Hubbard model must be considered.

9.3 Dynamical mean field approximation

9.3.1 Single-site effective model

Despite its apparent simplicity, the 1-band Hubbard model is a complicated fermionic many-body problem for which there is no known solution except in one dimension. The most interesting physics (such as the metal-insulator transition) occurs in the intermediate coupling regime where the kinetic and potential energy terms are both important and where perturbative approaches (either starting from the non-interacting or the atomic limit) fail. Even the numerical solution of this fermionic lattice problem problem poses great challenges. The dimension of the Hilbert space grows as $4^{n_{\text{sites}}}$, so the exact diagonalization of finite clusters is limited to $\lesssim 20$ sites. Path integral Monte Carlo simulations suffer from the *fermionic sign problem*, because paths which exchange particles will come with minus signs. Variational approaches such as DMRG are so far restricted to one-dimensional systems.

Thus in order to gain some insights into the properties of the Hubbard (and related) models, approximate simulation methods are needed. One such approximate approach which has proven to be very useful, is *dynamical mean field theory* (DMFT). Here, the lattice problem is replaced by an effective *single-site* problem plus a self-consistency condition. To appreciate this idea, it is useful to briefly recall the static mean-field approximation of the classical Ising model, which is illustrated in the left hand panel of Fig. 9.1. Here, one extracts one particular spin, S_0, from the lattice and replaces all the remaining degrees of freedom by an effective external magnetic field $h_{\text{eff}} = zJm$ (z is the coordination number and m the magnetization per site). The lattice system

$$H^{\text{Ising}} = -J \sum_{\langle ij \rangle} S_i S_j \tag{9.11}$$

is thus mapped to the single site effective model

$$H^{\text{Ising}}_{\text{eff}} = -h_{\text{eff}} S_0. \tag{9.12}$$

From this single-site model we can easily compute the magnetization

$$m_{\text{eff}} = \tanh(\beta h_{\text{eff}}), \tag{9.13}$$

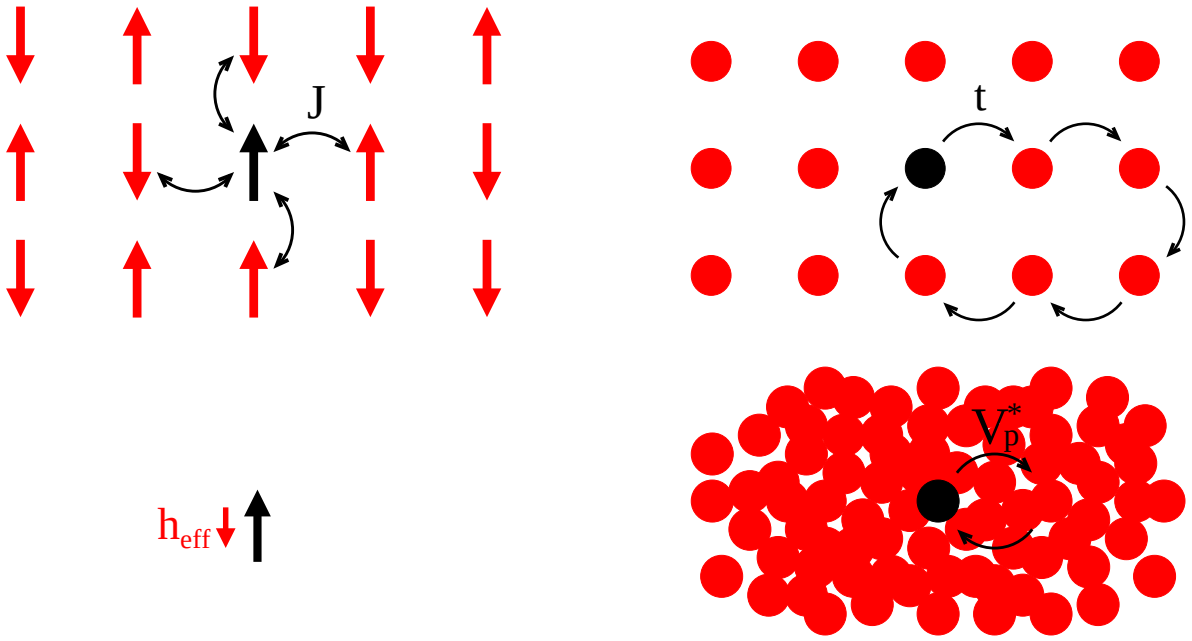

Figure 9.1: Mapping of the classical Ising model to a single-site effective model in static mean field theory (left panel) and mapping of the Hubbard model to a quantum impurity model in dynamical mean field theory (right panel).

and identifying the magnetization m of the lattice problem with the magnetization $m_{\rm eff}$ of the single site effective model gives the self-consistency condition

$$m \equiv m_{\rm eff} = \tanh(\beta z J m). \tag{9.14}$$

Returning now to the Hubbard model we may pursue a similar strategy, illustrated in the right hand panel of Fig. 9.1. We extract one particular site (the black one in the figure) from the lattice and replace the remaining degrees of freedom of the model by a bath of non-interaction fermions. The single-site effective problem in this case then becomes a *quantum impurity model* and processes where an electron hops from the black site into the lattice and after some excursion through the lattice returns to its original place are represented in the single-site effective model by transitions from the impurity into the bath and back. The amplitudes for such transitions are given by the hybridiztation parameters V of the impurity model.

So, in the single-site dynamical mean field approximation, the lattice problem

$$H^{\rm Hubbard} = -\mu \sum_{i\sigma} n_{i\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow} + t \sum_{\langle ij\rangle \sigma} a^\dagger_{i\sigma} a_{j\sigma} \tag{9.15}$$

is mapped to the single-site effective problem

$$H^{\rm Hubbard}_{\rm imp} = -\mu(n_\uparrow + n_\downarrow) + U n_\uparrow n_\downarrow + \sum_{p\sigma} (V^\sigma_p d^\dagger_\sigma a_{p\sigma} + h.c.) + \sum_{p\sigma} \epsilon_p a^\dagger_{p\sigma} a_{p\sigma}. \tag{9.16}$$

In this impurity Hamiltonian, the $d^\dagger$ create electrons in the impurity orbital (which represents a particular site of the lattice), $n = d^\dagger d$ and the $a_p^\dagger$ create bath states labeled by a quantum number p. The energy levels of the bath are ϵ_p.

9.3.2 Self-consistency condition

Solving the lattice problem essentially means computing the momentum-dependent Green's function $G(k, \omega)$. This function describes how electrons propagate through the lattice and many observables can be expressed in terms of $G(k, \omega)$. Similarly, solving the impurity problem means computing the impurity Green's function $G_{\text{imp}}(\omega)$, which has no momentum-dependence, because a single-site impurity models is zero-dimensional. The self-consistency condition, which fixes the parameters ϵ_p and V_p of the impurity model, demands that the local (momentum averaged) lattice Green's function is equal to the impurity Green's function,

$$\sum_k G(k, \omega) \equiv G_{\text{imp}}(\omega). \tag{9.17}$$

The solution must be found iteratively and these DMFT iterations involve as the essential approximation of the method a simplification of the momentum-dependence of the lattice self-energy. The self-energy describes the effect of interactions on the propagation of electrons. In the non-interacting model ($U = 0$) the Green's function is given by $G_{U=0}(k, \omega) = [\omega + \mu - \epsilon_k]^{-1}$ whereas the Green's function of the interacting model is given by $G(k, \omega) = [\omega + \mu - \epsilon_k - \Sigma(k, \omega)]^{-1}$. Therefore $\Sigma(k, \omega) = G^{-1}_{U=0}(k, \omega) - G(k, \omega)^{-1}$, and similarly the impurity self-energy is given by $\Sigma_{\text{imp}}(\omega) = G^{-1}_{\text{imp},U=0}(\omega) - G_{\text{imp}}(\omega)^{-1}$. The DMFT approximation amounts to identifying

$$\Sigma(k, \omega) \approx \Sigma_{\text{imp}}(\omega), \tag{9.18}$$

i.e. to ignoring the momentum dependence of the lattice self-energy. With this approximation we can rewrite Eq. (9.17) as

$$\sum_k [\omega + \mu - \epsilon_k - \Sigma_{\text{imp}}(\omega)]^{-1} \equiv G_{\text{imp}}(\omega). \tag{9.19}$$

Since $G_{\text{imp}}(\omega)$ and $\Sigma_{\text{imp}}(\omega)$ depend on the impurity model parameters ϵ_p and V_p we obtain a self-consistency condition for the impurity model.

9.3.3 DMFT self-consisteny loop

If a Monte Carlo method is used to solve the impurity model, then the bath degrees of freedom are integrated out and the self-consistency fixes the hybridization function $F_\sigma(-i\omega_n) = \sum_p \frac{|V_p^\sigma|^2}{i\omega_n - \epsilon_p}$ (hybridization expansion solver) or the "bath Green's function" $G_{0\sigma}(i\omega_n) = [i\omega_n + \mu - F_\sigma(-i\omega_n)]^{-1}$ (weak-coupling solver). Here, we have switched to Matsubara frequencies $\omega_n = (2n + 1)\pi/\beta$, because the Monte Carlo impurity solvers work in imaginary time. To start the calculation, one can use for example the analytically known F or G_0 corresponding to the non-interacting lattice model. The DMFT self-consistency loop then contains the following steps:

1. Solve the impurity problem, i. e. compute the Green's function $G_{\rm imp}(i\omega_n)$ for the given "bath" $G_0(i\omega_n)$.

