Quantum Monte Carlo

Lecture notes

Guglielmo Mazzola

SISSA
Trieste
Italy

Abstract

These notes cover Monte Carlo and several flavors of quantum Monte Carlo. Unlike many other sources, we treat lattice and continuous-space models together to highlight the foundations of the idea, which are model-independent. We focus on variational and projective QMC methods.

There are exercises that mostly require a numerical solution. Exercises marked with a (*) or (**) are expected to be harder.

Contents

Chapter 1

Many-Body Quantum Models

1.1 Introduction to Many-Body Hamiltonians

The study of many-body quantum systems is central to condensed matter physics, quantum chemistry, nuclear physics, and so on. These systems are characterized by the interactions between a large number of quantum particles, leading to complex phenomena that cannot be understood by studying individual particles in isolation. The fundamental challenge here is to solve the many-body Schrödinger equation, which describes the time evolution and ground states of such systems. In these note we will mostly deal with the **ground state** problem.

A general many-body Hamiltonian, $\hat{H}$, can be expressed using different types of variables, depending on the nature of the particles and the physical system under consideration. Common representations include quantum spins, fermionic particles, or particles in continuous space (i.e. bosons, electrons).

1.1.1 Quantum Spin Systems

Quantum spin systems are toy models ubiquitous in condensed matter physics and quantum information. Here, the fundamental degrees of freedom are quantum spins, typically represented by Pauli matrices. Notice that, which some small abuse of language, a spin-1/2 variable is a *qubit* in the language of quantum information.

The Trasverse Field Ising model. A canonical example is the Ising model with a transverse field (TFIM), which describes interacting spins on a lattice:

$$H = -\sum_{i,j=1}^{L} J_{i,j}\sigma_i^z\sigma_j^z - \Gamma\sum_{i=1}^{L}\sigma_i^x \tag{1.1}$$

where L is the number of lattice sites, $J_{i,j}$ represents the coupling strength between spins at sites i and j, and Γ is the strength of the transverse magnetic field. The operators σ_i^z and σ_i^x are the Pauli matrices acting on the spin at site i. The Pauli matrices are given by:

$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{1.2}$$

These matrices, along with the identity matrix, form a basis for 2×2 complex matrices. and satisfy the commutation relations:

$$\left[\sigma^\alpha, \sigma^\beta\right] = 2i\epsilon^{\alpha\beta\gamma}\sigma^\gamma \tag{1.3}$$

$$\{\sigma^\alpha, \sigma^\beta\} = 2\delta^{\alpha\beta}\mathbb{I} \tag{1.4}$$

where $\alpha, \beta, \gamma \in \{x, y, z\}$, $\epsilon^{\alpha\beta\gamma}$ is the Levi-Civita symbol, and $\mathbb{I}$ is the 2×2 identity matrix. The eigenstates of σ^z are typically denoted as $|\uparrow\rangle = \begin{pmatrix}1\\0\end{pmatrix}$ and $|\downarrow\rangle = \begin{pmatrix}0\\1\end{pmatrix}$. Notice that here we use the convention that the eigenvalues of the σ^z opererator are ± 1, not $\pm\hbar/2$.

The TFIM (with uniform couplings, J) provides the simplest example of a zero-temperature phase transition purely driven by quantum fluctuations between a paramagnet and a ferromagnet as seen, for example, in the Ising ferromagnet $LiHoF_4$. For this reason it is often used as a toy model for theory and simulations.

The Heisenberg Model. The Heisenberg model describes quantum spins interacting isotropically on a lattice. The Hamiltonian for the spin-1/2 Heisenberg model is given by:

$$H = \sum_{i,j=1}^{L} J_{i,j}\left(\sigma_i^x\sigma_j^x + \sigma_i^y\sigma_j^y + \sigma_i^z\sigma_j^z\right), \tag{1.5}$$

where L is the number of lattice sites, $J_{i,j}$ is the exchange interaction between spins at sites i and j, and σ_i^α with $\alpha = x, y, z$ are the Pauli matrices acting on site i.

Alternatively, the Hamiltonian can be expressed in terms of the ladder (raising and lowering) operators

$$\sigma^+ = \frac{1}{2}(\sigma^x + i\sigma^y) = \begin{pmatrix}0 & 1\\0 & 0\end{pmatrix}, \quad \sigma^- = \frac{1}{2}(\sigma^x - i\sigma^y) = \begin{pmatrix}0 & 0\\1 & 0\end{pmatrix}, \tag{1.6}$$

in which case the Hamiltonian becomes:

$$H = \sum_{i,j=1}^{L} J_{i,j}\left(2\sigma_i^+\sigma_j^- + 2\sigma_i^-\sigma_j^+ + \sigma_i^z\sigma_j^z\right). \tag{1.7}$$

This form is especially useful in numerical schemes. The operator $\sigma_i^+\sigma_j^- + \sigma_i^-\sigma_j^+$ describes spin-flip processes between sites i and j, and it exchanges spin states $|\uparrow\downarrow\rangle \leftrightarrow |\downarrow\uparrow\rangle$, and does nothing when both spins are equal. It is related to SWAP operator, defined as:

$$\mathrm{SWAP}_{ij} = \frac{1}{2}\left(\vec{\sigma}_i \cdot \vec{\sigma}_j + \mathbb{I}\right),$$

where $\vec{\sigma}_i \cdot \vec{\sigma}_j = \sigma_i^x\sigma_j^x + \sigma_i^y\sigma_j^y + \sigma_i^z\sigma_j^z$.

For a simple system with two spin-1/2 particles ($L = 2$), assuming isotropic coupling $J_{1,2} = J$, the Hamiltonian becomes:

$$H = J\left(\sigma_1^x\sigma_2^x + \sigma_1^y\sigma_2^y + \sigma_1^z\sigma_2^z\right). \tag{1.8}$$

In the basis $\{|\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle, |\downarrow\downarrow\rangle\}$, this Hamiltonian is represented by the matrix:

$$H = -J \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.9}$$

1.1.2 Fermionic and Bosonic Systems in Second Quantization

Another example of *discrete* quantum degrees of freedom are lattice fermions. In quantum chemistry and condensed matter physics, particularly when dealing with electrons, it is often convenient to use the formalism of second quantization. This approach describes the creation and annihilation of particles in specific quantum states (orbitals). For a system of interacting fermions, such as electrons in a molecule or solid, the Hamiltonian can be written as:

$$H = \sum_{p,k} h_{pk} c_p^\dagger c_k + \sum_{p,q,r,s} h_{pqrs} c_p^\dagger c_q^\dagger c_r c_s, \tag{1.10}$$

where $c_p^\dagger$ and c_p are the fermionic creation and destruction operators, which create (annihilate) a particle in the spin-orbital p. The coefficients h_{pk} and h_{pqrs} represent one-body and two-body interaction integrals, respectively, derived from the underlying physical interactions.

Fermions are particles that obey Fermi-Dirac statistics. They are characterized by half-integer spin (e.g., 1/2, 3/2, ...). The most crucial property of fermions is that their many-particle wavefunction must be antisymmetric with respect to the exchange of any two identical particles. This leads to the Pauli exclusion principle, which states that no two identical fermions can occupy the same quantum state simultaneously.

In second quantization, fermionic operators satisfy the following anti-commutation relations:

$$\{c_p c_k\} = c_p c_k + c_k c_p = 0 \tag{1.11}$$

$$\{c_p^\dagger a_k^\dagger\} = c_p^\dagger c_k^\dagger + c_k^\dagger c_p^\dagger = 0 \tag{1.12}$$

$$\{c_p c_k^\dagger\} = c_p c_k^\dagger + c_k^\dagger c_p = \delta_{pk} \tag{1.13}$$

where δ_{pk} is the Kronecker delta. These relations ensure the antisymmetric nature of the fermionic wavefunction.

Bosons are particles that obey Bose-Einstein statistics. They are characterized by integer spin (e.g., 0, 1, 2, ...). Unlike fermions, the many-particle wavefunction of identical bosons must be symmetric with respect to the exchange of any two identical particles. This means that multiple bosons can occupy the same quantum state, leading to phenomena like Bose-Einstein condensation.

In second quantization, bosonic operators satisfy the following commutation relations:

$$[b_p, b_k] = b_p b_k - b_k b_p = 0 \tag{1.14}$$

$$\left[b_p^\dagger, b_k^\dagger\right] = b_p^\dagger b_k^\dagger - b_k^\dagger b_p^\dagger = 0 \tag{1.15}$$

$$\left[b_p, b_k^\dagger\right] = b_p b_k^\dagger - b_k^\dagger b_p = \delta_{pk} \tag{1.16}$$

where $b_p^\dagger$ and b_p are bosonic creation and destruction operators.

1.1.3 Continuous Space Systems in First Quantization

For systems where particles move in continuous space, such as atoms, molecules, or nuclei, the Hamiltonian is often expressed in the first quantization formalism, using position and momentum operators. As an example, for a system of N electrons and N protons (e.g., a neutral atom or molecule), the non-relativistic Hamiltonian in atomic units is:

$$\hat{H} = \hat{H}_e + \hat{H}_p + \hat{H}_{ep}, \tag{1.17}$$

$$\hat{H}_e = -\frac{1}{2}\sum_{i=1}^{N} \nabla_i^2 + \frac{1}{2}\sum_{i\neq j}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}\,, \tag{1.18}$$

$$\hat{H}_p = -\frac{1}{2}\sum_{i=1}^{N} \nabla_i^2 + \frac{1}{2}\sum_{i\neq j}^{N} \frac{1}{|\mathbf{R}_i - \mathbf{R}_j|}\,, \tag{1.19}$$

$$\hat{H}_{ep} = -\sum_{i,j=1}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{R}_j|}\,, \tag{1.20}$$

in atomic units, where $\mathbf{r} = \{\mathbf{r}_1, \dots \mathbf{r}_N\}$ and $\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_N\}$, are the electron and proton coordinates, respectively. $\hat{H}_e$ describes the kinetic and Coulomb repulsion of electrons, $\hat{H}_p$ describes the kinetic and Coulomb repulsion of protons, and $\hat{H}_{ep}$ describes the electron-proton attraction.

When dealing with systems of identical particles in quantum mechanics, the indistinguishability principle dictates a specific symmetry requirement for their many-body wavefunctions. In the first-quantization representation, where particles are described by their individual coordinates, this is expressed through symmetrization or antisymmetrization. For a system of two identical **bosons**, with individual single-particle wavefunctions $\phi_a(\mathbf{r}_1)$ and $\phi_b(\mathbf{r}_2)$, the total two-particle wavefunction $\Psi_B(\mathbf{r}_1, \mathbf{r}_2)$ must be symmetric under the exchange of any two particles. This means that exchanging their labels (or coordinates) leaves the wavefunction unchanged:

$$\Psi_B(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}}\left(\phi_a(\mathbf{r}_1)\phi_b(\mathbf{r}_2) + \phi_a(\mathbf{r}_2)\phi_b(\mathbf{r}_1)\right)$$

This symmetric form ensures that if the particles are identical bosons, their joint probability density remains invariant upon exchange.
In contrast, for a system of two identical **fermions**, the total two-particle wavefunction $\Psi_F(\mathbf{r}_1, \mathbf{r}_2)$ must be antisymmetric under the exchange of any two particles. This implies that exchanging their labels results in a sign change of the wavefunction:

$$\Psi_F(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}}\left(\phi_a(\mathbf{r}_1)\phi_b(\mathbf{r}_2) - \phi_a(\mathbf{r}_2)\phi_b(\mathbf{r}_1)\right) \tag{1.21}$$

This antisymmetric form is directly linked to the Pauli Exclusion Principle, which states that no two identical fermions can occupy the exact same quantum state (if $\phi_a = \phi_b$ then $\Psi_F = 0$).
For systems with N identical particles, these principles generalize, requiring the many-body wavefunction to be fully symmetric (for bosons) or fully antisymmetric (for fermions) under

the exchange of any pair of particles, often constructed using permanents for bosons and determinants for fermions.

For N identical bosons occupying single-particle states $\phi_{i_1}, \phi_{i_2}, \ldots, \phi_{i_N}$, the normalized total wavefunction is given by the permanent:

$$\Psi_B(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \sum_P P\left[\phi_{i_1}(\mathbf{r}_1)\phi_{i_2}(\mathbf{r}_2)\cdots\phi_{i_N}(\mathbf{r}_N)\right] \tag{1.22}$$

where the sum is over all $N!$ permutations P of the particle coordinates, and $P[\ldots]$ means applying the permutation P to the set of particle coordinates. This can be more compactly written as a permanent of the matrix whose j-th row corresponds to particle $\mathbf{r}_j$ and k-th column to state ϕ_{i_k}:

$$\Psi_B(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \mathrm{Perm} \begin{pmatrix} \phi_{i_1}(\mathbf{r}_1) & \phi_{i_2}(\mathbf{r}_1) & \cdots & \phi_{i_N}(\mathbf{r}_1) \\ \phi_{i_1}(\mathbf{r}_2) & \phi_{i_2}(\mathbf{r}_2) & \cdots & \phi_{i_N}(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{i_1}(\mathbf{r}_N) & \phi_{i_2}(\mathbf{r}_N) & \cdots & \phi_{i_N}(\mathbf{r}_N) \end{pmatrix} \tag{1.23}$$

Conversely, for N identical fermions, the total wavefunction must be antisymmetric and is represented by the **Slater determinant**:

$$\Psi_F(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \det \begin{pmatrix} \phi_{i_1}(\mathbf{r}_1) & \phi_{i_2}(\mathbf{r}_1) & \cdots & \phi_{i_N}(\mathbf{r}_1) \\ \phi_{i_1}(\mathbf{r}_2) & \phi_{i_2}(\mathbf{r}_2) & \cdots & \phi_{i_N}(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{i_1}(\mathbf{r}_N) & \phi_{i_2}(\mathbf{r}_N) & \cdots & \phi_{i_N}(\mathbf{r}_N) \end{pmatrix} \tag{1.24}$$

The determinant naturally incorporates the antisymmetry property, ensuring that if any two particles are in the same single-particle state (i.e., if any two columns are identical), the determinant, and thus the wavefunction, vanishes.

For example, three identical fermions occupying the distinct orbitals ϕ_a, ϕ_b, ϕ_c, the total wavefunction $\Psi_F(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$ is the determinant of the same 3×3 matrix:

$$\begin{aligned}\Psi_F(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \frac{1}{\sqrt{6}}\Big[&\phi_a(\mathbf{r}_1)\phi_b(\mathbf{r}_2)\phi_c(\mathbf{r}_3) - \phi_a(\mathbf{r}_1)\phi_b(\mathbf{r}_3)\phi_c(\mathbf{r}_2) - \phi_a(\mathbf{r}_2)\phi_b(\mathbf{r}_1)\phi_c(\mathbf{r}_3) \\ &+\phi_a(\mathbf{r}_2)\phi_b(\mathbf{r}_3)\phi_c(\mathbf{r}_1) + \phi_a(\mathbf{r}_3)\phi_b(\mathbf{r}_1)\phi_c(\mathbf{r}_2) - \phi_a(\mathbf{r}_3)\phi_b(\mathbf{r}_2)\phi_c(\mathbf{r}_1)\Big]\end{aligned}$$

Finally notice that here we are considering *spinless fermions*, i.e. the spin degrees of freedom is included in the labels. The above notation could mean e.g. $\phi_a(\mathbf{r}) := \phi_{1s,\uparrow}(\mathbf{r})$, $\phi_b(\mathbf{r}) := \phi_{1s,\downarrow}(\mathbf{r})$ and so on, for a simple molecular case.

In quantum mechanics, particularly when dealing with many-body systems (systems with many particles), we often transition from a "first quantization" picture to a "second quantization" picture. This shift simplifies the description of systems where the number of particles can change, or where particles are indistinguishable.

1.2 Relationship between First and Second Quantization picture

In *first quantization*, we describe the state of a single particle using a wavefunction or an orbital. For a system of multiple particles, we can construct a many-body wavefunction as a product (or an antisymmetrized/symmetrized product for fermions/bosons) of single-particle orbitals.
Consider a set of orthonormal single-particle orbitals, denoted by $\phi_i(\mathbf{r})$, where i is an index labeling the orbital, and $\mathbf{r}$ represents the spatial coordinates (and possibly spin coordinates). These orbitals form a complete basis set, meaning any single-particle state can be expressed as a linear combination of these orbitals.

In *second quantization*, instead of explicitly writing down many-body wavefunctions, we work with operators that create or destroy particles in specific single-particle states (orbitals). The *creation operator*, denoted as $c_i^\dagger$ (or $b_i^\dagger$ for bosons), is a fundamental operator in this formalism.
The operator $c_i^\dagger$ is defined such that when it acts on a quantum state, it adds a particle to the system in the single-particle orbital $\phi_i(\mathbf{r})$. The subscript i in $c_i^\dagger$ directly refers to the i-th single-particle orbital from the first quantization picture, $\phi_i(\mathbf{r})$.

If we have a vacuum state (a state with no particles), denoted by $|\varnothing\rangle$ (notice that this is not the *zero* basis state of an Hilber space!), then applying $c_i^\dagger$ to it creates a state with one particle in orbital ϕ_i:

$$c_i^\dagger|\varnothing\rangle = |\phi_i\rangle$$

This $|\phi_i\rangle$ is a single-particle state where the particle occupies the orbital ϕ_i. The usual *Fock state* representation is

$$c_i^\dagger|\varnothing\rangle = |0, 0, \cdots, 1_i, \cdots, 0\rangle, \tag{1.25}$$

if we denote a Fock state with M fermionic modes as

$$|\psi\rangle = |f_{M-1}, \ldots, f_1, f_0\rangle \tag{1.26}$$

where $f_p = \{0, 1\}$ can only take binary values. If we apply $c_j^\dagger$ to a state that already has a particle in orbital ϕ_i, it creates a state with two particles, one in ϕ_i and one in ϕ_j:

$$c_j^\dagger c_i^\dagger|\varnothing\rangle = |\cdots, 1_j, \cdots, 1_i, \cdots\rangle$$

However, the order in which we *create* a particle in the Fock state matters:

$$c_1^\dagger c_2^\dagger\,|00\rangle = |11\rangle\,, \quad c_2^\dagger c_1^\dagger\,|00\rangle = -\,|11\rangle\,, \tag{1.27}$$

because in the second case, we create the fermion in the second orbital, and after that, we create the first in orbital one.[1]

Notice that, the encoding defines an *ordering* of the spin-orbital in the Fock state, and this is essential to satisfy the commutation relation, so in practical calculations, the ordering must be consistent along the steps.

[1] This purely fermionic constraint is not captured by spin-1/2 the operators σ_i^-, σ_j^-.

In a more general fashion, then (single-particle) fermionic creation and annihilation operators are formally defined as

$$\begin{aligned} c_p^\dagger \left| f_{M-1}, \ldots, f_0 \right\rangle &= \delta_{f_p,0} (-1)^{\sum_{i=0}^{p-1} f_i} \left| f_{M-1}, \ldots, f_p \oplus 1, \ldots, f_0 \right\rangle \\ c_p \left| f_{M-1}, \ldots, f_0 \right\rangle &= \delta_{f_p,1} (-1)^{\sum_{i=0}^{p-1} f_i} \left| f_{M-1}, \ldots, f_p \oplus 1, \ldots, f_0 \right\rangle , \end{aligned} \tag{1.28}$$

where the first delta, i.e. $\delta_{f_p,0}$ ensures that we can create a particle, only if the spin-orbital p is empty,[2] and the phase term $(-1)^{\sum_{i=0}^{p-1} f_i}$ restores the correct sign by counting the number of occupied modes *before* the orbital p. The spin-orbital occupation operator is given by

$$\hat{n}_i = c_i^\dagger c_i, \tag{1.29}$$

and counts the number of fermions in a given fermionic mode.

The *field operator*, often denoted as $\Psi^\dagger(\mathbf{r})$, is a more fundamental creation operator in second quantization. It creates a particle at a specific spatial (and spin)[3] point $\mathbf{r}$. $\Psi^\dagger(\mathbf{r})$ can be expressed as a linear combination of the orbital-specific creation operators $c_i^\dagger$ using the basis orbitals $\phi_i(\mathbf{r})$:

$$\Psi^\dagger(\mathbf{r}) = \sum_i \phi_i^*(\mathbf{r}) c_i^\dagger \tag{1.30}$$

Here, $\phi_i^*(\mathbf{r})$ is the complex conjugate of the orbital $\phi_i(\mathbf{r})$. This equation essentially states that creating a particle at position $\mathbf{r}$ is equivalent to creating particles in all possible orbitals, weighted by the amplitude of each orbital at that position.

Conversely, the creation operator $c_i^\dagger$ for a specific orbital $\phi_i(\mathbf{r})$ can be defined as an integral of the field operator $\Psi^\dagger(\mathbf{r})$ over all space:

$$c_i^\dagger = \int d\mathbf{r}\, \phi_i(\mathbf{r}) \Psi^\dagger(\mathbf{r}) \tag{1.31}$$

This integral definition shows that the creation operator for a specific orbital ϕ_i effectively "collects" the ability to create a particle at various points $\mathbf{r}$, weighted by the shape of the orbital $\phi_i(\mathbf{r})$. It projects the field operator onto the specific orbital ϕ_i. Notice that in case of orbitals defined as functions centered around a nucleus $\mathbf{R}$, there is an implicit dependence $\phi_i(\mathbf{r} - \mathbf{R})$.

Basically, the relations Eq.1.31 allows us to calculate the coefficient in the second quantized hamiltonian, in Eq. 1.10, (after having defined the spin-orbital functions), from the first-quantized one 1.17.

A general first-quantized Hamiltonian for a system of N identical particles can be written as:

$$\hat{H} = \hat{H}^{(1)} + \hat{H}^{(2)} = \sum_{i=1}^{N} \hat{h}_i + \frac{1}{2} \sum_{i \neq j}^{N} \hat{V}_{ij} \tag{1.32}$$

[2] $\oplus$ denotes addition modulo 2 and it is just a general notation to ensure that $1 \oplus 1 = 0$.

[3] We can incorporate the variable σ into the array $\mathbf{r}$

where $\hat{h}_i$ represents the one-body operators acting on the i-th particle (e.g., kinetic energy, external potential); $\hat{v}_{ij}$: Represents the two-body interaction operators between the i-th and j-th particles, like Coulomb interaction).

The one-body part of the Hamiltonian, $\hat{H}^{(1)} = \sum_{i=1}^{N} \hat{h}_i$, becomes:

$$\hat{H}^{(1)} = \int d\mathbf{r}\, \Psi^\dagger(\mathbf{r}) \hat{h}(\mathbf{r}) \Psi(\mathbf{r}), \tag{1.33}$$

where $\hat{h}(\mathbf{r})$ represents the single-particle Hamiltonian operator (e.g., $\hat{h}(\mathbf{r}) = -\frac{\hbar^2}{2m}\nabla^2 + V_{ext}(\mathbf{r})$).

The two-body part of the Hamiltonian, becomes:

$$\hat{H}^{(2)} = \frac{1}{2} \int d\mathbf{r}_1 d\mathbf{r}_2\, \Psi^\dagger(\mathbf{r}_1) \Psi^\dagger(\mathbf{r}_2) V(\mathbf{r}_1, \mathbf{r}_2) \Psi(\mathbf{r}_2) \Psi(\mathbf{r}_1) \tag{1.34}$$

Where $V(\mathbf{r}_1, \mathbf{r}_2)$ is the potential energy function describing the interaction between two particles at positions $\mathbf{r}_1$ and $\mathbf{r}_2$ (e.g., $V(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}$ for Coulomb interaction).

If we have a complete set of orthonormal single-particle orbitals $\phi_i(\mathbf{r})$, you can express the field operators in terms of the creation and annihilation operators for these orbitals

$$\Psi^\dagger(\mathbf{r}) = \sum_i \phi_i^*(\mathbf{r}) c_i^\dagger \tag{1.35}$$

$$\Psi(\mathbf{r}) = \sum_i \phi_i(\mathbf{r}) c_i \tag{1.36}$$

Substituting these into the second-quantized Hamiltonian yields the more common form in terms of $c_i^\dagger$ and c_j:

$$\hat{H} = \sum_{ij} \langle i | \hat{h} | j \rangle\, c_i^\dagger c_j + \frac{1}{2} \sum_{ijkl} \langle ij | V | kl \rangle\, c_i^\dagger c_j^\dagger c_l c_k \tag{1.37}$$

Where:

- $\langle i | \hat{h} | j \rangle = \int d\mathbf{r}\, \phi_i^*(\mathbf{r}) \hat{h}(\mathbf{r}) \phi_j(\mathbf{r})$ is the matrix element of the one-body operator in the orbital basis.
- $\langle ij | V | kl \rangle = \int d\mathbf{r}_1 d\mathbf{r}_2\, \phi_i^*(\mathbf{r}_1) \phi_j^*(\mathbf{r}_2) V(\mathbf{r}_1, \mathbf{r}_2) \phi_k(\mathbf{r}_1) \phi_l(\mathbf{r}_2)$ is the matrix element of the two-body interaction. Note the order of indices for the annihilation operators, which is crucial for fermionic systems due to anticommutation relations.

1.2.1 Example: from H_2 to a small Hubbard model

Here we show how the projection of the Hamiltonian using a basis set formally maps into a lattice model. We are going to *truncate* the many-body Hilbert space using few orbitals. As an example, let's consider a simple model for the H_2 molecule, with two electrons of opposite spins, that can occupy localized orbitals ϕ_1, ϕ_2 centered on the two atoms. The four allowed

states which give a zero-magnetization are:

$$|1\rangle = c^\dagger_{1,\uparrow} c^\dagger_{1,\downarrow} |\varnothing\rangle \tag{1.38}$$

$$|2\rangle = c^\dagger_{1,\uparrow} c^\dagger_{2,\downarrow} |\varnothing\rangle \tag{1.39}$$

$$|3\rangle = c^\dagger_{2,\uparrow} c^\dagger_{1,\downarrow} |\varnothing\rangle \tag{1.40}$$

$$|4\rangle = c^\dagger_{2,\uparrow} c^\dagger_{2,\downarrow} |\varnothing\rangle \,. \tag{1.41}$$

Notice that here we explicitely include the spin, $\sigma = \{\uparrow, \downarrow\}$ degree of freedom since we are talking about electrons (not general fermionic particles). To calculate the matrix elements $\langle i|H|j\rangle$, with $i, j = 1, 2, 3, 4$ we need to insert the definition of the creation operators, as in the Sect 1.2. We arrive (not shown) at an hamiltonian written in second quantization:

$$H = -t \sum_\sigma \left(c^\dagger_{1,\sigma} c_{2,\sigma} + c^\dagger_{2,\sigma} c_{1,\sigma} \right) + U \sum_{i=1,2} n_{i,\uparrow} n_{i,\downarrow} + V \sum_{\sigma,\sigma'} n_{1,\sigma} n_{2,\sigma'}, \tag{1.42}$$

where t, U, V are parameters that depends on the spacial functional form of the choosen orbital $\phi_i = \phi(\mathbf{r} - \mathbf{R}_i)$, and the distance $|\mathbf{R}_1 - \mathbf{R}_2|$. Their definition is

$$-t = \int d\mathbf{r} \phi^*(\mathbf{r} - \mathbf{R}_1) \left(-\frac{\nabla^2}{2} - \sum_{i=1,2} \frac{1}{|\mathbf{r} - \mathbf{R}_i|} \right) \phi(\mathbf{r} - \mathbf{R}_2) \tag{1.43}$$

$$U = \int d\mathbf{r} d\mathbf{r}' |\phi(\mathbf{r} - \mathbf{R}_i)|^2 \frac{1}{|\mathbf{r} - \mathbf{r}'|} |\phi(\mathbf{r}' - \mathbf{R}_i)|^2 \tag{1.44}$$

$$V = \int d\mathbf{r} d\mathbf{r}' |\phi(\mathbf{r} - \mathbf{R}_1)|^2 \frac{1}{|\mathbf{r} - \mathbf{r}'|} |\phi(\mathbf{r}' - \mathbf{R}_2)|^2. \tag{1.45}$$

Eq. 1.42 is a two-site Hubbard model when $V = 0$, namely when the two orbitals are very localized and the the atoms are distant.

1.3 The Challenge of Solving Many-Body Problems (exactly)

Solving exactly these many-body (i.e. many spins, many electrons, many orbitals) problems exactly is exponentially hard. The fundamental difficulty arises from the explosion of the Hilbert space of the composite system which yields a **memory problem**: the dimension of the Hilbert space grows exponentially with the number of particles. For instance, a system of N spin-1/2 particles has a Hilbert space of dimension 2^N. Even for a modest number of particles, say $N = 50$, 2^{50} is an astronomically large number, making it impossible to store or diagonalize the Hamiltonian matrix directly. Storing a quantum state of 50 spins (or qubits) with double-precision coefficients for each of the 2^{50} possible components requires 16 PB of memory. This exponential scaling means that classical computers cannot efficiently simulate general many-body quantum systems.

Furthermore, it has been shown that finding the ground state energy of a general quantum many-body Hamiltonian is a QMA-hard problem (Quantum Merlin-Arthur hard). This implies that even a quantum computer, while potentially offering speedups for certain quantum problems, cannot efficiently solve the general ground state problem for arbitrary Hamiltonians. This motivates approximate methods.

1.4 Matrix Representation of the Hamiltonian

For small systems, it is possible to solve the many-body Schrödinger equation exactly by constructing the Hamiltonian matrix in a chosen basis and then diagonalizing it. This method, known as exact diagonalization, is limited by the exponential growth of the Hilbert space, but it provides exact solutions for small systems and serves as a benchmark for approximate methods.

1.4.1 Exact diagonalization of the TFIM

Let's consider the transverse field Ising model for L sites with periodic boundary conditions ($J_{i,j}$ only non-zero for nearest neighbors, $J_{i,i+1} = J$ and $J_{L,1} = J$, and Γ constant):

$$H = -J \sum_{i=1}^{L} \sigma_i^z \sigma_{i+1}^z - \Gamma \sum_{i=1}^{L} \sigma_i^x \tag{1.46}$$

where $\sigma_{L+1}^z = \sigma_1^z$.

The Hilbert space for L spin-1/2 particles has a dimension of 2^L. A convenient basis is the computational basis, where each state is a tensor product of individual spin eigenstates of σ^z:

$$|s_1 s_2 \dots s_L\rangle = |s_1\rangle \otimes |s_2\rangle \otimes \cdots \otimes |s_L\rangle$$

where $s_i \in \{\uparrow, \downarrow\}$ or, equivalently, $s_i \in \{+1, -1\}$ representing the eigenvalue of σ_i^z. We can map these states to integers from 0 to $2^L - 1$. For example, for $L = 3$:

- $|\uparrow\uparrow\uparrow\rangle \leftrightarrow 0$ (if $\uparrow$ is 0, $\downarrow$ is 1)
- $|\uparrow\uparrow\downarrow\rangle \leftrightarrow 1$
- ...
- $|\downarrow\downarrow\downarrow\rangle \leftrightarrow 2^L - 1$

Let's use the convention where $|\uparrow\rangle$ corresponds to eigenvalue $+1$ for σ^z and $|\downarrow\rangle$ to -1. The action of the Pauli operators on these basis states is:

$$\begin{aligned}
\sigma_i^z |s_1 \dots s_i \dots s_L\rangle &= s_i |s_1 \dots s_i \dots s_L\rangle \\
\sigma_i^x |s_1 \dots \uparrow_i \dots s_L\rangle &= |s_1 \dots \downarrow_i \dots s_L\rangle \\
\sigma_i^x |s_1 \dots \downarrow_i \dots s_L\rangle &= |s_1 \dots \uparrow_i \dots s_L\rangle
\end{aligned}$$

where s_i is the eigenvalue of σ_i^z.

