11

Fermion ground state methods

In this chapter, we discuss the fixed-node and constrained-path Monte Carlo methods for computing the ground state properties of systems of interacting electrons. These methods are arguably the two most powerful ones presently available for doing such calculations, but they are approximate. By sacrificing exactness, they avoid the exponential scaling of the Monte Carlo errors with system size that typically accompanies the simulation of systems of interacting electrons. This exponential scaling is called the *Fermion sign problem*. After a few general comments about the sign problem, we outline both methods, noting points of similarity and difference, plus points of strength and weakness. We also discuss the constrained-phase method, an extension of the constrained-path method, which controls the phase problem that develops when the ground state wave function cannot be real.

11.1 Sign problem

The "sign problem" refers to the exponential increase of the Monte Carlo errors with increasing system size or decreasing temperature (e.g., Loh et al., 1990, 2005; Gubernatis and Zhang, 1994) that often accompanies a Markov chain simulation whose limiting distribution is not everywhere positive. Such a case generally arises in simulations of Fermion and frustrated quantum-spin systems. It seems so inherent to Monte Carlo simulations of Fermion systems that the phrase "the sign problem" to many seems almost synonymous with the phrase "the Fermion sign problem."

Explanations for the cause of the sign problem vary and are still debated. In this chapter, we choose to summarize two explanations that seem to connect the causes in ground state Fermion simulations in the continuum and on the lattice. The sign problem, of course, is not limited to ground state calculations or even to Fermion simulations. While the cause we discuss in Sections 11.2 and 11.3 focuses on the

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

low-lying states of diffusion-like operators, several topological pictures have been proposed (Muramatsu et al., 1992; Samson, 1993; Gubernatis and Zhang, 1994; Samson, 1995). Some of these discussions are done in the context of the zero- and finite-temperature determinant methods (Muramatsu et al., 1992; Gubernatis and Zhang, 1994). Others are done more analytically from a Feynman path-integral point of view (Samson, 1993, 1995). Some are for particles with statistics other than Fermions. The presentation of the sign problem in this chapter is appropriate for the Monte Carlo methods discussed in this chapter.

It is often difficult to know a priori whether a given Fermion or quantum-spin Monte Carlo simulation suffers from a sign problem. Experience, however, has shown that the simulations for certain classes of Hamiltonians are likely to have such a problem. For example, quantum-spin systems with antiferromagnetic exchange interactions are a good bet, unless the lattice is bipartite. In the latter case, the Marshall transformation (Marshall, 1955) might convert the simulation of Hamiltonians with nearest-neighbor exchange interactions to a sign-free one with a ground state wave function that must be positive (Section 5.2.4). For the same Hamiltonian on a "frustrated" lattice, such as the triangular or Kagome lattice, the transformation ceases to remove the sign problem. Simulations of spin models with nearest and next-nearest neighbor exchange interactions of opposite sign typically have a sign problem for any lattice structure.

For interacting electrons on a lattice, sign-problem-free situations are even rarer. Proofs exist that special Hamiltonians lack a sign problem (e.g., Wu and Zhang, 2005). Unfortunately, these Hamiltonians are unphysical. Hamiltonians with particle-hole symmetry at certain electron fillings usually lack a sign problem (Hirsch, 1985), as do most Hamiltonians in one dimension, if the hopping is nearest neighbor. In the latter case, the restricted motion along the line allows an unambiguous relabeling of the electrons that eliminates the exchange of electrons and the source of the sign problem. Certain types of spinless Fermion models can be simulated with the auxiliary field method after switching to a Majorana Fermion representation (Li et al., 2014). The idea here is that the formulation in terms of Majorana Fermions doubles the number of determinants, which eliminates the sign problem in a manner similar to the case of spinful Fermion models with particle-hole symmetry.

If the sign problem cannot be trumped by the choice of the Hamiltonian or its expression in an alternative basis, then the burden of dealing with the sign problem falls upon the algorithm. In the continuum, there are several Green's function methods that have no Fermion sign problem (Zhang and Kalos, 1991; Kalos and Pederiva, 2000), but unfortunately these methods do not appear to be suitable for general efficient use. The fixed-node and constrained-path methods fill a void by being well-benchmarked approximate methods that do not suffer from a sign

problem. Accepting the approximation enables us to perform simulations for modestly sized systems instead of struggling with an exact simulation for small systems.

In principle, the strategy for sampling from a mixed-sign distribution is straightforward. In regions where the distribution is positive, we sample normally. In regions where it is negative, we sample normally from the distribution's absolute value and change the sign of the observable. Expectation values become sign-weighted averages. More explicitly, if $p(x)$ is the distribution of interest and its sign function is $s(x) = p(x)/|p(x)|$, then the *sign-weighted average* is

$$\langle A \rangle = \frac{\int dx s(x) A(x) \, |p(x)|}{\int dx s(x) \, |p(x)|} = \frac{\langle sA \rangle_{|p|}}{\langle s \rangle_{|p|}}.$$

The subscript on the angular brackets indicates that the average is with respect to the distribution $|p(x)|$.

For many Markov chain Monte Carlo simulations, and in particular for the Monte Carlo power methods, the limiting distribution is unknown and typically not nonnegative. Particularly problematic is the situation when we do not know the location of the nodal surface that delineates the regions of phase space differing in sign, because this limits the utility of using sign-weighted averages. Without this knowledge, the weights of the configurations have mixed sign, and eventually about half of the configuration population has a positive sign and the other half a negative sign. This almost balanced mixture causes $\langle sA \rangle \approx 0$ and $\langle s \rangle \approx 0$, and the sign-weighted average becomes indeterminate (the error bar becomes much larger than the value of the observable). Various heuristic arguments (Section 5.4) and the experience from countless simulations demonstrate that these sign-weighted averages approach zero exponentially fast as the number of lattice sites or the inverse temperature increases. The associated relative errors also grow exponentially fast, leading to an algorithm that quickly becomes unusable. The irony of the Fermion sign problem is that in principle we do not need sign-weighted averages. We only need to solve for the wave function in the region corresponding to one sign. From this solution, we can use an antisymmetrization procedure to generate the remainder of the solution explicitly. The problem is that we do not know the boundaries of this region.

As discussed in the previous chapter, Monte Carlo power methods define a transition probability $P(C'|C)$ for generating the Markov chain from a Green's function $\langle C'|G|C \rangle = G(C', C) \equiv w(C', C)P(C'|C)$. While we can always define nonnegative transition probabilities, we cannot always at the same time define nonnegative weights $w(C', C)$. As also discussed in the previous chapter, these power methods usually importance sample via a guiding function $\psi_G(C) \equiv \langle C|\psi_G \rangle$ so the weights and transition probabilities actually used are defined relative to $\bar{G}(C', C) \equiv \psi_G(C')G(C', C)/\psi_G(C)$. For systems of frustrated quantum spins, a

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

typical situation is that the guiding function is positive, while the matrix elements of the Green's function have mixed signs. For systems of Fermions, the situation is often reversed with $G(C', C)$ being nonnegative but the guiding function at C' and C having opposite signs. On a transition from $|C\rangle$ to $|C'\rangle$, as defined by a positive $G(C', C)$, the importance-sampled Green's function $\bar{G}(C', C)$ can change sign because the transition produces an odd number of electron exchanges from $\psi_G(C)$ to $\psi_G(C')$.