2. Extract the self-energy of the impurity model: $\Sigma_{\rm imp}(i\omega_n) = G_0^{-1}(i\omega_n) - G_{\rm imp}^{-1}(i\omega_n)$.

3. Identify the lattice self-energy with the impurity self-energy ($\Sigma(k, i\omega_n) = \Sigma_{\rm imp}(i\omega_n)$) and compute the local lattice Green's function $G_{\rm loc}(i\omega_n) = \sum_k [i\omega_n + \mu - U/2 - \Sigma_{\rm imp}(i\omega_n)]^{-1}$.

4. Apply the DMFT self-consistency condition ($G_{\rm loc}(i\omega_n) = G_{\rm imp}(i\omega_n)$) and use this to define a new "bath" $G_{0,\rm new}^{-1}(i\omega_n) = G_{\rm loc}^{-1}(i\omega_n) + \Sigma_{\rm imp}(i\omega_n)$.

The computationally expensive step in this procedure is the solution of the impurity problem. Once the calculation has converged, the bath contains information about the topology of the lattice and tries to emulate the lattice as well as it can. The impurity, which exchanges electrons with the bath, will thus feel (at least to some extent) as if it were a site of a lattice. Obviously, not all the physics can be captured by a single-site impurity model. In particular spatial fluctuations, which become important in low-dimensional systems (such as the two-dimensional planes of high-temperature superconductors), are completely neglected. To remedy this deficiency, *cluster extensions* of dynamical mean field theory have been developed. In these calculations, a cluster of several sites is embedded in a self-consistently determined bath. Cluster DMFT allows to describe spatial correlations on the cluster exactly, while longer range correlations are treated on a mean-field level. However, the computational effort for solving the impurity problem grows rapidly with the number of sites or orbitals (see Chapter 10).

9.4 LDA+DMFT

Dynamical mean field calculations become exact in the limit of infinite dimension or coordination number, in the non-interacting limit ($U = 0$) and in the atomic limit ($t = 0$). The DMFT formalism can describe band-like behavior (quasi-particle peaks) *and* atomic-like behavior (Hubbard bands). It thus captures the competition between localization and delocalization which is a crucial ingredient of the physics of strongly correlated materials. In order to enable "ab-initio" simulations of real compounds, the dynamical mean field framework has been combined with density functional theory in the LDA approximation. The resulting machinery is called "LDA+DMFT". The idea is to use the Kohn-Sham eigenvalues ϵ_k^{KS} in the self-consistency equation (9.19) instead of the dispersion of some tight-binding model (for example Eqs. (9.8)-(9.10)). By doing so one encounters a problem: while density functional theory in the LDA approximation does not take into account all the interactions between d or f electrons it captures some of them. If we now explicitly describe the local interaction in the strongly correlated orbitals via some U-term in the impurity model, some interaction contribution appears twice and this "double counting" of interaction energy must be compensated by adding a *double counting correction* E_{DC} to the self-energy of the correlated orbitals. The self-consistency condition thus becomes

$$\sum_k [i\omega_n + \mu - \epsilon_k^{KS} - \Sigma_{\rm imp}(i\omega_n) - E_{DC}]^{-1} \equiv G_{\rm imp}(i\omega_n). \tag{9.20}$$

Nobody knows how to do this subtraction in a clean and consistent way. In practice, one uses formulas like $E_{DC} = U\langle n\rangle$ with $\langle n\rangle$ the average occupancy.

In actual material simulations, many bands will be considered, so that the Kohn-Sham eingenvalues form a matrix H_k^{LDA} in orbital space. Only the d- or f-orbitals will be explicitly treated in the impurity calculation and yield a self-energy. Thus $\Sigma_{\rm imp}$ will be a matrix of the same size as H_k^{LDA}, but the only non-zero elements will be in the block corresponding to the strongly correlated orbitals. Similarly the double-counting correction will be a diagonal matrix with non-zero elements only for the correlated orbitals. In the multi-orbital case, one can use for example $E_{DC} = \langle U\rangle\langle n_{\rm corr}\rangle$ with $\langle U\rangle$ the average of the interaction parameters of the (multiorbital) impurity problem and $\langle n_{\rm corr}\rangle$ the average occupancy of the correlated orbitals. This orbital-independent shift assures that the crystal-field splittings in the LDA band structure are preserved by the double counting correction. The chemical potential is then adjusted such that the correct total number of electrons in the correlated and uncorrelated orbitals is obtained.

For example in a LDA+DMFT simulation of Cerium, which has atomic configuration [Xe]$4f^1 5d^1 6s^2$, one may want to consider the $4f$, $5d$, $6s$ and $6p$ bands, which gives a H_k^{LDA} matrix of dimension 14+10+2+6=32. Intra-orbital and inter-orbital interactions are added to the 14 f-states and compensated by a "double-counting" shift of the chemical potential. A 14-orbital impurity problem is solved at each DMFT iteration yielding the self-energies for the f-orbitals. Correlations in the $5d$, $6s$ and $6p$ bands are treated on the LDA level and enter the calculation through the self-consistency loop

$$I_f\Big(\sum_k [i\omega_n + \mu - H_k^{LDA} - \Sigma_{\rm imp}(i\omega_n) I_f - E_{DC} I_f]^{-1}\Big) I_f \equiv G_{\rm imp}(i\omega_n), \tag{9.21}$$

which involves the inverse of 32×32 matrices. Here, I_f is a diagonal matrix which is 1 for the indices corresponding to correlated f-orbitals and zero otherwise.

Chapter 10

Diagrammatic Monte Carlo methods for impurity models

10.1 Introduction

An impurity model describes an atom or molecule embedded in some host or *bath*, with which it can exchange electrons. This exchange of electrons allows the impurity to make transitions between different quantum states, and leads to a non-trivial dynamics. Therefore, despite the zero dimensional nature (which makes impurity problems computationally much more tractable than fermionic lattice models), their numerical simulation remains a challenging task. Methods such as exact diagonalization or numerical RG, which explicitly treat a finite number of bath states, work well for single orbital models. However, because the number of bath states must be increased proportional to the number of orbitals, the computational effort grows exponentially with system size, and requires severe truncations of the bath already for two orbitals. Monte Carlo methods have the advantage that the bath is integrated out and thus the (infinite) size of the bath Hilbert space does not affect the simulation. While restricted to finite temperature, Monte Carlo methods are thus the method of choice for the solution of large multi-orbital or cluster impurity problems.

Over the last few years, significant progress has been made (both in terms of efficiency and flexibility) with the development of diagrammatic Monte Carlo techniques. This chapter provides an overview over two recently developed, complementary methods: (i) the weak-coupling approach, which scales favorably with system size and allows the efficient simulation of large impurity clusters, and (ii) the strong-coupling approach, which can handle impurity models with strong interactions.

For simplicity, we will focus on the single orbital Anderson impurity model (Fig. 10.1)

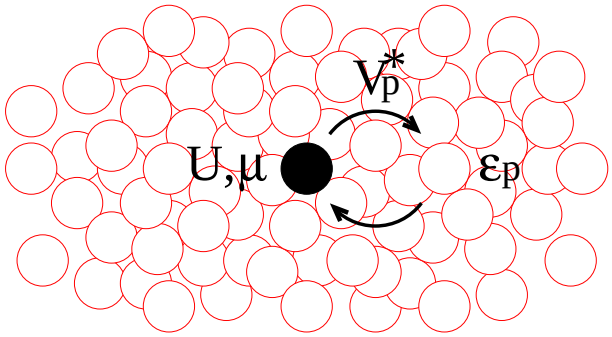

Figure 10.1: Schematic representation of a quantum impurity model. Spin up and down electrons on the impurity (black dot) interact with a repulsive energy U and can hop to non-interacting bath levels ϵ_p with transition amplitude V_p^*.

defined by the Hamiltonian $H = H_0 + H_U + H_{\rm bath} + H_{\rm mix}$ with

$$\begin{aligned} H_0 &= -(\mu - U/2)(n_\uparrow + n_\downarrow), && (10.1)\\ H_U &= U(n_\uparrow n_\downarrow - (n_\uparrow + n_\downarrow)/2), && (10.2)\\ H_{\rm bath} &= \sum_{\sigma,p} \epsilon_p a^\dagger_{p,\sigma} a_{p,\sigma}, && (10.3)\\ H_{\rm mix} &= \sum_{\sigma,p} (V_p^\sigma d^\dagger_\sigma a_{p,\sigma} + h.c.). && (10.4) \end{aligned}$$

Here, $H_0 + H_U \equiv H_{\rm loc}$ describes the impurity with creation operators $d^\dagger_\sigma$, $H_{\rm bath}$ a non-interacting bath of electrons (labeled by quantum numbers p) with creation operators $a^\dagger_{p,\sigma}$, while $H_{\rm mix}$ controls the exchange of electrons between the impurity and the bath. The transition amplitudes V_p^σ are called *hybridizations*.

The impurity model partition function Z is given by

$$Z = Tr\left[e^{-\beta H}\right], \tag{10.5}$$

with β the inverse temperature, and $Tr = Tr_d Tr_a$ the trace over the impurity and bath states. By "solving the impurity model" we essentially mean computing the impurity Green's function $(0 < \tau < \beta)$

$$g(\tau) = \langle T d(\tau) d^\dagger(0)\rangle = \frac{1}{Z} Tr\left[e^{-(\beta-\tau)H} d e^{-\tau H} d^\dagger\right], \tag{10.6}$$

which we choose to be positive.