To construct the Hamiltonian matrix $H_{mn} = \langle m| H |n\rangle$, we iterate through all 2^L basis states $|n\rangle$ and apply the Hamiltonian terms.

- **Ising term** ($-\sum J_{i,j}\sigma_i^z\sigma_j^z$): This term is diagonal in the σ^z basis. For a given basis state $|n\rangle = |s_1 \dots s_L\rangle$, the contribution to the diagonal element H_{nn} is $-J\sum_{i=1}^{L} s_i s_{i+1}$.
- **Transverse field term** ($-\Gamma \sum \sigma_i^x$): This term is off-diagonal. When σ_i^x acts on a state $|n\rangle$, it flips the i-th spin, transforming $|n\rangle$ into a new state $|n'\rangle$. The matrix element $\langle n'| (-\Gamma\sigma_i^x) |n\rangle$ will be $-\Gamma$. All other matrix elements for this term are zero.

By systematically calculating these matrix elements for all pairs of basis states, we can construct the full $2^L \times 2^L$ Hamiltonian matrix. Once constructed, standard numerical linear algebra routines (e.g., eigenvalue solvers) can be used to find the eigenvalues (energies) and eigenvectors (wavefunctions) of the system.

Exercise 1.4.1: Diagonalization of TFIM

Consider the transverse field Ising model with $L = 6$ sites, $J_{ij} = J = 1$, and $\Gamma = 1$.

1. Write the algorithm to construct the $2^L \times 2^L$ Hamiltonian matrix. *Note: you need a 'converter' from index to physical configuration.*
2. Diagonalize it to find the ground state energy per spin.
3. Plot the magnetization as a function of Γ and try to perform a size scaling analysis.

Exercise 1.4.2: Diagonalization of Heisemberg model

Consider the Heisenberg model with $L = 6$ sites, $J_{ij} = J = 1$.

1. Write the algorithm to construct the $2^L \times 2^L$ Hamiltonian matrix.
2. Diagonalize it to find the ground state energy per spin.
3. Re-do the code implementing diagonalization only in a fixed subspace, i.e total magnetization $= 0$. This will select only configuration with a given number of up(down) spins, so you will be able to reach larger systems.

Exercise 1.4.3: Diagonalization of H_2 model

Consider the second quantized hamiltonian for H_2, of Eq. 1.42.

1. Construct the 4×4 Hamiltonian matrix.
2. Diagonalize it to find the ground state in the truncated basis.
3. Vary the parameters t, U, V, and observe a qualitative change in the ground state.

1.4.2 Discretization of Continuous Space Systems

Before we used a spacial orbital basis set to tackle a continuos space model. However we can use another type of basis, directly the (discretized) positions. For particles in continuous space, the Hamiltonian involves differential operators (like the kinetic energy operator, $-\frac{1}{2}\nabla^2$). To solve these problems numerically, we may discretize the continuous space into a finite mesh of points. This transforms the differential equation into a matrix equation, allowing for numerical solution.

Let's consider the one-dimensional quantum harmonic oscillator as a simple example (this is a one-body quantum problem). In atomic units, its Hamiltonian is:

$$\hat{H} = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2 \tag{1.47}$$

To solve this numerically, we can discretize the spatial coordinate x. We define a finite window $[-L, L]$ and place M equally spaced grid points x_j within this window:

$$x_j = -L + (j-1)\Delta x, \quad \text{for } j = 1, \ldots, M$$

where $\Delta x = \frac{2L}{M-1}$ is the grid spacing.

A wavefunction $\psi(x)$ can then be represented by a vector of its values at the grid points: $\Psi = (\psi(x_1), \psi(x_2), \ldots, \psi(x_M))^T$. The potential energy term, $\frac{1}{2}x^2$, becomes a diagonal matrix in this representation: $V_{jk} = \frac{1}{2}x_j^2 \delta_{jk}$

The kinetic energy operator, $-\frac{1}{2}\frac{d^2}{dx^2}$, is more challenging. We can approximate the second derivative using finite differences. The central finite difference approximation for the second derivative at point x_j is:

$$\frac{d^2\psi}{dx^2}\Big|_{x_j} \approx \frac{\psi(x_{j+1}) - 2\psi(x_j) + \psi(x_{j-1})}{(\Delta x)^2}$$

This approximation leads to a tridiagonal matrix representation for the kinetic energy operator, often referred to as the Discrete Variable Representation (DVR) for the kinetic energy in a real-space grid:

$$T_{jk} = -\frac{1}{2(\Delta x)^2} \begin{cases} -2 & \text{if } j = k \\ 1 & \text{if } j = k \pm 1 \\ 0 & \text{otherwise} \end{cases}$$

The full Hamiltonian matrix is then the sum of the kinetic and potential energy matrices: $H_{jk} = T_{jk} + V_{jk}$. This $M \times M$ matrix can then be diagonalized to find the approximate energy eigenvalues and eigenvectors (which represent the wavefunction values on the grid). The accuracy of this method depends on the choice of M (number of grid points) and L (size of the spatial window). Increasing M and L generally improves accuracy but also increases the computational cost.

Now the problem comes when we consider N particles. If we consider a discrete lattice (per dimension) consisting of M discretized positions, then the Hilbert space becomes M^N dimensional. The lattice regularized continuos space provides an intuitive bridge between continuous space and lattice models (or second quantized chemistry, where a 'site' becomes a 'spin-orbital'.

Exercise 1.4.4: Lattice regularized model

Consider the harmonic oscillator above, Eq. 1.47.

1. Set reasonable L and M parameters.
2. Write the algorithm to construct the DVR Hamiltonian matrix.
3. Diagonalize it to find the ground state energy; compare it with the exact value 0.5.
4. Plot the calculated ground state (ground state vector) and compare with the analytical solution (pay attention to the normalization and the definition of the grid).
5. Perform a convergence analysis varying M.

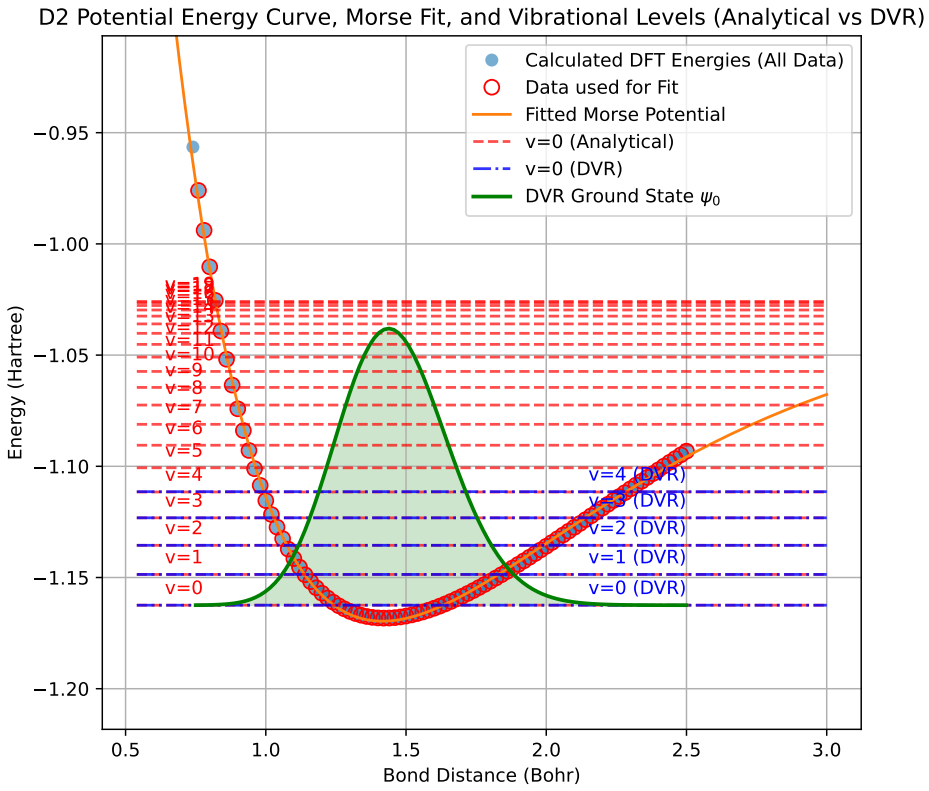


Figure 1.1: Potential energy surface of D_2 calculated from DFT with PBE functional and fit using a Morse potential in the range $[0.75 - 2.5]$ Bohr. Plotted are the energy levels calculated using analytical formulas (available for the Morse potentials) and the numerical DVR method (using the *sinc* version of the kinetic operator, see AppendixA). The ground state energy is -1.162451 Ha. The first three "vibrational" energies are 0.007194 Ha, 0.021027 Ha, and 0.034122 Ha. The reduced mass of the system is 1837.4711 atomic units (electron mass), since the mass of D is 2.016 a.m.u.

Exercise 1.4.5: Morse potential and zero point motion

Consider the (shifted) Morse potential, defined as

$$V(r) = D_e \left(1 - e^{-a(r-r_e)}\right)^2 + (E_{\text{diss}} - D_e)\,, \tag{1.48}$$

with parameters $D_e = 0.143783$ Ha, $a = 1.164822$ Bohr^{-1}, $r_e = 1.415847$ Bohr, $E_{\text{diss}} - 1.025862$ Ha. These parameters are fitted against DFT-PBE data, for the hydrogen molecule, in the interval $[0.75 - 2.5]$ Bohr so they are appropriate to fit well the bottom of the well (see Fig. 1.1), but yield a qualitatively wrong dissociation energy

1. Set reasonable L and M parameters.
2. Write the algorithm to construct the DVR Hamiltonian matrix (note that you need to use explicitely the reduced mass μ of the molecule).
3. Diagonalize it to find the ground state energy per spin. Compare this value with the exact value for the model, -1.162451 Ha, and with spectroscopic measurement (vibrational energy).
4. Plot the calculated ground state (ground state vector).
5. Perform a convergence analysis varying M.
6. Use the sinc version to discretize the kinetic operator, see Appendix A and test the convergence in this case.

1.5 The Power Method for Ground State Projection

An efficient and conceptually simple technique to extract the ground state of a quantum system is the *power method*. Given a Hermitian Hamiltonian H acting on a finite-dimensional Hilbert space and an initial trial state $|\psi_0\rangle$ that has non-zero overlap with the ground state, the method relies on repeated applications of the Hamiltonian to filter out higher-energy components.

The basic idea is to apply powers of a shifted Hamiltonian $(\Lambda - H)^n$ to $|\psi_0\rangle$, where Λ is a constant chosen such that $C > E_{\max}$, the largest eigenvalue of H. The sequence

$$|\psi_n\rangle = (\Lambda - H)^n |\psi_0\rangle \tag{1.49}$$

converges to the ground state $\left|\psi^{(GS)}\right\rangle$ as $n \to \infty$, up to normalization:

$$\lim_{n\to\infty} \frac{(\Lambda - H)^n |\psi_0\rangle}{\|(C - H)^n |\psi_0\rangle\|} = \left|\psi^{(GS)}\right\rangle . \tag{1.50}$$

To show this, let H be an Hamiltonian with eigenstates $\{|\phi_n\rangle\}$ and corresponding eigenvalues E_n, ordered as $E_0 < E_1 \leq E_2 \leq \cdots$ (notice that we assume a non-degenerate ground state). Any initial state $|\psi_0\rangle$ can be expanded in this basis as:

$$|\psi_0\rangle = \sum_n c_i |\phi_i\rangle\,, \quad \text{with } c_i = \langle\phi_i|\psi_0\rangle\,. \tag{1.51}$$

The action of the shifted hamiltonian $(\Lambda - H)$ on $|\psi_0\rangle$ is:

$$|\psi_n\rangle = \sum_i c_i (\Lambda - E_i)^n |\phi_i\rangle . \tag{1.52}$$

Since $\Lambda - E_0 > \Lambda - E_1 > \cdots$, the factor $(\Lambda - E_0)^n$ dominates as $n \to \infty$, and the contributions from higher-energy components decay exponentially faster:

$$|\psi_n\rangle \approx c_0 (\Lambda - E_0)^n |\phi_0\rangle + (\text{small corrections}). \tag{1.53}$$

After normalization, the resulting state converges to the ground state:

$$\lim_{n\to\infty} \frac{|\psi_n\rangle}{\| |\psi_n\rangle \|} = |\phi_0\rangle , \quad \text{if } c_0 \neq 0. \tag{1.54}$$

This shows that the power method together with the re-normalization acts as a *filter*, exponentially suppressing components of $|\psi_0\rangle$ corresponding to higher-energy eigenstates. Convergence is guaranteed provided the initial state has non-zero overlap with the ground state. This method is analogous to a projection operator in imaginary time:

$$\lim_{\tau\to\infty} e^{-\tau H} |\psi_0\rangle \propto \left|\psi^{(GS)}\right\rangle ,$$

but is simpler to implement as it avoids the exponential and instead uses matrix-vector multiplications.

In practice, the convergence rate depends on the spectral gap between the ground state and the first excited state. Moreover, normalization is necessary at each step to avoid numerical overflow. Notice that, this does not solve any of the exponential memory challenges of exact diagonalization, since the full matrix needs to be constructed. Moreover, since H is usually a *sparse* matrix, if one start from a initial sparse vector $|\psi_0\rangle$ the first iterations $H|\psi_n\rangle$ only involves a number of operations that scales with L, namely one could not even write the full matrix H, and just keep in memory the vector $|\psi_n\rangle$, spawning new components x' which are connected to the non-zero elements, x, of $|\psi_{n-1}\rangle$ by the matrix elements $\langle x'|H|x\rangle$.

The Lanczos method (not featured in these notes) cleverly uses the locality of the Hamiltonian to find the ground state in a more efficient way.

Exercise 1.5.1: Power method on the TFIM

Consider the same Ising and Heisemberg Hamiltonians solved with exact diagonalization.

1. Initialize the starting state ψ_0 with random components.
2. Apply to hamiltonian H as a matrix-vector operation and renormalize the output state.
3. Repeat until convergence, plotting the energy vs projection step.

Exercise 1.5.2: Smarter Power method on the TFIM

Re-do the projection above without storing the full matrix table.

1. Initialize the starting state with a *sparse* vector ψ_0 of random components.
2. Find the updated vector filling in / updating the coefficients relative to the basis state generated by the application of H.
3. Repeat until convergence, plotting the energy vs projection step.

1.6 The Variational Principle

The variational principle is a powerful method to approximate the ground state energy of a quantum system. Let us consider a Hamiltonian H with eigenstates $\{|\psi_n\rangle\}$ and corresponding eigenvalues E_n, such that

$$H|\psi_n\rangle = E_n|\psi_n\rangle, \quad E_0 < E_1 \leq E_2 \leq \cdots . \tag{1.55}$$

Let $\left|\psi^{(GS)}\right\rangle = |\psi_0\rangle$ denote the exact ground state with energy E_0, and let $|\psi_T\rangle$ be a normalized *trial* wavefunction, $\langle\psi_T|\psi_T\rangle = 1$. We define the variational energy as:

$$E_T = \langle\psi_T| H |\psi_T\rangle . \tag{1.56}$$

Expanding the trial wavefunction in the eigenbasis of H:

$$|\psi_T\rangle = \sum_n a_n |\psi_n\rangle, \quad \sum_n |a_n|^2 = 1, \tag{1.57}$$

we find:

$$E_T = \sum_n |a_n|^2 E_n \geq E_0 \tag{1.58}$$

Thus, the variational energy is always greater than the exact ground state, and becomes equal only when the trial state is the ground state.
We now analyze how close the trial wavefunction is to the true ground state in terms of its energy error. Let us define the error as

$$\epsilon = E_T - E_0 = \sum_n |a_n|^2 (E_n - E_0). \tag{1.59}$$

We separate the ground state contribution ($n = 0$) from the rest:

$$\epsilon = \sum_{n\neq 0} |a_n|^2 (E_n - E_0). \tag{1.60}$$

Let us introduce the *gap*, $\Delta = E_1 - E_0$, assuming a finite system where $\Delta > 0$. Since

$E_n \geq E_0 + \Delta$ for $n \geq 1$, we find $\epsilon \geq \Delta \sum_{n\neq 0} |a_n|^2$.
Define

$$\eta = 1 - |a_0|^2 = \sum_{n\neq 0} |a_n|^2, \tag{1.61}$$

which quantifies the deviation of $|\psi_T\rangle$ from the true ground state. Then we obtain:

$$\eta \leq \frac{\epsilon}{\Delta}. \tag{1.62}$$

This inequality tells us that in order to have an accurate approximation of the ground state, meaning $\eta = 1 - |a_0|^2 \ll 1$, it is needed that the energy error ϵ is much smaller than the gap Δ. That is:

$$\epsilon \ll \Delta \quad \Rightarrow \quad |\psi_T\rangle \approx \left|\psi^{(GS)}\right\rangle. \tag{1.63}$$

Unfortuntately, the gap is closing with N for most interesting systems, thus making this approach also harder and harder. We present some examples of trial wavefunctions in the next Sections.

1.6.1 Example: The Jastrow Wavefunction for Quantum Spin Systems

We start with the simplest case: quantum spin systems. Here, the Jastrow ansatz is a remarkably simple yet effective variational wavefunction that can incorporate correlations beyond the mean-field level.

In its general form, the Jastrow wavefunction assigns an amplitude to each spin configuration $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$:

$$\psi(\boldsymbol{\sigma}) = \exp\left(\sum_{i<j} W_{ij}\sigma_i\sigma_j\right). \tag{1.64}$$

where $\sigma_i = \pm 1$ are the eigenvalues of σ_i^z, and W_{ij} are symmetric variational parameters encoding the interaction strength between spin i and spin j. This form, depending on the sign and strength of the parameters Ws is able to penalize components in the wavefunction where spins are parallel(antiparallel). For instance, if the Hamiltonian, present a dominant ferromagnetic couploing $J >> \Gamma$, then the ground state is dominated by the configurations $\uparrow\uparrow \cdots \uparrow$ and $\downarrow\downarrow \cdots \downarrow$ and fluctuations around them characherized by few domain walls. Configurations like $\uparrow\downarrow\uparrow \cdots \uparrow$ are expected to have a vanishing amplitude. The Jastrow form can efficiently capture the dominant longitudinal (S_i^z–S_j^z) correlations present in the ground state.

1.6.2 Example: Trial Wavefunctions for H_2 in a Minimal Basis

We use the example of Sect. 1.2.1 The simplest approximation, usually done by chemists, is to construct delocalized bonding and antibonding orbitals:

$$\psi_b = \frac{1}{\sqrt{2}}(\phi_1 + \phi_2), \tag{1.65}$$

$$\psi_a = \frac{1}{\sqrt{2}}(\phi_1 - \phi_2), \tag{1.66}$$

where we avoided to use the traditional overlap factor $S = \langle\phi_1|\phi_2\rangle$ because we can assumed that the ϕ_i atomic orbitals were already ortogonalized.
We then define creation operators for these molecular orbitals:

$$d^\dagger_{b,\sigma} = \frac{1}{\sqrt{2}}\left(c^\dagger_{1,\sigma} + c^\dagger_{2,\sigma}\right), \tag{1.67}$$

$$d^\dagger_{a,\sigma} = \frac{1}{\sqrt{2}}\left(c^\dagger_{1,\sigma} - c^\dagger_{2,\sigma}\right). \tag{1.68}$$

The molecular orbital (MO) ansatz for the ground state is given by placing both electrons in the bonding orbital:

$$\left|\Psi^{(b)}_{\mathrm{MO}}\right\rangle = d^\dagger_{b,\uparrow} d^\dagger_{b,\downarrow} |\varnothing\rangle . \tag{1.69}$$

The spacial part of the wf is given by a Slater determinant with both electrons occupying the bonding MO (spin-singlet state):

$$\Psi^{(b)}_{\mathrm{MO}}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}}\left[\psi_b(\mathbf{r}_1)\psi_b(\mathbf{r}_2)\right], \tag{1.70}$$

While this is a reasonable trial state at short bond lengths, it becomes a poor approximation in the large separation limit $R \to \infty$. To see this, we expand the wavefunction back in terms of the atomic orbitals:

$$\left|\Psi^{(b)}_{\mathrm{MO}}\right\rangle \propto \left(c^\dagger_{1,\uparrow} + c^\dagger_{2,\uparrow}\right)\left(c^\dagger_{1,\downarrow} + c^\dagger_{2,\downarrow}\right)|\varnothing\rangle . \tag{1.71}$$

Expanding this product gives four terms:

$$\left|\Psi^{(b)}_{\mathrm{MO}}\right\rangle \propto c^\dagger_{1,\uparrow}c^\dagger_{1,\downarrow}|\varnothing\rangle + c^\dagger_{1,\uparrow}c^\dagger_{2,\downarrow}|\varnothing\rangle + c^\dagger_{2,\uparrow}c^\dagger_{1,\downarrow}|\varnothing\rangle + c^\dagger_{2,\uparrow}c^\dagger_{2,\downarrow}|\varnothing\rangle . \tag{1.72}$$

At large R, the physical ground state should correspond to one electron localized on each atom. However, the first and last terms above correspond to both electrons being on the same atom. It is physically wrong that all four configurations equally contribute at large distance. Therefore, the single SD ansatz fails to describe the correct physics in the dissociation limit.

A better approximation at large bond distance is the Heitler–London (HL) ansatz, which enforces that each electron remains on a different atom:

$$|\Psi_{\mathrm{HL}}\rangle = \frac{1}{\sqrt{2}}\left(c^\dagger_{1,\uparrow}c^\dagger_{2,\downarrow} + c^\dagger_{2,\uparrow}c^\dagger_{1,\downarrow}\right)|\varnothing\rangle . \tag{1.73}$$

This state avoids unphysical double occupancies and correctly reduces to two isolated hydrogen atoms when $R \to \infty$. In terms of our previous basis, this is $|\Psi_{\mathrm{HL}}\rangle = \frac{1}{\sqrt{2}}\left(|2\rangle + |3\rangle\right)$, where $|2\rangle, |3\rangle$ are defined in Sect 1.2.1.
Interestingly, the HL ansatz can be rewritten using molecular orbitals. Define the Slater determinant with both electrons in the antibonding orbital:

$$\left|\Psi^{(a)}_{\mathrm{MO}}\right\rangle = d^\dagger_{a,\uparrow} d^\dagger_{a,\downarrow} |\varnothing\rangle . \tag{1.74}$$

Then the HL wavefunction can be obtained (up to normalization) as:

$$|\Psi_{\mathrm{HL}}\rangle \propto \left|\Psi^{(b)}_{\mathrm{MO}}\right\rangle - \left|\Psi^{(a)}_{\mathrm{MO}}\right\rangle . \tag{1.75}$$

In general, the exact ground state of the H_2 molecule in this minimal basis is a linear combination of the bonding and antibonding determinants:

$$|\Psi_{\mathrm{exact}}\rangle = \alpha \left|\Psi^{(b)}_{\mathrm{MO}}\right\rangle + \beta \left|\Psi^{(a)}_{\mathrm{MO}}\right\rangle , \tag{1.76}$$

where the coefficients α, β are determined by minimizing the total energy. This combination interpolates between the two regimes.

Eq. 1.76 define a possible *trial state* in the basis of SDs. This is the language of chemists, where wfs are expanded as linear combination of SDs, and they have several methods to calculate these coefficients. Another way to expand the variational freedom is to use a larger set of starting atomic orbitals, $\phi_1, \phi_2, \phi_3, \cdots$ which can be centered on the same atoms, but having different 'shapes', i.e. $1s, 2s, 2p, ...$ in the language of atom center orbitals.

Alternatively, one could construct a variational wavefunction, without resorting to MOs, trying to reduce the amplitude of the configurations where two electrons sit at the same ions, i.e. something like this:

$$|\psi_T(\lambda)\rangle \propto \lambda c^\dagger_{1,\uparrow} c^\dagger_{1,\downarrow} |\varnothing\rangle + c^\dagger_{1,\uparrow} c^\dagger_{2,\downarrow} |\varnothing\rangle + c^\dagger_{2,\uparrow} c^\dagger_{1,\downarrow} |\varnothing\rangle + \lambda c^\dagger_{2,\uparrow} c^\dagger_{2,\downarrow} |\varnothing\rangle . \tag{1.77}$$

This kind of wavefunction is traditionally used in lattice models, see below.

1.6.3 Example: A Trial Wavefunction for the Hubbard Model

The simplest Hamiltonian in which electron-electron correlations play an essential role is the one-dimensional Hubbard model:

$$H = -t \sum_{\langle i,j\rangle,\sigma} \left(c^\dagger_{i\sigma} c_{j\sigma} + \mathrm{H.c.} \right) + U \sum_i n_{i\uparrow} n_{i\downarrow}, \tag{1.78}$$

where $c^\dagger_{i\sigma}$ ($c_{i\sigma}$) creates (annihilates) an electron with spin σ on site i, and $n_{i\sigma} = c^\dagger_{i\sigma} c_{i\sigma}$ is the corresponding number operator. The sum $\langle i, j\rangle$ runs over nearest-neighbor sites in a 1D lattice of length L, with (possibly) periodic boundary conditions. Physically one could see this arising from a chain of H atoms (where only one atomic orbital is in play), a sufficiently distant positions (see Sect. 1.2.1).

In this case, we would be at the particularly important regime of *half-filling*, where the number of electrons N equals the number of sites L. At $U = 0$, the electrons behave as a free Fermi gas and delocalize across the lattice. In the opposite limit $t = 0$, electrons are completely localized to minimize the interaction energy, and the ground state is a called 'Mott insulator'. Because of the fundamentally different nature of the ground states in the two limits $U = 0$ and $t = 0$, a *metal-insulator transition* is expected at intermediate coupling. [4]

[4] This is true in higher dimensions where we have a finite critical value $(U/t)_c$. In one dimension, the model is exactly solvable via the Bethe Ansatz, and the ground state is shown to be insulating for any $U > 0$.

We define an electron configuration $|x\rangle$ as a Fock state where all electron positions and spin projections are fixed. For example:

$$|x\rangle = |\uparrow, \uparrow\downarrow, 0, 0, \downarrow, \dots\rangle = c^\dagger_{1,\uparrow} c^\dagger_{2,\uparrow} c^\dagger_{2,\downarrow} c^\dagger_{5,\downarrow} \cdots |\varnothing\rangle \,, \tag{1.79}$$

where each site can be empty (0), singly occupied ($\uparrow$ or $\downarrow$), or doubly occupied ($\uparrow\downarrow$). Each $|x\rangle$ is a Slater determinant in position-spin space.
It is useful to define the number of doubly occupied sites in a configuration as:

$$D = \sum_i n_{i\uparrow} n_{i\downarrow}. \tag{1.80}$$

At $U = 0$, the ground state $|\Psi_0\rangle$ is a Fermi sea of non-interacting electrons:

$$|\Psi_0\rangle = \prod_{k \le k_F,\, \sigma} c^\dagger_{k,\sigma} |0\rangle = \sum_x \langle x|\Psi_0\rangle \, |x\rangle \,, \tag{1.81}$$

where the Fourier-transformed operators $c^\dagger_{k,\sigma}$ create momentum eigenstates, and the last equality expresses the state in the configuration basis.

However, this uncorrelated wavefunction is a poor approximation for $U \gg t$. In this limit, the average density of doubly occupied sites is $\langle D\rangle \approx 1/4$, which is energetically unfavorable when U is large. If one believe that, in the exact ground state for $U \neq 0$, these configurations are suppressed, then one can come up with a simple and powerful variational ansatz that captures the physics of a Mott insulator: the *Gutzwiller wavefunction*,

$$|\Psi_g\rangle = e^{-gD} |\Psi_0\rangle = \sum_x e^{-gD(x)} \langle x|\Psi_0\rangle \, |x\rangle \,, \tag{1.82}$$

where $g \ge 0$ is a variational parameter, and $D(x) = \langle x|D|x\rangle$ counts the number of doubly occupied sites in configuration $|x\rangle$.

This wavefunction penalizes configurations with large $D(x)$, so incorporating local correlations. In the limit $g \to \infty$, all configurations with $D > 0$ are completely projected out, leading to a variational state entirely composed of singly occupied sites, the correct qualitative behavior of a Mott insulator.

To summarize:

- At $U = 0$, the uncorrelated Fermi sea $|\Psi_0\rangle$ is exact and consists of a single determinant.
- At $U \gg t$, the Gutzwiller-projected wavefunction $|\Psi_g\rangle$ becomes a better approximation.
- The variational parameter g can be optimized to minimize the total energy $E_g = \langle\Psi_g| H |\Psi_g\rangle \,/\, \langle\Psi_g|\Psi_g\rangle$.

1.6.4 Example: Jastrow Wavefunction for Liquid Helium

Finally we consider an example also for continuos space models. Indeed, the Jastrow ansatz was originally developed to describe the ground-state properties of strongly interacting quantum fluids, with one of the most important applications being liquid Helium. The two stable

isotopes, ^{4}He and ^{3}He, exhibit markedly different quantum behavior due to their respective bosonic and fermionic statistics. Despite these differences, both can be effectively modeled using a many-body wavefunction incorporating Jastrow-type correlations.

At low temperatures, ^{4}He remains in a liquid state due to the dominance of zero-point motion over the weak van der Waals attraction between atoms. It exhibits superfluidity and is a paradigmatic example of a strongly correlated bosonic quantum fluid. On the other hand, ^{3}He, being fermionic, does not exhibit superfluidity unless cooled to much lower temperatures, and its ground state is more complex due to the antisymmetric nature of the wavefunction.

The central idea behind the Jastrow ansatz is to model the effects of two-body correlations induced by the interatomic repulsion at short distances and weak attraction at larger distances. The many-body wavefunction is written as:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \prod_{i<j} f(r_{ij}), \tag{1.83}$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between atoms i and j, and $f(r)$ is a correlation function to be optimized or parametrized.

In the case of ^{4}He, where the particles are bosons, this Jastrow form is fully symmetric under particle exchange, as required. A simple choice for the correlation function is:

$$f(r) = \exp\left[-\frac{1}{2}\left(\frac{b}{r}\right)^5\right], \tag{1.84}$$

which reflects the strong repulsion at short range[5]. This choice suppresses the amplitude of the wavefunction when two atoms are too close together, mimicking the effects of the hard-core interaction. In principle, one could also add multi-body correlations.

For ^{3}He, the situation is more involved due to the fermionic nature of the atoms. In this case, the total wavefunction must be antisymmetric under particle exchange. A common strategy is to combine the Jastrow factor with a Slater determinant built from single-particle orbitals:

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \prod_{i<j} f(r_{ij}) \cdot \det[\phi_k(\mathbf{r}_l)], \tag{1.85}$$

where $\{\phi_k\}$ are suitable single-particle orbitals (e.g., plane waves in a box or Bloch states in a periodic lattice). The Jastrow factor enhances the physical realism by introducing correlations beyond the mean-field Slater determinant description, while the determinant ensures the correct fermionic antisymmetry. In both cases, the Jastrow wavefunction captures essential features of the ground state, especially the avoidance of short-range overlaps due to strong repulsion.

1.7 The Plan for the Course

Now that we have introduce several ingredients we are in the position to set the key questions and problems we are going to tackle in the class:

[5]The exponent 5 is not casual. It stems from satisfying cusp conditions, see next Chapters.