The fixed-node and constrained-path methods typically scale as the cube of the system size. They eliminate the exponential error scaling by preventing walker movement across an approximate nodal surface of the sampling distribution. We first discuss the basics of the fixed-node method and then the basics of the constrained-path method. The details of the methods have important implications with respect to the estimation of expectation values. We discuss both methods simultaneously for mixed estimators and highlight the differences between the forward-walking estimator in the fixed-node method and the back-propagation estimator in the constrained-path method that often replaces the mixed estimator. After the discussion of mixed estimators we discuss the constrained-phase method, an extension of the constrained-path method, that controls the phase problem that develops when the ground state wave function cannot be real.

11.2 Fixed-node method

The fixed-node method was developed decades ago for diffusion Monte Carlo (Section 10.4.2) simulations of Fermions in the continuum (Anderson, 1975, 1976; Moskowitz et al., 1982; Reynolds et al., 1982). Associated with it is an interesting physical picture of what happens to the wave function when a sign problem exists. In this power method, a Monte Carlo algorithm evolves a configuration $C = (r_1, r_2, \ldots, r_N)$ of N labeled electron positions through phase space. When importance sampling is used (Section 10.4.3 and Section 11.1), the Monte Carlo sampling generates empirically a distribution function $f(C, \tau) = \psi(C, \tau)\psi_G(C)$ where $\psi_G(C)$ is the guiding (importance) function representing the ground state.[1] One can show (Hammond et al., 1994) that this distribution solves a partial-differential equation that defines the electron distribution function's deterministic imaginary-time dynamics

$$\frac{\partial f}{\partial \tau} = D\nabla^2 f - D\nabla \cdot \left[f F_Q\right] + [E_T - E_L]f,$$

[1] For simplicity, we temporarily drop the spin of each electron that is associated with each spatial position r_i.

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

where the "diffusion constant" is $D = \hbar^2/2m$, the local energy is (Section 9.3.1)

$$E_L = \langle C|H|\psi\rangle/\langle C|\psi\rangle \equiv H\psi(C)/\psi(C),$$

and the vector field F_Q is given by $F_Q = \nabla \ln|\psi|^2$. This field, called the *quantum force*, is a drift term appearing when the dynamics is importance sampled, that is, when the weight $f(C,\tau) = \psi(C,\tau)\psi_G(C)$.

The importance of this differential equation, which is not one that we are interested in solving numerically, is that it is well known that the spatial extremal eigenfunction of a diffusive differential equation is nodeless. The nodeless solution maximizes the effect of the Laplacian ∇^2 that is driving the diffusion. In continuum diffusion Monte Carlo simulations, we call the nodeless solution Bosonic. The first excited state, as all other eigenstates, must have nodes.[2] This first excited state is the Fermionic state we want. For large systems, the gap between the Bosonic and Fermionic eigenvalues becomes exponentially small and eventually lost in the noise of the simulation. Hence, in the continuum, the collapse of the Fermionic solution to the Bosonic solution causes the sign problem.

The continuum fixed-node method prevents the Bosonic collapse by imposing an infinite potential energy barrier at the Fermionic nodal surface, preventing any random walker from crossing it and thereby generating walkers of mixed sign. With these added potentials, the original Hamiltonian becomes some H_{eff}. The method exactly solves the Schrödinger equation for the effective Hamiltonian for the ground state $(E_{\text{eff}}, \psi_{\text{eff}})$ within this nodal interior. Because the location of the nodal surface is unknown, the method approximates the exact surface by using the nodal surface of a good approximation to the Fermionic ground state. If the method used the exact surface, it would generate the exact answer. The Fermionic guiding state $|\psi_G\rangle$ used in the importance sampling hopefully closely approximates the exact ground state and serves a dual purpose by also fixing the nodal surface. *The Monte Carlo solution inherits its Fermionic character from this guiding state.*

This sign problem picture, which is supported by continuum simulations, is enlightening but not as appropriate on a lattice as it is in the continuum. On a lattice, the picture changes as the connection with a diffusion-like equation of motion is less clear. Not surprisingly, the implementation of the fixed-node method changes with respect to defining a nodal surface in a discrete space and to the details of the effective Hamiltonian acting when a random walker tries to cross the surface.

In the *lattice fixed-node method* (Bemmel et al., 1994; ten Haaf et al., 1995), we use what we called the linear Green's function Monte Carlo method (Section 10.4.1). Sometimes the diffusion Monte Carlo method is used (Section 10.4.2).

[2] We recall that the lowest eigenpair of H is the maximal pair of $\exp(-\tau H)$.

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

We recall that the linear method has no Trotter approximation and iterates

$$\left|\psi'\right\rangle = \left[I - \tau\,(H - E_T)\right]|\psi\rangle\,.$$

In this iteration we represent the ground state $|\psi_0\rangle$ in a complete orthonormal basis of labeled electrons as

$$|\psi_0\rangle = \sum_C |C\rangle\langle C|\psi_0\rangle = \sum_C \psi_0(C)|C\rangle, \quad \psi_0(C) > 0, \tag{11.1}$$

and after we project the iteration onto the labeled-configuration basis, we obtain

$$\psi'(C') = \sum_C \langle C'|I - \tau(H - E_T)|C\rangle\psi(C) = \sum_C G(C', C)\psi(C).$$

When the iteration becomes stationary, that is, when the condition $\psi' = \psi_0$ is fulfilled, the Monte Carlo simulation is sampling from

$$P(C) = \frac{\psi_0(C)}{\sum_C \psi_0(C)}.$$

With importance sampling (Section 10.4.3)

$$\bar{\psi}'(C') = \sum_C \bar{G}(C', C)\bar{\psi}(C),$$

where $\bar{\psi}(C) = \psi_G(C)\psi(C)$. Importance sampling thus changes the transition probability from which we draw our samples. Almost always we take $\psi_G(C) = \psi_T(C)$. Sampling only in the region where $\psi_T(C)$ is positive resolves the issue. Making the matrix elements of $\bar{G}$ positive yields the transition probability for the Markov chain. In the continuum, G is generally easily made positive.[3]

A nodal surface for lattice problems cannot be precisely located. It is something not approachable by making successively smaller and smaller spatial steps that avoid crossing the surface. On a lattice, the sign problem morphs into something associated with *sign flips* as electrons hop between lattice points. Sign flips occur for several reasons. As in the continuum, the guiding state can change sign. In contrast to the continuum, where the kinetic energy is always negative, the hopping part of many lattice Hamiltonians has amplitudes of mixed signs. More generally, the Hamiltonian almost always connects lattice sites in different nodal regions.

We also recall that in the linear Green's function Monte Carlo method (Section 10.4) the condition for no sign problem was $\langle C|H - \sigma I|C'\rangle \leq 0$ for all C and C'.[4] If so, then a sign change occurs only if the walker moves from one nodal region

[3] In the continuum making the elements of G positive is a matter of making its diagonal elements negative.
[4] In what follows $\sigma = E_T$, and for notational simplicity we replace $H - \sigma I$ by H.