Diagrammatic Monte Carlo simulation relies on an expansion of the partition function into a series of diagrams and the stochastic sampling of (collections) of these diagrams. We represent the partition function as a sum (or, more precisely, integral) of configurations c with weight w_c,

$$Z = \sum_c w_c, \tag{10.7}$$

and implement a random walk $c_1 \to c_2 \to c_3 \to \ldots$ in configuration space in such a way that each configuration can be reached from any other in a finite number of steps (*ergodicity*) and that *detailed balance* is satisfied,

$$|w_{c_1}|p(c_1 \to c_{c_2}) = |w_2|p(c_2 \to c_1). \tag{10.8}$$

This assures that each configuration is visited with a probability proportional to $|w_c|$ and one can thus obtain an estimate for the Green's function from a finite number N of measurements:

$$g \approx \frac{\sum_{i=1}^{N} w_{c_i} g_{c_i}}{\sum_{i=1}^{N} w_{c_i}} = \frac{\sum_{i=1}^{N} |w_{c_i}| \mathrm{sign}_{c_i} g_{c_i}}{\sum_{i=1}^{N} |w_{c_i}| \mathrm{sign}_{c_i}} = \frac{\langle \mathrm{sign} \cdot g \rangle_{MC}}{\langle \mathrm{sign} \rangle_{MC}}. \tag{10.9}$$

The error on this estimate decreases like $1/\sqrt{N}$. If the average sign of the configurations is small and decreases exponentially with decreasing temperature, the algorithm suffers from a *sign problem*.

10.2 General recipe

The first step in the diagrammatic expansion is to rewrite the partition function as a time ordered exponential using some *interaction representation*. We split the Hamiltonian into two parts, $H = H_1 + H_2$ and define the time dependent operators in the interaction picture as $O(\tau) = e^{\tau H_1} O e^{-\tau H_1}$. We furthermore introduce the operator $A(\beta) = e^{\beta H_1} e^{-\beta H}$ and write the partition function as $Z = Tr[e^{-\beta H_1} A(\beta)]$. The operator $A(\beta)$ satisfies $dA/d\beta = e^{\beta H_1}(H_1 - H)e^{-\beta H} = -H_2(\beta)A(\beta)$ and can be expressed as $A(\beta) = T \exp[-\int_0^\beta d\tau H_2(\tau)]$.

In a second step, the time ordered exponential is expanded into a power series,

$$\begin{aligned} Z &= Tr\left[e^{-\beta H_1} T e^{-\int_0^\beta d\tau H_2(\tau)}\right] \\ &= \sum_{n=0}^{\infty} \int_0^\beta d\tau_1 \ldots \int_{\tau_{n-1}}^\beta d\tau_n Tr\left[e^{-(\beta-\tau_n)H_1}(-H_2)\ldots e^{-(\tau_2-\tau_1)H_1}(-H_2)e^{-\tau_1 H_1}\right], \end{aligned} \tag{10.10}$$

which is a representation of the partition function of the form (10.7), namely the sum of all configurations $c = \{\tau_1, \ldots, \tau_n\}$, $n = 0, 1, \ldots$, $\tau_i \in [0, \beta)$ with weight

$$w_c = Tr\left[e^{-(\beta-\tau_n)H_1}(-H_2)\ldots e^{-(\tau_2-\tau_1)H_1}(-H_2)e^{-\tau_1 H_1}\right] d\tau^n. \tag{10.11}$$

In the following we will discuss in detail two complementary diagrammatic Monte Carlo algorithms, namely

1. a *weak-coupling* approach, based on an expansion of Z in powers of the interaction U, and on an interaction representation in which the time evolution is determined by the *quadratic* part $H_0 + H_{\mathrm{bath}} + H_{\mathrm{mix}}$ of the Hamiltonian,

2. a *"strong-coupling"* approach, based on an expansion of Z in powers of the impurity-bath hybridization V, and an interaction representation in which the time evolution is determined by the *local* part $H_0 + H_U + H_{\mathrm{bath}}$ of the Hamiltonian.

10.3 Weak-coupling approach

The first diagrammatic impurity solver, proposed by Rubtsov *et al.* three years ago,[1] is based on an expansion in $H_2 = H_U$. Here, I we will consider a variant of the weak coupling approach, worked out very recently by Gull *et al.*,[2] which combines the weak-coupling expansion with an auxiliary field decomposition. This "continuous-time auxiliary field method" is an adaptation of an algorithm by Rombouts *et al.*[3] for lattice models (the first diagrammatic Monte Carlo algorithm for Fermions) and in some respects similar to the time-honored Hirsch-Fye algorithm.[4]

10.3.1 Monte Carlo configurations

Following Rombouts and collaborators, we define $H_2 = H_U - K/\beta$ and $H_1 = H - H_2 = H_0 + H_{\rm bath} + H_{\rm mix} + K/\beta$, with K some arbitrary constant. Equation (10.10) then gives the expression for the partition function after expansion in H_2, and (10.11) the weight of a configuration of n *"interaction vertices"*. At this stage, we expand our configuration space by decoupling each interaction vertex using the decoupling formula proposed by Rombouts,

$$-H_2 = K/\beta - U(n_\uparrow n_\downarrow - (n_\uparrow + n_\downarrow)/2) = \frac{K}{2\beta}\sum_{s=-1,1} e^{\gamma s(n_\uparrow - n_\downarrow)}, \qquad (10.12)$$

$$\cosh(\gamma) = 1 + (\beta U)/(2K). \qquad (10.13)$$

This formula can easily be verified by checking the four states $|0\rangle$, $|\uparrow\rangle$, $|\downarrow\rangle$, and $|\uparrow\downarrow\rangle$. The configuration space is now the collection of all possible *Ising spin configurations* on the imaginary time interval $[0,\beta)$: $c = \{\{\tau_1, s_1\}, \ldots, \{\tau_n, s_n\}\}$, $n = 0, 1, \ldots$, $\tau_i \in [0,\beta)$, $s_i = \pm 1$. These configurations have weight

$$w_c = Tr\Big[e^{-(\beta-\tau_n)H_1} e^{\gamma s_n (n_\uparrow - n_\downarrow)} \ldots e^{-(\tau_2-\tau_1)H_1} e^{\gamma s_1 (n_\uparrow - n_\downarrow)} e^{-\tau_1 H_1}\Big] \Big(\frac{K d\tau}{2\beta}\Big)^n. \qquad (10.14)$$

All the operators in the trace are quadratic in c and a, so we can first separate the spin components and then proceed to the analytical calculation of the trace. Introducing $H_1^\sigma = -\mu(n_\sigma - U/2) + \sum_p \epsilon_p a^\dagger_{p,\sigma} a_{p,\sigma} + \sum_p (V_{\sigma,p} c^\dagger_\sigma a_{p,\sigma} + h.c.)$, which is the Hamiltonian of the non-interacting impurity model, the trace in Eq. (10.14) becomes ($Z_{0,\sigma} = Tr[e^{-\beta H_1^\sigma}]$)

$$Tr\Big[\ldots\Big] = e^{-K} \prod_\sigma Tr\Big[e^{-(\beta-\tau_n)H_1^\sigma} e^{\gamma s_n \sigma n_\sigma} \ldots e^{-(\tau_2-\tau_1)H_1^\sigma} e^{\gamma s_1 \sigma n_\sigma} e^{-\tau_1 H_1^\sigma}\Big]. \qquad (10.15)$$

Using the identity $e^{\gamma s \sigma n_\sigma} = e^{\gamma s\sigma} c^\dagger_\sigma c_\sigma + c_\sigma c^\dagger_\sigma = e^{\gamma s\sigma} - (e^{\gamma s\sigma} - 1) c_\sigma c^\dagger_\sigma$, the trace factors can be expressed in terms of non-interacting impurity Green's functions g_0 and evaluated using Wick's theorem. For example, at first order, we find

$$Tr\Big[e^{-(\beta-\tau_1)H_1^\sigma}(e^{\gamma s\sigma} - (e^{\gamma s\sigma}-1)c_\sigma c^\dagger_\sigma) e^{-\tau_1 H_1}\Big] = Z_{0,\sigma}(e^{\gamma s\sigma} - g_{0\sigma}(0_+)(e^{\gamma s\sigma} - 1)). \qquad (10.16)$$

[1] A. N. Rubtsov, V. V. Savkin and A. I. Lichtenstein, Phys. Rev. B **72**, 035122 (2005).
[2] E. Gull, P. Werner, O. Parcollet and M. Troyer, Europhys. Lett. **82** 57003 (2008).
[3] S. M. A. Rombouts, K. Heyde, and N. Jachowicz, Phys. Rev. Lett. **82**, 4155 (1999).
[4] J. E. Hirsch and R. M. Fye, Phys. Rev. Lett. **56**, 2521 (1986).