- How do we compute in practice $E_T = \langle \psi_T | H | \psi_T \rangle / \langle \psi_T | \psi_T \rangle$?
- How do we optimize the energy?
- Can we go beyond the variatonal approach?
- Can we go beyond ground states?
- What models can be efficiently simulated?

Chapter 2

Classical Monte Carlo Sampling

Monte Carlo (MC) methods play a crucial role in computational physics, but also in technology, finance and so on. Learning MC opens therefore a lot of "doors" for several, seemingly disconnected areas. MC is essential for evaluating high-dimensional integrals. In our context, these arise when dealing with many-body wavefunctions that are not simple Slater determinants. In this chapter, we introduce the mathematical foundation of Monte Carlo, explain how random sampling techniques are used to compute physical observables, and explore their application in both classical and quantum systems. We also discuss practical implementation considerations and statistical tools for estimating uncertainty.

One of the primary motivations for using Monte Carlo techniques arises in quantum systems involving many interacting particles. For example, the Hartree-Fock approximation reduces the computation of the total energy of N electrons to the evaluation of one- and two-body integrals. This simplification brings down the dimensionality of the problem from $3N$ to at most six dimensions, making it computationally tractable.

However, when we introduce electron correlation effects through more accurate wavefunctions e.g., by multiplying the Slater determinant with a Jastrow factor,this simplification is lost. The energy expectation value:

$$E_T = \frac{\langle \psi_T | \, H \, | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle} \tag{2.1}$$

requires integration over all $3N$ coordinates, which is computationally infeasible using deterministic methods for large N. Monte Carlo methods provide a way to handle these high-dimensional integrals efficiently by sampling configurations probabilistically instead of deterministically computing integrals over all variables.

2.1 Probability and Central Limit Theorem

Here we give a brief reminder of some concept of probability theory. A random variable x is the "value" of an outcome (could be an experiment). This is not deterministic but occur with a give probability. For instance, the event i, produce and outcome x_i and occurs with probability p_i. The probabilities of the event are normalized

$$\sum_i p_i = 1. \tag{2.2}$$

For discrete random variables x with probabilities p_i, the mean and variance are:

$$\langle x \rangle = \sum_i p_i x_i, \tag{2.3}$$

$$\mathrm{var}(x) = \sum_i p_i (x_i - \langle x \rangle)^2. \tag{2.4}$$

The standard deviation $\sigma = \sqrt{\mathrm{var}(x)}$ provides a measure of uncertainty in x. In the continuous case, using a probability density $\rho(x)$:

$$\langle x \rangle = \int dx\, \rho(x) x, \tag{2.5}$$

$$\mathrm{var}(x) = \int dx\, \rho(x)(x - \langle x \rangle)^2. \tag{2.6}$$

In continuos case the normalization reads: $\int dx\, \rho(x) = 1$.

An essential result for statistical estimation is the Central Limit Theorem. Let $x_1, \ldots, x_N$ be i.i.d. random variables with mean $\langle x \rangle$ and variance σ^2 . The sample mean:

$$\bar{x}_N = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{2.7}$$

approaches a Gaussian distribution as $N \to \infty$:

$$\bar{x}_N \sim \mathcal{N}\left(\langle x \rangle, \frac{\sigma^2}{N} \right). \tag{2.8}$$

This implies that the statistical error on the mean decreases as $1/\sqrt{N}$, which justifies using sampling to estimate expectations.

2.2 Markov Chains

A Markov chain is a stochastic sequence x_n generated by:

$$x_{n+1} = F(x_n, \xi_n), \tag{2.9}$$

where ξ_n are i.i.d. random variables and F is a deterministic function. The future depends only on the present, not on the full past, but in a probabilistic way. We aim to sample from a target distribution $\pi(x)$. We can think that the randoma variable are distributed accordingly to an initial probability density $\rho_0(x)$

The evolution of the marginal distribution $\rho_n(x)$ is governed by the Master equation:

$$\rho_{n+1}(x') = \sum_x K(x'|x) \rho_n(x), \tag{2.10}$$

where $K(x'|x)$ is the transition probability.[1] This gives the probability to transition from state x to x', and its rows are normalized to 1:

$$\sum_{x'} K(x'|x) = 1. \tag{2.11}$$

[1]Sometimes we will also use the notation $K(x, x'), K(x \to x')$, or $K_{x,x'}$ (notice the reverse order x, x' compared to the notation $(x'|x)$).

So the probability of being at state x' is given by the convolution between the transition probability and the probability density at the previous step.
Markov chains can also be depicted visually in the form of a graph, with the state space $\mathcal{S}$ drawn as a collection of nodes and the edges representing the non-zero transition probabilities K_{ij}.

As an example, suppose that our space is given by nodes A, B, C, and D. and the transition probability is

$$K = \begin{pmatrix} 0.6 & 0.2 & 0.1 & 0.1 \\ 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.5 \\ 0.5 & 0.2 & 0.2 & 0.1 \end{pmatrix} \tag{2.12}$$

You can visualize the transitions using the following diagram:

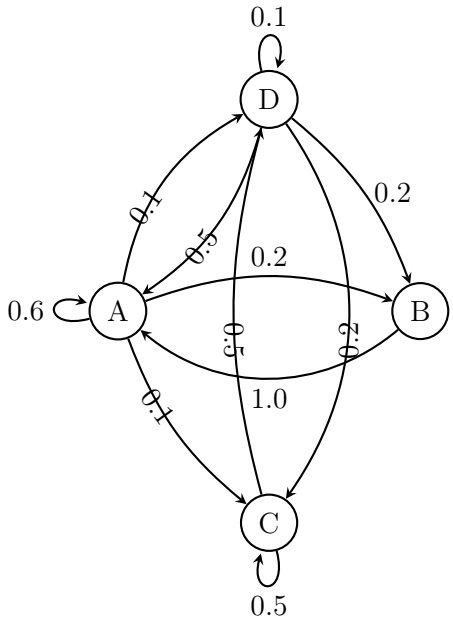


Figure 2.1: Transition graph for the example Markov chain.

In Markov Chain MC one typically starts from a definite state, e.g A at $n = 0$. For this pedagogical example we use explicitely write the probability density as vectors. Then:

$$\vec{\rho}_0 = (1, 0, 0, 0)$$

At $n = 1$, the distribution becomes:

$$\vec{\rho}_1^T = \vec{\rho}_0^T K = (0.6, 0.2, 0.1, 0.1)$$

Notice that we are using this 'strange' notation for matrix-vector multiplication, where we

apply the matrix from the right to a row vector,[2] If sampling from this vector results in state D, this becomes the current state. Repeating this generates a trajectory, for example:

$$A - D - A - C - D - A - B - A - C - ...$$

To calculate the probability distribution at $n = 2$, we use:

$$\vec{\rho}_2^T = \vec{\rho}_1^T K,$$

recursively. From a trajectory perspective, one needs to evaluate all the 'paths' that connect the inital state A from each of the final states x', given that there is an intermediate step. For example, the probability of being in state D at $n = 2$ is:

$$\begin{aligned}
\Pr(X_2 = D|X_0 = A) &= \sum_{j=1}^{4} \Pr(X_1 = j|X_0 = A) \cdot \Pr(X_2 = D|X_1 = j) \\
&= \sum_{j=1}^{4} K_{1j} \cdot K_{j4} \\
&= K_{11}K_{14} + K_{12}K_{24} + K_{13}K_{34} + K_{14}K_{44} \\
&= 0.6 \cdot 0.1 + 0.2 \cdot 0 + 0.1 \cdot 0.5 + 0.1 \cdot 0.1 \\
&= 0.06 + 0 + 0.05 + 0.01 = 0.12
\end{aligned}$$

From a linear algebra persepctive, we evolve distributions using repeated application or matrix-vectors operations:

$$\vec{\rho}_{n+1}^T = \vec{\rho}_n^T K \quad \text{and} \quad \vec{\rho}_{n+k}^T = \vec{\rho}_n^T K^k$$

A stationary distribution $\vec{\pi}$ satisfies:

$$\vec{\pi}^T = \vec{\pi}^T K, \quad \text{or} \quad \sum_i \pi_i K_{ij} = \pi_j \tag{2.13}$$

which implies it remains unchanged after a sufficent number of iterations. Notice therefore a fundamental distinction between the random variable x and its probability distribution $\rho(x)$: The evolution of the probability distribution is fully deterministic, while the dynamics of the variable x is not.

2.2.1 Example: Convergence to the Stationary Distribution

We now show numerically that the Markov chain defined above converges to its stationary distribution. Consider again the transition matrix of Eq. 2.12. Starting from:

$$\vec{\rho}_0 = (1, 0, 0, 0)$$

[2]The resulting vector $\vec{y}^T = \vec{x}^T K$ is given by

$$y_j = \sum_{i=1}^{N} x_i K_{ij}, \quad \text{for } j = 1, \ldots, N.$$

we compute:

$$\begin{aligned}\vec{\rho}_1^T &= (0.6, 0.2, 0.1, 0.1)\\ \vec{\rho}_2^T &= (0.68, 0.12, 0.09, 0.11)\\ \vec{\rho}_3^T &= (0.678, 0.128, 0.086, 0.108)\\ \vec{\rho}_4^T &= (0.6788, 0.1270, 0.0863, 0.1079)\end{aligned}$$

The distribution stabilizes after a few steps:

$$\vec{\pi} \approx (0.679, 0.127, 0.086, 0.108)$$

and satisfies approximately:

$$\vec{\pi}^T K \approx \vec{\pi}^T$$

thus confirming convergence to a stationary distribution.[3]

A natural question is asking under which conditions a K evolution leads to a stationary distribution $\bar{\rho}(x)$ such that:

$$\bar{\rho}(x') = \sum_x K(x'|x)\bar{\rho}(x). \tag{2.14}$$

The theory of Markov-Chains is pretty heavy, and we will leave theorems aside. We are going to rely on some physical intuitions. The reason why the above example led to a limiting distribution is because the Markov chain is

- *Irreducible*: It's possible to reach any state from any other state in a finite number of steps.
- *Aperiodic*: The system does not get trapped in cycles, i.e., the chain does not return to states only at fixed multiples of time steps.

These two properties ensure the chain is ergodic, which is enough to guarantee convergence to a stationary distribution. Notice that these are stricly valid for a finite space of dimension N.

In our example, the graph defined by K allows movement (possibly indirectly) between all states (irreducibility), and contains self-loops (i.e non-zero diagonal elements) breaking periodicity.
It is clear that, if K allows a limiting distribution π, then this is eigenvector with eigenvalue 1.

[3]To find the stationary distribution analytically, solve $\vec{\pi}^T K = \vec{\pi}^T$ with $\sum_i \pi_i = 1$. This leads to the system:

$$\begin{aligned}0.6\pi_1 + 1.0\pi_2 + 0.0\pi_3 + 0.5\pi_4 &= \pi_1\\ 0.2\pi_1 + 0.0\pi_2 + 0.0\pi_3 + 0.2\pi_4 &= \pi_2\\ 0.1\pi_1 + 0.0\pi_2 + 0.5\pi_3 + 0.2\pi_4 &= \pi_3\\ 0.1\pi_1 + 0.0\pi_2 + 0.5\pi_3 + 0.1\pi_4 &= \pi_4\\ \pi_1 + \pi_2 + \pi_3 + \pi_4 &= 1\end{aligned}$$

Solving this yields $\vec{\pi} \approx (0.679, 0.127, 0.086, 0.108)$.

2.2.2 Detail Balance

The example featured above is quite artificial, but there is a stronger criteria that allows one to assess the existence of a limiting distribution. A sufficient (but not necessary) condition is detailed balance:

$$K(x'|x)\pi(x) = K(x|x')\pi(x'). \tag{2.15}$$

This expresses a balance between forward and reverse transitions, analogous to equilibrium in physical systems. Detailed balance implies stationarity, and it can be shown by substitution into the Master equation.

Exercise 2.2.1: Implementation of Markov Chain trajectories

Write a code that implements the Markov chain, and check using independent measurements that you reach the stationary limiting distribution. You can simply output an histogram of visited states, since the space is just four-dimensional.

Exercise 2.2.2: Violation of Detail balance

Show that the example of Eq. 2.12, does not respect the detailed balance property.

2.2.3 Spectral Convergence and Autocorrelation Time

Let K be the transition matrix of the Markov process. Assume K is diagonalizable and let its eigenvalues be $1 = \lambda_0 > |\lambda_1| \geq |\lambda_2| \geq \cdots$. The right eigenvector corresponding to $\lambda_0 = 1$ is the stationary distribution $\pi(x)$.
Suppose the initial distribution can be expanded in terms of the eigenvectors v_i:

$$\rho_0 = \sum_i a_i v_i, \tag{2.16}$$

with $v_0 = \pi$ and $a_0 = 1$ (due to normalization). Then after n steps:

$$\rho_n = K^n \rho_0 = \sum_i a_i \lambda_i^n v_i. \tag{2.17}$$

As $n \to \infty$, all components with $|\lambda_i| < 1$ decay exponentially fast:

$$\rho_n \to a_0 v_0 = \pi. \tag{2.18}$$

The convergence rate is governed by the second-largest eigenvalue λ_1:

$$\|\rho_n - \pi\| \sim |\lambda_1|^n. \tag{2.19}$$

The gap $\delta = 1 - |\lambda_1|$ determines how quickly the Markov chain equilibrates. Notice the formal similarity with the Power Method.
The mixing time t_{mix}, defined as the time needed for the Markov chain to be ϵ-close to π, can be bounded as:

$$(\delta^{-1} - 1) \ln\left(\frac{1}{2\epsilon}\right) \leq t_{\text{mix}} \leq \delta^{-1} \ln\left(\frac{1}{\epsilon \min_s \pi(s)}\right). \tag{2.20}$$

Basically, this means that the mixing time scales as $1/\delta$.

Constructing K is infeasible in large systems. Instead, the autocorrelation time τ must be measured empirically. τ tells how many steps are needed before samples become effectively uncorrelated. For an observable f, its autocorrelation function decays as:

$$C_f(n) = \langle f(x_0)f(x_n)\rangle - \langle f\rangle^2 \sim e^{-n/\tau}. \tag{2.21}$$

Then the statistical error scales as:

$$\sigma^2_{\text{eff}} \approx 2\tau\frac{\sigma^2}{N}. \tag{2.22}$$

The equilibration time is the number of steps before observables stabilize; these early samples are discarded (burn-in).

2.3 The Metropolis Algorithm

The Metropolis algorithm constructs a Markov chain whose stationary distribution is a given target distribution $\pi(x)$. Notice that the working hypothesis is that we *known* what probability we need to sample from, e.g. $\pi(x) \propto e^{-\beta V(x)}$, but we don't know how to generate configurations sampled exactly as that.
The Metropolis Markov Chain is defined by a transition probability of the form:

$$K(x'|x) = T(x'|x)A(x'|x) \tag{2.23}$$

where:

- $T(x'|x)$ is the **proposal** or *trial* probability, giving the probability of proposing a move from x to x';
- $A(x'|x)$ is the **acceptance** probability, which decides whether the proposed move is accepted.

To ensure that $\pi(x)$ is the stationary distribution, $A(x'|x)$ is chosen to satisfy the *detailed balance condition*:

$$\pi(x)T(x'|x)A(x'|x) = \pi(x')T(x|x')A(x|x'). \tag{2.24}$$

A common choice that satisfies detailed balance is:

$$A(x'|x) = \min\left(1, \frac{\pi(x')T(x|x')}{\pi(x)T(x'|x)}\right) \tag{2.25}$$

If $T(x'|x)$ is symmetric (i.e., $T(x'|x) = T(x|x')$), then the acceptance probability simplifies to:

$$\boxed{A(x'|x) = \min\left(1, \frac{\pi(x')}{\pi(x)}\right)} \tag{2.26}$$

One of the most important aspect of the algorithm is that it does not require the *normalization* of the probability distribution, since only the ratio $\pi(x')/\pi(x)$ enters in the evaluation.
The pseudocode of the Metropolis algorithm with symmetric proposal is:

Input: Initial state $x = x_0$

For loop: $n < M_{\text{steps}}$

- Propose $x' \sim T(x'|x)$
- Compute acceptance ratio $r = \frac{\pi(x')}{\pi(x)}$
- Accept with probability $A = \min(1, r)$
  - if accepted: $x_{n+1} = x'$
  - else: $x_{n+1} = x$
- Compute functions of x_n, i.e. $f_n = f(x_n)$.

Note that $A(x|x)$ can be zero, but $K(x|x)$ remains finite because it includes the total probability of all rejected moves:

$$K(x|x) = 1 - \sum_{x' \neq x} T(x'|x)A(x'|x). \tag{2.27}$$

Operativelly, one compute the acceptance by draing a random uniform number z between 0 and 1, and check if $z < r$. The acceptance move allows us to move from a configuration with high probability toward ones less probable.

There is a tradeoff in choosing the trial distribution $T(x'|x)$. Proposing a state x' very close to x increases the acceptance probability, since $\pi(x') \approx \pi(x)$, but leads to small changes and thus high autocorrelation in the samples. Conversely, proposing distant moves lowers acceptance but may explore the space more rapidly if accepted.

A common empirical strategy is to tune the proposal distribution such that the acceptance rate is around 50%. Although the ideal scenario would be 100% acceptance and zero autocorrelation, in practice this is not achievable. When a move is rejected, the chain stays at x, which still contributes to averages. Rejections increase the number of times the chain remains in the same state, slowing exploration but maintaining correctness.

2.4 Error Estimation and Binning

Because MCMC samples are correlated, naive error estimates underestimate uncertainties. A standard method is binning analysis:

1. Divide data (originally an array of M_{steps} values) into M bins of size n_{bin}.
2. Compute bin averages $\bar{f}_i$.
3. Estimate variance:
$$\sigma^2 = \frac{1}{M-1} \sum_{i=1}^{M} (\bar{f}_i - \bar{f})^2. \tag{2.28}$$
4. The standard error on the average is $\sigma/\sqrt{M}$.

As n increases, correlations within bins vanish, and the estimate becomes reliable. In other words, we are treating the binned value as the independent samples, and there are M of them, not M_{steps}.

2.5 Importance Sampling

The concepts introduced so far find a natural and powerful application in the context of statistical physics. onsider a system of L classical spins, with configuration denoted by

$$x = (\sigma_1, \sigma_2, \dots, \sigma_L), \tag{2.29}$$

where each spin $\sigma_i \in \{-1, +1\}$. The energy of a configuration is given by a function $E(x)$, for example the Hamiltonian of the classical Ising model:

$$E(x) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_i \sigma_i, \tag{2.30}$$

where $\langle i, j \rangle$ denotes nearest-neighbor pairs, J is the interaction strength, and h is the external magnetic field.

Our goal is to compute thermal expectation values of observables $f(x)$ under the Boltzmann distribution:

$$\langle f \rangle = \frac{1}{Z} \sum_x f(x) e^{-\beta E(x)}, \tag{2.31}$$

where the partition function is

$$Z = \sum_x e^{-\beta E(x)}, \tag{2.32}$$

and $\beta = 1/(k_B T)$ is the inverse temperature. Notice that computing Eq. 2.32 is impossible in reasonably large systems, since it is a N-large summation, and the number of configurations are $N = 2^L$. On the constrary, computing the energy $E(x)$ of a configuration is efficient because it involves a number of operations that scales with L. We need to escape the direct summation way to compute both integrals.

A naive approach would be to sample configurations x uniformly and compute the average:

$$\langle f \rangle \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i) e^{-\beta E(x_i)}, \qquad x_i \sim 1/N \tag{2.33}$$

but this method is highly inefficient because most randomly chosen configurations contribute negligibly due to their large energy (if the temperature is low enough). If one is interested to predict the low-temperature phases, there is vanishing probability to pick randomly such configurations, e.g. $x = (1, 1, 1, \cdots)$. This leads to high variance and poor convergence.

A more efficient strategy is *importance sampling*: we sample configurations x according to the target distribution

$$\pi(x) = \frac{e^{-\beta E(x)}}{Z}, \tag{2.34}$$

and estimate the average as

$$\langle f \rangle \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i), \qquad x_i \sim \pi(x). \tag{2.35}$$

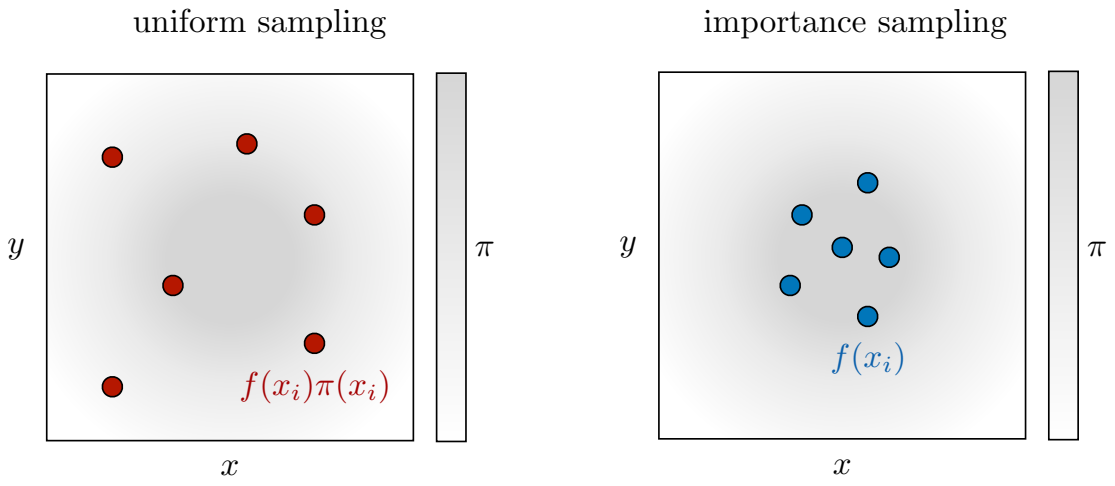


Figure 2.2: Left: we generate samples, x_i, uniformly in the space, each sample contributes to the sum, $\langle f \rangle$, with $f(x_i)e^{-\beta E(x_i)}/Z$; most of these contribution vanish due to the probability weight. Right: we generate samples distributed already as $\pi(x) = e^{-\beta E(x)}/Z$; now each sample contributes with $f(x_i)$.

Notice that now, the sum over x_i is understood as x_i distributed accordingly to $\pi(x)$. This drastically reduces the variance and improves convergence. This motivates our need to use Markov Chain to sample from a target $\pi(x)$.

The Metropolis algorithm is particularly well-suited for importance sampling in this context. It generates a Markov chain whose stationary distribution is precisely the Boltzmann distribution $\pi(x)$. Crucially, the algorithm does not require knowledge of the partition function Z, since it only needs the ratio of probabilities:

$$\frac{\pi(x')}{\pi(x)} = e^{-\beta(E(x')-E(x))}. \tag{2.36}$$

In the Metropolis algorithm, a common choice for the trial[4] distribution $T(x'|x)$ is to propose a new configuration x' by flipping a single spin at a randomly chosen site. This ensures the Markov chain explores the configuration space in small, local steps.

Once the Markov chain reaches equilibrium (after an initial thermalization period), we collect N samples $\{x_1, x_2, \ldots, x_N\}$ from the chain and compute:

$$\langle f \rangle \approx \frac{1}{N} \sum_{i=1}^{N} f(x_i). \tag{2.37}$$

The associated statistical error can be estimated using standard techniques such as binning analysis, which accounts for the autocorrelation between samples. Obviously, the proposal move needs to accomodate the symmetries of the Hamiltonian. A single spin flip will never be accepted in an Heisenberg model simulation. In that case, the spin exchange move will make sense.

[4]Do not confuse this T with the T of temperature.

Exercise 2.5.1: Critical Slowing Down in the 2D Ising Model

Consider the classical Ising model on a two-dimensional square lattice of size $L \times L$ with periodic boundary conditions. The energy of a spin configuration $x = (\sigma_1, \ldots, \sigma_{L^2})$ is given by:

$$E(x) = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j, \tag{2.38}$$

where $\sigma_i \in \{-1, +1\}$ and the sum is over nearest-neighbor pairs $\langle i, j \rangle$. Assume $J = 1$ and $\beta = 1/T$ is the inverse temperature. Start with small L, i.e. 4×4, then try to increase the size.

1. Implement the Metropolis algorithm using a local spin-flip proposal (i.e., $T(x'|x)$ consists of flipping a randomly chosen spin).
2. Measure the average magnetization $\langle M \rangle$ and energy $\langle E \rangle$ for several values of temperature T, particularly near the critical temperature $T_c \approx 2.269$.
3. Plot the autocorrelation function of the magnetization time series and extract the autocorrelation time τ_{corr} as a function of T.
4. Discuss how the efficiency of the Metropolis algorithm degrades near T_c. This phenomenon is known as *critical slowing down*.
5. **The Art of Debug.** Start small. When in doubt remember that you can exactly compute the expectation values and Z, if 2^L is still decently small.

2.6 Cluster Updates for Spin Systems

Local-update Monte Carlo methods such as the Metropolis alg orithm suffer from *critical slowing down* near second-order phase transitions. In particular, for models like the 2D Ising model close to the critical temperature T_c, spin configurations develop large correlated domains. A local spin-flip move becomes inefficient because it is unlikley to reverse the magnetization of a large cluster by flipping one spin at time, creating many domain walls. The system evolves slowly, leading to large autocorrelation times. A good thing about MC is that the pseudodynamics we follow does not need to be close to the physical dynamics of the system. If we are clever, we can therefore solve critical slow downs.

One great example of this idea are *cluster algorithms*. These non-local[5] update schemes flip entire clusters of spins simultaneously, allowing the simulation to decorrelate much faster, even near T_c. One of the most well-known cluster algorithms is the *Wolff algorithm*, which we describe below.

The Wolff algorithm (Fig 2.3) constructs a cluster of aligned spins starting from a randomly chosen seed spin and grows the cluster by adding neighboring aligned spins with a certain probability. Once the cluster is built, all spins in the cluster are flipped. This method satisfies

[5]The algorithm is however local in the sense that the information to proceed constructing the cluster are local.

detailed balance and samples from the correct Boltzmann distribution $\pi(x) \propto e^{-\beta E(x)}$ without needing to know the partition function.

Let x be a spin configuration, and define a new configuration x' by flipping a cluster C:

- Choose a random site i on the lattice as the *seed* of the cluster.
- Initialize the cluster $C = \{i\}$.
- For each neighbor j of any spin in C:
  - If $\sigma_j = \sigma_i$ and $j \notin C$, add j to C with probability $p_{\text{add}} = 1 - e^{-2\beta J}$.
- Repeat until no more spins are added.
- Flip all spins in C: $\sigma_k \to -\sigma_k$ for all $k \in C$.

The Wolff algorithm preserves detailed balance and ergodicity. The acceptance of a cluster flip is always 1 (since the Bolzmann acceptance probability has been transfered into the proposal step), so the update is rejection-free.
The probability of constructing a given cluster from x and flipping it is equal to the probability of constructing the same cluster from x' and flipping it back, thus satisfying:

$$\pi(x)K(x'|x) = \pi(x')K(x|x'). \tag{2.39}$$

Close to criticality this algorithm is many orders of magnitudes (a factor L^2) better than the local update methods. They allow to observe many full-volume spin-reversal along the same Markov Chain, while the single spin-flip trajectory would be trapped in one of the basins (all-up spins + fluctuations, or the opposite).

In principle, one can restrict the growth of the cluster to a specific subregion of the lattice or to specific directions. For example, in anisotropic systems where the coupling J is much stronger along one direction (e.g., the z-axis) and weaker or disordered in orthogonal directions (e.g., x, y-axis), it may be more efficient to grow clusters only along the strongly interacting direction.

In this modified Wolff algorithm, since the cluster construction does not sample all bonds surrounding the flipped cluster, the update no longer automatically satisfies detailed balance. Therefore, an acceptance step must be reintroduced.
The idea is to compute the energy change ΔE only from the bonds crossing the boundary of the cluster that were not considered during the cluster growth (i.e., those not "tested"). These are the only bonds whose energy contributions differ between the old and proposed configurations.

The acceptance probability is then:

$$A = \min\left(1, e^{-\beta \Delta E_{\text{boundary}}}\right) \tag{2.40}$$

where $\Delta E_{\text{boundary}}$ is the change in energy across these external bonds.

This technique is particularly useful when simulating disordered systems. Indeed, in disordered couplings and low temperatures, the clusters still may grow as to include all spins.

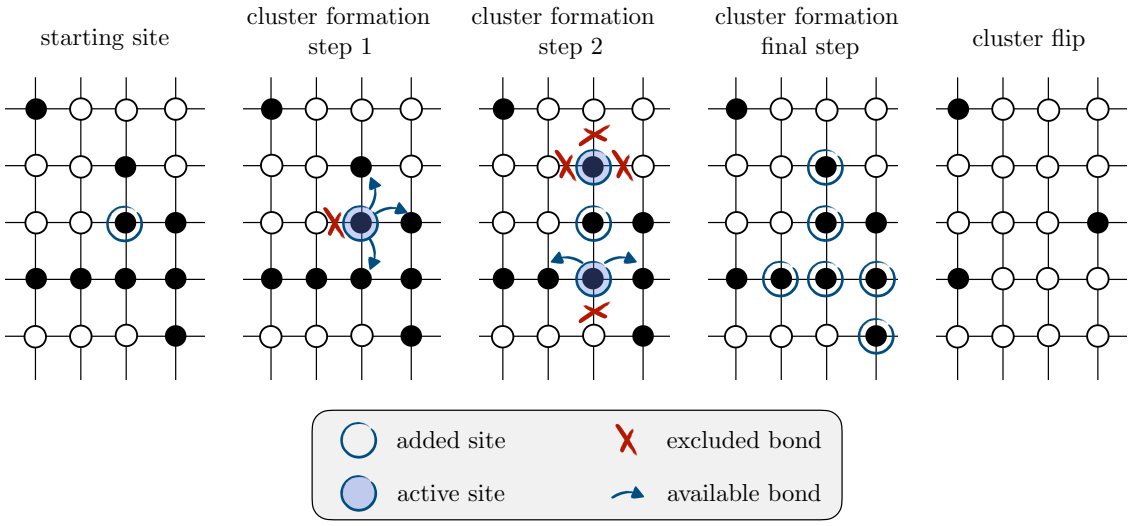


Figure 2.3: Steps of the Wolff algorithm. Notice that at each active site there may be excluded bond to be added (if you try to add to the cluster an opposite spin) or available bonds (linking a parallel spin). Available bonds are added only with a probability. A suitable data structure can simplify the book-keeping of the active sites at each cluster formation step.

This is unwanted if the model is disordered, as one never flips single (or a small group of) spins as to accomodate frustrations.

Exercise 2.6.1: Solving the Critical Slowing Down in the 2D Ising Model (*)

Consider the same model of Exercise 2.5.1.

1. Implement the Metropolis algorithm using the cluster spin-flip proposal (i.e., $T(x'|x)$ of the Wolff Algorithm.
2. Compare the results from Exercise 2.5.1.
3. Compare the autocorrelation times. A simple plot of the average magnetization as a function of iteration is enough.
4. Devise a fair way to compare the convergence of the two methods (single vs cluster flips) as a function of the MC iteration. Is the computational speed-up surviving?