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

to another. More generally, when configuration changes cause $\langle C'|H|C\rangle\psi_T(C)\psi_T(C') > 0$, the lattice fixed-node method (Bemmel et al., 1994; ten Haaf et al., 1995) prevents sign flips by replacing the original Hamiltonian with an effective one created from the original by truncating the off-diagonal terms and adding terms to the diagonal.[5] The off-diagonal elements of this effective Hamiltonian are

$$\langle C'|H_{\text{eff}}|C\rangle = \begin{cases} \langle C'|H|C\rangle & \text{if } \langle C'|H|C\rangle\psi_T(C')\psi_T(C) < 0 \\ 0 & \text{otherwise,} \end{cases} \tag{11.2}$$

while the diagonal elements are

$$\langle C|H_{\text{eff}}|C\rangle = \langle C|H|C\rangle + \langle C|V_{\text{sf}}|C\rangle. \tag{11.3}$$

The last term in this equation is a spin-flip potential that partially corrects for the contribution of the transitions left of H_{eff}. This potential has only diagonal elements and is defined by

$$\langle C|V_{\text{sf}}|C\rangle = \sum_{C'}{}^{\text{sf}} \langle C|H|C'\rangle\psi_T(C')/\psi_T(C). \tag{11.4}$$

For the diagonal elements, the superscript "sf" on the summation sign signifies that the summation over C' is over all configurations C for which $\langle C'|H|C\rangle\psi_T(C)\psi_T(C') > 0$. On the diagonal of the effective Hamiltonian, a positive potential energy replaces a hop (an off-diagonal term) that would flip a sign. The Monte Carlo power method exactly solves for the ground state $(E_{\text{eff}}, \psi_{\text{eff}})$ of this effective Hamiltonian. An important feature of this method is that the ground state of H_{eff} is, as we will show, an upper bound of the exact ground state energy.

Several key characteristics of the lattice fixed-node method are: (1) It accommodates sign problems other than the Fermion sign problem; (2) the exact ground state energy is obtained only if $\psi_T(C) = \psi_0(C)$ is exact (Section 11.4); and (3) the effective ground state energy is variational, that is, $E_{\text{eff}} \geq E_0$.

Whether there is a sign problem or not, efficient use of Green's function methods requires trial states we can easily evaluate for different configurations. This is possible for a number of interesting many-body wave functions such as those discussed in Section 9.2. The ability to use a variety of trial states is a strength of this method. Best results are obtained if the variational Monte Carlo method, a good approximate theory, or some other wave function optimization method first refines this trial state. We note that refining a trial state to produce a truly excellent estimate of the ground state energy may produce a trial state less effective than desired in estimating other observables.

[5] Recall from Section 10.4 the conditions for the positivity of $G = I - \tau(H - E_T I)$.

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

11.3 Constrained-path method

In the constrained-path method (Zhang et al., 1995, 1997; Carlson et al., 1999), we represent the ground state $|\psi_0\rangle$ as a sum of Slater determinants $|\psi_0\rangle = \sum_\phi c_\phi |\phi\rangle$ with $c_\phi > 0$ instead of as a sum over spatial configurations (11.1). This difference and its implications distinguish the constrained-path method from the diffusion Monte Carlo fixed-node method (Section 10.4.2). In many other respects, they are formally the same.

The constrained-path method was developed for projection problems where the Trotter approximation and Hubbard-Stratonovich transformation are used on the exponential of the Hamiltonian to enable the projection of one Slater determinant onto another (Appendix F). The resulting random walk is in the space of Slater determinants $|\phi\rangle$, and the Monte Carlo sampling from the distribution of the Hubbard-Stratonovich fields generates a distribution $f(\phi, \tau)$ that estimates the imaginary-time projection to the ground state.

Analogous to the continuum problem, a deterministic diffusion-like partial differential equation

$$\frac{\partial f}{\partial \tau} = \frac{1}{2} D(\phi) f + \left[\nabla_\phi V_1(\phi)\right] \cdot \nabla_\phi f + V_2(\phi) f \tag{11.5}$$

for the weight $f(\phi, \tau)$ exists when the Hubbard-Stratonovich fields are Gaussian distributed (Fahy and Hamann, 1991). Here the diffusion operator D and the drift V_1 and branching V_2 operators have explicit representations in terms of the operators of the Hamiltonian.

What is special about this equation is its invariance under the Slater determinant parity transformation $|\phi\rangle \rightarrow -|\phi\rangle$ (Fahy and Hamann, 1990, 1991; Hamann and Fahy, 1990). This symmetry enables the classification of its Slater determinant eigenfunctions as having an odd or even parity. The lowest eigenstate of this partial differential equation has even parity, that is, $f_0(\phi, \tau) = f_0(-\phi, \tau)$. As it is a sum of Slater determinants, it is naturally Fermionic in configuration space.

In terms of this distribution, the imaginary-time-dependent wave function is

$$|\psi(\tau)\rangle = \sum_\phi f(\phi, \tau) |\phi\rangle, \tag{11.6}$$

where the sum is over all normalized Slater determinants, and expectation values become

$$\langle A(\tau)\rangle = \frac{\langle \psi(\tau)|A|\psi(\tau)\rangle}{\langle \psi(\tau)|\psi(\tau)\rangle} = \frac{\sum_\phi \sum_{\phi'} f(\phi, \tau) f(\phi', \tau) \langle \phi|A|\phi'\rangle}{\sum_\phi \sum_{\phi'} f(\phi, \tau) f(\phi', \tau) \langle \phi|\phi'\rangle}. \tag{11.7}$$

Fahy and Hamann (1990, 1991) showed that the lowest eigenstate of (11.5) has even parity. As we have done for other discussions of projection methods, let us

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

make the eigenfunction expansion

$$f(\phi, \tau) = \sum_i c_i f_i(\phi) \exp(-E_i \tau).$$

As $\tau \to \infty$, we see that the even-parity lowest eigenstate dominates this expansion and also dominates the integrands of the numerator and denominator in (11.7). In this limit, because $|\psi\rangle$ and $-|\psi\rangle$ are different points in the manifold of Slater determinants, the even-parity $f(\phi, \tau)$ contributions in (11.6) must cancel out to yield the lowest odd-parity eigenstate. Thus, the integrals in (11.7) are dominated by the lowest odd-parity eigenfunction. A statistical evaluation of these integrals ultimately lacks sufficient accuracy to capture the difference in the different parity contributions.

This "diffusion in the space of Slater determinants" approach was developed when the main ground state method for lattices was the determinant method discussed in Appendix I. The approach presents another interesting perspective on the sign problem. It was also trying to reconcile the similar nature of the sign problem in diffusion Monte Carlo in the continuum and the sign problem in the determinant method. The continuum method was at the time believed not to be using Hubbard-Stratonovich fields. We now know that the continuum diffusion Monte Carlo method is an auxiliary-field method with a Gaussian Hubbard-Stratonovich transformation applied to the exponential of the kinetic energy (Zhang et al., 1997; Section 10.7). In the determinant method, the transformation is applied to the exponential of the potential energy. The main difference between the methods is that one executes a random walk in configuration space and the other executes one in Slater determinant space.