For n spins, this expression generalizes to

$$Tr\Big[e^{-(\beta-\tau_n)H_1^\sigma}e^{\gamma s_n\sigma n_\sigma}\dots e^{-(\tau_2-\tau_1)H_1^\sigma}e^{\gamma s_1\sigma n_\sigma}e^{-\tau_1 H_1^\sigma}\Big] = Z_{0,\sigma}\det N_\sigma^{-1}(\{s_i,\tau_i\}), \quad (10.17)$$

where N_σ is a $(n\times n)$ matrix defined by the location of the decoupled interaction vertices, the spin orientations, and the non-interaction Green's functions:

$$N_\sigma^{-1}(\{s_i,\tau_i\}) \;\equiv\; e^{\Gamma_\sigma} - G_{0\sigma}\Big(e^{\Gamma_\sigma}-I\Big). \quad (10.18)$$

The notation is $e^{\Gamma_\sigma}\equiv\mathrm{diag}(e^{\gamma\sigma s_1},\dots,e^{\gamma\sigma s_n})$, $(G_{0\sigma})_{i,j}=g_{0\sigma}(\tau_i-\tau_j)$ for $i\neq j$, $(G_{0\sigma})_{i,i}=g_{0\sigma}(0_+)$. Combining Eqs. (10.14), (10.15), (10.17) and (10.18) we thus obtain the following weight for configuration $c=\{\{\tau_1,s_1\},\dots,\{\tau_n,s_n\}\}$:

$$w_c = e^{-K}\Big(\frac{Kd\tau}{2\beta}\Big)^n\prod_\sigma Z_{0\sigma}\det N_\sigma^{-1}(\{s_i,\tau_i\}). \quad (10.19)$$

10.3.2 Sampling procedure and detailed balance

For ergodicity it is sufficient to insert/remove spins with random orientation at random times, because this allows in principle to generate all possible configurations. Furthermore, the random walk in configuration space must satisfy the detailed balance condition (10.8). Splitting the probability to move from configuration c_i to configuration c_j into a probability to *propose* the move and a probability to *accept* it,

$$p(c_i\to c_j) = p^{\mathrm{prop}}(c_i\to c_j)p^{\mathrm{acc}}(c_i\to c_j), \quad (10.20)$$

we arrive at the condition

$$\frac{p^{\mathrm{acc}}(c_i\to c_j)}{p^{\mathrm{acc}}(c_j\to c_i)} = \frac{p^{\mathrm{prop}}(c_j\to c_i)}{p^{\mathrm{prop}}(c_i\to c_j)}\frac{|w(c_j)|}{|w(c_i)|}. \quad (10.21)$$

There is some flexibility in choosing the proposal probabilities. A reasonable choice for the insertion/removal of a spin is the following (illustrated in Fig. 10.2):

- *Insertion*
  Pick a random time in $[0,\beta)$ and a random direction for the new spin:
  $p^{\mathrm{prop}}(n\to n+1)=(1/2)(d\tau/\beta)$,
- *Removal*
  Pick a random spin: $p^{\mathrm{prop}}(n+1\to n)=1/(n+1)$.

For this choice, the ratio of acceptance probabilities becomes

$$\frac{p^{\mathrm{acc}}(n\to n+1)}{p^{\mathrm{acc}}(n+1\to n)} = \frac{K}{n+1}\prod_{\sigma=\uparrow,\downarrow}\frac{|\det(N_\sigma^{(n+1)})^{-1}|}{|\det(N_\sigma^{(n)})^{-1}|}, \quad (10.22)$$

and the random walk can thus be implemented for example on the basis of the Metropolis algorithm, *i.e.* the proposed move from n to $n\pm1$ is accepted with probability

$$\min\left[1,\frac{p^{\mathrm{acc}}(n\to n\pm1)}{p^{\mathrm{acc}}(n\pm1\to n)}\right]. \quad (10.23)$$

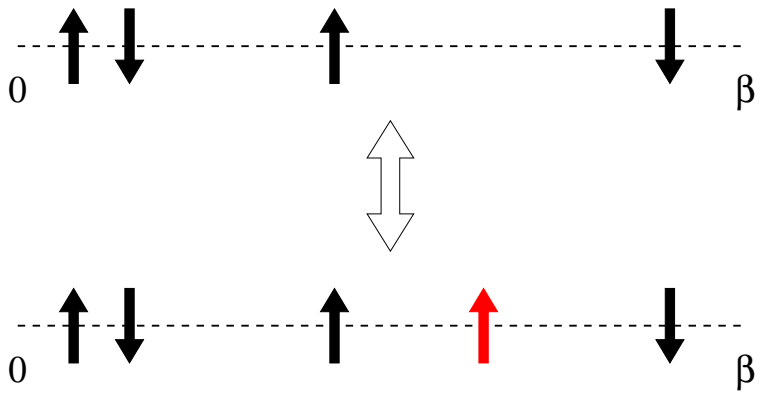

Figure 10.2: Local update in the continuous-time auxiliary field method. The dashed line represents the imaginary time interval $[0, \beta)$. We increase the perturbation order by adding a spin with random orientation at a random time. The perturbation order is decreased by removing a randomly chosen spin.

10.3.3 Determinant ratios and fast matrix updates

From Eq. (10.22) it follows that each update requires the calculation of a ratio of two determinants. Computing the determinant of a matrix of size $(n \times n)$ is an $O(n^3)$ operation (LU decomposition). The important thing to realize is that each insertion or removal of a spin merely changes one row and one column of the matrix N_σ^{-1}. We will now show that it is therefore possible to evaluate the ratio in Eq. (10.22) in a time $O(n^2)$ (insertion) or $O(1)$ (removal).

The objects which are stored and manipulated during the simulation are, besides the lists of the times $\{\tau_i\}$ and spins $\{s_i\}$, the matrices $N_\sigma = (e^{\Gamma_\sigma} - G_{0\sigma}(e^{\Gamma_\sigma} - I))^{-1}$. Inserting a spin adds a new row and column to N_σ^{-1}. We define the blocks (omitting the σ index)

$$(N^{(n+1)})^{-1} = \begin{pmatrix} (N^{(n)})^{-1} & Q \\ R & S \end{pmatrix}, \; N^{(n+1)} = \begin{pmatrix} \tilde{P} & \tilde{Q} \\ \tilde{R} & \tilde{S} \end{pmatrix}, \tag{10.24}$$

where Q, R, S denote $(n \times 1)$, $(1 \times n)$, and (1×1) matrices, respectively, which contain the contribution of the added spin. The determinant ratio needed for the acceptance/rejection probability is then given by

$$\frac{\det(N^{(n+1)})^{-1}}{\det(N^{(n)})^{-1}} = \frac{1}{\det \tilde{S}} = S - [R][N^{(n)}Q]. \tag{10.25}$$

As we store $N^{(n)}$, computing the acceptance/rejection probability of an insertion move is an $O(n^2)$ operation. If the move is accepted, the new matrix $N^{(n+1)}$ is computed out

of $N^{(n)}, Q, R$, and S, also in a time $O(n^2)$:

$$\tilde{S} = (S - [R][N^{(n)}Q])^{-1}, \tag{10.26}$$

$$\tilde{Q} = -[N^{(n)}Q]\tilde{S}, \tag{10.27}$$

$$\tilde{R} = -\tilde{S}[RN^{(n)}], \tag{10.28}$$

$$\tilde{P} = N^{(n)} + [N^{(n)}Q]\tilde{S}[RN^{(n)}]. \tag{10.29}$$

It follows from Eq. (10.25) that the calculation of the determinant ratio for removing a spin is $O(1)$, since it is just element $\tilde{S}$, and from the above formulas we also immediately find the elements of the reduced matrix:

$$N^{(n)} = \tilde{P} - \frac{[\tilde{Q}][\tilde{R}]}{\tilde{S}}. \tag{10.30}$$

10.3.4 Measurement of the Green's function

To compute the contribution of a configuration c to the Green's function measurement (10.6), we insert a creation operator $d^\dagger$ at time 0 and an annihilation operator d at time τ,

$$g^c_\sigma(\tau) = \frac{1}{w_c} Tr\Big[e^{-(\beta-\tau_n)H_1} e^{\gamma s_n(n_\uparrow - n_\downarrow)} \dots e^{-(\tau_{k+1}-\tau)H_1} d_\sigma e^{-(\tau-\tau_k)H_1} \dots e^{\gamma s_1(n_\uparrow - n_\downarrow)} e^{-\tau_1 H_1} d^\dagger_\sigma\Big] \Big(\frac{K d\tau}{2\beta}\Big)^n. \tag{10.31}$$

with w_c given in Eq. (10.14). The same steps as in section 10.3.1 (Wick's theorem) then lead to the expression

$$\begin{aligned} g^c_\sigma(\tau) &= \frac{1}{\det N^{-1}_\sigma \det N^{-1}_{\bar{\sigma}}} \det N^{-1}_{\bar{\sigma}} \det \begin{pmatrix} (N^{(n)}_\sigma)^{-1} & [g_{0\sigma}(\tau_i)] \\ -[g_{0\sigma}(\tau-\tau_j)(e^{\Gamma_{\sigma j}}-1)] & g_{0\sigma}(\tau) \end{pmatrix} \\ &= g_{0\sigma}(\tau) + [g_{0\sigma}(\tau-\tau_j)(e^{\Gamma_{\sigma j}}-1)] N^{(n)}_\sigma [g_{0\sigma}(\tau_i)]. \end{aligned} \tag{10.32}$$

The second equality follows from Eq. (10.25) and square brackets denote vectors of length n. To avoid unnecessary and time consuming summations during the Monte Carlo simulations, we only accumulate the quantity

$$S_\sigma(\tilde{\tau}) \equiv \sum_{k=1}^{n} \delta(\tilde{\tau} - \tau_k) \sum_{l=1}^{n} \big[(e^{\Gamma_\sigma} - I) N_\sigma\big]_{kl} g_{0\sigma}(\tau_l), \tag{10.33}$$

binning the time points $\tilde{\tau}$ on a fine grid. After the simulation is completed, the Green's function is computed as

$$g_\sigma(\tau) = g_{0\sigma}(\tau) + \int_0^\beta d\tilde{\tau} g_{0\sigma}(\tau - \tilde{\tau}) \big\langle S_\sigma(\tilde{\tau}) \big\rangle_{MC}. \tag{10.34}$$

10.3.5 Expansion order and role of the parameter K

It follows from Eq. (10.10) that

$$\begin{aligned}\langle -H_2\rangle &= \frac{1}{\beta}\int_0^\beta d\tau \langle -H(\tau)\rangle \\ &= \frac{1}{\beta}\frac{1}{Z}\sum_{n=0}^{\infty}\frac{n+1}{(n+1)!}\int_0^\beta d\tau\int_0^\beta d\tau_1\ldots\int_0^\beta d\tau_n Tr\Big[e^{-\beta H_1}T(-H_2(\tau))(-H_2(\tau_n))\ldots(-H_2(\tau_1))\Big] \\ &= \frac{1}{\beta}\frac{1}{Z}\sum_c n(c)w_c = \frac{1}{\beta}\langle n\rangle, \end{aligned} \tag{10.35}$$

and because $\langle -H_2\rangle = K/\beta - U\langle n_\uparrow n_\downarrow - (n_\uparrow + n_\downarrow)/2\rangle$ we conclude that the average perturbation order $\langle n\rangle$ is related to the parameter K and the potential energy by

$$\langle n\rangle = K - \beta U\langle n_\uparrow n_\downarrow - (n_\uparrow + n_\downarrow)/2\rangle. \tag{10.36}$$

Increasing K leads to a higher perturbation order (and thus slower matrix updates), but through Eq. (10.13) also to a smaller value of γ and thus to less polarization of the auxiliary spins. A K of the order 1 appears to work well. We also learn from Eq. (10.36) that the average perturbation order grows essentially proportional to U (as expected for a weak-coupling method), and proportional to inverse temperature.