2.7 Metropolis Sampling in Continuous Space

In classical statistical mechanics, the canonical ensemble is defined by the Boltzmann distribution:

$$\pi(x) = \frac{1}{Z} e^{-\beta V(x)}, \tag{2.41}$$

where $V(x)$ is the potential energy of the system, $\beta = 1/k_B T$ is the inverse temperature, and $Z = \int dx\, e^{-\beta V(x)}$ is the partition function. Our goal is to generate configurations x distributed according to $\pi(x)$ using a Markov chain with transition kernel $K(x'|x)$ satisfying detailed balance:

$$\pi(x)K(x'|x) = \pi(x')K(x|x'). \tag{2.42}$$

2.7.1 Simple Metropolis Proposals

In continuous spaces, the simplest Metropolis algorithm begins by proposing a new configuration x' via a random displacement from the current position x. For example, one may use:

- **Uniform proposal:** $x' = x + \delta$, with δ drawn uniformly from $[-\epsilon, \epsilon]$.
- **Gaussian proposal:** $x' = x + \eta$, where $\eta \sim \mathcal{N}(0, \sigma^2)$.

The proposal distribution $K(x'|x)$ is symmetric in both cases. Therefore, the Metropolis acceptance probability reduces to:

$$A(x \to x') = \min\left[1, \frac{\pi(x')}{\pi(x)}\right] = \min\left[1, e^{-\beta[V(x')-V(x)]}\right]. \tag{2.43}$$

The choice of ϵ (for uniform) or σ (for Gaussian) directly affects the acceptance rate. These are hyperparameters tuned to ensure efficient sampling; values that are too small lead to slow exploration, while values that are too large result in low acceptance.

2.7.2 Langevin Dynamics

A better approach uses information about the gradient of the potential. Langevin dynamics introduces a drift term that biases proposals toward lower-energy regions:

$$\dot{x}(t) = f(x) + \eta(t), \tag{2.44}$$

where $f(x) = -\nabla V(x)$ is the deterministic force and $\eta(t)$ is Gaussian white noise with:

$$\langle \eta(t) \rangle = 0, \quad \langle \eta(t)\eta(t') \rangle = 2k_B T \delta(t - t'). \tag{2.45}$$

To numerically integrate this stochastic differential equation, we discretize time into steps of length Δ. Over one time interval (t_n, t_{n+1}), the increment in position is:

$$x_{n+1} - x_n = \int_{t_n}^{t_{n+1}} dt\, f(x(t)) + \int_{t_n}^{t_{n+1}} dt\, \eta(t). \tag{2.46}$$

The first term is approximated using the Euler method as $\Delta f(x_n)$. For the second term, we use the fact that the integral of white noise over a finite time interval is a Gaussian random variable:

$$\int_{t_n}^{t_{n+1}} dt\, \eta(t) = \sqrt{2\Delta k_B T} z_n, \tag{2.47}$$

where z_n is a standard normal variable: $z_n \sim \mathcal{N}(0,1)$. Notice that now the amplitude of the random component is proportional to $\sqrt{\Delta}$. To prove this, let's use $1 =< z_n >^2$, and check that the above definition is compatible with this.

$$< z_n >^2 = \frac{1}{2\Delta k_B T} \int_{t_n}^{t_{n+1}} dt \int_{t_n}^{t_{n+1}} dt' \langle \eta(t)\eta(t') \rangle = \tag{2.48}$$

$$= \frac{2k_B T}{2\Delta k_B T} \int_{t_n}^{t_{n+1}} dt \int_{t_n}^{t_{n+1}} dt' \delta(t-t') = \frac{1}{\Delta} \int_{t_n}^{t_{n+1}} dt = 1 \tag{2.49}$$

Thus, the full *discretized Langevin* equation becomes:

$$\boxed{x_{n+1} = x_n + \Delta f(x_n) + \sqrt{2\Delta k_B T} z_n} \tag{2.50}$$

Notice that now, the only parameter that controls the sampling is the time-step Δ. This can be seen as a hyperparameter for the proposal. However, as we are going to demonstrate, in the limit $\Delta \to 0$, the fixed point of the distribution generated by the Langevin is the Boltzmann distribution, without need for acceptance.

Metropolis-Langevin. Let us consider still the case of finite Δ and cosider this as a trial proposal for a Metropolis. Here, the resulting transition kernel generally does not satisfy detailed balance. Let us write directly in d dimensions, because

$$\mathbf{x}' = \mathbf{x} - \Delta\, \mathbf{f}(\mathbf{x}) + \sqrt{2\Delta\, k_B T}\, \boldsymbol{\eta}, \tag{2.51}$$

where $\mathbf{f}(\mathbf{x}) = -\nabla U(\mathbf{x})$ is the force vector and $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$ is a d-dimensional standard normal vector, the transition kernel is given by:

$$K(\mathbf{x}'|\mathbf{x}) = \frac{1}{(4\pi\Delta k_B T)^{d/2}} \exp\left(-\frac{\|\mathbf{x}' - \mathbf{x} + \Delta \mathbf{f}(\mathbf{x})\|^2}{4\Delta k_B T} \right). \tag{2.52}$$

This proposal does not satisfy detailed balance with respect to the equilibrium distribution $\pi(\mathbf{x}) \propto \exp[-U(\mathbf{x})/k_B T]$. To correct for this, a Metropolis-Hastings acceptance step is introduced, with acceptance probability:

$$\alpha(\mathbf{x} \to \mathbf{x}') = \min\left\{ 1, \frac{\pi(\mathbf{x}')K(\mathbf{x}|\mathbf{x}')}{\pi(\mathbf{x})K(\mathbf{x}'|\mathbf{x})} \right\}. \tag{2.53}$$

Substituting in the expressions for π and K, we obtain:

$$\alpha(\mathbf{x} \to \mathbf{x}') = \min\left\{ 1, \exp\left[-\frac{U(\mathbf{x}') - U(\mathbf{x})}{k_B T} \right.\right. \tag{2.54}$$

$$\left.\left. + \frac{1}{4\Delta k_B T} \left(\|\mathbf{x} - \mathbf{x}' + \Delta \mathbf{f}(\mathbf{x}')\|^2 - \|\mathbf{x}' - \mathbf{x} + \Delta \mathbf{f}(\mathbf{x})\|^2 \right) \right] \right\}. \tag{2.55}$$

Expanding the norms and simplifying using scalar products gives:

$$\alpha(\mathbf{x} \to \mathbf{x}') = \min \left\{ 1, \exp \left[-\frac{U(\mathbf{x}') - U(\mathbf{x})}{k_B T} \right. \right. \tag{2.56}$$

$$\left. \left. + \frac{1}{2k_B T} \left((\mathbf{x}' - \mathbf{x}) \cdot [\mathbf{f}(\mathbf{x}) + \mathbf{f}(\mathbf{x}')] + \frac{\Delta}{2} \left(\|\mathbf{f}(\mathbf{x}')\|^2 - \|\mathbf{f}(\mathbf{x})\|^2 \right) \right) \right] \right\}. \tag{2.57}$$

Exercise 2.7.1: Metropolis vs Langevin sampling

Consider an harmonic classical potential $V = \frac{1}{2}\omega^2 x^2$, at a finite temperature T.

1. Use the Metropolis sampling to evaluate the expectation values of $\langle V \rangle$ and $\langle x \rangle$.
2. Use the Langevin dynamics and check that it converges to the exact values in the limit $\Delta \to 0$.

2.7.3 From the Langevin Dynamics to Fokker-Planck equation

The discretized Langevin equation induces a stochastic map from x_n to x_{n+1}, which defines a transition kernel $K(x'|x)$ describing the conditional probability of reaching configuration x' starting from x after one time step. Since $z \sim \mathcal{N}(0,1)$,

$$K(x'|x) = \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \delta(x' - x - \Delta f(x) - \sqrt{2\Delta k_B T}\, z) \tag{2.58}$$

This kernel represents the probability density of making a transition from x to x' in one time step, by considering the probability of extracting a random displacement z. We can perform (see Appendix B) the integral finding

$$K(x'|x) = \frac{1}{\sqrt{4\pi\Delta k_B T}} \exp\left[-\frac{(x' - x - \Delta f(x))^2}{4\Delta k_B T} \right] \tag{2.59}$$

In Appendix B we further prove that this lead to the Fokker-Planck equation that is a differential equation for the evolution of the probability $\rho(x,t)$. This is useful to prove convergence to the Boltzmann distribution and the draw interesting comparison between quantum mechanics and classical statistical mechanics. The computation is not straightforward and several texts use different methods. Here we choose the simplest way, without use of Fourier transform, or using stochastic calculus. The final result reads,

$$\boxed{\frac{\partial \rho(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[f(x)\rho(x,t) \right] + k_B T \frac{\partial^2 \rho(x,t)}{\partial x^2}} \tag{2.60}$$

The stationary solution $\rho(x) = \pi(x)$ is easily verified by direct substitution[6], confirming that the discretized Langevin dynamics converges to the correct Boltzmann equilibrium.

[6] setting $\frac{\partial \rho(x,t)}{\partial t} = 0$, and $\pi(x) = \exp[-V(x)/k_B T]$

2.7.4 From the Fokker-Planck to the Schrödinger Equations

The Fokker-Planck equation can be formally recasted into the Schrödinger equation. This connection is obtained by plugging this ansatz in Eq. 2.60:

$$P(x,t) = \psi_0(x)\,\Phi(x,t) \tag{2.61}$$

where $\psi_0(x) = \sqrt{P_0(x)} = \frac{1}{\sqrt{Z}} e^{-V(x)/2k_BT}$, which, being $\int dx\, \psi_0^2(x) = 1$, is a normalized quantum state.
Indeed, inserting this into the Fokker-Planck equation we obtain an equation for $\Phi(x,t)$ which is an imaginary-time Schrödinger Eq. with an effective Hamiltonian $\mathcal{H}$:

$$\frac{\partial}{\partial t}\Phi(x,t) = -\mathcal{H}\Phi(x,t) = -\left[-k_BT\frac{\partial^2}{\partial x^2} + \mathcal{V}(x)\right]\Phi(x,t) \tag{2.62}$$

where

$$\mathcal{V}(x) = \frac{1}{4k_BT}\left(\frac{\partial V(x)}{\partial x}\right)^2 - \frac{1}{2}\frac{\partial^2 V(x)}{\partial x^2}. \tag{2.63}$$

Notice that the minima of the effective potential, are the minima of the original potential $V(x)$, but the effective potentials has more minima and saddle points. The ground state of this effective Hamiltonian can be found exactly and is simply $\psi_0(x)$. Moreover, the corresponding ground state energy is $E_0 = 0$ (notice that the effective hamiltonian spectrum is not defined up to a constant)[7].
Therefore, the full solution of the Schrödinger equation, and thus of the corresponding Fokker-Planck equation, can be written in closed form by expanding the initial condition as:

$$\frac{P(x,t=0)}{\psi_0(x)} = \Phi(x,t=0) = \sum_n a_n\psi_n(x), \tag{2.64}$$

with coefficients

$$a_n = \int dx\, \psi_n(x)\frac{P(x,t=0)}{\psi_0(x)}, \tag{2.65}$$

implying $a_0 = 1$ from the normalization condition on $P(x,t=0)$.
The time evolution of the probability distribution is then given by:

$$P(x,t) = \sum_n a_n\psi_n(x)\psi_0(x)e^{-E_nt}, \tag{2.66}$$

which shows that, for large times t, $P(x,t)$ converges exponentially fast to the stationary equilibrium distribution $\psi_0^2(x)$.

The characteristic equilibration time τ is given by the inverse of the energy gap:

$$\tau = \frac{1}{E_1 - E_0} = \frac{1}{E_1}. \tag{2.67}$$

[7] If you use $V(x) = \omega^2x^2/2$ then you will notice that the second term in $\mathcal{V}(x)$, the 'hessian' term, will provide a negative shift that will cancel out the zero-point-motion energy that a 'normal' quantum harmonic oscillator would have.

This formally justifies that the use of the Langevin dynamics as a sampler for the Boltzmann distribution by iterating the discretized equation for simulation times much larger than the correlation time τ. This pathway can be further explored to establish more connection between classical statistical mechanics and quantum formalism, but we will not go further here.

2.8 Exponential Divergence of Autocorrelation Time in Disordered Systems

In many disordered or frustrated statistical systems, Markov Chain Monte Carlo simulations suffer from extremely long autocorrelation times (cfn. Sect 2.2.3), especially at low temperatures. When the free-energy landscape contains many deep local minima separated by barriers, the system becomes trapped in metastable states. Escaping from these states requires crossing free-energy barriers ΔF that typically grow with system size, leading to an exponential suppression of the transition probability:

$$\tau \sim \exp(\Delta F / k_B T) \sim \exp\Big(cL^{\theta}\Big),$$

where L is the system size, c is a constant, and $\theta > 0$ is a barrier exponent.

A well-studied example is the two-dimensional $\pm J$ Ising spin glass, with Hamiltonian

$$E(\{s_i\}) = -\sum_{\langle i,j \rangle} J_{ij} s_i s_j, \quad s_i = \pm 1, \quad J_{ij} \in \{-1, +1\} \text{ randomly assigned.} \tag{2.68}$$

At low temperatures, the system has a highly degenerate and rugged energy landscape, leading to a proliferation of metastable states. The transition between these configurations becomes extremely slow with increasing system size, and local MCMC methods like Metropolis sampling are quickly rendered ineffective. Only recently the Houdayer cluster move have has been introduced effectively, yet it is applicable only in 2D systems.

Another archetypal case is the Sherrington-Kirkpatrick (SK) model, a fully connected spin glass described by

$$E(\{s_i\}) = -\sum_{i<j} J_{ij} s_i s_j, \quad J_{ij} \sim \mathcal{N}(0, 1/L). \tag{2.69}$$

This model also exhibits a glassy phase with exponentially many low-energy states. Below the spin glass transition temperature, the system becomes dynamically arrested. The autocorrelation time grows as $\tau \sim \exp(cL)$, where L is the number of spins.

Similar behavior is observed in random constraint satisfaction problems, such as random k-SAT and graph coloring near the satisfiability threshold. In these models, the solution space fractures into clusters separated by large Hamming distances. Local search algorithms struggle to navigate between clusters, and equilibration times diverge exponentially with system size.

These difficulties are not limited to discrete spin systems. In continuous-space models, such as those describing classical particles in disordered media or structural glasses, similar

challenges arise. For example, in a Lennard-Jones mixture used to model glass-forming liquids, particles experience an amorphous potential landscape with deep wells and barriers.

These challenges have motivated the development of enhanced sampling techniques such as umbrella sampling, parallel tempering (replica exchange), metadynamics, and path sampling methods. While such techniques can mitigate the severity of critical slowing down or glassy dynamics, they completely eliminate the exponential scaling in deeply frustrated or strongly first-order systems.

Exercise 2.8.1: Mixing times of Metropolis for the SK model (*)

Consider the SK model of E 2.69. Notice that the couplings strenght j_{ij} descrease with $1/\sqrt{L}$ because the model has *all-to-all* connectivity, so its total energy would grow more than linearly with L. Set a temperature of $T = 1$, in energy scale.

- Construct the full $K_{i,j}$ matrix for the local update Metropolis proposal strategy $K_{ij} = T_{ij}A_{ij}$.
- Diagonalize it and find the gap δ.
- Perform a scaling analysis for system sizes between about 4 to 8.
- Show that it decreases exponential with 2^{-cL}, with a c marginally better than 1.
- Keep the system's size fixed, e.g. $L = 6$ and now vary the temperature. How does the plot looks like?
- (Bonus) Repeat and compared with a different choice of the transition matrix T_{ij}, the uniform distribution, namely, choosing a j site at random, irrespective of i. How does it compared with the local updates? Check that it is better at large temperatures.

2.9 A bit of complexity theory

In this subsection we provide a simple overview of complexity classes that may be useful in general. This is important to internalize the obvious, yet sometimes forgotten, statement that *also purely classical problem can be exponentially hard to solve.* The exposition here is colloquial, as more rigorous definitions exists. The runtime (the worst-case scenario) determine the complexity class of a problem. If we consider classical computation, very common classes are P and NP:

- a problem belongs to the P class if a polynomial (in L) algorithm can solve the worst-case instances. These problems are considered *easy* from the complexity perspective, although the order of the polynomial function can make a great difference in practice.
- The NP class include computational **decision** problems for which any given "yes"-solution can be verified in polynomial time by a deterministic machine. Notice that optimization problems are not stricly speaking decision problems, as their output is a function value. They can be recasted as decision problems if we include also a threeshold value, i.e. E_c. The corresponding verifiable statement, in poly time, is wether the

found solution E_{found} is lower than E_c. In this sense, also optimization problem can be included as member of complexity classes.

Notice that NP decision problems could, in theory, be solved in polynomial time on a hypothetical nondeterministic computer, but for which no polynomial-time algorithm is currently known on standard deterministic machines. Such a nondeterministic device can conceptually explore both branches of an if statement at once, though it is unable to recombine them afterward. An equivalent picture is to imagine a system with an exponential number of processors that operate independently without any communication.[8] Furthermore, for a problem to belong to NP, any proposed solution must be verifiable on a classical deterministic computer within polynomial time.

- NP-hard problems include classes of problems which are at least as hard as the hardest problems in NP. It may sound strange but the NP-hard class is not fully included in NP. NP-hard problems which are in NP, are called NP-complete. All NP-complete problems are also NP-hard, but the opposite is not true. Consequently, solving an NP-hard problem efficiently would amount to solving any NP problem, and discovering a polynomial-time algorithm for even one of them would revolutionize computing and undermine the security assumptions of conventional cryptographic systems.

  Notice again that this labeling is valid only if we consider the worst-case scenario. Think about the classical ferromagnetic model: here it is easy to guess the ground state, yet it is a special instance of an Ising spin-glass Hamiltonian.

However, it should be noted that for problems with a certain structure, there may exist better methods than brute force search and correspondingly better quantum methods. It is important to remember that everything mentioned above applies when seeking an exact solution. In practice, though, sometimes one can settle for a good solution, even if it's not the true global minimum. Something good is still better than nothing.

In such cases, one can turn to approximation algorithms, which aim to find these near-optimal solutions. These solutions must provide provable guarantees regarding their closeness to the optimal solution. But how can we quantitatively define if a solution is near-optimal? Here, the approximation ratio can be used, which is the ratio between the returned solution and the optimal one. An approximation algorithm can thus provide the following guarantee: it provides a solution that cannot be worse than a given approximation ratio. This defines the APX class: the set of NP optimization problems that allow polynomial-time approximation algorithms with an approximation ratio bounded by a constant.

[8] this statement could induce some into thinking that a quantum computer could solve NP problem in polynomial time.

Chapter 3

Variational Monte Carlo

In this Chapter we show how to actually compute the energy of correlated trial state, such as those presented in Sect. 1.6.

3.1 Local Energy and Zero Variance Property

Variational Monte Carlo (VMC) is a stochastic method for computing ground-state properties of quantum systems using a trial[1] wavefunction $\psi(x)$, where x denotes a configuration in Hilbert space. The configuration x can represent, for example, a spin state on a lattice or a set of continuous particle coordinates.
The central quantity of interest is the variational energy,

$$E_{\text{var}} = \frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle}, \tag{3.1}$$

which provides an upper bound to the ground-state energy of the system due to the variational principle. Note that we implicitely assume that $\psi(x)$ depends on a set of parameters (this will be the focus of the next section). This quantity can be conveniently rewritten in terms of an expectation value over the probability distribution defined by the square modulus of the wavefunction:

$$E_{\text{var}} = \sum_x \frac{|\psi(x)|^2}{\langle\psi|\psi\rangle} \cdot \frac{\langle x|\hat{H}|\psi\rangle}{\psi(x)} \equiv \sum_x P(x) E_L(x), \tag{3.2}$$

where we define:

- $P(x) = \frac{|\psi(x)|^2}{\langle\psi|\psi\rangle}$ as the normalized probability distribution,
- $E_L(x) = \frac{\langle x|\hat{H}|\psi\rangle}{\psi(x)}$ as the *local energy*.

This formulation allows us to evaluate E_{var} through a Monte Carlo sampling of N configurations x distributed according to $P(x)$, and averaging the corresponding values of $E_L(x)$.

[1] Previously also denoted as ψ_T.

Crucially, if we use Metropolis, we do not need to work with normalized wavefunctions. Indeed, the exponentially large summation of the righthand side of Eq. 3.2 can be replaced by the MC summation we have seen in the previous Chapter 2. Operationally we have

$$\boxed{E_{\text{var}} \approx \frac{1}{N}\sum_{i=1}^{N} E_L(x_i), \quad \text{with} \;\; x_i \sim P(x)} \tag{3.3}$$

A particularly important property of the local energy is the *zero variance principle*: if the trial wavefunction ψ is an exact eigenstate of the Hamiltonian $\hat{H}$, with eigenvalue E, then $E_L(x)$ becomes independent of x and equal to the corresponding eigenvalue:

$$E_L(x) = \frac{\langle x|\hat{H}|\psi\rangle}{\langle x|\psi\rangle} = \frac{E\langle x|\psi\rangle}{\langle x|\psi\rangle} = E. \tag{3.4}$$

Therefore, the random variable $E_L(x)$ is independent of the configuration x, which immediately implies that its variance is zero. The average of $E_L(x)$ also coincides with the exact eigenvalue E. Clearly, the closer the variational state $\psi(x)$ is to an exact eigenstate, the smaller the variance of $E_L(x)$ will be. This observation is very important, as a reduced variance leads to smaller statistical fluctuations and better numerical efficiency.

Moreover, the average of the square of the local energy corresponds to the exact quantum expectation value of $\hat{H}^2$. That is,

$$\langle E_L(x)^2\rangle = \sum_x P(x)E_L(x)^2 = \frac{\langle\psi|\hat{H}^2|\psi\rangle}{\langle\psi|\psi\rangle}, \tag{3.5}$$

Thus, the variance of the local energy is:

$$\text{var}(E_L) = \langle E_L(x)^2\rangle - \langle E_L(x)\rangle^2 = \frac{\langle\psi|(\hat{H}-E)^2|\psi\rangle}{\langle\psi|\psi\rangle}. \tag{3.6}$$

This is exactly the quantum variance of the Hamiltonian with respect to the variational state ψ. Hence, the smaller the variance, the closer the wavefunction is to an exact eigenstate.
Instead of minimizing the energy directly, it is sometimes advantageous to minimize the variance of the local energy as a function of the variational parameters. In theory, looking a the variance makes sense because we know the target value to reach: zero, while we don't know a priori the exact energy. In practice, one should observe a reduction of the variance, as the energy also decreases.

3.1.1 Lattice Models

In lattice systems, such as quantum spin models or lattice fermion models, the Hilbert space is discrete. A configuration is denoted by $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_L)$, representing the spin states at each site of the system.
The Hamiltonian $\hat{H}$ in these models usually contains terms that act locally or on small clusters of neighboring spins. In this setting, the local energy can be evaluated as a sum

over configurations $\boldsymbol{\sigma}'$ that are connected to the current configuration $\boldsymbol{\sigma}$ by non-zero matrix elements of the Hamiltonian:

$$E_L(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} \frac{\langle \boldsymbol{\sigma}|\hat{H}|\boldsymbol{\sigma}'\rangle \psi(\boldsymbol{\sigma}')}{\psi(\boldsymbol{\sigma})}. \tag{3.7}$$

Because most Hamiltonians of interest are local, this sum only involves a small number of terms, making the computation efficient. Notice that this way, we can make use of the definition of the Hamiltonian in terms of its matrix elements.
More generally, the same strategy applies to the computation of the expectation value of any operator $\hat{A}$. Defining the corresponding *local estimator* as

$$A_L(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} \frac{\langle \boldsymbol{\sigma}|\hat{A}|\boldsymbol{\sigma}'\rangle \psi(\boldsymbol{\sigma}')}{\psi(\boldsymbol{\sigma})}, \tag{3.8}$$

one can estimate $\langle \hat{A} \rangle$ as the average of $A_L(\boldsymbol{\sigma})$ over the sampled configurations. This generalization is crucial for measuring observables other than the energy.

Exercise 3.1.1: Hopping particle on a lattice

Consider a spinless particle that can hop on a lattice made of L sites. The hopping amplitude is parametrized by t, and there is also a linear potential. This system is a lattice discretized version of a particle. The Hamiltonian is

$$H = -t \sum_{i=1}^{L-1} (c_{i+1}^\dagger c_i + h.c.) + V \sum_{i=1}^{L} (i-1) c_i^\dagger c_i. \tag{3.9}$$

The particle cannot move outside the range $[1, L]$. Consider a trial function of the form

$$\psi_T(i) \propto e^{-gi}, \tag{3.10}$$

and system parameters $V/|t| = 10$, $L = 30$.

1. Write the local energy.
2. Write a VMC code that calculate the expectation value of the energy at fixed parameter g.
3. Find the optimal value of g with simple grid search.

Exercise 3.1.2: Heisenberg model in 1D (*)

Consider the one-dimensional Heisenberg model with an even number L of sites:

$$H = J \sum_{i=1}^{L} \vec{\sigma}_i \cdot \vec{\sigma}_{i+1} \tag{3.11}$$

with periodic boundary conditions (PBC) $\vec{\sigma}_{i+L} = \vec{\sigma}_i$. A variational wavefunction for the ground state of H can be defined in the basis of configurations $\{s_i^z, \cdots, c_L^z\}$, where each site i has a definite value of $s_i^z = \pm 1$, and the total spin projection on the z-axis vanishes: $\sum_i s_i^z = 0$. In this basis, the wavefunction can be written as:

$$\psi(x(s_i^z, \cdots, c_L^z) = \mathrm{Sign}(x) \times \exp\left(\alpha \sum_{i \neq j} v_{ij}\, s_i^z s_j^z\right) \tag{3.12}$$

where: α is a variational parameter, $v_{ij} = \ln\left(d_{ij}^2\right)$ is the spin Jastrow factor, which depends on the chord distance d_{ij} between two sites, defined as:

$$d_{ij} = 2 \sin\left(\frac{\pi |i-j|}{L}\right),$$

$\mathrm{Sign}(s_i^z, \cdots, c_L^z) = (-1)^P$, where $P = \sum_{i=1}^{L/2} \left(S_{2i}^z + 1\right)/2$ is the so-called Marshall sign, determined by the number of up spins on even sites.

1. Write down the matrix element of the hamiltonian and the local energy.
2. Write a code that perform a VMC for a system of $L = 100$.
3. Find the best variational parameter α.
4. The exact result for an infinite chain is $-4 \times J \times 0.44314718$. Can you provide a finite size converge study?

3.1.2 Continuous Systems

In systems with particles in continuous space, the Hilbert space is infinite-dimensional, and configurations are denoted by $\boldsymbol{x} = (\vec{x}_1, \ldots, \vec{x}_L)$, where $\vec{x}_i \in \mathbb{R}^d$ is the position of particle i.

A typical Hamiltonian takes the form:

$$\hat{H} = -\sum_{i=1}^{L} \frac{\hbar^2}{2m} \nabla_i^2 + V(\boldsymbol{x}), \tag{3.13}$$

where $V(\boldsymbol{x})$ is the potential energy, which may include both external and interaction terms.

The corresponding local energy reads:

$$E_L(\boldsymbol{x}) = -\sum_{i=1}^{L} \frac{\hbar^2}{2m} \frac{\nabla_i^2 \psi(\boldsymbol{x})}{\psi(\boldsymbol{x})} + V(\boldsymbol{x}). \tag{3.14}$$

This expression involves second derivatives of the wavefunction with respect to spatial coordinates, which can be cumbersome and error-prone to compute analytically. In modern implementations, these derivatives are often evaluated using *automatic differentiation*, which provides a flexible and accurate way to obtain gradients and Laplacians, especially when the wavefunction is expressed as a computational graph (e.g., using neural networks).

Exercise 3.1.3: Local energy an harmonic oscillator

Consider the following Hamiltonian

$$H = -\frac{d^2}{dx^2} + x^2, \tag{3.15}$$

which has ground state energy of $E_0 = 1$, and ground state wf, $\psi_0(x) = \frac{1}{\pi^{1/4}} e^{-x^2/2}$. Suppose that we don't know this but guess simply the functional form of normalized the ground-state,

$$\psi_T(x) = \frac{\sqrt{\theta}}{\pi^{1/4}} e^{-\theta^2 x^2/2}. \tag{3.16}$$

1. Write analytically the local energy $E_L(x;\theta)$.
2. Show that the local energy is constant when the ψ_T becomes the exact gound state.
3. (bonus) Write a VMC Metropolis code and numerically check that the variance of the energy decreases as $\theta \to 1$.

3.1.3 Electronic systems: Jastrow and Cusp Conditions

Let us consider a trial wavefunction $\psi(\boldsymbol{x})$ is introduced, where a configuration is defined as $\boldsymbol{x} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_L)$, with $\boldsymbol{r}_i \in \mathbb{R}^d$ representing the spatial coordinate of particle i. To distinguish spatial and spin degrees of freedom, we explicitly write the configuration of a single particle as $x = (\boldsymbol{r}, \sigma)$. A widely used class of variational wavefunctions takes the form of a product between a symmetric correlation factor, known as the Jastrow factor, and a symmetrization or antisymmetrization component. For fermionic systems, the latter is typically a Slater determinant ensuring antisymmetry under particle exchange:

$$\psi(\boldsymbol{x}) = \mathcal{D}(\boldsymbol{x}) \times \exp\left(-\sum_{i<j} u(r_{ij}) \right), \tag{3.17}$$

where $\mathcal{D}(\boldsymbol{x})$ denotes a Slater determinant and $u(r_{ij})$ is a two-body correlation function, often chosen to include variational parameters. A popular choice for $u(r)$ is the Padé form, which captures short-range correlations effectively. This structure is also suitable for bosonic systems, such as liquid ^{4}He, by replacing the Slater determinant with a fully symmetric function (e.g., a constant or a permanent).

In fermionic systems with equal numbers of spin-up and spin-down electrons, the Slater determinant is often decomposed into a product of two determinants, one for each spin species. That is, the Slater-Jastrow wavefunction is written as

$$\psi(\boldsymbol{x}) = \det\left[\phi_k^{\uparrow}(\boldsymbol{r}_i)\right] \cdot \det\left[\phi_k^{\downarrow}(\boldsymbol{r}_j)\right] \cdot e^{-U(\boldsymbol{x})}, \tag{3.18}$$

where the orbitals ϕ_k^{σ} are spin-dependent, $1 \leq i \leq N/2$ while, $N/2+1 \leq j \leq N$ (assuming the first half of the electrons to be the up spins). $U(\boldsymbol{r})$ is the total Jastrow exponent.

The *local energy* is constant if and only if ψ is an exact eigenfunction of the Hamiltonian. In practice, however, the trial wavefunction is approximate, and E_L fluctuates, especially near coalescence points where particles approach each other or a nucleus. Known exact properties of the wavefunction can be imposed here. The*Kato cusp conditions*, which dictates the behavior of the wavefunction ψ as an electron approaches a nucleus or another electron. The **electron-nucleus** cusp is

$$\left.\frac{\partial \psi}{\partial r_{iI}}\right|_{r_{iI}\to 0} = -Z_I \psi, \tag{3.19}$$

where $r_{iI} = |\boldsymbol{r}_i - \boldsymbol{R}_I|$ and Z_I is the atomic number of nucleus I. The derivation is in Appendix C.

The cusp conditions also put constrains on the shape of the exact ground state when two electrons become close. There are two version depending if the electrons have parallel or anti-parallel spins. In the first case, the Pauli exclusion principle is satisfied by the spin part of the total wavefunction, such that the spacial part can be finite at the coalescence point. This leads to a cusp-condition formally equivalent to the electron-nucleus one. The parallel spin case is a more complicated to derive (see Appendix C).