The analysis leading to the diffusion equation (11.5) assumes that the auxiliary fields are Gaussian distributed. The details of the diffusion equation depend on the distribution of Hubbard-Stratonovich fields, and generally, we use discrete fields. In any case, the advantage of this formal analysis was insight and not an efficient algorithm. However, the analysis did lead to a proposal in Slater determinant space called the *positive projection method* (Fahy and Hamann, 1990) as an analog to the fixed-node approximation in configuration space. The positive projection method became the springboard for the constrained-path method. If $\psi < 0$, the positive projection method assumes $f(\psi) = 0$, which is equivalent to assuming $V_2(\psi)$ in (11.5) is infinity. Under these conditions the lowest eigenstate has odd parity.

At its core the constrained-path method is a power method of the diffusion Monte Carlo type. The key operator in this method is $\exp(-\Delta\tau H)$, which is the same key operator as in the zero- and finite-temperature determinant methods (Chapter 7 and Appendix I). We recall that in these chapters an operator B represented $\exp(-\Delta\tau H)$ after the Trotter approximation and the Hubbard-Stratonovich transformation, and

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

the propagation was by a noninteracting Hamiltonian that depended on a time-dependent auxiliary field. We can represent one step in this propagation as

$$\left|\psi'\right\rangle = \sum_x P(x)B(x)|\psi\rangle.$$

Here, B is an operator and $P(x)$ is the distribution of the auxiliary field x.

In the constrained-path method, each walker $|\phi\rangle$ is a Slater determinant: $|\phi\rangle = a^\dagger_{N_\sigma}\cdots a^\dagger_2 a^\dagger_1|0\rangle$, where $a^\dagger_i = \sum_j c^\dagger_j \Phi_{ji}$ is defined by the $N \times N^\sigma$ matrix Φ. By Thouless's theorem (Appendix F), the propagation of this state $e^{-\Delta\tau c^\dagger M c}|\phi\rangle$ via a noninteracting Fermion Hamiltonian is to a new Slater determinant $\left|\phi'\right\rangle$, where $\left|\phi'\right\rangle = b^\dagger_{N_\sigma}\cdots b^\dagger_2 b^\dagger_1|0\rangle$, and $b^\dagger_i = \sum_j c^\dagger_j \Phi'_{ji}$ is defined by the $N \times N^\sigma$ matrix $\Phi' = \exp(-\Delta\tau M)\Phi = B\Phi$. Here B is a matrix.[6]

To define a way to implement the positive projection method, we first recall that in the ground state determinant method (Appendix I) our expectation values were of the form

$$\langle A\rangle = \frac{\sum_x P(x)\langle\psi_T|B_L(x_L)\cdots B_{L/2-1}(x_{L/2-1})AB_{L/2}(x_{L/2})\cdots B_1(x_1)|\psi_T\rangle}{\sum_x P(x)\langle\psi_T|B_L(x_L)B_{L-1}(x_{L-1})\cdots B_1(x_1)|\psi_T\rangle}.$$

Here $x = (x_1, x_2, \ldots, x_L)$ represents a configuration of Hubbard-Stratonovich fields, and $P(x)$ is the joint distribution function associated with these fields (Section 7.1.1 and Appendix G). Constraining the sign of the terms in the denominator creates the sign of the numerator. This constraint creates a nodal surface equal to the hyperplane perpendicular to some constraining wave function $|\psi_C\rangle$, which might be one or a sum of a few Slater determinants. The better $|\psi_C\rangle$ approximates the ground state, the closer the hyperplane becomes to the exact nodal surface of f. The positive projection method imposes the L conditions

$$\begin{aligned}
\langle\psi_C|\,B_1\,|\psi_T\rangle &> 0,\\
\langle\psi_C|\,B_2B_1\,|\psi_T\rangle &> 0,\\
&\vdots\\
\langle\psi_C|\,B_{L/2}B_{l-1}\cdots B_1\,|\psi_T\rangle &> 0,\\
\langle\psi_T|\,B_LB_{L-1}\cdots B_{L/2+1}\,|\psi_C\rangle &> 0,\\
&\vdots\\
\langle\psi_T|\,B_L\,|\psi_C\rangle &> 0.
\end{aligned}$$

Only configurations of Hubbard-Stratonovich fields that satisfy all L constraints are used. For Gaussian-distributed Hubbard-Stratonovich fields these conditions create an impenetrable barrier for the random walkers.

[6] We are overloading notation to curb the proliferation of notation. The context clearly determines whether B is an operator or a matrix.

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

The constrained-path method (Zhang et al., 1995, 1997; Carlson et al., 1999) makes the positive projection suggestion more practical by two modifications. The first makes the projection open ended,

$$\begin{aligned} \langle \psi_C | B_1 | \psi_T \rangle &> 0, \\ \langle \psi_C | B_2 B_1 | \psi_T \rangle &> 0, \\ &\vdots \end{aligned}$$

that is, it becomes the power method instead of the fixed-length projection in the determinant method. The second modification eliminates any random walker violating $\langle \phi | \psi(\tau) \rangle > 0$; that is, it does not look at what might happen in the future. In practice, we choose the constraining state $|\psi_C\rangle$ to be the guiding state $|\psi_G\rangle$, which in turn is just $|\psi_T\rangle$. The constraint is now something easily imposed.

Asymptotically, the constrained-path method samples from the imaginary-time-independent distribution $\pi(\phi) = c_\phi / \sum_\phi c_\phi$ with $c_\phi > 0$. The decomposition of $|\psi_0\rangle$ in terms of a set of Slater determinants Ω is not unique. We could have just as well written $|\psi_0\rangle = \sum_{\Omega'} d_\phi |\phi\rangle$ with $d_\phi > 0$ or $|\psi_0\rangle = \sum_{\Omega''} f_\phi |\phi\rangle$ with $f_\phi > 0$, where Ω' and Ω'' are different sets of Slater determinants. This flexibility arises because the set of Slater determinants is not orthonormal, $\langle \phi' | \phi \rangle \neq \delta_{\phi\phi'}$, and is overcomplete, $\sum_\phi |\phi\rangle\langle\phi| \neq I$.

Because the random walks in the constrained-path method occur in a different basis than the random walks in the fixed-node method (Fig. 11.1), we need a different type of nodal constraint. In this different basis, transitions from a region where the overlap $\langle \psi_T | \phi(\tau) \rangle$ is positive to one where it is negative cause the sign problem. The regions with overlaps of different signs are not physically distinguishable, and the propagation thus mixes degenerate bases caused by the plus-minus symmetry. We need a procedure to prevent this mixing from happening indiscriminately. Breaking the plus-minus symmetry by constraining each walker to the region where $\langle \psi_T | \phi(\tau) \rangle > 0$ prevents such mixing.