10.4 "Strong coupling" approach - expansion in the impurity-bath hybridization

The second diagrammatic method, which is in many ways complementary to the weak-coupling approach, is based on an expansion of the partition function in powers of the impurity-bath hybridization V. This method has also been developed very recently, first for the Anderson impurity model[5] and then in a "matrix" formulation which allows to treat arbitrary impurity models.[6]

10.4.1 Monte Carlo configurations

Here, we decompose the Hamiltonian as $H_2 = H_{\rm mix}$ and $H_1 = H - H_2 = H_0 + H_U + H_{\rm bath}$. Since $H_2 \equiv H_2^{d^\dagger} + H_2^d = \sum_{\sigma,p} V_p^\sigma d_\sigma^\dagger a_{p,\sigma} + \sum_{\sigma,p'} V_{p'}^{\sigma *} d_\sigma a_{p,\sigma}^\dagger$ has two terms, corresponding to electrons hopping from the bath to the impurity and from the impurity back to the bath, only even perturbation orders contribute to Eq. (10.10). Furthermore, at perturbation order $2n$ only the $(2n)!/(n!)^2$ terms corresponding to n creation operators $d^\dagger$ and n annihilation operators d will contribute. We can therefore write the partition

[5] P. Werner, A. Comanac, L. de' Medici, M. Troyer and A. J. Millis, Phys. Rev. Lett. **97**, 076405 (2006).

[6] P. Werner and A. J. Millis, Phys. Rev. B **74**, 155107 (2006); K. Haule, Phys. Rev. B **75**, 155113 (2007).

function as a sum over configurations $c = \{\tau_1, \ldots, \tau_n; \tau'_1, \ldots, \tau'_n\}$:

$$Z = \sum_{n=0}^{\infty} \int_0^\beta d\tau_1 \ldots \int_{\tau_{n-1}}^\beta d\tau_n \int_0^\beta d\tau'_1 \ldots \int_{\tau'_{n-1}}^\beta d\tau'_n Tr\Big[e^{-\beta H_1} T H_2^d(\tau_n) H_2^{d^\dagger}(\tau'_n) \ldots H_2^d(\tau_1) H_2^{d^\dagger}(\tau'_1)\Big]. \tag{10.37}$$

Since the time evolution of the Anderson model (given by H_1) does not rotate the spin, there is an additional constraint, namely that both for spin up and spin down, there is an equal number of creation and annihilation operators. Taking this into account and writing out the expressions for H_2^d and $H_2^{d^\dagger}$ explicitly, we find

$$\begin{aligned} Z = \sum_{\{n_\sigma\}} \prod_\sigma \int_0^\beta d\tau_1^\sigma \ldots \int_{\tau_{n_\sigma-1}^\sigma}^\beta d\tau_{n_\sigma}^\sigma \int_0^\beta d\tau_1'^\sigma \ldots \int_{\tau'^\sigma_{n_\sigma-1}}^\beta d\tau'^\sigma_{n_\sigma} \\ \times Tr\Big[e^{-\beta H_1} T \prod_\sigma \sum_{p_1,\ldots,p_{n_\sigma}} \sum_{p'_1,\ldots,p'_{n_\sigma}} V_{p_1}^\sigma V_{p'_1}^{\sigma *} \ldots V_{p_{n_\sigma}}^\sigma V_{p'_{n_\sigma}}^{\sigma *} \\ d_\sigma(\tau_{n_\sigma}^\sigma) a^\dagger_{\sigma,p_{n_\sigma}}(\tau_{n_\sigma}^\sigma) a_{\sigma,p'_{n_\sigma}}(\tau'^\sigma_{n_\sigma}) d^\dagger_\sigma(\tau'^\sigma_{n_\sigma}) \ldots d_\sigma(\tau_1^\sigma) a^\dagger_{\sigma,p_1}(\tau_1^\sigma) a_{\sigma,p'_1}(\tau_1'^\sigma) d^\dagger_\sigma(\tau_1'^\sigma)\Big]. \end{aligned} \tag{10.38}$$

Now, because the d and a operate on different spaces and H_1 does not mix the impurity and bath states, we can separate the bath and the impurity and write

$$\begin{aligned} Z &= Z_{\rm bath} \sum_{\{n_\sigma\}} \prod_\sigma \int_0^\beta d\tau_1^\sigma \ldots \int_{\tau_{n_\sigma-1}^\sigma}^\beta d\tau_{n_\sigma}^\sigma \int_0^\beta d\tau_1'^\sigma \ldots \int_{\tau'^\sigma_{n_\sigma-1}}^\beta d\tau'^\sigma_{n_\sigma} \\ &\times Tr_d\Big[e^{-\beta H_{\rm loc}} T \prod_\sigma d_\sigma(\tau_{n_\sigma}^\sigma) d^\dagger_\sigma(\tau'^\sigma_{n_\sigma}) \ldots d_\sigma(\tau_1^\sigma) d^\dagger_\sigma(\tau_1'^\sigma)\Big] \\ &\times \frac{1}{Z_{\rm bath}} Tr_a\Big[e^{-\beta H_{\rm bath}} T \prod_\sigma \sum_{p_1,\ldots,p_{n_\sigma}} \sum_{p'_1,\ldots,p'_{n_\sigma}} V_{p_1}^\sigma V_{p'_1}^{\sigma *} \ldots V_{p_{n_\sigma}}^\sigma V_{p'_{n_\sigma}}^{\sigma *} \\ &\qquad a^\dagger_{\sigma,p_{n_\sigma}}(\tau_{n_\sigma}^\sigma) a_{\sigma,p'_{n_\sigma}}(\tau'^\sigma_{n_\sigma}) \ldots a^\dagger_{\sigma,p_1}(\tau_1^\sigma) a_{\sigma,p'_1}(\tau_1'^\sigma)\Big], \end{aligned} \tag{10.39}$$

where $Z_{\rm bath} = Tr_a e^{-\beta H_{\rm bath}}$, and $H_{\rm loc} = H_0 + H_U$. Since the bath is non-interacting, there is a Wick theorem for the bath and $Tr_a[\ldots]$ can be expressed as the determinant of some matrix, whose size is equal to the perturbation order. To find the elements of this matrix, it is useful to consider the lowest perturbation order, $n_\sigma = 1$, $n_{\bar\sigma} = 0$. In this case

$$\begin{aligned} &\sum_{p_1} \sum_{p'_1} V_{p_1}^\sigma V_{p'_1}^{\sigma *} \frac{1}{Z_{\rm bath}} Tr_a\Big[e^{-\beta H_{\rm bath}} T a^\dagger_{\sigma,p_1}(\tau_1^\sigma) a_{\sigma,p'_1}(\tau_1'^\sigma)\Big] \\ &= \sum_{p_1} \frac{|V_{p_1}^\sigma|^2}{e^{-\epsilon_{p_1}\beta}+1} \begin{cases} e^{-\epsilon_{p_1}(\beta-(\tau_1^\sigma-\tau_1'^\sigma))} & \tau_1^\sigma > \tau_1'^\sigma \\ -e^{-\epsilon_{p_1}(\tau_1'^\sigma-\tau_1^\sigma)} & \tau_1^\sigma < \tau_1'^\sigma \end{cases}. \end{aligned} \tag{10.40}$$