An operational way to derive the cusp conditions is by avoiding divergence in the local energy. We assume a functional for the trial function

$$\psi(\boldsymbol{r}) = e^{-u(r)} f(\boldsymbol{r}), \tag{3.20}$$

where $u(r)$ is a Jastrow correlation factor depending only on the interelectronic distance $r = |\boldsymbol{r}_i - \boldsymbol{r}_j|$, and $f(\boldsymbol{r})$ is the antisymmetric part.

- For two electrons with **antiparallel spins**, the wavefunction f is finite at coalescence, and the cusp condition is
$$\left.\frac{du}{dr_{ij}}\right|_{r_{ij}\to 0} = -\frac{1}{2}. \tag{3.21}$$
- For two electrons with **parallel spins**, the Pauli exclusion principle enforces a node at coalescence, i.e., $f \to 0$ as $r_{ij} \to 0$. The wavefunction vanishes linearly, and the cusp condition becomes
$$\left.\frac{du}{dr_{ij}}\right|_{r_{ij}\to 0} = -\frac{1}{4}. \tag{3.22}$$

Exercise 3.1.4: VMC for the Helium atom (*/)**

Consider the hamiltonian for the Helium atom

$$\hat{H} = -\frac{1}{2}\nabla_1^2 - \frac{1}{2}\nabla_2^2 - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{r_{12}}. \tag{3.23}$$

We can define the (spacial) trial state $\psi_T(\boldsymbol{r}_1, \boldsymbol{r}_2)$, parametrized with a, b, c real numbers,

$$\psi_T(\boldsymbol{r}_1, \boldsymbol{r}_2) = \exp\left[-a \cdot (r_1 + r_2) + \frac{r_{12}}{c\,(1 + b r_{12})}\right]. \tag{3.24}$$

Notice that in this case the two electrons are distinguishable because of opposite spins and there is no need for the determinant. Alternatively, you can see the one-body part of the Jastrow as the product of two Slater-like single-particle orbitals $\phi(r) = e^{-ar}$ (see Eq. 3.18.

1. Use the cusp conditions to fix a and c, leaving only b as real variational parameter.
2. (**)Write a VMC code and find the variational energy as a function of b, and compare with the exact value of -2.903 Ha.

   Hint: you can use some symbolic calculus software to calculate E_L: it is a bit cumbersome.

3.2 Optimization of the Wavefunction

The quality of the variational energy depends critically on the choice of the trial wavefunction $\psi(x; \{\alpha_k\})$, which is typically parametrized by a set of variational parameters $\{\alpha_k\}$. The goal is to find the set of parameters that minimizes E_{var}, ideally reaching the ground-state energy.

In traditional approaches, the number of parameters was small, but in modern variational ansätze, such as neural quantum states (NQS), the number of parameters can reach thousands or even millions. This makes optimization both challenging and computationally demanding.

One difficulty arises when trying to evaluate differences between two independently estimated stochastic quantities. When the difference becomes small, statistical noise may dominate, obscuring the optimization direction. More sophisticated techniques are therefore required.

3.2.1 Correlated Sampling

To mitigate noise in comparing closely related wavefunctions, one can use *correlated sampling*. This method allows estimating energy differences more accurately by reusing the same sample of configurations.

Let $\psi_0(x)$ be a reference wavefunction, and consider a new wavefunction $\psi_1(x)$. If we sample configurations x according to $|\psi_0(x)|^2$, we can reweight the estimates using:

$$E_{\text{var}}[\psi] \approx \frac{\sum_{x\sim|\psi_0|^2} w(x) E_L(x)}{\sum_{x\sim|\psi_0|^2} w(x)}, \quad w(x) = \left|\frac{\psi_1(x)}{\psi_0(x)}\right|^2. \tag{3.25}$$

This allows us to estimate the energy and other observables of a slightly modified wavefunction without generating a new set of samples. For instance $\psi_0 = \psi(\theta)$, while $\psi_1 = \psi(\theta + \delta\theta)$.

3.2.2 Gradients

An efficient alternative to comparing energy values is to compute the *gradient* of the energy with respect to the variational parameters. This approach provides a direct way to perform optimization using gradient descent or its variants.

We introduce the *logarithmic derivative* of the wavefunction with respect to the parameter α_k:

$$O_k(x) = \frac{1}{\psi(x)}\frac{\partial\psi(x)}{\partial\alpha_k} = \frac{\partial \log\psi(x)}{\partial\alpha_k}. \tag{3.26}$$

The gradient of the variational energy is given by:

$$f_k = -2\,\mathrm{Re}\left\{\langle (E_L - \langle E_L\rangle)(O_k - \langle O_k\rangle)\rangle\right\}. \tag{3.27}$$

It can be evaluated as a Monte Carlo average. During each iteration of the simulation, we compute and store $E_L(x)$ and $O_k(x)$ for each configuration x_i.

$$f_k = -2\,\mathrm{Re}\left\{\sum_i (E_L(x_i) - \langle E_L\rangle)(O_k(x_i) - \langle O_k\rangle)\right\}, \qquad \text{with} \quad x_i \sim P(x). \tag{3.28}$$

Proof: (with real valued wf):

$$\partial_k\left(\frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle}\right) = 2\frac{\langle\partial_k\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle} - 2\frac{\langle\partial_k\psi|\psi\rangle\langle\psi|\hat{H}|\psi\rangle}{|\langle\psi|\psi\rangle|^2} \tag{3.29}$$

$$= 2\left(\frac{\langle\psi|O_k\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle} - E_{\mathrm{var}}\frac{\langle\psi|O_k|\psi\rangle}{\langle\psi|\psi\rangle}\right) \tag{3.30}$$

$$= 2\left(\sum_i O_k(x_i)E_L(x_i) - \sum_i E_L(x_i)\sum_i O_k(x_i)\right), \tag{3.31}$$

where we have used the fact that OH is Hermitian operator, and the definition of $E_{\mathrm{var}} = \langle E_L(x)\rangle_{P(x)}$.[2] Notice also the the $\psi(x)$ is not assumed normalized, hence the explicit normalization, so we need to consider the variation of the norm as a function of θ_k.

Once that we have the force we can update the parameters using

$$\theta_k = \theta_k + \delta\theta_k = \theta_k + \Delta f_k, \tag{3.34}$$

[2]This expression is equivalent to the one above:

$$\langle (E_{_L} - \langle E_{_L}\rangle)(O - \langle O\rangle)\rangle = \langle E_{_L}O - E_{_L}\langle O\rangle - \langle E_{_L}\rangle O + \langle E_{_L}\rangle\langle O\rangle\rangle \tag{3.32}$$

$$= \langle E_{_L}O\rangle - \langle E_{_L}\rangle\langle O\rangle - \langle E_{_L}\rangle\langle O\rangle + \langle E_{_L}\rangle\langle O\rangle = \langle E_{_L}O\rangle - \langle E_{_L}\rangle\langle O\rangle. \tag{3.33}$$

where Δ is a hyperparameter that determines how quickly we perfom the gradient descent.

Note: While computing the derivative of the log of a function seems scary, it is actually simple for the form of the wavefunctions that we will use in these notes, which are Jastrow (or simular) functions. Computing the derivative of parameters that are inside in the Slater-Determinant would be much more cumbersome.

3.2.3 Natural Gradients a.k.a Stochastic Reconfiguration

The basic gradient descent method can be improved by preconditioning the gradient using information about the geometry of the variational manifold. This leads to the *stochastic reconfiguration* (SR) method, also known as *quantum natural gradient descent*.

The idea is to use the *quantum geometric tensor* (QGT), defined as:

$$S_{k\ell} = \langle (O_k - \langle O_k \rangle)^* (O_\ell - \langle O_\ell \rangle) \rangle, \tag{3.35}$$

to rescale the gradient directions. The update rule becomes:

$$\sum_\ell S_{k\ell}\, \delta\theta_\ell = \Delta f_k. \tag{3.36}$$

In most implementations, it is not necessary to construct or store the full matrix $S_{k\ell}$. Instead, iterative solvers can be used that only require the action of S on a vector. This allows the method to scale efficiently with the number of parameters.

One practical issue is that the matrix S can become ill-conditioned due to statistical noise, making the solution of the linear system unstable. To address this, a regularization term is added to the diagonal:

$$S_{kk} \to S_{kk}(1 + \epsilon), \tag{3.37}$$

where ϵ is a small positive parameter (e.g., 10^{-3}). This procedure improves numerical stability and ensures that the optimization step is well-behaved.

Exercise 3.2.1: VMC optimization with gradients

Consider the TFIM hamiltonian in 1D with PBC:

$$H = -\sum_{i}^{L} \sigma_i^z \sigma_{i+1}^z - \sum_{i}^{L} \sigma_i^x, \tag{3.38}$$

and define the ansatz

$$\psi_T(s_1, \cdots, s_L) = \exp\left(\theta_1 \sum_i s_i s_{i+1} + \theta_2 \sum_i s_i s_{i+2}\right), \tag{3.39}$$

where $s_i \in \{-1, +1\}$ represents the spin z-value at site i, and θ_1, θ_2 are variational parameters encoding nearest and next-nearest neighbor interactions, respectively.

1. Write a VMC code with gradient optimization to find the best variational energy, use $L = 20$. Reasonable values for parameters are $\Delta = 0.1 - 0.01$, and order of 1000s MC sweeps (include also thermalization sweeps). You need to identify and evaluate O_1 and O_2, and accumulate the necessary averages.

2. Plot the energy vs optimization steps and compare with the exact diagonalization value, -25.491.

3. Implement also the natural gradient descent, computing also the covariance matrix (2×2 in this case). Use a reasonable value of $\epsilon = 0.001$. Converge should occur within 100 steps. You can use the function `delta = np.linalg.solve(S, f)`, where `delta` is a two-dimensional array that contains the parameter shift $\delta\theta/\Delta$.

3.3 Restricted Boltzmann Machines

In the last decade we have seen the convergence of fields that were previously seen distant: quantum Monte Carlo and Machine Learning (ML). Carleo and Troyer first realized that a ML model, being a *function* could be also described to represent a *wave-function*. The first and simplest ML model is the Restricted Boltzmann Machine (RBM). The RBM is defined by a set of *visible* units $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)$, representing physical degrees of freedom (e.g., spin-1/2 variables with $\sigma_j \in \{-1, +1\}$), and a set of *hidden* units $\mathbf{h} = (h_1, \dots, h_M)$, also binary-valued (see Fig. 3.1).

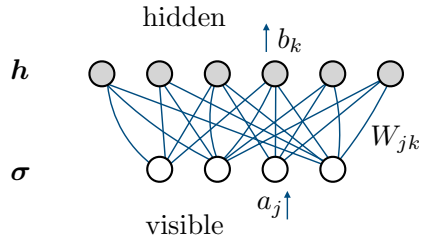


Figure 3.1: RBM network with visible and hidden neurons. Notice the absence of intra-layer connections.

The joint energy of a configuration of visible and hidden units is given by:

$$E(\boldsymbol{\sigma}, \mathbf{h}) = -\sum_{j=1}^{N} a_j \sigma_j - \sum_{k=1}^{M} b_k h_k - \sum_{j=1}^{N}\sum_{k=1}^{M} W_{jk}\sigma_j h_k, \tag{3.40}$$

where a_j, b_k, and W_{jk} are variational parameters: visible biases, hidden biases, and connection weights, respectively.

The RBM wavefunction is obtained by marginalizing over the hidden degrees of freedom:

$$\Psi(\boldsymbol{\sigma}) = \sum_{\mathbf{h}} e^{-E(\boldsymbol{\sigma}, \mathbf{h})}. \tag{3.41}$$

Due to the bipartite structure of the RBM, the sum over hidden variables can be performed exactly:

$$\Psi(\boldsymbol{\sigma}) = \exp\left(\sum_{j=1}^{N} a_j \sigma_j\right) \prod_{k=1}^{M} \sum_{h_k = \pm 1} \exp\left(h_k \left[b_k + \sum_{j=1}^{N} W_{jk}\sigma_j\right]\right) \tag{3.42}$$

$$= \exp\left(\sum_{j=1}^{N} a_j \sigma_j\right) \prod_{k=1}^{M} 2\cosh\left(b_k + \sum_{j=1}^{N} W_{jk}\sigma_j\right). \tag{3.43}$$

Thus, the final visible-only form of the variational wavefunction becomes:

$$\Psi(\boldsymbol{\sigma}) = \exp\left(\sum_{j} a_j \sigma_j\right) \prod_{k} 2\cosh\left(b_k + \sum_{j} W_{jk}\sigma_j\right). \tag{3.44}$$

This RBM ansatz can be interpreted as a compact representation of a correlated quantum

state. The product of cosh terms introduces multi-spin correlations of arbitrary range and order. In particular, the hidden units mediate effective interactions between visible spins in a compact, efficiently parameterized form.

In VMC, we require the logarithmic derivatives of the wavefunction with respect to each variational parameter:

$$\mathcal{O}_\theta(\boldsymbol{\sigma}) = \frac{1}{\Psi(\boldsymbol{\sigma})}\frac{\partial \Psi(\boldsymbol{\sigma})}{\partial \theta} = \frac{\partial \log \Psi(\boldsymbol{\sigma})}{\partial \theta}. \tag{3.45}$$

From the log-wavefunction:

$$\log \Psi(\boldsymbol{\sigma}) = \sum_j a_j \sigma_j + \sum_k \log \left[2 \cosh \left(b_k + \sum_j W_{jk} \sigma_j \right) \right], \tag{3.46}$$

we derive the following:

- Visible bias a_j:

$$\mathcal{O}_{a_j}(\boldsymbol{\sigma}) = \sigma_j \tag{3.47}$$

- Hidden bias b_k:

$$\mathcal{O}_{b_k}(\boldsymbol{\sigma}) = \tanh \left(b_k + \sum_j W_{jk} \sigma_j \right) \tag{3.48}$$

- Weight W_{jk}:

$$\mathcal{O}_{W_{jk}}(\boldsymbol{\sigma}) = \sigma_j \cdot \tanh \left(b_k + \sum_i W_{ik} \sigma_i \right) \tag{3.49}$$

These quantities are used in estimating the gradient of the energy with respect to the variational parameters or in constructing the quantum Fisher matrix for stochastic reconfiguration (called natural gradient in the ML jargon).

Finally, notice that the simple architecture described above can only describe real-valued and positive-valued states. Extension can be made and as of today, neural-network based ansatzes are the most accurate for many-body quantum spin systems. These are outside the scope of these notes.

Efficient Sampling: Use of Lookup Tables

Computing the full wavefunction ratio from scratch at ach Metropolis iteration can be expensive for large systems. To accelerate this, we store a *lookup table* of the effective arguments of the cosh, which we denote by:

$$\theta_k(\boldsymbol{\sigma}) = b_k + \sum_j W_{jk} \sigma_j. \tag{3.50}$$

Notice that the usage of look-up tables is recommended also for the Jastrow function of the previous exercises.

These quantities appear repeatedly in both the wavefunction and its derivatives:

$$\Psi(\boldsymbol{\sigma}) \propto \prod_k \cosh\left(\theta_k(\boldsymbol{\sigma})\right), \quad \mathcal{O}_{b_k} = \tanh\left(\theta_k\right), \quad \mathcal{O}_{W_{jk}} = \sigma_j \tanh\left(\theta_k\right). \tag{3.51}$$

When a single spin σ_s is flipped, from $\sigma_s \to -\sigma_s$, the configuration changes from $\boldsymbol{\sigma}^{(s)} \to \boldsymbol{\sigma}^{(\bar{s})}$, and each θ_k updates as:

$$\theta_k^{'} = \theta_k - 2W_{sk}\sigma_s. \tag{3.52}$$

This rule is derived by observing that the spin flip changes $\sigma_s \to -\sigma_s$, which affects only the terms in the sum over j where $j = s$. Hence, instead of recomputing all θ_k from scratch, we can efficiently update them using the cached values.
Using this strategy, both the Metropolis acceptance ratio and the logarithmic derivatives can be computed saving a factor N.

In implementation, one typically:

- Initializes θ_k at the start of sampling.
- Proposes a spin flip, computes the updated θ_k' using Eq. (3.52), and calculates the acceptance probability.
- If the move is accepted, overwrite the current θ_k values with the updated ones.

Exercise 3.3.1: VMC with a RBM (**)

Consider the same TFIM of Exercise 3.2.1 but the RBM of Eq. 3.44 as ansatz. Choose $M = \alpha N$, start with $\alpha = 1$ and $L = 8$ to not have too many parameters W's.

1. Write the VMC code with the RBM ansatz. Start with random and small parameters distributed gaussianly or uniformly in the range [-0.01,0.01].
2. Find the best variational energy and compare with the Jastrow case (using the same L, obviously). Try to increase α.

Chapter 4

Diffusion and Green's Function Monte Carlo

In this Chapter we describe the simplest idea to go beyond the variational method. We refer to the continuos space version, named Diffusion Monte Carlo. The lattice version of the method is usually dubbed Green's function Monte Carlo.

4.1 Imaginary-time projection

Consider a quantum system with Hamiltonian $\hat{H}$, acting on a Hilbert space of states $\phi_i(\boldsymbol{r})$, which are eigenfunctions of the Hamiltonian:

$$\hat{H}\phi_i(\boldsymbol{r}) = E_i\phi_i(\boldsymbol{r}), \tag{4.1}$$

where $E_0 < E_1 < E_2 < \ldots$, and $\boldsymbol{r} \in \mathbb{R}^{3N}$ denotes the positions of all N particles in continuous space.
Given an initial wavefunction $\psi_{\text{init}}(\boldsymbol{r})$ that admits the decomposition

$$\psi_{\text{init}}(\boldsymbol{r}) = \sum_i c_i\phi_i(\boldsymbol{r}), \tag{4.2}$$

we define an imaginary-time evolution:

$$\boxed{-\frac{\partial}{\partial\tau}\Psi(\boldsymbol{r},\tau) = \left(\hat{H} - E_T\right)\Psi(\boldsymbol{r},\tau)} \tag{4.3}$$

with formal solution

$$\Psi(\boldsymbol{r},\tau) = e^{-\tau(\hat{H}-E_T)}\psi_T(\boldsymbol{r}) = \sum_i c_i e^{-\tau(E_i-E_T)}\phi_i(\boldsymbol{r}). \tag{4.4}$$

As $\tau \to \infty$, excited-state components decay exponentially faster than the ground state, yielding

$$\Psi(\boldsymbol{r},\tau) \xrightarrow[\tau\to\infty]{} c_0\, e^{-\tau(E_0-E_T)}\phi_0(\boldsymbol{r}). \tag{4.5}$$

The parameter E_T is a *constant energy offset* introduced to prevent uncontrolled growth or decay of the norm. This factor does not affect the spatial form of the projected ground state $\phi_0(\boldsymbol{r})$, only its overall normalization.

The imaginary-time propagation thus filters out the ground state from any initial state ψ_{init} that has a non-zero overlap with ϕ_0, exponentially suppressing all contributions from higher-energy eigenstates.
This is the equivalent in continuos-time of the power method of Sect. 1.5. However, the aim is to have a non-exponetially scaling version of it, thus avoiding to work with full wavefunctions.

Let's us consider a generic wavefunction $\Psi(\boldsymbol{r}, \tau)$. Its imaginary-time evolution (of a imaginary-time Δ_τ, not necessarily small) can be written in integral form using the Green's function $G(\boldsymbol{r}, \boldsymbol{r}', \Delta_\tau)$:

$$\Psi(\boldsymbol{r}, \tau + \Delta_\tau) = \int d\boldsymbol{r}'\, G(\boldsymbol{r}, \boldsymbol{r}', \Delta_\tau)\, \Psi(\boldsymbol{r}', \tau). \tag{4.6}$$

The Green's function is defined as the matrix element (in coordinate basis) of the imaginary-time evolution operator,

$$G(\boldsymbol{r}, \boldsymbol{r}', \Delta_\tau) = \langle \boldsymbol{r} | e^{-\Delta_\tau (\hat{H} - E_T)} | \boldsymbol{r}' \rangle. \tag{4.7}$$

We now write the Hamiltonian as the sum of a kinetic and a potential term:

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^{N} \nabla_i^2 + V(\boldsymbol{r}), \tag{4.8}$$

where $V(\boldsymbol{r})$ is a general potential energy acting on the full configuration $\boldsymbol{r} \in \mathbb{R}^{3N}$. For this purpose we do not need to specify the different terms.

We now analyze its form in two limiting cases.

- *Purely kinetic case:* neglecting the potential, the Hamiltonian becomes $\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2$, and the Green's function reduces to a free-particle propagator:

  $$G_{\text{kin}}(\boldsymbol{r}, \boldsymbol{r}', \Delta_\tau) = \left(\frac{1}{2\pi \Delta_\tau} \right)^{\frac{3N}{2}} \exp\left(-\frac{|\boldsymbol{r} - \boldsymbol{r}'|^2}{2\Delta_\tau} \right). \tag{4.9}$$

  This corresponds to a diffusion kernel in configuration space. This can be derived directly inserting plane-waves into Eq. 4.7 or by using the results we derived in Sect. 2.7.3 with the following substutions $K \to G_{\text{kin}}$; $f = 0$; $\Delta \to \Delta_\tau$, $k_B T \to 1/2$, namely recognize that the imaginary time Schrodinger equation is formally the heat-equation, or a Fokker-Planck without drift.

- *Purely potential case:* when the kinetic term is neglected and $\hat{H} = V(\boldsymbol{r})$, the Green's function becomes diagonal in real space:

$$G_{\text{pot}}(\boldsymbol{r}, \boldsymbol{r}', \Delta_\tau) = \delta(\boldsymbol{r} - \boldsymbol{r}') \exp\left[-\Delta_\tau \left(V(\boldsymbol{r}) - E_T\right)\right]. \tag{4.10}$$

  This leads to a local exponential rescaling of the wavefunction amplitude. For regions where $V(\boldsymbol{r}) > E_T$, the wavefunction decays, while it grows where $V(\boldsymbol{r}) < E_T$.

These two limiting forms describe, respectively, *diffusion* and *amplitude modulation*, and will be used to construct an approximate propagator valid for small Δ_τ.

4.2 Trotter Formulas

We now introduce the Lie–Trotter product formula, which allows us to approximate the exponential of a sum of non-commuting operators. The exponential of a sum of operators $\hat{A} + \hat{B}$ can be expressed using the *Lie product formula*:

$$e^{\tau(\hat{A}+\hat{B})} = \lim_{n\to\infty} \left(e^{\frac{\tau}{n}\hat{A}} e^{\frac{\tau}{n}\hat{B}}\right)^n. \tag{4.11}$$

This motivates the first-order *Lie–Trotter approximation*[1], which holds for small time steps $\Delta_\tau = \tau/n$:

$$e^{\Delta_\tau(\hat{A}+\hat{B})} \approx e^{\Delta_\tau \hat{A}} e^{\Delta_\tau \hat{B}} + \mathcal{O}(\Delta_\tau^2). \tag{4.12}$$

In the context of imaginary-time evolution, we identify:

$$\hat{A} = -\frac{1}{2}\sum_{i=1}^{N} \nabla_i^2, \quad \hat{B} = V(\boldsymbol{r}) - E_T,$$

so that $\hat{H} - E_T = \hat{A} + \hat{B}$.

A more accurate second-order formula, known as the symmetric Trotter–Suzuki decomposition, is given by:

$$e^{-\Delta_\tau(\hat{A}+\hat{B})} \approx e^{-\frac{1}{2}\Delta_\tau \hat{B}} e^{-\Delta_\tau \hat{A}} e^{-\frac{1}{2}\Delta_\tau \hat{B}} + \mathcal{O}(\Delta_\tau^3). \tag{4.13}$$

This formula requires the same number of steps, plus one additional for $\hat{B}$, but achieves a higher accuracy in Δ_τ.[2]

[1] These formulas remain valid when $\hat{A}$ and $\hat{B}$ are *unbounded self-adjoint operators*, such as the kinetic and potential energy operators in quantum mechanics, under suitable domain conditions. For example, in the case $\hat{H} = \hat{T} + \hat{V}$, with $\hat{T} = -\frac{1}{2}\sum_i \nabla_i^2$ and $\hat{V} = V(\boldsymbol{r})$, the Lie product formula is rigorously valid on a dense subspace of the Hilbert space where both operators act. However, the finite-time Lie–Trotter approximation is more delicate: for it to be accurate, the wavefunction (on which the real or imaginary-time propagation is applied to) must have sufficient regularity.

[2] the $B/2$ terms at the boundaries merge into a full B step, except for the initial and final half-step.

Using the definiton of the kinetic and potential propagators in real space found above we get Putting these together, we obtain the full Green's function as:

$$G(\boldsymbol{r}, \boldsymbol{r}', \Delta) \approx G_{\text{pot}}(\boldsymbol{r}, \Delta_\tau/2) \cdot G_{\text{kin}}(\boldsymbol{r}, \boldsymbol{r}', \Delta_\tau) \cdot G_{\text{pot}}(\boldsymbol{r}', \Delta_\tau/2), \tag{4.14}$$

more explicitly,

$$G(\boldsymbol{r}, \boldsymbol{r}', \Delta_\tau) \approx \left(\frac{1}{2\pi\Delta_\tau}\right)^{\frac{3N}{2}} \exp\left[-\frac{(\boldsymbol{r}-\boldsymbol{r}')^2}{2\Delta_\tau}\right] \exp\left[-\frac{\Delta_\tau}{2}\left(V(\boldsymbol{r}) + V(\boldsymbol{r}') - 2E_T\right)\right]. \tag{4.15}$$

4.3 Implementing the Projection Stochastically: Simple Algorithms

After splitting the total imaginary time τ into n intervals of size Δ_τ, the time-evolved wavefunction is expressed as an n-fold convolution of Green's functions:

$$\Psi(\boldsymbol{r}_n, \tau) = \int d\boldsymbol{r}_{n-1} \cdots \int d\boldsymbol{r}_1 \int d\boldsymbol{r}_0\, G(\boldsymbol{r}_n, \boldsymbol{r}_{n-1}, \Delta_\tau) \cdots G(\boldsymbol{r}_2, \boldsymbol{r}_1, \Delta_\tau)\, G(\boldsymbol{r}_1, \boldsymbol{r}_0, \Delta_\tau)\psi_{\text{init}}(\boldsymbol{r}_0, 0), \tag{4.16}$$

where $\psi_T(\boldsymbol{r}_0, 0)$ is our initial state the we want to (imaginary)time-evolve, and the accumulated error is of order Δ_τ^2. The whole idea behind DMC is to perform this multi-dimensional integral stochastically.

4.3.1 Single Walker DMC

We introduce the concept of a *walker*: which is a pair of a multi-particle configuration $\boldsymbol{r}_k$ and a weight ω_k that evolves in imaginary time: $(\boldsymbol{r}_k, \omega_k)$. The index k denote the imaginary-time slice. At the beginning the configuration is sampled from $\psi_T(\boldsymbol{r}_0)$, and the weight is $\omega_0 = 1$. We can separately consider the action of the two Green functions, to a configuration:

- *Diffusion step*: The walker is displaced according to the kinetic part of the Green's function. This corresponds to a random Gaussian displacement:

  $$\boldsymbol{r}_{k+1} = \boldsymbol{r}_k + \boldsymbol{\eta}, \quad \text{with } \eta_i \sim \mathcal{N}(0, \Delta_\tau) \text{ where } 0 \le i \le 3N.$$

  We have already seen in Sect. 2.7.3 that the kinetic Green function applied is equivalent a diffusive dynamics.

- *Branching step*: The weight ω_k of the walker is updated based on the potential energy at the current and previous positions. According to the Green's function,

  $$\omega_{k+1} = \omega_k \times \exp\left[-\frac{\Delta_\tau}{2}\left(V(\boldsymbol{r}_{k+1}) + V(\boldsymbol{r}_k) - 2E_T\right)\right].$$

Thus, each walker propagates by random displacement and accumulates a weight depending on the local potential.

Now formally, the final wavefunction (that tends toward the ground state ϕ_0, could be reconstructed from the weighted distribution of the walkers at the final propagation time $k = n$,

$$\phi_0(\boldsymbol{r}) \leftarrow \Psi(\boldsymbol{r}, \tau = n\Delta\tau) = \langle \omega_n \delta(\boldsymbol{r}_n - \boldsymbol{r}) \rangle, \tag{4.17}$$

and the average is understood as average over independent dynamics of duration n steps. This follows from Eq. 4.16.
However, it is extremely inefficient to calculate properties from the wavefunction alone, so one is usually interested in the average value of the energy. We are going to provide a method to calculate the energy even for the single-walker method later, but before let's discuss a (small) improvement of the naive algorithm.

4.3.2 Multi-Walker DMC

The main problem of the above method is that the weight at step n wildly fluctuates as it is the multiplication of n factors, and is bound to diverge either to zero or infinity.

Now imaginary-time projection is obtained by evolving an ensemble of walkers $\{\boldsymbol{r}_k^{(i)}, \omega_k^{(i)}\}_{i=1}^{N_w}$, where N_w is the number of walkers at time slice k.

Instead of keeping explicit weights for each walker, the *many-walker formulation* simulates the effect of the potential via stochastic replication or deletion of walkers (Fig 4.1). This process is often called **branching**, and replaces the continuous weight update with a discrete population change:

- After the diffusion step $\boldsymbol{r}_{k+1}^{(i)} = \boldsymbol{r}_k^{(i)} + \boldsymbol{\eta}^{(i)}$, each walker's weight factor is computed:

$$w^{(i)} = \exp\left[-\frac{\Delta_\tau}{2}\left(V(\boldsymbol{r}_{k+1}^{(i)}) + V(\boldsymbol{r}_k^{(i)}) - 2E_T\right)\right].$$

- The number of copies $M^{(i)}$ of walker i to retain at step $k+1$ is determined by stochastic rounding:

$$M^{(i)} = \text{int}\left[\omega^{(i)} + z^{(i)}\right], \quad z^{(i)} \sim \mathcal{U}[0, 1).$$

  Walkers with $M^{(i)} = 0$ are removed; those with $M^{(i)} > 1$ are duplicated accordingly.[3]

However, the problem here is that the number of walker can explode. To stabilize the population size N_w, the reference energy E_T is dynamically adjusted during the evolution, ensuring that the average walker number remains approximately constant. This is also how the naive algorithm aim to evaluate the ground state energy. This is clearly inefficient and may work only for the simplest hamiltonians.

To summarize the problems of the DMC algorithms as presented so far are

- $V(\boldsymbol{r}) - E_T$ may be unbonded in some region, causing wild fluctuation of the weights, and therefore the walker population.

[3]This choice reflects that a walker can be spawned or killed probabilistically. If the weight is vanishing the survival probability is also vanishing, however, walkers with $\omega < 1$ can still survive.

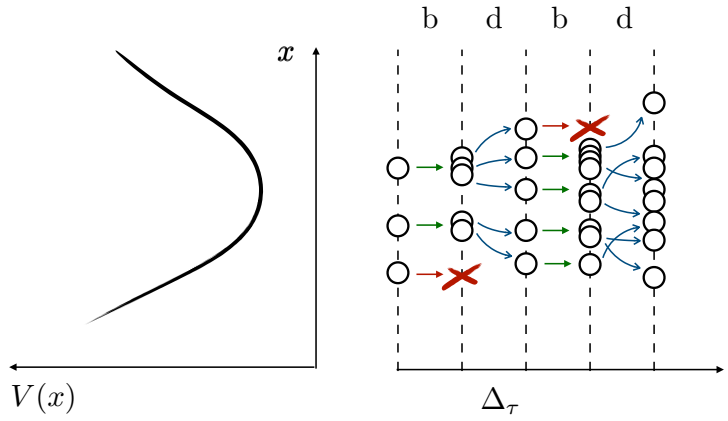


Figure 4.1: Some steps of the multi-walker DMC algorithm, with branching (b) and diffusion (d) steps.