Instead of applying the constraint to the entire population of walkers at once, we apply the constraint to one random walker at a time: If $|\psi(\tau)\rangle = \sum_\phi c_\phi |\phi(\tau)\rangle$ with $c_\phi > 0$, then $\langle \psi_T | \psi(\tau) \rangle > 0$ if $\langle \psi_T | \phi(\tau) \rangle > 0$ for all $|\phi(\tau)\rangle$. For a given step, $|\psi'\rangle = \sum_x P(x) B(x) |\psi\rangle$, so we sample an x from $P(x)$ and then generate a new walker $|\phi'\rangle = B(x)|\phi\rangle$. If $\langle \psi_T | \phi' \rangle > 0$, we keep the walker. If not, we eliminate it.

The constrained-path approximation is not a fixed-node approximation. While the chosen $|\psi_T\rangle$ may constrain a $|\phi(\tau)\rangle$ to be a member of a "positive" set, this $|\phi(\tau)\rangle$ may overlap a member of a "negative" set. In contrast to the fixed-node method, the constrained-path method thus does not separate the basis into orthogonal sets. One consequence is that the constrained-path method can sometimes produce the exact solution even if the constraining wave function is approximate

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

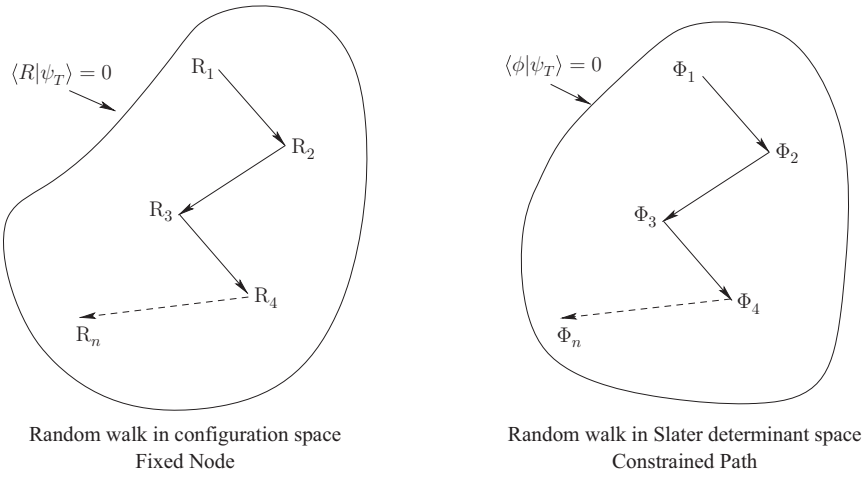

Figure 11.1 Schematic representation of the difference between the random walk in the fixed-node and constrained-path methods. In the fixed-node method, the walk is from point to point $R = (r_1, r_2, \ldots, r_n)$ in configuration space and is within the nodal surface $\langle R|\psi_T\rangle = 0$. In the constrained-path method the walk is from point to point in the space of Slater determinants and is within the nodal surface $\langle \phi|\psi_T\rangle = 0$.

and at any given imaginary time has the incorrect nodal surface in configuration space. The eigenpair $(E_{\text{eff}}, \psi_{\text{eff}})$ satisfies

$$E_v = E_{\text{eff}} = \frac{\langle \psi_{\text{eff}}|H|\psi_{\text{eff}}\rangle}{\langle \psi_{\text{eff}}|\psi_{\text{eff}}\rangle}.$$

Some key characteristics of the constrained-path method are (1) the nodal surface of $\langle \psi_T|\phi(\tau)\rangle = 0$ is not the same as that of $\langle \psi_{\text{eff}}|\phi(\tau)\rangle = 0$; (2) sometimes an approximate nodal surface $\langle \psi_T|\phi(\tau)\rangle = 0$ produces the exact ground state energy; and (3) the exact solution is found in the noninteracting case.

11.4 Estimators

11.4.1 Mixed estimator

In the chapter on the power method (Section 10.5), we discussed three estimators of the energy: the variational estimate E_v, the mixed estimate E_m, and the growth estimate E_g. In the fixed-node method, the three are equivalent and provide an upper bound to the ground state energy E_0. In the constrained-path method, the relation of the three among themselves is unclear (Carlson et al., 1999). Only the variational estimator is always an upper bound.

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

In the fixed-node method, because $\psi(C,\tau) = \bar{\psi}(C,\tau)/\psi_T(C)$, we clearly have

$$E_v = \frac{\sum_C \psi(C,\tau)(H\psi(C,\tau))}{\sum_C \psi^2(C,\tau)} \geq E_0.$$

At large τ, the walkers are distributed with a probability density $\psi_T(C)\psi(C)$, and both $\psi_T(C)$ and $\psi(C)$ go to zero linearly near the nodal surface. Because the Hamiltonian and the constraint are local operators, this implies that the growth and mixed estimators are equal:

$$E_g = E_m = \frac{\sum_C \psi_T(C)E_L(C)\psi(C)}{\sum_C \psi_T(C)\psi(C)}.$$

The constrained propagator is identical to the exact one except near the nodal surface, and the constraint discards any contributions to the ground state that are orthogonal to $|\psi_T\rangle$ and $|\psi\rangle$. Accordingly, the nodal region gives no net contribution to either $\langle\psi_T|H|\psi\rangle$ or $\langle\psi|H|\psi\rangle$. Therefore the variational energy is identical to the growth and mixed estimators. The mixed energy is the one more efficiently computed.

We can demonstrate that the variational energy defined by the effective fixed-node Hamiltonian is an upper bound to the ground state energy E_0 (ten Haaf et al., 1995). To do this, we need to show that the ground state of H_{eff} is an upper bound to the true ground state energy. First, we write the effective and actual Hamiltonians as

$$H_{\text{eff}} = H_{\text{tr}} + V_{\text{sf}},$$

$$H = H_{\text{tr}} + H_{\text{sf}},$$

and then define the configuration space matrix elements of H_{tr}, H_{sf}.

To define these elements, we first use the first of the above equations and note that (11.3) implies that the diagonal elements of H_{tr} equal the diagonal elements of H_{eff}. Then, from (11.2), we see that the off-diagonal elements of H_{eff} equal the off-diagonal elements of H. Hence, because V_{sf} is diagonal, the off-diagonal elements of H_{tr} equal those of H. Because the diagonal elements of the truncated Hamiltonian H_{tr} are $\langle C|H|C\rangle$, it follows from the second of the aforementioned equations that the diagonal elements of H_{sf} must be zero. Finally, because the off-diagonal elements of H_{tr} are zero, this same equation implies that the off-diagonal elements of H_{sf} equal those of H.