Note that $Z_{\rm bath} = \prod_\sigma \prod_p (e^{-\epsilon_p\beta}+1)$. Introducing the β-antiperiodic *hybridization function*

$$F_\sigma(\tau) = \sum_p \frac{|V_p|^2}{e^{-\epsilon_p\beta}+1} \begin{cases} e^{-\epsilon_p(\beta-\tau)} & \tau > 0 \\ -e^{-\epsilon_p(-\tau)} & \tau < 0 \end{cases}, \quad F_\sigma(-i\omega_n) = \sum_p \frac{|V_p^\sigma|^2}{i\omega_n - \epsilon_p}, \tag{10.41}$$

which is related to the non-interacting Green's function $G_{0\sigma}$ of Section 10.3 by $F_\sigma(-i\omega_n) = i\omega_n + \mu - U/2 - G_{0\sigma}(i\omega_n)^{-1}$, the first order result becomes $F_\sigma(\tau_1^\sigma - \tau_1'^\sigma)$. For higher orders, one obtains

$$\frac{1}{Z_{\rm bath}} Tr_a \Big[e^{-\beta H_{\rm bath}} T \prod_\sigma \sum_{p_1,\ldots,p_{n_\sigma}} \sum_{p'_1,\ldots,p'_{n_\sigma}} V^\sigma_{p_1} V^{\sigma *}_{p'_1} \ldots V^\sigma_{p_{n_\sigma}} V^{\sigma *}_{p'_{n_\sigma}} a^\dagger_{\sigma,p_{n_\sigma}}(\tau^\sigma_{n_\sigma}) a_{\sigma,p'_{n_\sigma}}(\tau'^\sigma_{n_\sigma}) \ldots a^\dagger_{\sigma,p_1}(\tau^\sigma_1) a_{\sigma,p'_1}(\tau'^\sigma_1) \Big] = \prod_\sigma \det M_\sigma^{-1}, \quad (10.42)$$

where M_σ^{-1} is a $(n_\sigma \times n_\sigma)$ matrix with elements

$$M_\sigma^{-1}(i,j) = F_\sigma(\tau_i^\sigma - \tau_j'^\sigma). \quad (10.43)$$

In the hybridization expansion method, the configuration space consists of all sequences $c = \{\tau_1^\uparrow, \ldots, \tau_{n_\uparrow}^\uparrow; \tau_1'^\uparrow, \ldots, \tau_{n_\uparrow}'^\uparrow | \tau_1^\downarrow, \ldots, \tau_{n_\downarrow}^\downarrow; \tau_1'^\downarrow, \ldots, \tau_{n_\downarrow}'^\downarrow\}$, of $n_\uparrow$ creation and annihilation operators for spin up ($n_\uparrow = 0, 1, \ldots$), and $n_\downarrow$ creation and annihilation operators for spin down ($n_\downarrow = 0, 1, \ldots$). The weight of this configuration is

$$\begin{aligned} w_c &= Z_{\rm bath} Tr_d \Big[e^{-\beta H_{\rm loc}} T \prod_\sigma d_\sigma(\tau^\sigma_{n_\sigma}) d^\dagger_\sigma(\tau'^\sigma_{n_\sigma}) \ldots d_\sigma(\tau^\sigma_1) d^\dagger_\sigma(\tau'^\sigma_1) \Big] \\ &\quad \times \prod_\sigma \det M_\sigma^{-1}(\tau^\sigma_1, \ldots, \tau^\sigma_{n_\sigma}; \tau'^\sigma_1, \ldots, \tau'^\sigma_{n_\sigma}) (d\tau)^{2n_\sigma}. \end{aligned} \quad (10.44)$$

The trace factor represents the contribution of the impurity, which fluctuates between different quantum states, as electrons hop in and out. The determinants resum all the bath evolutions which are compatible with the given sequence of transitions.

To evaluate the trace factor, we use the eigenbasis of $H_{\rm loc}$, which is $|0\rangle$ (energy $E_0 = 0$), $|\uparrow\rangle$, $|\downarrow\rangle$ (energy $E_1 = -\mu$) and $|\uparrow\downarrow\rangle$ (energy $E_2 = U - 2\mu$). In this basis, the time evolution operator $e^{-\tau H_{\rm loc}} = {\rm diag}(e^{-\tau E_0}, e^{-\tau E_1}, e^{-\tau E_1}, e^{-\tau E_2})$ is diagonal while the operators d_σ and $d_\sigma^\dagger$ will produce transitions between eigenstates with amplitude ± 1.

Because the time evolution does not flip the spin, the creation and annihilation operators for given spin have to alternate. This allows us to separate the operators for spin up from those for spin down and to depict the time evolution by a *collection of segments* (each segment representing a time interval in which an electron of spin up or down resides on the impurity). At each time, the eigenstate of the impurity follows immediately from the segment representation and we can easily compute the trace factor as

$$Tr_d \Big[e^{-\beta H_{\rm loc}} T \prod_\sigma d_\sigma(\tau^\sigma_{n_\sigma}) d^\dagger_\sigma(\tau'^\sigma_{n_\sigma}) \ldots d_\sigma(\tau^\sigma_1) d^\dagger_\sigma(\tau'^\sigma_1) \Big] = \exp \Big[\mu(l_\uparrow + l_\downarrow) - U l_{\rm overlap} \Big], \quad (10.45)$$

with l_σ the total "length" of the segments for spin σ and $l_{\rm overlap}$ the total length of the overlap between up and down segments. The lower panel of Fig. 10.3 shows a configuration with 3 segments for spin up and two segments for spin down; the time intervals where segments overlap, indicated by gray rectangles, correspond to a doubly occupied impurity and cost a repulsion energy U.

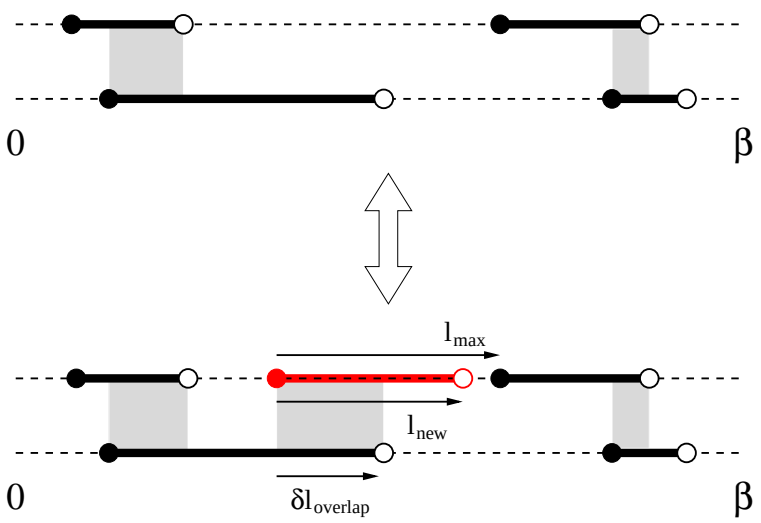

Figure 10.3: Local update in the "segment" picture. The two segment configurations correspond to spin up and spin down. Each segment depicts a time interval in which an electron of the corresponding spin resides on the impurity (the end points are the locations of the operators $d^\dagger$ and d). We increase the perturbation order by adding a segment or anti-segment of random length for random spin. The perturbation order is decreased by removing a randomly chosen segment.

10.4.2 Sampling procedure and detailed balance

For ergodicity, it is sufficient to insert and remove pairs of creation and annihilation operators (segments or anti-segments) for spin up and down. One possible strategy for inserting a segment is the following: we pick a random time in $[0, \beta)$ for the creation operator. If it falls on an existing segment, the impurity is already occupied and the move is rejected. If it falls on an empty space, we compute $l_{\max}$, the length from this position to the next segment (in the direction of increasing τ). If there are no segments, $l_{\max} = \beta$. The position of the new annihilation operator is then chosen randomly in this interval of length $l_{\max}$ (see Fig. 10.3). If we propose to remove a randomly chosen segment for this spin, then the proposal probabilities are

$$p^{\text{prop}}(n_\sigma \to n_\sigma + 1) = \frac{d\tau}{\beta}\frac{d\tau}{l_{\max}}, \tag{10.46}$$

$$p^{\text{prop}}(n_\sigma + 1 \to n_\sigma) = \frac{1}{n_\sigma + 1}, \tag{10.47}$$

and the ratio of acceptance probabilities therefore becomes

$$\frac{p^{\text{acc}}(n_\sigma \to n_\sigma + 1)}{p^{\text{acc}}(n_\sigma + 1 \to n_\sigma)} = \frac{\beta l_{\max}}{n_\sigma + 1} e^{\mu l_{\text{new}} - U\delta l_{\text{overlap}}} \frac{|\det(M_\sigma^{(n_\sigma+1)})^{-1}|}{|\det(M_\sigma^{(n_\sigma)})^{-1}|}. \tag{10.48}$$

Here, l_{new} is the length of the new segment, and $\delta l_{\text{overlap}}$ the change in the overlap. Again, we compute the ratio of determinants using the fast update formulas discussed in Section 10.3.

10.4.3 Measurement of the Green's function

The strategy is to create configurations which contribute to the Green's function measurement by decoupling the bath from a given pair of creation and annihilation operators in c. The idea is to write

$$g(\tau) = \frac{1}{Z}\sum_c w_c^{d(\tau)d^\dagger(0)} = \frac{1}{Z}\sum_c w_c^{(\tau,0)} \frac{w_c^{d(\tau)d^\dagger(0)}}{w_c^{(\tau,0)}}, \tag{10.49}$$

where $w_c^{d(\tau)d^\dagger(0)}$ denotes the weight of configuration c with and additional operator $d^\dagger(0)$ and $d(\tau)$ in the trace factor, and $w_c^{(\tau,0)}$ the complete weight corresponding to the enlarged operator sequence (including enlarged hybridization determinants). Since the trace factors of both weights are identical, and $\det M_c^{-1}$ is a minor of $\det(M_c^{(\tau,0)})^{-1}$, we find

$$\frac{w_c^{d(\tau)d^\dagger(0)}}{w_c^{(\tau,0)}} = \frac{\det M_c^{-1}}{\det(M_c^{(\tau,0)})^{-1}} = (M_c^{(\tau,0)})_{j,i}, \tag{10.50}$$

with i and j denoting the row and column corresponding to the new operators $d^\dagger$ and d in the enlarged $(M_c^{(\tau,0)})^{-1}$. To transform the sum over c into a sum over configurations $\tilde{c} = \{c, \tau_i, \tau_j'\}$, the new operators must be free to be anywhere on the imaginary time interval, which (due to translational invariance) yields a factor $\frac{1}{\beta}\Delta(\tau, \tau_i - \tau_j')$, with

$$\Delta(\tau,\tau') = \begin{cases} \delta(\tau-\tau') & \tau' > 0 \\ -\delta(\tau-\tau'-\beta) & \tau' < 0 \end{cases}. \tag{10.51}$$

Hence, the measurement formula for the Green's function becomes

$$g(\tau) = \frac{1}{Z}\sum_{\tilde{c}} w_{\tilde{c}} \sum_{i,j} \frac{1}{\beta}\Delta(\tau, \tau_i - \tau_j')(M_{\tilde{c}})_{j,i} = \Big\langle \sum_{i,j} \frac{1}{\beta}\Delta(\tau, \tau_i - \tau_j') M_{j,i} \Big\rangle_{MC}. \tag{10.52}$$

Note that if we let all the integrals run from 0 to β, there is a factor $1/(n!)^2$ in w_c and $1/((n+1)!)^2$ in $w_{\tilde{c}}$, with n the size of M_c. Changing from a sum over c to a sum over $\tilde{c}$ therefore adds a factor $(n+1)^2$ if we restrict the measurement to a specific pair of $d^\dagger$ and d. Equivalently, we can sum over all the $(n+1)^2$ pairs of operators in the enlarged configuration.