- Calculating the energy as parameter that keeps the population stable is very unstable as well.
- So far, the formalism only works for positive defined wavefunctions.

4.4 Sign-Problem in Fermionic DMC and Fixed Node Approximation

4.4.1 1D Example

The diffusion Monte Carlo method evolves a wavefunction $\Psi(\boldsymbol{r}, \tau)$ in imaginary time toward the ground state $\phi_0(\boldsymbol{r})$ of the Hamiltonian. However, for fermionic systems, the true ground state is antisymmetric and therefore changes sign in configuration space. This introduces the notorious *sign problem*.

To illustrate this, consider a one-dimensional system with a real-valued fermionic ground state $\phi_0^{(\mathrm{F})}(x)$, which energy $E_0^{(\mathrm{F})}$ that has regions where it is positive and regions where it is negative (see Fig. 4.2). Notice that the ground-state of the same hamiltonian H, without nodes, $\phi_0^{(\mathrm{B})}(x)$, would have energy $E_0^{(\mathrm{B})} < E_0^{(\mathrm{F})}$ We can define the positive and negative parts of a wavefunction, $\psi(x)$ as:

$$\psi_+(x) = \frac{1}{2}\left(|\psi(x)| + \psi(x)\right), \qquad \psi_-(x) = \frac{1}{2}\left(|\psi(x)| - \psi(x)\right). \tag{4.18}$$

Clearly, $\psi(x) = \psi_+(x) + \psi_-(x)$, and these components are non-overlapping and non-negative (i.e., $\psi_+(x) \geq 0$, $\psi_-(x) \geq 0$).

Due to the linearity of the Schrödinger equation, one might hope to propagate ψ_+ and ψ_- independently and the substract the solutions:

$$\psi(x, \tau) = \psi_+(x, \tau) - \psi_-(x, \tau), \tag{4.19}$$

each obeying its own imaginary-time evolution under the same Hamiltonian.

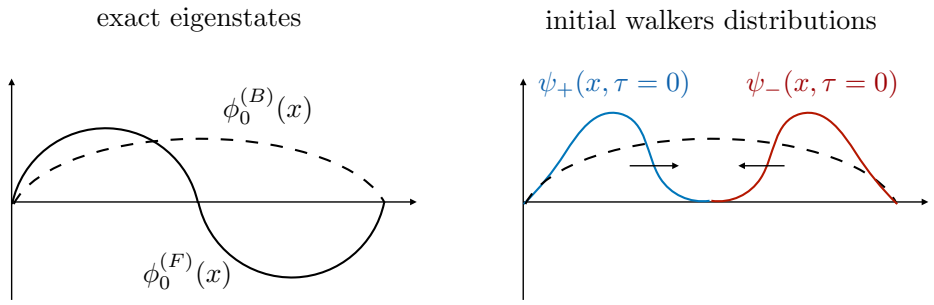


Figure 4.2: The shape of the wavefunctions discussed in the text.

Using the same derivation of Sect. 4.1, and assuming the expansion $\psi_\pm(x,\tau=0) = c^B\phi_0^{(\mathrm{B})}(x) \pm c^F\phi_0^{(\mathrm{F})}(x)$, we can see that the signal

$$\frac{|\psi_+ - \psi_-|}{|\psi_+ + \psi_-|} \propto e^{-\tau(E_0^{(B)} - E_0^{(F)})} \tag{4.20}$$

vanishes exponentially with the energy gap between the node-less and the wanted solution, and this is size-extensive.

4.4.2 Fixing the Nodal Structure

Algorithmically what happens is that the "positive" walkers, initialized in the ψ_+ region freely diffuse toward the negative region (of the exact ground state) and the opposite.
In principle, if we knew the nodal surface (that in 1D is just a point) we could constrain the positive and negative walkers to stay in their assigned regions. In this case, one could guess the nodal point by symmetry but in the general case it is not possible. Notice that the nodes of the many-body wavefunctions lies on a $3N-1$ dimensional surface (or $dN-1$ in arbitrary dimension d). The surface of the points in which the $3N$ (spinless) electrons coincide is instead $dN-d$ dimensional, so the coincidence and the nodal surface become equal only at $d=1$, and this is way one-dimensional fermionic problems can be solved exactly. The nodal surface must pass through the coincidence point but has a larger degree of freedom.

The Fixed Node approximation is assuming a pre-definite nodal surface (i.e the one coming from a Slater Determinant) and absorb walkers of a given sign that cross it. It is possible to demonstrate that the fixed node approximation is variational.

4.5 Importance Sampling DMC

The algorithm above misses the important zero-variance property that VMC has. We aim to have a more stable algorithm, and we need a different object to sample from. We define

the importance-sampled function:

$$f(\boldsymbol{r},\tau) = \psi_T(\boldsymbol{r})\psi(\boldsymbol{r},\tau). \tag{4.21}$$

Notice that it has the dimension of a wavefunction squared, that is, a more reasonable distribution to sample from. Again the trial function $\psi_T(\boldsymbol{r})$ is a fixed state, that could be the final state of a VMC simulation, and does not depend on time. It is possible to see that the imaginary time equation

$$-\frac{\partial}{\partial\tau}\psi(\boldsymbol{r},\tau) = \hat{H}\psi(\boldsymbol{r},\tau) = -\frac{1}{2}\nabla^2\psi(\boldsymbol{r},\tau) + V\psi(\boldsymbol{r},\tau), \tag{4.22}$$

(where with ∇^2 is the 3N-dimensional ∇_i^2 operators acting on all coordinates), can be recasted as an equation for f, not without some cumbersome calculations (that we reproduce in Appendix D). The equation reads

$$\boxed{-\frac{\partial}{\partial\tau}f(\boldsymbol{r},\tau) = -\frac{1}{2}\nabla^2 f(\boldsymbol{r},\tau) + \nabla(\mathbf{v}f) + E_L f(\boldsymbol{r},\tau)}, \tag{4.23}$$

where we define the so-called *quantum-velocity*,

$$\mathbf{v}(\boldsymbol{r}) = \frac{\nabla\psi_T(\boldsymbol{r})}{\psi_T(\boldsymbol{r})}, \tag{4.24}$$

(where with ∇ is the 3N-dimensional ∇_i operator acting on all coordinates)

Eq. 4.23 (without the last term) has the form of the Fokker-Plank equation with drift of Sect 2.7.3, which implies a drift-diffusion (dd) dynamics, given by Green's function

$$G_{\mathrm{dd}}(\boldsymbol{r},\boldsymbol{r}',\Delta_\tau) = \left(\frac{1}{2\pi\Delta_\tau}\right)^{\frac{3N}{2}} \exp\left(-\frac{|\boldsymbol{r}-\boldsymbol{r}'-\Delta_\tau\mathbf{v}(\boldsymbol{r})|^2}{2\Delta_\tau}\right). \tag{4.25}$$

This is valid only for short-time and corresponds to a Langevin dynamics with a force,

$$\boldsymbol{r}' = \boldsymbol{r} + \Delta_\tau\mathbf{v}(\boldsymbol{r}) + \boldsymbol{\eta}, \quad \text{with } \eta_i \sim \mathcal{N}(0,\Delta_\tau) \text{ where } 0 \le i \le 3N \quad \text{or} \tag{4.26}$$

$$\boldsymbol{r}' = \boldsymbol{r} + \Delta_\tau\mathbf{v}(\boldsymbol{r}) + \sqrt{\Delta_\tau}\,\boldsymbol{z}, \quad \text{with } z_i \sim \mathrm{unif}[0,1) \text{ where } 0 \le i \le 3N \tag{4.27}$$

The *branching* (b) Green's function becomes instead

$$G_{\mathrm{b}}(\boldsymbol{r},\boldsymbol{r}',\Delta_\tau) = \delta(\boldsymbol{r}-\boldsymbol{r}')\exp\left[-\Delta_\tau\left(E_L(\boldsymbol{r}) - E_T\right)\right], \tag{4.28}$$

where we have re-introduced the energy-shift E_T. We observe that

- The branching term features $E_L(\boldsymbol{r})$, rather than $V(\boldsymbol{r})$, a much smoother function, leading to greatly reduced statistical fluctations in the number of walkers.
- The drift increases the density of walkers when $\psi_T(\boldsymbol{r})$ is large, therefore automatically enforces the fixed-node constraints, as the walkers are repelled from the nodal surface. However, since we made the short-time approximation implicitely when we associate the Gaussian propagator to the Fokker-Plank-like equation 4.23, the walker *can* cross the node for finite Δ_τ.[4]

[4] To cure this we can introduce an acceptance step to the proposal induced by the Langevin dynamics to have an umbiased sampling. Notice that proposal is not symmetric because of the drift. This needs to be considered in the acceptance, see Eq. 2.53.

4.5.1 Computing the energy

Another advantage of this formalism is that allows us to compute the energy as average of the local energy as in VMC. Indeed we have that,

$$E_0 = \frac{\langle\phi_0|\hat{H}|\phi_0\rangle}{\langle\phi_0|\phi_0\rangle} = \lim_{\tau\to\infty} \frac{\langle\psi_T e^{-\tau/2\hat{H}}|\hat{H}|e^{-\tau/2\hat{H}}\psi_T\rangle}{\langle\psi_T e^{-\tau/2\hat{H}}|e^{-\tau/2\hat{H}}\psi_T\rangle} \tag{4.29}$$

$$= \lim_{\tau\to\infty} \frac{\langle\psi_T e^{-\tau\hat{H}}|\hat{H}|\psi_T\rangle}{\langle\psi_T e^{-\tau/\hat{H}}|\psi_T\rangle} = \frac{\langle\phi_0|\hat{H}|\psi_T\rangle}{\langle\phi_0|\psi_T\rangle} \tag{4.30}$$

$$= \lim_{\tau\to\infty} \frac{\int d\boldsymbol{r} f(\boldsymbol{r},\tau)E_L(\boldsymbol{r})}{\int d\boldsymbol{r} f(\boldsymbol{r},\tau)} \approx \frac{1}{M_{\rm smp}} \sum_{m=1}^{M_{\rm smp}} E_L(\boldsymbol{r}_m), \tag{4.31}$$

where $\boldsymbol{r}_m$ are sampled from *mixed* distribution $f(\boldsymbol{r},\tau)$.

This holds only for the energy. For a generic operator $\hat{A}$ this is not correct, however is can be shown that the following correction works in the limit $\phi_0 \approx \psi_T$:

$$\frac{\langle\phi_0|\hat{A}|\phi_0\rangle}{\langle\phi_0|\phi_0\rangle} \approx 2\frac{\langle\phi_0|\hat{A}|\psi_T\rangle}{\langle\phi_0|\psi_T\rangle} - \frac{\langle\psi_T|\hat{A}|\psi_T\rangle}{\langle\psi_T|\psi_T\rangle}. \tag{4.32}$$

4.5.2 Single-Walker Implementation

Now that we have found a much better way to perform the DMC algorithm we can return as well to the single-walker implementation. In this case we need to keep track of the weights ω_k, as the diffusion progress.
Instead of performing independent simulations of n steps, one can have a longer diffusion process, of lenght $\mathcal{N} >> \ell$, and consider all the possible sequence of lenght ℓ that one can find inside. For instance one could consider that the walker at step i is equilibrated and distributed according to f, and re-assign weight $\omega_i = 1$, and perform a projection up until step $\ell + i$. In practice, the full DMC simulation of lenght $\mathcal{N}$ is performed, and the

- local energies $E_L(\boldsymbol{r}_i) \equiv e_i$
- weights $b(\boldsymbol{r}_i) = \exp\left[-\Delta_\tau \left(E_L(\boldsymbol{r}_i) - E_T\right)\right] \equiv b_i$,

are stored in a file, with $0 \le i \le \mathcal{N}$.

Then the ground state energy can be read by post-processing the file and computing

$$E_0 = \lim_{\ell\to\infty} \frac{\sum_{i>i_0}^{\mathcal{N}} \mathcal{G}_i^\ell e_i}{\sum_{i>i_0}^{\mathcal{N}} \mathcal{G}_i^\ell}, \quad \text{with} \;\; \mathcal{G}_i^\ell = \prod_{j=1}^{\ell} b_{i-j}. \tag{4.33}$$

The error bars needs to be performed using the binning method.

Exercise 4.5.1: Single-walker DMC for the Helium atom (/***)**

Consider the Hamiltonian and trial wavefunction of Exercise 3.1.4.

1. Write a single-walker importance sampling code, implementing the drift-diffusion for the walker position dynamics, with e_i, and b_i as output.
2. Postprocess to read-out the ground state energy. Use a reasonable value for ℓ. A too large ℓ will cause instabilities.
3. (Bonus) Implement the Metropolis acceptance correction to reduce the time step error.
4. Compare with the VMC best energy and th exact value. Is the final DMC depending on the trial wave function parameter's choice?

4.6 Green's Function Monte Carlo

The counterpart of DMC for lattice models is called *Green's function MC* (GFMC). We are not going to reintroduce the whole idea of stochastically sampling the imaginary-time evolution, bus simply adapting the concept to the discrete world. The algorithm is exacly the stochastic implementation of the power method of Sect. 1.5. We are going to directly introduce the importance sampling variant. Notice that now we are in a discrete basis set. Given a configuration $|x\rangle$, we denote its connected elements by $|x'\rangle$. These are the L configurations for which the matrix element

$$H_{x,x'} = \langle x'|H|x\rangle \tag{4.34}$$

is non zero, $|H_{x,x'}| \neq 0$.

The discrete evolution of the walker, x in imaginary time follows a stochastic process generated by the Green's function,

$$G_{x,x'} = \langle x'|\Lambda 1 - H|x\rangle = \Lambda \tag{4.35}$$

The Green function matrix elements must be positive definite to have the meaning of a transition probability, in case this is not possible, then we have again a *sign problem.* A sufficiently large value of Λ is introduced to have always positive diagonal elements.
We then introduce the *importance sampling* GFMC:

$$\bar{G}_{x,x'} = \frac{\psi_T(x')}{\psi_T(x)} G_{x,x'} \tag{4.36}$$

The transition probability matrix used to update a walker is given by

$$p_{x,x'} = \bar{G}_{x,x'}/b_x \tag{4.37}$$

$$b_x = \sum_{x'} \bar{G}_{x,x'} = \Lambda - E_L(x) > 0 \tag{4.38}$$

The normalization factor b_x is the weight of the configuration x, and therefore the object $p_{x,x'}$, being positive-definite by hypotesis, has the meaning of a transition probability. Notice that we can numerically compute $\bar{G}_{x,x'}$ because we can evaluate to numerical precision the ratio $\psi_T(x')/\psi_T(x)$.

In practice, at configuration x, we need to select one of the connected configuration x' with probability $p_{x,x'}$. Operatively, this is simply done by computing and stacking all the $p_{x,x'}$'s (which sum gives 1) and extract a random number z between 0 and 1 and see in which interval it falls. For instance, suppose that $a, b, \cdots$ are the connected elements to x. We ideally divide the segment [0,1) as it follows

$$[0, p_{x,x}) \bigcup [p_{x,x}, p_{x,x} + p_{x,a}) \bigcup [p_{x,x} + p_{x,a}, p_{x,x} + p_{x,a} + p_{x,b}) \cdots . \tag{4.39}$$

The single-walker implementation directly follows the recipe of Sect. 4.5.2 but with these new ingredients.

Exercise 4.6.1: GFMC for the TFIM (**)

Consider the TFIM of Exercise 3.2.1.

1. Write a single-walker GFMC code that implements the walk in imaginary time. Use $L = 20$ to begin with.
2. Postprocess the e_x and b_x value to obtain the ground state energy.

Chapter 5

Path Integral Monte Carlo

In this Chapter we are going to present QMC algorithms based on the path-integral formalism of quantum mechanics. These are used to simulate finite-temperature systems but also ground states. As usual we will cover both the continuos and the discrete model worlds.

5.1 Finite Temperature Formalism

The equilibrium properties of a quantum system at temperature $T = 1/\beta$ are encoded in the density matrix:

$$\hat{\rho} = e^{-\beta \hat{H}}, \tag{5.1}$$

where $\hat{H}$ is the Hamiltonian operator. The partition function is given by

$$Z = \mathrm{Tr}(e^{-\beta \hat{H}}). \tag{5.2}$$

Let $\{\phi_i\}$ denote the complete orthonormal set of eigenstates of $\hat{H}$, such that

$$\hat{H}|\phi_i\rangle = E_i|\phi_i\rangle. \tag{5.3}$$

In this basis, the density matrix can be written as

$$\hat{\rho} = \sum_i e^{-\beta E_i} |\phi_i\rangle\langle\phi_i|, \tag{5.4}$$

and the partition function becomes

$$Z = \sum_i e^{-\beta E_i}. \tag{5.5}$$

If we instead use the coordinate basis $\{|x\rangle\}$, where x denotes a configuration of the system, the matrix elements of the density operator are given by

$$\rho(x, x') = \langle x|e^{-\beta \hat{H}}|x'\rangle. \tag{5.6}$$

In the case of discrete coordinates, the trace becomes a sum:

$$Z = \sum_x \rho(x, x), \tag{5.7}$$

while for continuous degrees of freedom, it becomes an integral:

$$Z = \int dx\, \rho(x, x). \tag{5.8}$$

The expectation value of an operator $\hat{O}$ at finite temperature is given by

$$\langle \hat{O} \rangle = \frac{1}{Z} \mathrm{Tr} \left(\hat{O} e^{-\beta \hat{H}} \right), \tag{5.9}$$

which in the basis of the eigenstates becomes

$$\langle \hat{O} \rangle = \frac{1}{Z} \sum_i \langle \phi_i | \hat{O} | \phi_i \rangle e^{-\beta E_i} \tag{5.10}$$

whereas, in the coordinate basis, this becomes

$$\langle \hat{O} \rangle = \frac{1}{Z} \int dx\, \langle x | \hat{O} e^{-\beta \hat{H}} | x \rangle, \tag{5.11}$$

or, when $\hat{O}$ is diagonal in the coordinate basis,

$$\langle \hat{O} \rangle = \frac{1}{Z} \int dx\, O(x)\, \rho(x, x). \tag{5.12}$$

In general, the Hamiltonian may consist of several non-commuting parts,

$$\hat{H} = \hat{H}_1 + \hat{H}_2 + \hat{H}_Q, \tag{5.13}$$

where $\hat{H}_1$, $\hat{H}_2$, and $\hat{H}_Q$ do not commute with each other. In the following, we will explicitly develop the formalism needed to handle such cases directly in the basis chosen.

5.2 Path Integrals in Continuos Variable Models

We start from the continuos variable case since we have most of the ingredients already prepared in Chapter 4. Let's start from the density matrix,

$$\hat{\rho}(\boldsymbol{r}, \boldsymbol{r'}; \beta) = \langle \boldsymbol{r} | e^{-\beta \hat{H}} | \boldsymbol{r'} \rangle = \sum_i \phi^*(\boldsymbol{r}) \phi(\boldsymbol{r'}) e^{-\beta E_i}, \tag{5.14}$$

where we explicitely indicate the parameter β, and we used Eq 5.4 for the last step. However, given that we do not know[1] in principle the eigenstates of the Hamiltonian, we need to try to calculate the matrix element directly in the coordinate basis. Gladly we already encountered

[1] This representation is useful to get exact results from exact diagonalization of small systems.

this object in the previous Chapter, see Eq. 4.7, namely the Green's function $G(\boldsymbol{r}, \boldsymbol{r}')$.

We observe that the β parameters can be understood as an *imaginary time* finite interval, therefore we split it in P smaller intervals[2] of size $\Delta_\tau = \beta/P = 1/(PT)$. The matrix element between two configurations $\boldsymbol{r} \equiv \boldsymbol{r}_0$ and $\boldsymbol{r'} \equiv \boldsymbol{r}_P$ can be written in the same we wrote Eq. 4.16,

$$\rho(\boldsymbol{r}_0, \boldsymbol{r}_P; \beta) = \int d\boldsymbol{r}_1 \int d\boldsymbol{r}_2 \, \cdots \int d\boldsymbol{r}_{P-1} \, \rho(\boldsymbol{r}_0, \boldsymbol{r}_1, \Delta_\tau)\rho(\boldsymbol{r}_1, \boldsymbol{r}_2, \Delta_\tau) \cdots \rho(\boldsymbol{r}_{P-1}, \boldsymbol{r}_P, \Delta_\tau), \tag{5.15}$$

with the difference that, this time we do not integrate over $\boldsymbol{r}_0$ since it is a fixed input. Performing the Trace simply means to identify $\boldsymbol{r}_0 = \boldsymbol{r}_0$ and integrate over this missing coordinate, hence

$$Z_\beta = \int d\boldsymbol{r}_0 \int d\boldsymbol{r}_1 \int d\boldsymbol{r}_2 \, \cdots \int d\boldsymbol{r}_{P-1} \, \rho(\boldsymbol{r}_0, \boldsymbol{r}_1, \Delta_\tau)\rho(\boldsymbol{r}_1, \boldsymbol{r}_2, \Delta_\tau) \cdots \rho(\boldsymbol{r}_{P-1}, \boldsymbol{r}_0, \Delta_\tau). \tag{5.16}$$

So far, this is still exact. The reason why we introduce this format it because we hope to compute the *short-time* density matrix $\rho(\boldsymbol{r}_k, \boldsymbol{r}_{k+1}, \Delta_\tau)$, with $0 \leq k \leq P-1$, using the Trotter-Lie approximation, thus separaring the two non-commuting parts of $\hat{H}$.

Inserting the results obtained in Chapter 4 we readily obtain

$$Z \approx \left(\frac{m}{2\pi\Delta_\tau}\right)^{\frac{3NP}{2}} \int \prod_{j=0}^{P-1} d\boldsymbol{r}_j \, \exp\left[-\sum_{j=0}^{P-1}\left(\frac{m}{2\Delta_\tau}(\boldsymbol{r}_{j+1} - \boldsymbol{r}_j)^2 + \Delta_\tau \, V(\boldsymbol{r}_j)\right)\right], \tag{5.17}$$

where we assumed that the position $\boldsymbol{r}$, collects all the $3N$ system's coordinates. In arbistrary dimensions one simply replaces $3 \to d$. Notice that the only gain we make in rewriting the partition function adding $P-1$ new replicas is that we can perform the multi-dimensional integral using MC.

5.2.1 Ring-polymer isomorphism

In the path integral formulation of quantum statistical mechanics, the thermal properties of a particle are obtained by summing over all possible paths $\boldsymbol{r}(\tau)$ the particle can take in imaginary time $\tau \in [0, \beta$, where $\beta = 1/(k_B T)$. A *path* refers to the projection of trajectory of the particle in configuration space, parameterized by imaginary time. The quantum partition function is expressed as a path integral over these trajectories, weighted by the Euclidean action:

$$Z = \int \mathcal{D}[\boldsymbol{r}(\tau)] \, e^{-S[\boldsymbol{r}(\tau)]}, \tag{5.18}$$

where the action is defined as

$$S[\boldsymbol{r}(\tau)] = \int_0^\beta \left[\frac{m}{2}\left(\frac{d\boldsymbol{r}}{d\tau}\right)^2 + V(\boldsymbol{r}(\tau))\right] d\tau, \tag{5.19}$$

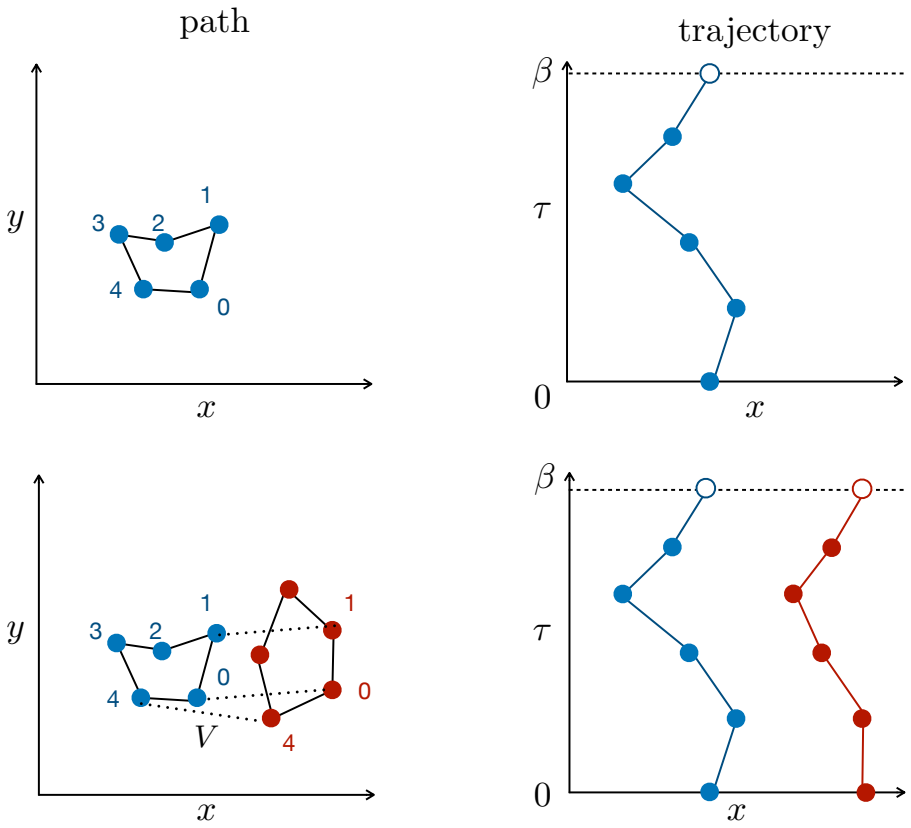


Figure 5.1: Path and trajectories in imaginary time for a single and two particles systems. The time is discretized using 5 steps for visualization sake. The inter-particle potential V acts only at *equal-times*. To be concrete: it would not be a problem in this sketch if the position of the left particle at imaginary time 1 overlaps with the replica number 5 of the right one. The trajectory plot is also called *world-line* representation. Notice the periodic-boundary conditions in imaginary-time.

and we ideally denote a trajectiory as a continuos-time object.

To evaluate the path integral numerically, we discretized imaginary time into P slices of width $\Delta_\tau = \beta/P$, and approximate the path $\boldsymbol{r}(\tau)$ by a set of P coordinates $\{\boldsymbol{r}_0, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_{P-1}\}$. These coordinates are interpreted as the positions of P replicas (or "beads") of the particle at successive imaginary time points. The discrete action given by

$$S_P[\{\boldsymbol{r}_j\}] = \sum_{j=0}^{P-1} \left[\frac{m}{2\Delta_\tau} (\boldsymbol{r}_{j+1} - \boldsymbol{r}_j)^2 + \Delta_\tau \, V(\boldsymbol{r}_j) \right], \tag{5.20}$$

and periodic boundary conditions $\boldsymbol{r}_P = \boldsymbol{r}_0$ to reflect the trace in the partition function.

[2] we use the symbol Δ_τ to maintain consistency with the previous Chapter. The β of this Chapter is the τ of DMC.

The discretized path integral has a compelling classical analogy: it maps exactly to a classical system of P particles connected in a ring by harmonic springs. Each bead $\boldsymbol{r}_j$ represents the configuration of the quantum particle at the imaginary time slice $\tau_j = j\Delta_\tau$, and the harmonic terms $(\boldsymbol{r}_{j+1} - \boldsymbol{r}_j)^2$ arise from the kinetic energy operator. The potential energy $V(\boldsymbol{r}_j)$ is evaluated independently at each bead, as if each bead "feels" the external potential separately. This structure is known as the *ring-polymer*.

It's important to distinguish between *imaginary time* τ and the *simulation time* used during the Monte Carlo process. Imaginary time serves as a label indexing the beads in the ring-polymer, each bead corresponds to a different point in the discretized path in imaginary time. In contrast, simulation time refers to the sequence of configurations generated during the Monte Carlo sampling process. In PIMC, we use suitable Monte Carlo moves to explore the configuration space of paths. The goal of this sampling is to generate an ensemble of ring-polymers whose statistical distribution reproduces the quantum thermal density matrix.

From the classical isomorphism, we gain an intuitive picture of quantum fluctuations. At high temperatures (small β), the spring constants between beads are stiff, and the ring collapses into a tight cluster resembling a classical particle. At low temperatures (large β), the springs become softer, allowing the ring to spread out in configuration space. This spread corresponds to quantum delocalization: the particle effectively explores a wider region of space due to its quantum nature. The width of the ring reflects the extent of the particle's zero-point motion and thermal fluctuations.

5.2.2 A Simple Energy Estimator

Physical observables are computed as ensemble averages over the space of discretized paths. For the total energy E, one commonly used approach is the *thermodynamic estimator*, which is derived by taking the derivative of the partition function with respect to the inverse temperature β. Specifically, the total energy is given by:

$$E = -\frac{\partial}{\partial\beta} \ln Z. \tag{5.21}$$

Evaluating this expression using the discretized path integral form yields the thermodynamic estimator for a single particle in d dimensions:

$$E_{\text{thermo}} = \left\langle \frac{dN}{2\Delta_\tau} - \frac{m}{2P} \sum_{j=0}^{P-1} \frac{(\boldsymbol{r}_{j+1} - \boldsymbol{r}_j)^2}{\Delta_\tau^2} + \frac{1}{P} \sum_{j=0}^{P-1} V(\boldsymbol{r}_j) \right\rangle, \tag{5.22}$$

where the angle brackets denote an ensemble average over ring-polymer configurations sampled during the Monte Carlo simulation.

The estimator consists of two contributions: a kinetic term (involving the squared distances between adjacent beads) and a potential term (averaged over all beads). This estimator is not very efficient at the interesting limit of small Δ_τ, as the first term clearly diverges as $1/\Delta_\tau$ and this needs to be cancelled by the also diverging second term. The*virial estimator*

provide a better performance, but its derivation is pretty cumbersome and is not featured here[3].

5.2.3 Monte Carlo Sampling and Path Updates

Efficient sampling of the path space is essential for the convergence the simulations. The goal is to construct a Markov chain whose stationary distribution is proportional to the discretized path integral weight e^{-S_P}. Various Monte Carlo move types are employed to explore the configuration space of ring polymers, ranging from simple local updates to more global proposals.

The most basic type of Monte Carlo move is the **local** bead displacement. In this move, a single bead $\boldsymbol{r}_j$ is randomly displaced to a new position $\boldsymbol{r}'_j = \boldsymbol{r}_j + \delta$, where δ is drawn from a uniform or Gaussian distribution centered at zero. The Metropolis acceptance probability for this move is given by

$$A = \min\left(1, \exp\left[-\Delta S_j\right]\right), \tag{5.23}$$

where $\Delta S_j = S_P[\ldots, \boldsymbol{r}'_j, \ldots] - S_P[\ldots, \boldsymbol{r}_j, \ldots]$ is the change in the discretized action resulting from the displacement of bead j. Only the terms in the action involving $\boldsymbol{r}_{j-1}, \boldsymbol{r}_j, \boldsymbol{r}_{j+1}$ need to be recomputed.