Now we compute the energy difference between H and H_{eff}, relative to an arbitrary state $|\psi\rangle$. We find

$$\Delta E = \langle\psi|\,(H_{\text{eff}} - H)\,|\psi\rangle = \langle\psi|\,(V_{\text{sf}} - H_{\text{sf}})\,|\psi\rangle\,,$$

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

which on using the matrix elements of $V_{\rm sf}$ and $H_{\rm sf}$ equals

$$\Delta E = \sum_C \psi(C) \left[\langle C| V_{\rm sf} |C\rangle \, \psi(C) - \sum_{C'} \langle C| H_{\rm sf} \left|C'\right\rangle \psi(C') \right].$$

We next rewrite this expression in terms of matrix elements of H,

$$\Delta E = \sum_C \psi(C) \left[\sum_{C'}{}^{\rm sf} \langle C|H|C'\rangle \frac{\psi_T(C')}{\psi_T(C)} \psi(C) - \sum_{C'}{}^{\rm sf} \langle C|H|C'\rangle \psi(C') \right]. \quad (11.8)$$

Denoting $s(C, C')$ as the sign of the matrix element $\langle C|H|C'\rangle$ and noting that all terms in the summations satisfy the condition

$$\langle C|H|C'\rangle \psi_T(C) \psi_T(C') < 0,$$

we obtain

$$\Delta E = \sum_{C,C'}{}^{\rm sf} |\langle C|H|C'\rangle| \left| \psi(C) \sqrt{\left| \frac{\psi_T(C')}{\psi_T(C)} \right|} - s(C, C') \psi(C') \sqrt{\left| \frac{\psi_T(C)}{\psi_T(C')} \right|} \right|^2,$$

an expression that is obviously nonnegative for any state $|\psi\rangle$. The ground state energy of $H_{\rm eff}$ is therefore an upper bound for the ground state energy of the original Hamiltonian H. Hence, $E_{\rm eff} \geq E_0$.

The lattice fixed-node method has several other important properties. First, we can easily show that $H|\psi_T\rangle = H_{\rm eff}|\psi_T\rangle$. This equality ensures that the fixed-node Green's function Monte Carlo procedure improves the energy of the trial state:

$$\frac{\langle \psi_T| H |\psi_T\rangle}{\langle \psi_T|\psi_T\rangle} = \frac{\langle \psi_T| H_{\rm eff} |\psi_T\rangle}{\langle \psi_T|\psi_T\rangle} \geq E_{\rm eff} = \frac{\langle \psi_{\rm eff}|H|\psi_{\rm eff}\rangle}{\langle \psi_{\rm eff}|\psi_{\rm eff}\rangle} \geq E_0.$$

Next, we can show that if the exact ground state is used for the trial state, then the ground state energy of the effective Hamiltonian equals the true ground state energy. Thus, by varying the trial state, we can reach the true ground state energy.

There is an important difference between the lattice and continuum fixed-node methods. In the continuum, only the sign of the trial state matters. If the nodal surface is correct, we obtain the exact energy regardless of the magnitude of the trial state. On the lattice, it follows from (11.8) that when $\Delta E = 0$, we obtain the exact energy if

$$\frac{\psi(C')}{\psi(C)} = \frac{\psi_0(C')}{\psi_0(C)}$$

for all sign-flipping configuration pairs (C, C'). Thus, the sign and relative magnitude of the trial state in configurations connected by sign flips must be correct.

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

In the constrained-path method, while the variational energy constructed from $|\psi(\tau)\rangle$ satisfies

$$E_v = \frac{\langle \psi_{\text{eff}}(\tau)|H|\psi_{\text{eff}}(\tau)\rangle}{\langle \psi_{\text{eff}}(\tau)|\psi_{\text{eff}}(\tau)\rangle},$$

the relative ranking of E_v, E_m, and E_g is unclear because the constraint discards configurations that are orthogonal to $|\psi_T\rangle$, but these discarded configurations are not necessarily orthogonal to $|\psi\rangle$ (Carlson et al., 1999).

On general grounds, if the constrained-path method defines a Markov process with a stationary distribution such that

$$H_{\text{eff}}|\psi(\tau)\rangle = E_g|\psi_{\text{eff}}(\tau)\rangle,$$

where $H_{\text{eff}} = H + \delta H$, then

$$\frac{\langle \psi(\tau)|H|\psi(\tau)\rangle}{\langle \psi(\tau)|\psi(\tau)\rangle} = E_m + \delta E_m \geq E_0,$$

with

$$\delta E_m = \frac{\langle \psi_T|\delta H|\psi(\tau)\rangle}{\langle \psi_T|\psi(\tau)\rangle} - \frac{\langle \psi(\tau)|\delta H|\psi(\tau)\rangle}{\langle \psi(\tau)|\psi(\tau)\rangle}.$$

Only when $\delta E_m \leq 0$ is E_m an upper bound to the ground state energy. The mixed estimator for the energy is generally not an upper bound to the ground state energy. In general, the growth and mixed estimators are not equal and only the variational estimator is an upper bound.[7]

11.4.2 Forward walking and back propagation

The mixed estimator provides accurate estimates of observables that commute with the Hamiltonian. For those that do not, it can provide poor estimates. For example, the mixed estimator predicts the total energy well but generally gives poor estimates for the kinetic and the potential energies. For observables that do not commute with the Hamiltonian we need to use other estimators. For these cases the simplest estimator is the *extrapolated estimator*

$$\langle A \rangle \approx \langle A \rangle_{\text{ex}} = 2\langle A \rangle_m - \langle A \rangle_v. \tag{11.9}$$

If the deviation of the trial state from the true ground state is significant, then this estimator also provides poor estimates. Fortunately, more reliable estimators exist. We touched upon these in Section 10.5. For the fixed-node method we discuss only

[7] We note that other estimators exist so that the constrained-path method can yield an upper bound to the ground state energy (Carlson et al., 1999).

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

what is called the *forward (or future) walking method* (Hammond et al., 1994). In discussing it, we distinguish between cases where the observable is diagonal in the simulation basis and cases where it is not (Nightingale, 1999).

We start with the *diagonal* case. We want

$$\langle A\rangle = \frac{\langle\psi_0|A|\psi_0\rangle}{\langle\psi_0|\psi_0\rangle} = \frac{\sum_C A(C)\psi_0^2(C)}{\sum_C \psi_0^2(C)}.$$

The problem is that when the Markov chain is stationary, it samples from $\psi_T(C)\psi_0(C)$ instead of $\psi_0^2(C)$. To sample from the ground state distribution the strategy is to estimate a weight $w_0(C)$ so that

$$\langle A\rangle = \frac{\sum_C A(C)\psi_0^2(C)}{\sum_C \psi_0^2(C)} = \frac{\sum_C w_0(C)A(C)\psi_T(C)\psi_0(C)}{\sum_C w_0(C)\psi_T(C)\psi_0(C)}.$$

Clearly, we need $w_0(C) \propto \psi_0(C)/\psi_T(C)$.