10.4.4 Generalization - Matrix formalism

It is obvious from the derivation in Section 10.4.1 that the hybridization expansion formalism is applicable to general classes of impurity models. Because the trace factor in the weight (10.44) is computed exactly, H_{loc} can contain essentially arbitrary interactions (e. g. spin-exchange terms in multi-orbital models), degrees of freedom (e. g. spins in Kondo-lattice models) or constraints (e. g. no double occupancy in the t-J model).

For multi-orbital impurity models with density-density interaction, the segment formalism is still applicable: we have now a collection of segments for each *flavor* α (orbital, spin) and the trace factor can still be computed from the length of the segments

(chemical potential contribution) and the overlaps between segments of different flavor (interaction terms).

If $H_{\rm loc}$ is not diagonal in the occupation number basis defined by the $d^\dagger_\alpha$, the calculation of $Tr_d\big[e^{-\beta H_{\rm loc}} T \prod_\alpha d_\alpha(\tau^\alpha_{n_\alpha}) d^\dagger_\alpha(\tau'^\alpha_{n_\alpha}) \dots d_\sigma(\tau^\alpha_1) d^\dagger_\alpha(\tau'^\alpha_1)\big]$ becomes more involved. We now have to compute the trace explicitly in some basis of $H_{\rm loc}$ – for example the eigenbasis, in which the time evolution operators $e^{-H_{\rm loc}\tau}$ become diagonal. The operators d_α and $d^\dagger_\alpha$ are expressed as matrices in this eigenbasis, and the evaluation of the trace factor thus involves the multiplication of matrices whose size is equal the size of the Hilbert space of $H_{\rm loc}$. Since the dimension of the Hilbert space grows *exponentially* with the number of flavors, the calculation of the trace factor becomes the computational bottleneck of the simulation, and the matrix formalism is therefore restricted to a relatively small number of flavors ($\lesssim 10$).

An important point, explained in the paper by Haule, is the use of *conserved quantum numbers* (typically particle number for spin up and spin down, momentum, ...). If the eigenstates of $H_{\rm loc}$ are grouped according to these quantum numbers, the operator matrices will acquire a sparse block structure, because for example $d^\dagger_{\uparrow,q}$ will connect the states corresponding to quantum numbers $m = \{n_\uparrow, n_\downarrow, K\}$ to those corresponding to $m' = \{n_\uparrow + 1, n_\downarrow, K + q\}$ (if they exist). Checking the compatibility of the operator sequence with a given starting block furthermore allows one to find the (potentially) contributing quantum number sectors without any matrix multiplications. The evaluation of the trace is thus reduced to a block matrix multiplication of the form

$$\sum_{\text{contr.}m} Tr_m \Big[\dots (O)_{m'',m'} (e^{-(\tau'-\tau)H_{\rm loc}})_{m'} (O)_{m',m} (e^{-\tau H_{\rm loc}})_m \Big]. \tag{10.53}$$

10.5 Comparison between the two approaches

The weak and "strong" coupling methods are in many ways complementary and their respective strengths/weaknesses result from the scaling of the computational effort with interaction strength and system size. For the Anderson impurity model considered in these notes, the U dependence of the average perturbation order is shown in Fig. 10.4 (these are dynamical mean field theory calculations for a one-band Hubbard model).[7] In the weak-coupling algorithms, where the average perturbation order is related to the potential energy, one finds a roughly linear increase of the perturbation order with U. In the hybridization-expansion method, the average perturbation order is related to the kinetic energy, and decreases as the interaction strength increases. Thus, in single site models with only density density interactions (where the evaluation of the trace factor in Eq. (10.44) is cheap), the hybridization expansion method beats the weak coupling method in the regime of strong correlations.

For more complicated models, which require the matrix formalism discussed in section 10.4.4, the hybridization expansion method scales exponentially with system size, and can only be applied to relatively small systems.[8] Here, the weak-coupling approach

[7] E. Gull, P. Werner, A. J. Millis, and M. Troyer, Phys. Rev. B **76**, 235123 (2007).

[8] While the calculation of the trace over atomic states is time consuming, it also yields useful information about the system (histogram of visited states).

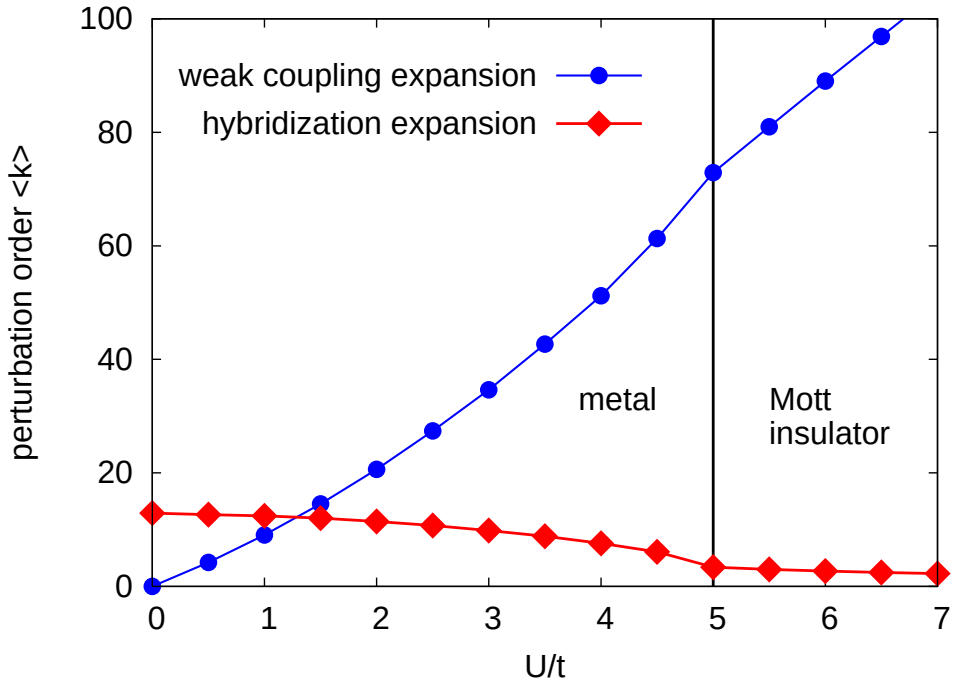

Figure 10.4: Average perturbation order for the weak-coupling and strong coupling (hybridization expansion) algorithm. These results correspond to the DMFT solution of the one-band Hubbard model with semi-circular density of states of bandwidth $4t$, and temperature $\beta = 1/T = 30$. The bath is therefore different for each data point.

– if applicable – becomes the method of choice. Table 10.1 gives a summary of the different scalings (assuming diagonal hybridization) and indicates which solver is appropriate for which type of problem.

solver	scaling		use for
weak-coupling	β^3	L^3	impurity clusters with density-density interactions and hopping
hybridization expansion (segment formulation)	β^3	L	single site multi-orbital models with density-density interaction
hybridization expansion (matrix formulation)	β	$\exp(L)$	single site multi-orbital models with general U_{ijkl}

Table 10.1: Scaling of the different impurity solvers with inverse temperature and system size.

Appendix A

Numerical methods

A.1 Numerical root solvers

The purpose of a root solver is to find a solution (a root) to the equation

$$f(x) = 0, \tag{A.1}$$

or in general to a multi-dimensional equation

$$\vec{f}(\vec{x}) = 0. \tag{A.2}$$

Numerical root solvers should be well known from the numerics courses and we will just review three simple root solvers here. Keep in mind that in any serious calculation it is usually best to use a well optimized and tested library function over a hand-coded root solver.