Given that the system is purely classical we can use all range of **molecular dynamics** methods, with a thermostat. For instance we can use a simple overdamped Langevin dynamics as described in Sect 2.7.2, which for bead j, takes the form

$$\boldsymbol{r}'_j = \boldsymbol{r}_j - \gamma \nabla_{\boldsymbol{r}_j} S_P + \sqrt{2\gamma}\,\eta, \tag{5.24}$$

where γ is a MD step-size parameter, $\nabla_{\boldsymbol{r}_j} S_P$ is the local gradient of the action, and η is a Gaussian noise term with zero mean and unit variance.

To improve the efficiency of sampling long-wavelength fluctuations in the path, global updates such as the **bisection algorithm** are used. This method leverages the exact sampling of the free-particle (kinetic) path integral and updates multiple beads simultaneously.

In the bisection algorithm, a path segment between two fixed endpoints $\boldsymbol{r}_j$ and $\boldsymbol{r}_{j+\ell}$ is updated by recursively sampling intermediate points using the exact free-particle propagator. The total imaginary time between the endpoints is $\Delta = \ell \Delta_\tau$.

The midpoint $\boldsymbol{r}_{j+\ell/2}$, assuming ℓ is even, is sampled from the conditional probability distribution:

$$\boldsymbol{r}_{j+\ell/2} \sim \mathcal{N}\left(\frac{\boldsymbol{r}_j + \boldsymbol{r}_{j+\ell}}{2}, \frac{\Delta}{4m}\right), \tag{5.25}$$

where the variance $\sigma^2 = \Delta/(4m) = \ell\Delta_\tau/(4m)$ takes into account that the imaginary time step that separates the mid-point from the extrema is $\ell\Delta_\tau/2$, and is bigger than Δ_τ.
Once that this mid-point has been proposed, one then does the same recursively until all the

[3] See *Path integrals in the theory of condensed helium*, D. Ceperley, 1995

beads have been re-placed. In absence of a potential, this new path configuration would be accepted with probability 1. Once the full segment is proposed using the kinetic propagator, the move is accepted or rejected based on the change in potential energy along the segment:

$$A = \min\left(1, \exp\left[-\Delta S_V\right]\right), \tag{5.26}$$

where $\Delta S_V = \sum_{\text{segment}} \Delta_\tau \left[V(\boldsymbol{r}_j^{\text{new}}) - V(\boldsymbol{r}_j^{\text{old}})\right]$.

5.2.4 Path Integral Ground State

The Path Integral Ground State (PIGS) method, also known as Variational Path Integral (VPI), is a zero-temperature quantum Monte Carlo technique that projects the exact ground state from a trial wavefunction $\psi_T(\boldsymbol{r})$ using imaginary-time evolution. In contrast to finite-temperature path integral methods like PIMC, where paths are closed and periodic, PIGS paths are *open*, with boundary conditions determined by ψ_T.

PIGS is formally based on the imaginary-time projection of DMC:

$$\psi_0 \propto \lim_{\tau\to\infty} e^{-\tau\hat{H}}\psi_T. \tag{5.27}$$

To compute expectation values, the operator is symmetrically applied to both sides of $\hat{O}$:

$$\langle\hat{O}\rangle = \frac{\langle\psi_T|e^{-\tau\hat{H}}\,\hat{O}\,e^{-\tau\hat{H}}|\psi_T\rangle}{\langle\psi_T|e^{-2\tau\hat{H}}|\psi_T\rangle}. \tag{5.28}$$

This structure is mathematically equivalent to a finite-temperature path integral at inverse temperature $\beta = 2\tau$, but with the crucial difference that the ends of the path are fixed by ψ_T, not joined cyclically.

In the path integral language, this corresponds to an open polymer-like path $\{\boldsymbol{r}_0, \ldots, \boldsymbol{r}_P\}$, where the endpoints are weighted by the trial wavefunction: $\psi_T(\boldsymbol{r}_0)$ and $\psi_T(\boldsymbol{r}_P)$. Equivalently, there is an additional part of the asction,$-\ln\psi_T$ acting on the extremal beads.
Observables are typically measured at or near the center of the path, where the effects of the trial wavefunction at the endpoints have been sufficiently projected out. In the limit of large projection time τ, the central configurations reflect pure ground-state physics, independent of ψ_T.

Monte Carlo sampling in PIGS is performed similarly to PIMC, using local or global updates such as bead displacements and bisection. The main difference lies in the treatment of the endpoints: proposals that move $\boldsymbol{r}_0$ or $\boldsymbol{r}_P$ must include the ratio of trial wavefunction amplitudes to maintain detailed balance:

$$A = \min\left(1, \frac{\psi_T(\boldsymbol{r}_0')\psi_T(\boldsymbol{r}_P')}{\psi_T(\boldsymbol{r}_0)\psi_T(\boldsymbol{r}_P)} e^{-\Delta S}\right), \tag{5.29}$$

where ΔS is the change in the internal path action.

Although PIGS is formally exact in the limit $\tau \to \infty$, in practice only moderate projection times are needed if ψ_T has reasonable overlap with the ground state.

5.2.5 Conceptual Difference Between DMC and PIMC

We are in the position now to observe the difference in two algorithms that stems from the same idea: imaginary-time evolution.

In **DMC**, the system is represented by a large population of *walkers*, each corresponding to a single configuration in real space. These walkers are propagated forward in small imaginary-time steps, using a stochastic process. Here the algorithm evolves the system in real (Monte Carlo) time steps that correspond directly to evolution in imaginary time. Ground-state properties emerge from the long-time asymptotic behavior of the walker distribution.

In contrast, **PIMC** represents a single quantum system as an extended object in imaginary time: a closed *path* or *ring polymer* consisting of P time slices. All imaginary-time configurations are carried along simultaneously as part of the path. In this formulation, imaginary time is treated as an additional coordinate, and the *Monte Carlo simulation time* is distinct from imaginary time. The simulation explores the space of full paths through sampling updates.
This key difference reflects a broader conceptual split: DMC treats quantum dynamics as an evolving population, whereas PIMC treats it as a statistical ensemble over full imaginary-time trajectories.

5.3 Implementing Symmetries in Real-Space PIMC: Bosons and Permutations

What we described so far did not take into account quantum symmetry, but would be valid for "Boltzmannion", i.e. simulating nuclear quantum effects in atomic systems. The situation becomes more complicated for simulating bosons and fermions. In real-space path integral Monte Carlo, quantum indistinguishability is encoded directly into the configuration space by summing over all permutations of particle labels. For bosons, this implies that the thermal density matrix must be fully symmetric under exchange of any two particles. This fundamental symmetry plays a central role in describing phenomena such as Bose-Einstein condensation and superfluidity, and must be incorporated explicitly in the PIMC sampling.

The symmetrized partition function for N bosons is expressed as:

$$Z_B = \frac{1}{N!} \sum_{\mathcal{P}} \int d\boldsymbol{R}\, \rho(\boldsymbol{R}, \mathcal{P}\boldsymbol{R}; \beta), \tag{5.30}$$

where $\boldsymbol{R} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$ denotes the coordinates of all particles, β is the inverse temperature, and $\mathcal{P}$ is an element of the permutation group of N objects. Notice that here we changed notation from the Sections above, because of the need to explicitly define particles, i.e. they cannot be "hidden" inside the general multi-particle multi-coordinated array $\boldsymbol{r}$ used before. The key feature of this expression is that the density matrix connects each configuration $\boldsymbol{R}$ at imaginary time $\tau = 0$ to a permuted configuration $\mathcal{P}\boldsymbol{R}$ at $\tau = \beta$.

This implies that in the path integral representation, particle worldlines are not required to be closed individually, but rather can form permutation cycles. Each particle i starts at a

position $\boldsymbol{r}_i(0)$ and ends at the position $\boldsymbol{r}_{\mathcal{P}(i)}(\beta)$. The set of all worldlines therefore consists of one or more cyclic chains of polymers, corresponding to the decomposition of $\mathcal{P}$ into disjoint cycles.

This topological structure has a direct physical interpretation. At high temperatures (small β), only the identity permutation contributes significantly, and each particle forms an isolated, closed ring polymer. As the temperature is lowered, permutation cycles involving two or more particles become statistically relevant. At very low temperatures, macroscopic exchange cycles can appear.

In this picture, a path represents not just the thermal fluctuations of a single particle, but a space-time entangled configuration of several indistinguishable particles whose identities are only defined modulo permutation. This is a direct reflection of the quantum mechanical principle that the many-body wavefunction for bosons must be symmetric under exchange.

Incorporating these permutations into the Monte Carlo sampling requires moves that change the topology of the path. Since the path is discretized into P time slices, permutations affect the connectivity at the boundary between $\tau = 0$ and $\tau = \beta$.
A naive method would simply to sample possible permutation and close the paths accordingly without moving the beads, however these update would suffer from a very-high rejection.
A more clever way is the following (schematic):

- Pick an imaginary time window, of say, p slices, from index i to j, $p = j - i$.
- Find all the permutations (restricting a range of $N_{\max}$ particle to exchange, usually it is 2,3 or 4). Let's call this set $\mathcal{S}$.
- Compute the weight of each permutation $\mathcal{P}$ using for instance the free-particle density-matrix
$$w(\mathcal{P}) = \frac{\rho(\boldsymbol{R}_i, \mathcal{P}\boldsymbol{R}_j; p\Delta_\tau)}{\sum_{\mathcal{P}' \in \mathcal{S}} \rho(\boldsymbol{R}_i, \mathcal{P}'\boldsymbol{R}_j; p\Delta_\tau)} \tag{5.31}$$
- Select a permutation $\mathcal{P}_k$ using this relative probability table.

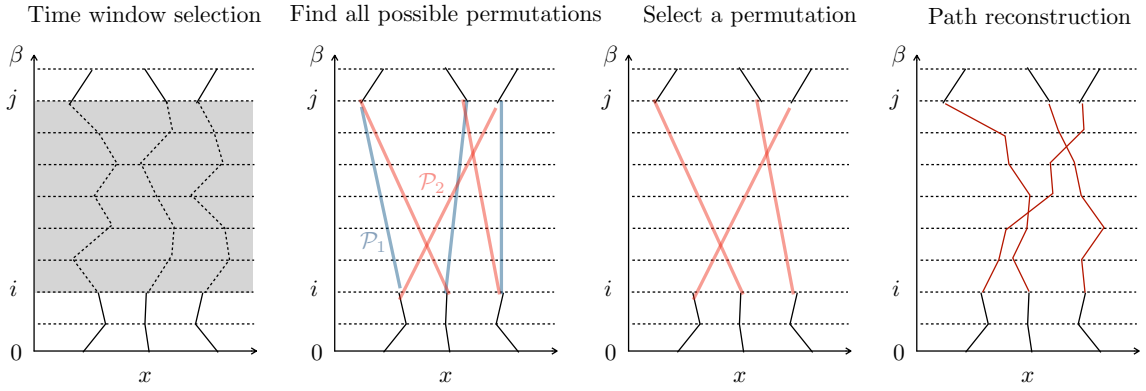


Figure 5.2: Sketch of the permutation sampling algorithm. For sake of visualization we only show two permutations among the all possible ones.

- Try to construct paths that connect the chosen endpoints using e.g bisection.
- Accept or reject using the difference in the actions.

The proposal step is shown in Fig. 5.2.

5.3.1 The Sign Problem in Continuos Space PIMC

With fermions the partition function is

$$Z_F = \frac{1}{N!}\sum_{\mathcal{P}} \text{sign}[\mathcal{P}] \int d\boldsymbol{R}\, \rho(\boldsymbol{R}, \mathcal{P}\boldsymbol{R}; \beta), \tag{5.32}$$

where $\text{sign}[\mathcal{P}] = +1\,(-1)$ for even (odd) permutations. The alternating signs in the sum cause large cancellations, which are at the core of the so-called *fermion sign problem*.

To perform Monte Carlo sampling, one often defines a modified (bosonic) partition function

$$Z_0 = \int dX\, |W(X)|, \tag{5.33}$$

where $W(X)$ is the configuration weight and X denotes the full set of paths in imaginary time. Configurations are sampled according to $|W(X)|$, and expectation values are reconstructed as

$$\langle \hat{A} \rangle = \frac{\langle \hat{A} S(X) \rangle_0}{\langle S(X) \rangle_0}, \qquad S(X) = \frac{W(X)}{|W(X)|}, \tag{5.34}$$

where $\langle \cdot \rangle_0$ denotes the average with respect to the absolute-value measure.
The *average sign*

$$S = \langle S(X) \rangle_0 = \frac{Z_F}{Z_0} \tag{5.35}$$

quantifies the severity of the problem. Since Z_F and Z_0 can be written in terms of free energies,

$$Z_F = e^{-\beta F_F}, \qquad Z_0 = e^{-\beta F_0}, \tag{5.36}$$

we obtain

$$S = \frac{Z_F}{Z_0} = e^{-\beta(F_F - F_0)}. \tag{5.37}$$

The difference $\Delta F = F_F - F_0$ is an extensive quantity, i.e. $\Delta F \propto N$, leading to

$$S \sim e^{-\beta N \Delta f}, \tag{5.38}$$

where $\Delta f = \Delta F / N$ is the free-energy density difference between the fermionic and bosonic systems.

Hence, the average sign decays *exponentially* with both inverse temperature β and the number of particles N. As a consequence, the statistical uncertainty in Monte Carlo estimators,

$$\frac{\Delta A}{A} \sim \frac{1}{S\sqrt{M}}, \tag{5.39}$$

grows exponentially unless the number of samples M increases accordingly. This exponential decay of S constitutes the essence of the fermion sign problem: although the sampling scheme is formally exact, the efficiency vanishes exponentially with system size or low temperature.

Fixed-node approximation. One approximation is to never allow paths crossing the nodal surface. This is similar to the FN-DMC idea.

5.4 Path Integrals in Spin Models

The derivation of path-integral formula in lattice models is a bit more cumbersome. Let's start with a simple case.

5.4.1 A Single Quantum Spin

Consider a single spin-$\frac{1}{2}$ particle with Hamiltonian

$$\hat{H} = -\gamma\hat{\sigma}^x - h\hat{\sigma}^z, \tag{5.40}$$

where $\hat{\sigma}^x, \hat{\sigma}^z$ are Pauli matrices, and $\gamma \geq 0, h$ are constants. The partition function at inverse temperature β is

$$Z = \mathrm{Tr}\left(e^{-\beta\hat{H}}\right). \tag{5.41}$$

Since $\hat{\sigma}^x$ and $\hat{\sigma}^z$ do not commute, we apply the Trotter decomposition.
Divide the imaginary time β into P slices of size $\Delta_\tau = \beta/P$, so that

$$e^{-\beta\hat{H}} = \left(e^{-\Delta_\tau\hat{H}}\right)^P \approx \left(e^{\Delta_\tau\gamma\hat{\sigma}^x}e^{\Delta_\tau h\hat{\sigma}^z}\right)^P + \mathcal{O}(\Delta_\tau^2), \tag{5.42}$$

To handle the non-commuting operators carefully, we insert resolutions of identity in the $\hat{\sigma}^z$ eigenbasis between every exponential factor.
Each Trotter step consists of two exponentials of non commuting operators, so for P steps, there are $2P$ operators, requiring $2P$ resolutions of identity:

$$\mathbb{I} = \sum_{s=\pm1} |s\rangle\langle s|, \tag{5.43}$$

where $\hat{\sigma}^z|s\rangle = s|s\rangle$, and $s = \pm1$. So the partition function reads,

$$\begin{aligned} Z = \sum_{\{s_1=\pm1; s_2=\pm1,\cdots\}} &\langle s_1|e^{\Delta_\tau\gamma\sigma^x}|s_2\rangle\langle s_2|e^{\Delta_\tau h\hat{\sigma}^z}|s_3\rangle\cdots\langle s_k|e^{\Delta_\tau\gamma\sigma^x}|s_{k+1}\rangle\langle s_{k+1}|e^{\Delta_\tau h\hat{\sigma}^z}|s_{k+2}\rangle \\ &\cdots\langle s_{2P-1}|e^{\Delta_\tau\gamma\sigma^x}|s_{2P}\rangle\langle s_{2P}|e^{\Delta_\tau h\hat{\sigma}^z}|s_1\rangle \end{aligned} \tag{5.44}$$

where the sequence of states[4] along imaginary time are $s_1, s_2, \ldots, s_{2P}$.

Since $\hat{\sigma}^z$ is diagonal in the $|s\rangle$ basis, we have

$$\langle s|e^{\Delta_\tau h\hat{\sigma}^z}|s'\rangle = e^{\Delta_\tau hs}\delta_{s,s'}, \tag{5.45}$$

[4] we use the index k, because we reserve i, j as spacial indexes.

therefore the actual number of slices is P, rather than $2P$. We could have done it at the beginning by inspecting the matrix element of $\langle s|e^{\Delta_\tau \gamma \hat{\sigma}^x} e^{\Delta_\tau h \hat{\sigma}^z}|s'\rangle = \langle s|e^{\Delta_\tau \gamma \hat{\sigma}^x}|s'\rangle e^{\Delta_\tau h s}$, but this way is more systematic and useful for hamiltonians beyond the TFIM.

The off-diagonal operator $e^{\Delta_\tau \gamma \hat{\sigma}^x}$ is a bit more cumbersome. The exponential of the Pauli-x matrix, can be computed exactly by Taylor expansion of the exponential and using the special properties of the Pauli matrices[5]

$$e^{\Delta_\tau \gamma \hat{\sigma}^x} = \cosh(\Delta_\tau \gamma)\, \mathbb{I} + \sinh(\Delta_\tau \gamma)\, \hat{\sigma}^x = \begin{pmatrix} \cosh(\Delta_\tau \gamma) & \sinh(\Delta_\tau \gamma) \\ \sinh(\Delta_\tau \gamma) & \cosh(\Delta_\tau \gamma) \end{pmatrix}. \tag{5.46}$$

Therefore,

$$\langle s|e^{\Delta_\tau \gamma \hat{\sigma}^x}|s'\rangle = \begin{cases} \cosh(\Delta_\tau \gamma), & s = s', \\ \sinh(\Delta_\tau \gamma), & s \neq s'. \end{cases} \tag{5.47}$$

Now we want to write this in a more compact form, and since the are two cases that needs to be considered, we could use the product ss' to perfom implicitely this if statement. We aim to write,

$$\langle s_k|e^{\Delta_\tau \gamma \hat{\sigma}^x}|s_{k+1}\rangle \propto e^{K s_k s_{k+1}}. \tag{5.48}$$

To find the value of K we use[6]

$$\frac{\cosh(\Delta_\tau \gamma)}{\sinh(\Delta_\tau \gamma)} = \coth(\Delta_\tau \gamma) = e^{2K} \implies K = \frac{1}{2} \log \coth(\Delta_\tau \gamma) = -\frac{1}{2} \log \tanh(\Delta_\tau \gamma). \tag{5.49}$$

Putting everything together, the partition function is

$$Z = \sum_{\{s_k = \pm 1\}} \prod_{i=1}^{P} \langle s_k|e^{\Delta_\tau \gamma \hat{\sigma}^x}|s_{k+1}\rangle \cdot e^{\Delta_\tau h s_i}, \tag{5.50}$$

$$\propto \sum_{\{s_k = \pm 1\}} e^{\Delta_\tau \left(\sum_{i=1}^{P} [J_\perp s_k s_{k+1} + h s_k] \right)}, \tag{5.51}$$

with periodic boundary $s_{P+1} = s_1$, and we defined

$$J_\perp = K/\Delta_\tau = -\frac{1}{2\Delta_\tau} \log \tanh(\Delta_\tau \gamma) > 0. \tag{5.52}$$

We can better rewrite the the partition function as

$$\boxed{Z \propto Z_P = \sum_{s_1 \pm 1} \sum_{s_2 \pm 1} \cdots \sum_{s_P \pm 1} e^{-\Delta_\tau H_P(s_1, s_2, \cdots, s_P)} := \sum_s W(s)} \tag{5.53}$$

[5] In this case $(\sigma^x)^2 = \mathbb{I}$, $(\sigma^x)^3 = \sigma^x$, etc..

[6] We aim to express the matrix element in the form: $\langle s|e^{\Delta_\tau \gamma \hat{\sigma}^x}|s'\rangle = C \cdot e^{K s s'}$ for some constants C and K, and $s, s' \in \{\pm 1\}$. This means $\cosh(\Delta_\tau \gamma) = C e^K$ and $\sinh(\Delta_\tau \gamma) = C e^{-K}$. We find K by dividing the two equations, we would find A by multiplying them: $A^2 = \cosh(\Delta_\tau \gamma) \cdot \sinh(\Delta_\tau \gamma) \quad \Rightarrow \quad A = \sqrt{\cosh(\Delta_\tau \gamma) \sinh(\Delta_\tau \gamma)}$. However, in the context of Monte Carlo, the constant A is not important as it would simply provide an non-essential re-normalization of the Z.

where[7],

$$H_P = -\sum_{k=1}^{P} \left(J_\perp s_k s_{k+1} + h s_k \right). \tag{5.54}$$

The quantum partition function can be written as an sum over the extended classical configurations $s = (s_1, s_2, ...)$, each of them contributing by a weight $W(s)$. Hence, this a *classical effective partition function* of a ferromagnetic 1D Ising model on a chain of length P with nearest-neighbor coupling $J_\perp > 0$. The fact that the interactions in imaginary time are ferromagnetic is important. For a more immediate understanding let us rewrite it as a function[8] of T,

$$J_\perp = -\frac{PT}{2} \log \tanh \left(\frac{\gamma}{PT} \right) > 0, \tag{5.55}$$

because $1/\Delta_\tau = PT$. This means that the ferromagnetic coupling decreases as γ increases. While it increases with increasing T. The size of the *quantum fluctuations* is proportional to the number of domain walls in the imaginary time direction, therefore is correctly following the expected behaviour in γ, and T.
Again, the classical isomorphism of a quantum system to a higher dimensional classical one is evident. In practice, by simulating with MC a classical 1D chain we can calculate finite temperature quantum properties of one quantum spin.

5.4.2 Evaluation of the off-diagonal operator

While computing diagonal operator is straightforward, given that the PIMC samples configurations already in the diagonal basis, the expectation value of $\langle \hat{\sigma}^x \rangle$ is more complicated. Intuitively we could expect that it can be evaluated tracing the number of domain walls in imaginary time, however we can be more precise[9]. Let's start with the definition

$$\langle \hat{\sigma}^x \rangle = \frac{\mathrm{Tr}\left[\hat{\sigma}^x e^{-\beta \hat{H}} \right]}{\mathrm{Tr}\left[e^{-\beta \hat{H}} \right]}. \tag{5.56}$$

We can re-do all the path-integral decomposition, but this time we have to consider the new matrix element $\langle s_k | \hat{\sigma}^x e^{\tau\gamma\hat{\sigma}^x} | s_{k+1} \rangle$ in the $\hat{\sigma}^z$ basis. Following the same steps above, we arrive at the following expression

$$\langle \hat{\sigma}^x \rangle = \sum_s W(s) \frac{1}{P} \sum_{k=1}^{P} \frac{\langle s_k | \hat{\sigma}^x e^{\tau\gamma\hat{\sigma}^x} | s_{k+1} \rangle}{\langle s_k | e^{\tau\gamma\hat{\sigma}^x} | s_{k+1} \rangle} \tag{5.57}$$

Using the same strategy above we can find that

$$\langle s_k | \hat{\sigma}^x e^{\tau\gamma\hat{\sigma}^x} | s_{k+1} \rangle = \begin{cases} \sinh(\tau\gamma), & \text{if } s_k = s_{k+1}, \\ \cosh(\tau\gamma), & \text{if } s_k \neq s_{k+1}. \end{cases} \tag{5.58}$$

[7] we extracted the minus sign to match the standard $-\beta H$ exponent.

[8] remember that the function $-\log\tanh(x)$ is always positive and "looks" like an $1/x$ function (just to get qualitatively its limits).

[9] We follow Krzakala et al, https://doi.org/10.1103/PhysRevB.78.134428

Therefore we can write in compact form

$$\frac{\langle s_k|\hat{\sigma}^x e^{\tau\gamma\hat{\sigma}^x}|s_{k+1}\rangle}{\langle s_k|e^{\tau\gamma\hat{\sigma}^x}|s_{k+1}\rangle} = \tanh(\tau\gamma)^{s_k s_{k+1}}. \tag{5.59}$$

Therefore the estimator we are looking for is

$$\langle\hat{\sigma}^x\rangle = \sum_s W(s)\frac{1}{P}\sum_{k=1}^{P}\tanh(\tau\gamma)^{s_k s_{k+1}} = \Big\langle\frac{1}{P}\sum_{k=1}^{P}\tanh(\tau\gamma)^{s_k s_{k+1}}\Big\rangle_{\text{PIMC}}. \tag{5.60}$$

Therefore we need to simply accumulate the average of the quantity in brakets $<\cdots>_{\text{PIMC}}$, along the PIMC simulation.

Exercise 5.4.1: PIMC of One Spin-1/2

Consider the Hamiltonian of Eq. 5.40. It's eigenvalues of the Hamiltonian are:

$$E_\pm = \pm\sqrt{\gamma^2 + h^2} \tag{5.61}$$

The partition function at inverse temperature β is:

$$Z = \text{Tr}\left(e^{-\beta\hat{H}}\right) = 2\cosh\left(\beta\sqrt{\gamma^2+h^2}\right) \tag{5.62}$$

The thermal average of the energy is:

$$\langle\hat{H}\rangle = -\frac{\partial}{\partial\beta}\log Z = -\sqrt{\gamma^2+h^2}\tanh\left(\beta\sqrt{\gamma^2+h^2}\right) \tag{5.63}$$

The thermal expectation values of the spin components are:

$$\langle\hat{\sigma}^x\rangle = \frac{\gamma}{\sqrt{\gamma^2+h^2}}\tanh\left(\beta\sqrt{\gamma^2+h^2}\right) \tag{5.64}$$

$$\langle\hat{\sigma}^z\rangle = \frac{h}{\sqrt{\gamma^2+h^2}}\tanh\left(\beta\sqrt{\gamma^2+h^2}\right). \tag{5.65}$$

1. Perform a PIMC simulation using the classical isomorphism to an effective 1D classical system, use a value of about $P = 64, 128$.
2. Use the cluster update to improve the efficiency.
3. Compare the calculated value for the energy and other expectation values with the exact ones.

5.4.3 Multidimensional Generalization

The generalization to the case of more than one spin is straighforward. In this case the added term $J_{ij}\hat{\sigma}_i^z\hat{\sigma}_j^z$ simply contributes to the diagonal part of the exponent, while the form of the off-diagonal operator remains the same. We can simply add the subscript s_k^i to denote the

value of the spin at imaginary-time slice k and spacial position $1 \leq i \leq L$. The *classical* Hamiltonian becomes

$$H_P = -\sum_{k=1}^{P}\left(\sum_{i=1}^{L} J_\perp s_k^i s_{k+1}^i + \sum_{i=1}^{L} h s_k^i + \sum_{i=1,j<i}^{L} J_{ij} s_k^i s_k^j\right). \tag{5.66}$$

Notice that the physical interaction J_{ij} couples the spins at equal-time, while the effective $J_\perp$ only occurs between the same spacial site i but at different imaginary times.

5.5 The Loop Algorithm

We show now how to generalize the path-integral contruction that we did for the ising model, and directly build-in an efficient cluster update that respect symmetries of a more general Hamiltonian when updating the world-lines.
Indeed, suppose that we are interested in the Heisenberg model instead. In this case, the world lines cannot break in imaginary time. Flipping just one spin in the extended lattice would not work.

5.5.1 The XXZ Model

We start by considering the XXZ model for two spin-$\frac{1}{2}$ particles. The Hamiltonian is given by

$$\hat{H} = -J_{xy}\left(\hat{\sigma}_1^x\hat{\sigma}_2^x + \hat{\sigma}_1^y\hat{\sigma}_2^y\right) - J_z\,\hat{\sigma}_1^z\hat{\sigma}_2^z, \tag{5.67}$$

where $\hat{\sigma}_i^{x,y,z}$ are the Pauli spin operators acting on site $i = 1, 2$. The parameters J_{xy} and J_z control the strength of the spin-spin interaction in the xy plane and along the z axis, respectively. A positive J_z corresponds to ferromagnetic coupling in the z direction.

In this basis, the non-zero matrix elements of the Hamiltonian $\hat{H}$ are[10]

$$\langle\uparrow\uparrow|\hat{H}|\uparrow\uparrow\rangle = -J_z, \qquad \langle\downarrow\downarrow|\hat{H}|\downarrow\downarrow\rangle = -J_z, \tag{5.68}$$

$$\langle\uparrow\downarrow|\hat{H}|\uparrow\downarrow\rangle = +J_z \qquad \langle\downarrow\uparrow|\hat{H}|\downarrow\uparrow\rangle = +J_z, \tag{5.69}$$

$$\langle\uparrow\downarrow|\hat{H}|\downarrow\uparrow\rangle = -2J_{xy}, \quad \langle\downarrow\uparrow|\hat{H}|\uparrow\downarrow\rangle = -2J_{xy}. \tag{5.70}$$

Now, we expand for small Δ_τ the exponential and we find the corresponding non-zero matrix elements of the operator $1 - \Delta_\tau\hat{H}$

$$W_1 = \langle\uparrow\uparrow|1-\Delta_\tau\hat{H}|\uparrow\uparrow\rangle = 1+\Delta_\tau J_z, \quad \langle\downarrow\downarrow|1-\Delta_\tau\hat{H}|\downarrow\downarrow\rangle = 1+\Delta_\tau J_z, \tag{5.71}$$

$$W_2 = \langle\uparrow\downarrow|1-\Delta_\tau\hat{H}|\uparrow\downarrow\rangle = 1-\Delta_\tau J_z, \quad \langle\downarrow\uparrow|1-\Delta_\tau\hat{H}|\downarrow\uparrow\rangle = 1-\Delta_\tau J_z, \tag{5.72}$$

$$W_3 = \langle\uparrow\downarrow|1-\Delta_\tau\hat{H}|\downarrow\uparrow\rangle = 2\Delta_\tau J_{xy}, \quad \langle\downarrow\uparrow|1-\Delta_\tau\hat{H}|\uparrow\downarrow\rangle = 2\Delta_\tau J_{xy}, \tag{5.73}$$

where the last row features the off-diagonal matrix elements, and we always assume that J_{xy} is positive, or can be made positive in a bipartite lattice.

[10] we neglect the factor $1/4 = 1/2 \times 1/2$.