To estimate this weight we first note that for large L

$$\langle A\rangle = \frac{\langle\psi_T|\, G^L A\,|\psi_0\rangle}{\langle\psi_T|\, G^L\,|\psi_0\rangle}.$$

With multiple insertions of unity, we transform this expression into

$$\langle A\rangle = \frac{\sum_{C_0,C_1,\ldots,C_L} \psi_T(C_L)G(C_L,C_{L-1})\cdots G(C_1,C_0)A(C_0)\psi_0(C_0)}{\sum_{C_0,C_1,\ldots,C_L} \psi_T(C_L)G(C_L,C_{L-1})\cdots G(C_1,C_0)\psi_0(C_0)}.$$

The next step is writing the relevant quantities in terms of their importance sampled values. Doing so we obtain

$$\langle A\rangle = \frac{\sum_{C_0,C_1,\ldots,C_L} \bar{G}(C_L,C_{L-1})\cdots \bar{G}(C_1,C_0)A(C_0)\bar{\psi}_0(C_0)}{\sum_{C_0,C_1,\ldots,C_L} \bar{G}(C_L,C_{L-1})\cdots \bar{G}(C_1,C_0)\bar{\psi}_0(C_0)},$$

where

$$\bar{G}(C',C) = \frac{\langle C'|\psi_T\rangle\langle C'|G|C\rangle}{\langle C|\psi_T\rangle} \tag{11.10}$$

is the modified transfer matrix (because of importance sampling). We now have a useful result. We know that the simulation eventually reaches a step 0 at which it starts to sample from the stationary importance sampled distribution $\bar{\psi}_0(C) = \psi_T(C)\psi_0(C)$. At this step we record the value of $A(C_0)$ plus the weight and state of each walker. We also know that for each walker at 0 the Markov chain of length L generates a population of descendants that sample a stationary distribution approximately equal to $\psi_0(C)$. The weights of the descendants thus give us the needed estimate of $w_0(C)$.

Some walkers at 0, however, have no descendants at L because of the killings in the stochastic reconfigurations (Section 10.3). Because of the branching, others

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

might have multiple descendants. Consequently, to estimate $\langle A \rangle$, we need to record the identity of a walker's parents at each step along the L-segments. We obtain the contribution of a walker at 0 to the numerator by multiplying its weight by its value of A at 0 and by the weights of its descendants at L. Not all walkers at 0 contribute to the value of the numerator. We accumulate the estimate of the denominator similarly.

Estimates of *nondiagonal* operators are more involved and potentially very difficult (Nightingale, 1999). For starters, we now have

$$\langle A \rangle = \frac{\sum_{C_0, C_1, \ldots, C_{L+1}} \bar{G}(C_{L+1}, C_L) \cdots \bar{G}(C_2, C_1) A(C_1, C_0) \bar{\psi}_0(C_0)}{\sum_{C_0, C_1, \ldots, C_{L+1}} \bar{G}(C_{L+1}, C_L) \cdots \bar{G}(C_1, C_0) \bar{\psi}_0(C_0)}, \tag{11.11}$$

which we can immediately write more conveniently as

$$\langle A \rangle = \frac{\sum_{C_0, C_1, \ldots, C_{L+1}} \bar{G}(C_{L+1}, C_L) \cdots \bar{G}(C_2, C_1) \bar{G}(C_1, C_0) \dfrac{A(C_1, C_0)}{\bar{G}(C_1, C_0)} \bar{\psi}_0(C_0)}{\sum_{C_0, C_1, \ldots, C_{L+1}} \bar{G}(C_{L+1}, C_L) \cdots \bar{G}(C_1, C_0) \bar{\psi}_0(C_0)}.$$

The fraction in the numerator of the last equation replaces the $A(C_0)$ in (11.11), and we execute the sampling procedure remembering a walker's parent just as we did in the diagonal case. Things work well as long as $A(C_1, C_0)$ is zero when $\bar{G}(C_1, C_0)$ is and vice versa. Generally this is not the case, and other techniques, which we do not describe, are needed (Nightingale, 1999). These techniques are much more involved than the previously mentioned.

For the constrained-path method, the mixed and extrapolated estimators have the same range of utility as they do in the fixed-node method. We move beyond these estimators with the same strategy we used for the fixed-node method, employing what is called the *back-propagation method* (Zhang et al., 1997) as the analog of the forward-walking method. In contrast to the forward-walking method, we do not need separate strategies for the diagonal and off-diagonal observables. Here, we calculate expectation values as sums of products of Green's functions in the same manner as we do for the determinant methods (Section 7.4 and Appendix I). As a consequence, we enjoy here the same ease and flexibility in the implementation of measurements as we do there. This ease is a positive feature of the constrained-path method.

In back-propagation, at step 0, we save the walkers (weights and Slater determinants) and then propagate the population via the B matrices according to the standard algorithm, noting the identity of parents and additionally saving the Hubbard-Stratonovich fields for each walker at each step. For each walker reaching L, we now back-propagate $\langle \psi_T |$ via

$$(((\langle \psi_T | B_L)\, B_{L-1})\, B_{L-2} \cdots)$$

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

until we arrive back at step 0, say, with walker $\langle\phi'|$. For each step of the back-propagation, we do not sample new Hubbard-Stratonovich fields, but rather we reuse those recorded in the forward steps but in reverse order. More explicitly, at the first back-propagation step, we use the fields for the last forward-propagation step; at the second back-propagation step, we use the fields for the next-to-last forward propagation step; etc. In the backward direction, we perform matrix stabilizations, but we do not apply the constrained-path condition or stochastically reconfigure. Each walker arriving at L has already passed the constrained-path test and the entire population was likely reconfigured several times. For each returning walker $\langle\phi'|$ we calculate $\langle\phi'|A|\phi\rangle$ for it and its parent $|\phi\rangle$ and then multiply this value times the product of the parent's weight at 0 and the descendant's weight at L. Not all parents contribute, only those with descendants do. Monte Carlo measurements are generally made periodically (Section 3.9). We can conveniently choose L to be a multiple of this period. After we complete the back-propagation measurement, we return to the population at L and use it to make the next L steps.

11.5 The algorithms

An implementation of the lattice fixed-node method follows the basic steps of an implementation of a standard power method. We refer the reader to the chapter on power methods (Chapter 10) for details. The key difference from a standard power method is that the fixed-node method uses H_{eff} instead of the original Hamiltonian H to avoid the sign problem. An implementation of the constrained-path method also uses the basic steps of a power method algorithm. However, it has additional Fermion-specific steps that resemble those in the zero-temperature determinant method (Appendix I). The way in which the sign problem is eliminated resembles the positive projection procedure.

Algorithm 39 gives the details of an implementation of the constrained-path method (Zhang et al., 1997). We simplify the discussion by excluding spin. As we noted in our discussion of the determinant methods (Chapter 7), we add spin by requiring the overlaps and Slater determinants to be products of a spin-up and a spin-down part. Other adjustments are fairly obvious.

In the algorithm, we suppose we have a stack of random walkers, each characterized by a weight w, a Slater determinant state $|\phi\rangle$, and an overlap integral $\mathcal{O}_T(\phi) = \langle\psi_T|\phi\rangle$. We represent the Slater determinant by an $N_{\text{states}} \times N_{\text{electrons}}$ matrix Φ (Appendix I). N_{states} is the number of orbitals and $N_{\text{electrons}} = N_{\text{up}} + N_{\text{down}}$ is the number of electrons. For the single-band Hubbard model N_{states} is the number of lattice sites, while for the periodic Anderson model, it is twice the number of lattice sites. It follows from Appendices F and H that $\mathcal{O}_T(\phi) = \det(\Phi_T^T\Phi)$.

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

Algorithm 39 Constrained-path Monte Carlo method.