A.1.1 The Newton and secant methods

The Newton method is one of best known root solvers, however it is not guaranteed to converge. The key idea is to start from a guess x_0, linearize the equation around that guess

$$f(x_0) + (x - x_0)f'(x_0) = 0 \tag{A.3}$$

and solve this linearized equation to obtain a better estimate x_1. Iterating this procedure we obtain the **Newton method**:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{A.4}$$

If the derivative f' is not known analytically, as is the case in our shooting problems, we can estimate it from the difference of the last two points:

$$f'(x_n) \approx \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}} \tag{A.5}$$

Substituting this into the Newton method (A.4) we obtain the **secant method**:

$$x_{n+1} = x_n - (x_n - x_{n-1})\frac{f(x_n)}{f(x_n) - f(x_{n-1})}. \tag{A.6}$$

The Newton method can easily be generalized to higher dimensional equations, by defining the matrix of derivatives

$$A_{ij}(\vec{x}) = \frac{\partial f_i(\vec{x})}{\partial x_j} \tag{A.7}$$

to obtain the **higher dimensional Newton method**

$$\vec{x}_{n+1} = \vec{x}_n - A^{-1}\vec{f}(\vec{x}) \tag{A.8}$$

If the derivatives $A_{ij}(\vec{x})$ are not known analytically they can be estimated through finite differences:

$$A_{ij}(\vec{x}) = \frac{f_i(\vec{x} + h_j\vec{e_j}) - f_i(\vec{x})}{h_j} \qquad \text{with} \qquad h_j \approx x_j\sqrt{\varepsilon} \tag{A.9}$$

where ε is the machine precision (about 10^{-16} for double precision floating point numbers on most machines).

A.1.2 The bisection method and regula falsi

Both the bisection method and the regula falsi require two starting values x_0 and x_1 surrounding the root, with $f(x_0) < 0$ and $f(x_1) > 0$ so that under the assumption of a continuous function f there exists at least one root between x_0 and x_1.

The **bisection method** performs the following iteration

1. define a mid-point $x_m = (x_0 + x_1)/2$.
2. if $\mathrm{sign} f(x_m) = \mathrm{sign} f(x_0)$ replace $x_0 \leftarrow x_m$ otherwise replace $x_1 \leftarrow x_m$

until a root is found.

The **regula falsi** works in a similar fashion:

1. estimate the function f by a straight line from x_0 to x_1 and calculate the root of this linearized function: $x_2 = (f(x_0)x_1 - f(x_1)x_0)/(f(x_1) - f(x_0)$
2. if $\mathrm{sign} f(x_2) = \mathrm{sign} f(x_0)$ replace $x_0 \leftarrow x_2$ otherwise replace $x_1 \leftarrow x_2$

In contrast to the Newton method, both of these two methods will always find a root.

A.1.3 Optimizing a function

These root solvers can also be used for finding an extremum (minimum or maximum) of a function $f(\vec{x})$, by looking a root of

$$\nabla f(\vec{x}) = 0. \tag{A.10}$$

While this is efficient for one-dimensional problems, but better algorithms exist.

In the following discussion we assume, without loss of generality, that we want to minimize a function. The simplest algorithm for a multi-dimensional optimization is

steepest descent, which always looks for a minimum along the direction of steepest gradient: starting from an initial guess $\vec{x}_n$ a one-dimensional minimization is applied to determine the value of λ which minimizes

$$f(\vec{x}_n + \lambda \nabla f(\vec{x}_n)) \tag{A.11}$$

and then the next guess $\vec{x}_{n+1}$ is determined as

$$\vec{x}_{n+1} = \vec{x}_n + \lambda \nabla f(\vec{x}_n) \tag{A.12}$$

While this method is simple it can be very inefficient if the "landscape" of the function f resembles a long and narrow valley: the one-dimensional minimization will mainly improve the estimate transverse to the valley but takes a long time to traverse down the valley to the minimum. A better method is the **conjugate gradient** algorithm which approximates the function locally by a paraboloid and uses the minimum of this paraboloid as the next guess. This algorithm can find the minimuim of a long and narrow parabolic valley in one iteration! For this and other, even better, algorithms we recommend the use of **library functions**.

One final word of warning is that all of these minimizers will only find a **local minimum**. Whether this local minimum is also the global minimum can never be decided by purely numerically. A necessary but never sufficient check is thus to start the minimization not only from one initial guess but to try many initial points and check for consistency in the minimum found.

A.2 The Lanczos algorithm

Sparse matrices with only $O(N)$ non-zero elements are very common in scientific simulations. We have already encountered them in the winter semester when we discretized partial differential equations. Now we have reduced the transfer matrix of the Ising model to a sparse matrix product. We will later see that also the quantum mechanical Hamilton operators in lattice models are sparse.

The importance of sparsity becomes obvious when considering the cost of matrix operations as listed in table A.1. For large N the sparsity leads to memory and time savings of several orders of magnitude.

Here we will discuss the iterative calculation of a few of the extreme eigenvalues of a matrix by the Lanczos algorithm. Similar methods can be used to solve sparse linear systems of equations.

To motivate the Lanczos algorithms we will first take a look at the power method for a matrix A. Starting from a random initial vector u_1 we calculate the sequence

$$u_{n+1} = \frac{Au_n}{||Au_n||}, \tag{A.13}$$

which converges to the eigenvector of the largest eigenvalue of the matrix A. The Lanczos algorithm optimizes this crude power method.

Table A.1: Time and memory complexity for operations on sparse and dense $N \times N$ matrices

operation	time	memory
storage		
dense matrix	—	N^2
sparse matrix	—	$\mathrm{O}(N)$
matrix-vector multiplication		
dense matrix	$\mathrm{O}(N^2)$	$\mathrm{O}(N^2)$
sparse matrix	$\mathrm{O}(N)$	$\mathrm{O}(N)$
matrix-matrix multiplication		
dense matrix	$\mathrm{O}(N^{\ln 7/\ln 2})$	$\mathrm{O}(N^2)$
sparse matrix	$\mathrm{O}(N) \ldots \mathrm{O}(N^2)$	$\mathrm{O}(N) \ldots \mathrm{O}(N^2)$
all eigen values and vectors		
dense matrix	$\mathrm{O}(N^3)$	$\mathrm{O}(N^2)$
sparse matrix (iterative)	$\mathrm{O}(N^2)$	$\mathrm{O}(N^2)$
some eigen values and vectors		
dense matrix (iterative)	$\mathrm{O}(N^2)$	$\mathrm{O}(N^2)$
sparse matrix (iterative)	$\mathrm{O}(N)$	$\mathrm{O}(N)$

Lanczos iterations

The Lanczos algorithm builds a basis $\{v_1, v_2, \ldots, v_M\}$ for the Krylov-subspace $K_M = span\{u_1, u_2, \ldots, u_M\}$, which is constructed by M iterations of equation (A.13). This is done by the following iterations:

$$\beta_{n+1} v_{n+1} = A v_n - \alpha_n v_n - \beta_n v_{n-1}, \tag{A.14}$$

where

$$\alpha_n = v_n^\dagger A v_n, \qquad \beta_n = |v_n^\dagger A v_{n-1}|. \tag{A.15}$$

As the orthogonality condition

$$v_i^\dagger v_j = \delta_{ij} \tag{A.16}$$

does not determine the phases of the basis vectors, the β_i can be chosen to be real and positive. As can be seen, we only need to keep three vectors of size N in memory, which makes the Lanczos algorithm very efficient, when compared to dense matrix eigensolvers which require storage of order N^2.

In the Krylov basis the matrix A is tridiagonal

$$T^{(n)} := \begin{bmatrix} \alpha_1 & \beta_2 & 0 & \cdots & 0 \\ \beta_2 & \alpha_2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \beta_n \\ 0 & \cdots & 0 & \beta_n & \alpha_n \end{bmatrix}. \tag{A.17}$$

The eigenvalues $\{\tau_1, \ldots, \tau_M\}$ of T are good approximations of the eigenvalues of A. The extreme eigenvalues converge very fast. Thus $M \ll N$ iterations are sufficient to obtain the extreme eigenvalues.

Eigenvectors

It is no problem to compute the eigenvectors of T. They are however given in the Krylov basis $\{v_1, v_2, \ldots, v_M\}$. To obtain the eigenvectors in the original basis we need to perform a basis transformation.

Due to memory constraints we usually do not store all the v_i, but only the last three vectors. To transform the eigenvector to the original basis we have to do the Lanczos iterations a second time. Starting from the same initial vector v_1 we construct the vectors v_i iteratively and perform the basis transformation as we go along.

Roundoff errors and ghosts

In exact arithmetic the vectors $\{v_i\}$ are orthogonal and the Lanczos iterations stop after at most $N-1$ steps. The eigenvalues of T are then the exact eigenvalues of A.

Roundoff errors in finite precision cause a loss of orthogonality. There are two ways to deal with that:

- Reorthogonalization of the vectors after every step. This requires storing all of the vectors $\{v_i\}$ and is memory intensive.
- Control of the effects of roundoff.

We will discuss the second solution as it is faster and needs less memory. The main effect of roundoff errors is that the matrix T contains extra spurious eigenvalues, called "ghosts". These ghosts are not real eigenvalues of A. However they converge towards real eigenvalues of A over time and increase their multiplicities.

A simple criterion distinguishes ghosts from real eigenvalues. Ghosts are caused by roundoff errors. Thus they do not depend on on the starting vector v_1. As a consequence these ghosts are also eigenvalues of the matrix $\tilde{T}$, which can be obtained from T by deleting the first row and column:

$$\tilde{T}^{(n)} := \begin{bmatrix} \alpha_2 & \beta_3 & 0 & \cdots & 0 \\ \beta_3 & \alpha_3 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \beta_n \\ 0 & \cdots & 0 & \beta_n & \alpha_n \end{bmatrix}. \tag{A.18}$$

From these arguments we derive the following heuristic criterion to distinguish ghosts from real eigenvalues:

- All multiple eigenvalues are real, but their multiplicities might be too large.
- All single eigenvalues of T which are *not* eigenvalues of $\tilde{T}$ are also real.

Numerically stable and efficient implementations of the Lanczos algorithm can be obtained from netlib or fromhttp://www.comp-phys.org/software/ietl/ .