This last step is needed because we are going to use the path-integral construction. We can directly step-up to a one-dimensional problem. Here we use a different partition for the two non-commuting sets of operators. We consider the spin-$\frac{1}{2}$ XXZ model on a one-dimensional chain of L sites with periodic boundary conditions. The Hamiltonian is given by

$$\hat{H} = -\sum_{i=1}^{L} \left[J_{xy} \left(\hat{\sigma}_i^x \hat{\sigma}_{i+1}^x + \hat{\sigma}_i^y \hat{\sigma}_{i+1}^y \right) + J_z \, \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z \right], \tag{5.74}$$

where $\hat{\sigma}_i^{x,y,z}$ are Pauli matrices acting on site i. To perform the path integral construction using Trotter-Suzuki decomposition, we split the Hamiltonian into two commuting parts:

$$\hat{H} = \hat{H}_{\text{even}} + \hat{H}_{\text{odd}}, \tag{5.75}$$

where

$$\hat{H}_{\text{even}} = \sum_{i \in \text{even}} \hat{h}_{i,i+1}, \tag{5.76}$$

$$\hat{H}_{\text{odd}} = \sum_{i \in \text{odd}} \hat{h}_{i,i+1}, \tag{5.77}$$

and the two-site bond Hamiltonian is

$$\hat{h}_{i,i+1} = -J_{xy} \left(\hat{\sigma}_i^x \hat{\sigma}_{i+1}^x + \hat{\sigma}_i^y \hat{\sigma}_{i+1}^y \right) - J_z \, \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z. \tag{5.78}$$

This decomposition ensures that all terms in $\hat{H}_{\text{even}}$ commute with each other, and likewise for $\hat{H}_{\text{odd}}$, since they act on disjoint bonds. The Trotter decomposition of the partition function

$$Z = \text{Tr} \left(e^{-\beta \hat{H}} \right)$$

is performed by alternating the even and odd operators at each time slice:

$$Z \approx \text{Tr} \left[\left(e^{-\Delta_\tau \hat{H}_{\text{even}}} e^{-\Delta_\tau \hat{H}_{\text{odd}}} \right)^P \right], \tag{5.79}$$

with $\beta = P\Delta_\tau$. This results in a path integral representation defined on a $(1+1)$-dimensional checkerboard lattice, where the two-site interactions $\hat{h}_{i,i+1}$ live on alternating plaquettes in spacetime.

We now insert complete sets of states in the $\hat{\sigma}^z$ basis between each exponential factor in the Trotter decomposition. This leads to a sum over spin configurations (replica configurations) defined on a $(1+1)$-dimensional checkerboard lattice.

Let $s_i^t \in \{\uparrow, \downarrow\}$ denote the spin at site i and imaginary time slice t. A full configuration is denoted by $\{s\} = \{s_i^t\}$, with $i = 1, \ldots, L$ and $t = 1, \ldots, P$, and we impose periodic boundary conditions in time: $s_i^{P+1} \equiv s_i^1$.

The partition function becomes

$$Z \approx \sum_{\{s\}} \prod_{t=1}^{P} \langle s^{t+1} | e^{-\Delta_\tau \hat{H}_{\text{even}}} e^{-\Delta_\tau \hat{H}_{\text{odd}}} | s^t \rangle, \tag{5.80}$$

where $|s^t\rangle$ denotes the spin configuration at time slice t, and the matrix element factorizes over the bonds due to the decomposition into commuting two-site terms. Each operator $e^{-\Delta_\tau \hat{H}_{\text{even/odd}}}$ acts non-trivially only on disjoint pairs of neighboring sites.

Thus, the total weight of a configuration $\{s\}$ is given by a product of local plaquette weights:

$$Z = \sum_{\{s\}} \prod_{p \in \text{plaquettes}} W_p(s_p), \tag{5.81}$$

where p labels the plaquettes of the checkerboard lattice, s_p is the local spin configuration on plaquette p, and $W_p(s_p)$ is the corresponding matrix element of $e^{-\Delta_\tau \hat{h}_{i,i+1}}$ for that bond and time slice.

Now, there are only six possible plaquette (4-spins) configurations (assuming we start from a valid full $L \times P$ configuration), and their weights are exacly the ones already found in Eq 5.68.

5.5.2 Graph representation and the loop break-ups

Label the six non-zero plaquette (vertex) types by

$$1^{\pm},\ 2^{\pm},\ 3^{\pm},$$

where the index distinguishes the structural class of the vertex and the sign indicates the spin orientation (+ for ↑, − for ↓). Their local weights are

$$W_{1^\pm} = 1 + \Delta_\tau J_z, \qquad W_{2^\pm} = 1 - \Delta_\tau J_z, \qquad W_{3^\pm} = 2\Delta_\tau J_{xy}.$$

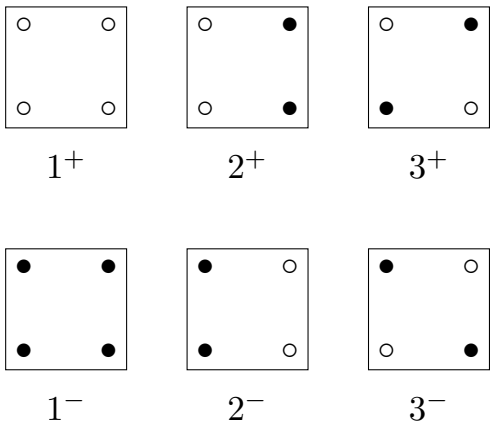
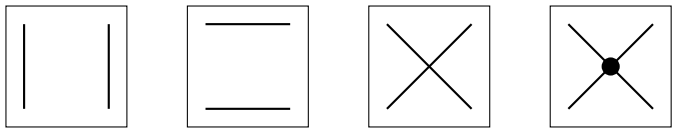


Figure 5.3: The six plaquette (vertex) configurations $1^\pm, 2^\pm, 3^\pm$. Empty circles denote ↑ spins, filled circles ↓. The two spins at the bottom/(top) correspond to imaginary time $t/(t+1)$.

Breakups weights. For each plaquette type i^σ introduce non-negative connection weights $w_{i^\sigma j^\rho} \geq 0$ describing the transition probability of type i^σ to a neighbouring plaquette of type j^ρ. The defining condition is

$$W_{i^\sigma} = \sum_{j^\rho} w_{i^\sigma j^\rho}, \tag{5.82}$$

where the sum runs over all neighbour types j^ρ that can be reached. The breakups can be visualized graphically in Fig. 5.4.



Vertical (V) Horizontal (H) Cross (X) Freezing (F)

Figure 5.4: Breakup (graph) types used in the loop algorithm: vertical (V), horizontal (H), cross (X), and freezing (F). The freezing breakup is represented like the cross but with a central circle, denoting that the loop is not allowed to flip across this plaquette.

The vertical breakup would allow the transition between plaquettes 1 and 2 (V = w_{12}), the horizontal breakup perform the transition from 2 to 3 (H = w_{23}), the cross breakup performs 1 to 3 (C = w_{13}); the frezing plaquettes simply flips all spins so it is (F = w_{ii}). Essentially, these are the generalization to a single spin flip for the ising model case. In these notes we neglect the plaquette of type 4 which would feature two spin up(down) at the bottom and two spins down(up) at the top. This is not allowed by the specific hamiltonian Eq. refeq:H:xxz.
The last thing we need to do is to explicitely assing a numerical value to the weights, given that they need to satisfy Eq. 5.82 We need to find solutions of

$$w_{11} + w_{12} + w_{13} = W_1 = 1 + \Delta_\tau J_z \tag{5.83}$$

$$w_{22} + w_{12} + w_{23} = W_2 = 1 - \Delta_\tau J_z \tag{5.84}$$

$$w_{33} + w_{13} + w_{23} = W_3 = 2\Delta_\tau J_{xy}. \tag{5.85}$$

$$\tag{5.86}$$

There are many of them and we try to minimize the weight of the freezing to decrease the autocorrelation time. We need to pay attention to not put too many weights to zero to allow for ergodic moves as well. In general, the best solution depends on the values of J_z and J_{xy}. It is possible to see that for the ferromagnetic isotropic model $J_z = J_{xy} = J$ a simple solution is

$$w_{12} = 1 - \Delta_\tau J; \quad w_{13} = 2\Delta_\tau J, \tag{5.87}$$

which satisfy the constraints. It is possible to convince ourselve that the algorithm is ergodic (notice that a loop configuration could be a worldline configuration as there is not horizontal breakup that would create a "circle" in space-time).

In the general case, a different choice for the weights must be done, including also the freezing plaquette (when $|J_z| > |J_{xy}|$.

Sampling rule (on-the-fly traversal). We sum-up the algorithm:

1. Pick a random starting leg (i, t) (or equivalently a plaquette of type i and an entrance leg).

2. At the current plaquette of type p sample the outgoing connection (i.e. choose the next plaquette type q) with probability

$$\Pr(\text{go to } q \mid \text{at } p) \;=\; \frac{w_{pq}}{W_p}. \tag{5.88}$$

   Operationally this means: sample which exit leg (or local breakup) to use according to the relative w_{pq}; follow that leg to the adjacent plaquette, which will have some type q.

3. Move to that neighbouring plaquette of type q, and repeat the same rule there (choose a continuation according to the row $w_{q,\bullet}$ and probabilities w_{qr}/W_q).

4. Stop when the traversal returns to the initial starting leg (closed loop). Flip the entire closed loop with the chosen probability (e.g. 1/2).

5.5.3 The Kandel–Domany Graph Framework

We now place the loop algorithm construction within the more general and elegant *Kandel–Domany (KD) framework* for cluster updates.

In the KD framework the partition function of a classical model is rewritten as a sum over configurations C and auxiliary "graph" variables G:

$$Z = \sum_C \sum_G W(C, G), \tag{5.89}$$

with the requirement that

$$\sum_G W(C, G) = W(C),$$

where $W(C) = e^{-\beta E[C]}$ is the original Boltzmann weight. One then samples first the graph G given C using probability

$$P_C(G) = \frac{W(C, G)}{W(C)},$$

and subsequently a new configuration C' given the fixed graph G, in a way that satisfies detailed balance.

A key simplification arises if the joint weight factorizes as

$$W(C, G) = \Delta(C, G)\, V(G), \tag{5.90}$$

where $\Delta(C, G) = 1$ if the graph G is compatible with configuration C, and zero otherwise, and $V(G) \geq 0$ is a graph-only weight. In that case, the update $C \to C'$ (while keeping G fixed) can be done simply by flipping clusters with probability 1/2 (heat-bath) or always (Metropolis), greatly simplifying the implementation.

This framework extends naturally to quantum (world-line) systems. One treats the world-line configurations as extended configurations C, and defines allowed local graph elements G (such as vertical, horizontal, cross, and sometimes freezing graphs) over plaquettes, such that the world lines remain intact and valid after applying G, i.e., $\Delta(C, G)$ enforces world-line continuity.

The weights $V(G)$ are chosen so that summing over all compatible graphs reproduces the plaquette weights $W_p(s_p)$, exactly as done above . Thus the KD framework provides a principled derivation of the breakup-decomposition used in loop algorithms: one picks a graph basis G, defines $\Delta(C, G)$, and solves for graph-weights $V(G)$ such that

$$W(C) = \sum_G \Delta(C, G)\, V(G).$$

5.5.4 Complexity of the Sign Problem in Lattice Models

Having introduced the average-sign ratio in and shown that it decays exponentially with system size and inverse temperature (see previous section), we now turn to a deeper complexity-theoretic argument originally given by Troyer and Wiese in 2004, which asserts that a *generic* solution of the sign problem would imply an unlikely collapse of complexity classes.

Consider a quantum lattice model whose partition function in a path-integral or world-line representation takes the form

$$Z = \sum_{c \in \Omega} W(c),$$

where the set of configurations Ω may include negative weights $W(c) < 0$. One may define the "bosonic version" of the model by replacing each weight by its absolute value,

$$Z_0 = \sum_{c \in \Omega} |W(c)|.$$

Then Monte Carlo sampling with weights $|W(c)|$ yields an average sign

$$\langle s(c) \rangle = \frac{Z}{Z_0}.$$

Troyer and Wiese show that if there existed a general algorithm that solves the sign problem (i.e., evaluates thermal averages in polynomial time *despite* the sign cancellations), then one could solve any problem in the class NP in polynomial time.

The argument proceeds as follows. They start from a known NP-complete decision problem: for example, given a classical three-dimensional Ising spin glass with Hamiltonian

$$H_{\text{glass}} = -\sum_{\langle j,k \rangle} J_{jk}\, \sigma_j\, \sigma_k,$$

with $\sigma_j = \pm 1$ and couplings $J_{jk} = 0, \pm J$. As mentioned in previous Chapter, the rigorous decision version of this optimization problem is to ask if the groundstate is lower than a threshold E_0.

Troyer and Wiese map this classical system to a quantum system with a sign problem. They replace the classical Ising spins by quantum spins, but chose the mapping $\sigma \to \hat{\sigma}^x$.

The random signs of the couplings are mapped to random signs of the off-diagonal matrix elements which cause a sign problem. The related bosonic model is the ferromagnet with all couplings $J_j k \leq 0$ that can be solved efficiently by the Loop Algorithm.

We find therefore a simple example in which the bosonic system is efficientlysolvable, while the fermionic system can be mapped to a NP-complete problem.

Appendix A

Sinc DVR Kinetic Energy Matrix

Here we use explicitely the mass μ, which can be the reduced mass for the diatomic system D_2. The goal is to represent the kinetic energy operator

$$\hat{T} = -\frac{\hbar^2}{2\mu}\frac{d^2}{dr^2}$$

on a uniform spatial grid using the DVR method, but using a better approximation. In the Sinc DVR (also known as the Fourier Grid Hamiltonian), we use a uniform grid with spacing Δr and total number of points N.

In the DVR approach, the wavefunction is approximated by its values at grid points $\psi(r_i)$. In the Sinc DVR, the basis functions are plane waves:

$$\phi_n(r) = \frac{1}{\sqrt{L}}e^{ik_n r}, \quad k_n = \frac{2\pi n}{L}, \quad n = -\frac{N}{2}, \ldots, \frac{N}{2} - 1$$

where $L = N\Delta r$ is the total length of the spatial domain, and periodic boundary conditions are assumed. In this basis, the kinetic energy operator is diagonal:

$$\langle \phi_n | \hat{T} | \phi_n \rangle = \frac{\hbar^2 k_n^2}{2\mu}$$

The DVR basis functions in real space are sinc functions centered at each grid point:

$$\chi_j(r) = \text{sinc}\left(\frac{\pi(r - r_j)}{\Delta r}\right)$$

To obtain the kinetic energy matrix in the position (DVR) basis, we compute:

$$T_{ij} = \left\langle \chi_i \left| -\frac{\hbar^2}{2\mu}\frac{d^2}{dr^2} \right| \chi_j \right\rangle$$

This can be evaluated analytically using Fourier analysis, resulting in the following expression for the kinetic energy matrix elements:

$$T_{ij} = \frac{\hbar^2}{2\mu(\Delta r)^2} \begin{cases} \frac{\pi^2}{3} & \text{if } i = j \\ \frac{2(-1)^{i-j}}{(i-j)^2} & \text{if } i \neq j \end{cases}$$

This result comes from projecting the kinetic operator onto a truncated Fourier basis and transforming back to the position representation using the orthonormality and completeness of the sinc functions.

Appendix B

Details of The Fokker Planck calculations

Transition probability $K(x'|x)$

We want to compute the integral

$$\int \frac{dz}{\sqrt{2\pi}}\, e^{-z^2/2}\, \delta\left(x' - x - \Delta f(x) - \sqrt{2\Delta k_B T}\, z\right) \tag{B.1}$$

Let us define

$$y = x' - x - \Delta f(x) \tag{B.2}$$

so that the delta function becomes

$$\delta\left(y - \sqrt{2\Delta k_B T}\, z\right) \tag{B.3}$$

Using the change-of-variable identity for the delta function,

$$\delta(az - b) = \frac{1}{|a|}\, \delta\left(z - \frac{b}{a}\right), \tag{B.4}$$

we obtain

$$\delta\left(y - \sqrt{2\Delta k_B T}\, z\right) = \frac{1}{\sqrt{2\Delta k_B T}}\, \delta\left(z - \frac{y}{\sqrt{2\Delta k_B T}}\right) \tag{B.5}$$

Substituting back into the integral, we have

$$\int \frac{dz}{\sqrt{2\pi}}\, e^{-z^2/2} \cdot \frac{1}{\sqrt{2\Delta k_B T}}\, \delta\left(z - \frac{y}{\sqrt{2\Delta k_B T}}\right) \tag{B.6}$$

Using the shifting property of the delta function, the integral evaluates to

$$\frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2\Delta k_B T}}\, \exp\left[-\frac{1}{2}\left(\frac{y}{\sqrt{2\Delta k_B T}}\right)^2\right] \tag{B.7}$$

Substituting $y = x' - x - \Delta f(x)$, we obtain the final result

$$\int \frac{dz}{\sqrt{2\pi}}\, e^{-z^2/2}\, \delta\left(x' - x - \Delta f(x) - \sqrt{2\Delta k_B T}\, z\right) = \frac{1}{\sqrt{4\pi\Delta k_B T}}\, \exp\left[-\frac{(x' - x - \Delta f(x))^2}{4\Delta k_B T}\right] \tag{B.8}$$

From Master equation to Fokker-Planck

We start from the Master equation for the probability distribution

$$P_{n+1}(x') = \int dx\, K(x'|x) P_n(x), \tag{B.9}$$

Now we use the symbol $P(x,t)$, rather than ρ_n just for convenience, we also use $k_B = 1$. We have therefore two integrals:

$$P(x', t+\Delta) = \int dx\, K(x'|x) P(x,t) = \int dx \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \delta(x' - x - \Delta f(x) - \sqrt{2\Delta T}\, z) P(x,t). \tag{B.10}$$

We perform the integral over x, recalling that $\int dx\, \delta(f(x)) = 1/|\partial_x f(x)|_{f(x)=0}$, and let's denote derivatives over space as $\partial_x f = f'$, we obtain:

$$P(x', t+\Delta) = \int dz\, \frac{e^{-z^2/2}}{\sqrt{2\pi}} \frac{1}{|1 + \Delta f'(x')|} P\left(x' - \Delta f(x') - \sqrt{2\Delta T}\, z\right). \tag{B.11}$$

Notice that, we have also replaced x with x' as the point where we evaluate the force, so nothing depends anymore on x. Now we expand the P up to second order (around the displacement $\Delta f(x') - \sqrt{2\Delta T} z$), because there is a $\sqrt{\Delta}$ term:

$$P\left(x' - \Delta f(x') - \sqrt{2\Delta T}\, z\right) \approx P(x') - P'(x') \left[\Delta f - \sqrt{2\Delta T} z\right] + \frac{P''}{2} \left[(\Delta f)^2 + 2\Delta T z^2\right] \tag{B.12}$$

we also expand $1/|1 + \Delta f'(x')| \approx 1 - \Delta f'(x')$, and put everything together

$$P(x', t+\Delta) = \int dz\, \frac{e^{-z^2/2}}{\sqrt{2\pi}} \left(1 - \Delta f'(x')\right) \left[P(x') - P'(x') \left[\Delta f - \sqrt{2\Delta T} z\right] + \frac{P''}{2} \left[(\Delta f)^2 + 2\Delta T z^2\right]\right]. \tag{B.13}$$

We now retain terms up the linear order in Δ:

$$P(x', t+\Delta) \approx \int dz\, \frac{e^{-z^2/2}}{\sqrt{2\pi}} \left[P(x') - \Delta f\ P'(x') - \Delta f'\ P(x') - P'(x)\sqrt{2\Delta T} z + P''(x') \Delta T z^2\right]. \tag{B.14}$$

Now, only few terms depend on z. We use that

$$\int_{-\infty}^{\infty} dz\, \frac{e^{-z^2/2}}{\sqrt{2\pi}} = 1, \quad \int_{-\infty}^{\infty} dz\, \frac{e^{-z^2/2}}{\sqrt{2\pi}} z = 0, \quad , \int_{-\infty}^{\infty} dz\, \frac{e^{-z^2/2}}{\sqrt{2\pi}} z^2 = 1, \tag{B.15}$$

so, the term with $\sqrt{\Delta}$-dependence vanish, while the other terms get multiplied by 1. We arrive to

$$P(x', t+\Delta) \approx P(x') - \Delta f\ P'(x') - \Delta f'\ P(x') + P''(x') \Delta T = P(x') + \Delta \left[-(f\ P(x'))' + P''(x') T\right] \tag{B.16}$$

So we have

$$\frac{P(x, t+\Delta) - P(x,t)}{\Delta} \approx -\frac{\partial}{\partial x} \left[f P(x,t)\right] + T \frac{\partial^2}{\partial x^2} P(x,t), \tag{B.17}$$

where we re-named the symbol x' to x, since only one appeared since several passages. This leads to the continuous-time Fokker-Planck equation in the limit $\Delta \to 0$:

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x} \left[f(x) P(x,t)\right] + T \frac{\partial^2 P(x,t)}{\partial x^2}. \tag{B.18}$$

Appendix C

Cusp Conditions

Electron-ion cusp

Consider the Hamiltonian for an electron near a nucleus located at $\boldsymbol{R}_I$:

$$\hat{H} = -\frac{1}{2}\nabla^2 - \frac{Z_I}{r}, \tag{C.1}$$

where $r = |\boldsymbol{r} - \boldsymbol{R}_I|$. We consider only the electron that is appraching the nucleus, so this is an effective two-body problem. We use radial coordinates centered on the nucleus. Near the nucleus, we assume that $\psi(\boldsymbol{r})$ depends only on r, so the Laplacian becomes

$$\nabla^2\psi = \frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\psi}{dr}\right). \tag{C.2}$$

The Schrödinger equation in the radial cordinate reads:

$$-\frac{1}{2}\left(\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\psi}{dr}\right)\right) - \frac{Z_I}{r}\psi = E\psi. \tag{C.3}$$

We now analyze the behavior as $r \to 0$. Assume a regular Taylor expansion:

$$\psi(r) = \psi(0) + r\psi'(0) + \mathcal{O}(r^2).$$

Inserting into the Laplacian gives:

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\psi'(r)\right) \approx \frac{1}{r^2}\frac{d}{dr}\left(r^2\psi'(0)\right) = \frac{1}{r^2}(2r\psi'(0)) = \frac{2\psi'(0)}{r}.$$

Substituting into the Schrödinger equation, we collect singular terms at leading order:

$$-\frac{\psi'(0)}{r} - \frac{Z_I\psi(0)}{r} + \mathcal{O}(1) = E\psi(0).$$

For the left-hand side to remain finite as $r \to 0$, the $1/r$ singularity must cancel:

$$\psi'(0) = -Z_I\psi(0), \tag{C.4}$$

which yields the Kato cusp condition as required.

Antiparallel-spin cusp condition

We consider only the bare two-electrons problem, neglecting all the rest of the interactions.

Near coalescence, we consider only the relative motion. In atomic units, the reduced mass of the two-electron system is $\mu = \frac{1}{2}$, and the effective Hamiltonian in the relative coordinate $\boldsymbol{r} = \boldsymbol{r}_i - \boldsymbol{r}_j$ becomes:

$$\hat{H}_{\text{rel}} = -\nabla^2 + \frac{1}{r}. \tag{C.5}$$

due to the reduced mass of the system, which is 1/2. Now we derive the cusp condition by imposing the absence of divergence of the local energy, following the derivation of Foulker e.t. al 2001.

We assume a functional for the trial function

$$\psi(\boldsymbol{r}) = e^{-u(r)} f(\boldsymbol{r}), \tag{C.6}$$

where $u(r)$ is a Jastrow correlation factor depending only on the interelectronic distance $r = |\boldsymbol{r}_i - \boldsymbol{r}_j|$, and $f(\boldsymbol{r})$ is a smooth, differentiable function as $r \to 0$. For antiparallel-spin electrons, the wavefunction does not vanish at coalescence, so we assume $f(\boldsymbol{r})$ is finite and non-zero near $r = 0$. The local energy is

$$E_L(\boldsymbol{r}) = \frac{H e^{-u(r)} f(\boldsymbol{r})}{e^{-u(r)} f(\boldsymbol{r})} \tag{C.7}$$

We compute the Laplacian:

$$\begin{aligned} \nabla^2 \psi &= \nabla^2 \left(e^{-u(r)} f(\boldsymbol{r}) \right) \\ &= e^{-u(r)} \left[\left(u'^2 - u'' - \frac{2}{r} u' \right) f - 2u' \frac{\hat{\boldsymbol{r}} \cdot \nabla f}{f} + \nabla^2 f \right], \end{aligned} \tag{C.8}$$

where primes denote derivatives with respect to r.

Substituting into the expression for the local energy:

$$\begin{aligned} E_L &= -\frac{\nabla^2 \psi}{\psi} + \frac{1}{r} \\ &= -\left(u'^2 - u'' - \frac{2}{r} u' \right) - 2u' \frac{\hat{\boldsymbol{r}} \cdot \nabla f}{f} + \frac{\nabla^2 f}{f} + \frac{1}{r}. \end{aligned} \tag{C.9}$$

In the limit $r \to 0$, we assume f, f', and $\nabla^2 f$ remain finite. The leading singular terms are:

$$\frac{2u'(r)}{r} + \frac{1}{r}. \tag{C.10}$$

To avoid a divergence in the local energy, this term must vanish:

$$\frac{2u'(0)}{r} + \frac{1}{r} = 0 \quad \Rightarrow \quad 2u'(0) + 1 = 0. \tag{C.11}$$

Thus, the cusp condition is:

$$\left. \frac{du}{dr} \right|_{r \to 0} = -\frac{1}{2} \tag{C.12}$$

This is the electron-electron cusp condition for antiparallel-spin electrons.

Parallel-spin cusp condition

We now derive the cusp condition for two electrons with parallel spins. Now it is the spacial part of the wavefunction that is antisymmetric, so $f(\boldsymbol{r}) \to 0$ as $r \to 0$. In the simplest case, we assume

$$f(\boldsymbol{r}) = \boldsymbol{a} \cdot \boldsymbol{r} = \nabla f \cdot \boldsymbol{r}. \tag{C.13}$$

Now we substitute this new expression into Eq. C.9, and find that the singular terms that now

$$\frac{2u'(r)}{r} + 2u'(r)\frac{\hat{\boldsymbol{r}} \cdot \nabla f}{\boldsymbol{r} \cdot \nabla f} + \frac{1}{r} = 0. \tag{C.14}$$

This equation simplifies to

$$\frac{4u'(r)}{r} + \frac{1}{r} = 0, \tag{C.15}$$

that yields

$$u' = -\frac{1}{4}. \tag{C.16}$$

Appendix D

Derivation of Importance Sampling DMC

We start from the gradient of the $\psi(\boldsymbol{r}, \tau)$ (which we simply write as ψ.

$$\nabla\psi = \nabla\left(\frac{f}{\psi_T}\right) = \frac{\nabla f}{\psi_T} - \frac{f\,\nabla\psi_T}{\psi_T^2}. \tag{D.1}$$

Now we compute the Laplacian $\nabla^2\psi = \nabla\cdot\nabla\psi$ by taking the divergence of the above expression:

$$\nabla^2\psi = \nabla\cdot\left(\frac{\nabla f}{\psi_T} - \frac{f\,\nabla\psi_T}{\psi_T^2}\right) = \nabla\cdot\left(\frac{\nabla f}{\psi_T}\right) - \nabla\cdot\left(\frac{f\,\nabla\psi_T}{\psi_T^2}\right) \tag{D.2}$$

We evaluate each term separately:

$$\nabla\cdot\left(\frac{\nabla f}{\psi_T}\right) = \frac{\nabla^2 f}{\psi_T} - \frac{\nabla f\cdot\nabla\psi_T}{\psi_T^2} \tag{D.3}$$

and

$$\nabla\cdot\left(\frac{f\,\nabla\psi_T}{\psi_T^2}\right) = \frac{\nabla f\cdot\nabla\psi_T}{\psi_T^2} + \frac{f\,\nabla^2\psi_T}{\psi_T^2} - 2\frac{f\,|\nabla\psi_T|^2}{\psi_T^3} \tag{D.4}$$

Putting everything together,

$$\nabla^2\psi = \frac{\nabla^2 f}{\psi_T} - \frac{\nabla f\cdot\nabla\psi_T}{\psi_T^2} - \frac{\nabla f\cdot\nabla\psi_T}{\psi_T^2} - \frac{f\,\nabla^2\psi_T}{\psi_T^2} + 2\frac{f\,|\nabla\psi_T|^2}{\psi_T^3} \tag{D.5}$$

which simplifies to

$$\nabla^2\psi = \frac{\nabla^2 f}{\psi_T} - 2\frac{\nabla f\cdot\nabla\psi_T}{\psi_T^2} - \frac{f\,\nabla^2\psi_T}{\psi_T^2} + 2\frac{f\,|\nabla\psi_T|^2}{\psi_T^3} \tag{D.6}$$

Multiplying both sides by ψ_T, we get

$$\psi_T \nabla^2 \psi = \nabla^2 f - 2\frac{\nabla f \cdot \nabla \psi_T}{\psi_T} - f\frac{\nabla^2 \psi_T}{\psi_T} + 2f\frac{|\nabla \psi_T|^2}{\psi_T^2} \tag{D.7}$$

Therefore the right-hand-side of the Schroedinger equation, written as a function of f, reads:

$$\psi_T \hat{H} \psi = \psi_T \left(-\frac{1}{2}\nabla^2 \psi + V\psi \right) \tag{D.8}$$

$$= -\frac{1}{2}\left(\nabla^2 f - 2\frac{\nabla f \cdot \nabla \psi_T}{\psi_T} - f\frac{\nabla^2 \psi_T}{\psi_T} + 2f\frac{|\nabla \psi_T|^2}{\psi_T^2} \right) + Vf \tag{D.9}$$

$$= -\frac{1}{2}\nabla^2 f + \frac{\nabla f \cdot \nabla \psi_T}{\psi_T} + \frac{f}{2}\frac{\nabla^2 \psi_T}{\psi_T} - f\frac{|\nabla \psi_T|^2}{\psi_T^2} + Vf \tag{D.10}$$

Now we use the following trick:

$$\frac{f}{2}\frac{\nabla^2 \psi_T}{\psi_T} = f\frac{\nabla^2 \psi_T}{\psi_T} - \frac{f}{2}\frac{\nabla^2 \psi_T}{\psi_T} \tag{D.11}$$

This allows us to identify the local energy, so we arrive to

$$\psi_T \hat{H} \psi = -\frac{1}{2}\nabla^2 f + \frac{\nabla f \cdot \nabla \psi_T}{\psi_T} + f\frac{\nabla^2 \psi_T}{\psi_T} - f\frac{|\nabla \psi_T|^2}{\psi_T^2} + E_L f \tag{D.12}$$

This seems very complicated by we can actually see that

$$+\frac{\nabla f \cdot \nabla \psi_T}{\psi_T} + f\frac{\nabla^2 \psi_T}{\psi_T} - f\frac{|\nabla \psi_T|^2}{\psi_T^2} = \nabla(\mathbf{v}f), \tag{D.13}$$

where

$$\mathbf{v} = \frac{\nabla \psi_T}{\psi_T}. \tag{D.14}$$

To prove it, just calculate

$$\nabla \cdot (\mathbf{v}f) = \nabla \cdot \left(f\frac{\nabla \psi_T}{\psi_T} \right) = \nabla f \cdot \frac{\nabla \psi_T}{\psi_T} + f\nabla \cdot \left(\frac{\nabla \psi_T}{\psi_T} \right) \tag{D.15}$$

$$\nabla \cdot \left(\frac{\nabla \psi_T}{\psi_T} \right) = \frac{\nabla^2 \psi_T}{\psi_T} - \frac{|\nabla \psi_T|^2}{\psi_T^2} \tag{D.16}$$

$$\nabla \cdot (\mathbf{v}f) = \frac{\nabla f \cdot \nabla \psi_T}{\psi_T} + f\frac{\nabla^2 \psi_T}{\psi_T} - f\frac{|\nabla \psi_T|^2}{\psi_T^2} \tag{D.17}$$

So the imaginary-time equation as a function of f reads

$$-\frac{\partial}{\partial \tau} f = -\frac{1}{2}\nabla^2 f + \nabla(\mathbf{v}f) + E_L f \tag{D.18}$$