Input: A trial state $|\psi_T\rangle$, described by the matrix Ψ_T, and an estimate of the ground state energy E_T.
Initialize a stack of N walkers described by the matrix $\Phi = \Psi_T$, a unit weight w, and a unit overlap integral $\mathcal{O}_T$;
repeat
  repeat
    Pop a walker off the stack ;
    Update $w \leftarrow we^{-\Delta\tau E_T}$;
    Compute $\Phi' = B_{K/2}\Phi$;
    Compute $\mathcal{O}'_T = \mathcal{O}_T(\Phi')$ and $[\Psi_T^T\Phi']^{-1}$;
    if $\mathcal{O}'_T < 0$ **then**
      Cycle to the next walker ;
    end if
    Update $\Phi \leftarrow \Phi'$, $w \leftarrow w\mathcal{O}'_T/\mathcal{O}_T$, $\mathcal{O}_T \leftarrow \mathcal{O}'_T$;
    Use Algorithm 40 to propagate the walker via B_V ;
    if $\mathcal{O}'_T < 0$ **then**
      Cycle to the next walker ;
    end if
    Compute $\Phi' = B_{K/2}\Phi$;
    Compute $\mathcal{O}'_T = \mathcal{O}_T(\phi')$ and $[\Psi_T^T\Phi']^{-1}$;
    if $\mathcal{O}'_T < 0$ **then**
      Cycle to the next walker ;
    end if
    Update $\Phi \leftarrow \Phi'$, $w \leftarrow w\mathcal{O}'_T/\mathcal{O}_T$, $\mathcal{O}_T \leftarrow \mathcal{O}'_T$;
    Push the walker onto the new stack ;
  until the stack is empty.
  If equilibrated, periodically estimate physical quantities ;
  Periodically re-orthogonalize the columns of Φ ;
  Periodically stochastically reconfigure the new stack to restore population size to N (Section 10.3) ;
  Make the new stack the old one ;
until an adequate number of measurements is collected.
Compute final averages and estimate their statistical errors ;
return the averages and their error estimates.

The main task in the algorithm is repeating the step that projects the population of walkers onto a distribution estimating the ground state. In Algorithm 39 we are using a higher-order symmetric Trotter approximation (5.13) to do this projection,

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

$$\exp[-\Delta\tau H] \approx \exp[-\tfrac{1}{2}\Delta\tau K]\exp[-\Delta\tau V]\exp[-\tfrac{1}{2}\Delta\tau K],$$

$$\Phi' = B_{K/2}B_V B_{K/2}\Phi.$$

In the above, $H = K + V$ and the matrices $B_{K/2}$ and B_V represent the half-step and full-step imaginary-time propagators for the hopping and potential energies (Section 7.2). In practice, we modify Algorithm 39 slightly to reduce the number of matrix-matrix multiplications by noting that

$$\begin{aligned}\Phi' &= \left(B_{K/2}B_V B_{K/2}\right)\left(B_{K/2}B_V B_{K/2}\right)\cdots\left(B_{K/2}B_V B_{K/2}\right)\Phi \\ &= B_{K/2}B_V\, B_K B_V\, \cdots B_K B_V B_{K/2}\Phi\end{aligned}$$

and adjusting the algorithm so that when we start a new measurement period, we first multiply Φ by $B_{K/2}$ and then again at the end of the period, just before we make measurements.

We execute the propagation by B_K and $B_{K/2}$ in procedures that do several things. Each does the matrix-matrix multiplication. With the resulting matrix Φ', each calculates the overlap matrix $\Psi_T^T\Phi'$ where the superscript T denotes transposition. Next, each uses standard LU-factorization[8] software to compute $\mathcal{O}_T = \det\Psi_T^T\Phi'$ and the inverse $[\Psi_T^T\Phi']^{-1}$. Then, they test if $\mathcal{O}_T < 0$. If so, that walker is not placed on the new stack. For each walker, we also compute the inverse of its overlap matrix because each multiplication by B_V follows a multiplication by $B_{K/2}$ or B_K. As we now discuss, we need this inverse at the start of the procedure that does the B_V propagation.

Algorithm 40 describes the propagation of a walker by B_V. This procedure is the one most computationally intensive. It borrows some of the algebra developed for the zero-temperature determinant method (Appendix I). There, as here, the Hubbard-Stratonovich transformation casts B_V into the form $\prod_i P(x_i)b_V(x_i)$, that is, the propagation factors into propagations for each Hubbard-Stratonovich field. As in the determinant method, the matrix representing $b_V(x_i)$ is a unit matrix except for the i-th diagonal element: $x_i = \pm 1$ and $P(x_i) = \frac{1}{2}$ (Section 7.2).

For a given i, we want to produce a Φ' from Φ. To do so, we importance sample a Hubbard-Stratonovich field x'. With importance sampling (Section 10.4.3), we select x' from $\mathcal{O}_T(\Phi')P(x)/\mathcal{O}_T(\Phi)$, but we do not yet know Φ'. We note that we can write

$$\Phi' = \sum_x \mathcal{O}_T\left[\Phi(x)\right]P(x)/\mathcal{O}_T(\Phi)b_V(x)\Phi \tag{11.12}$$

$$= N\left[p_1 b_V(+1) + p_2 b_V(-1)\right]\Phi, \tag{11.13}$$

[8] This factorization expresses a matrix as a product of a unit lower-triangular matrix L and an upper-triangular matrix U. The determinant of the given matrix is simply the product of the diagonal elements of U.

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

Algorithm 40 Constrained-path Monte Carlo method: propagation by B_V.

Input: A walker and the inverse of its overlap matrix $[\Psi_T^T\Phi]^{-1}$.
  for $i = 1$ to N_{states} **do**
    Compute $G_{ii} = 1 - [\Phi[\Psi_T^T\Phi]^{-1}\Psi_T^T]_{ii}$;
    Using (7.43) compute $\mathcal{C}(1) = \mathcal{O}_T(1)/\mathcal{O}_T$ and $\mathcal{C}(-1) = \mathcal{O}_T(-1)/\mathcal{O}_T$;
    Using (11.14) and (11.16) compute p_1, p_2 ;
    $p_1 = \max(0, p_1)$;
    $p_2 = \max(0, p_2)$;
    $N = p_1 + p_1$;
    if $N = 0$ **then**
      Set $\mathcal{O}_T < 0$;
      return the negatively weighted walker.
    end if
    if $p_1/N < \xi$ **then**
      $x' = 1$;
    else
      $x' = -1$;
    end if
    Update $w \leftarrow Nw$, $\mathcal{O}_T(\phi) \leftarrow [\mathcal{O}_T(x')/\mathcal{O}_T(\phi)]\mathcal{O}_T(\phi)$, and $\Phi \leftarrow b_V(x')\Phi$;
    Using (7.45), update $[\Psi_T^T\Phi]^{-1} \leftarrow [\Psi_T^T\Phi']^{-1}$;
    Save the Hubbard-Stratonovich field for this walker, this step, and this site ;
  end for
  Record the identity of the walker's parent ;
  return the updated walker.

where

$$p_1 = \frac{P(1)}{N}\frac{\mathcal{O}_T(1)}{\mathcal{O}_T}, \quad p_2 = \frac{P(-1)}{N}\frac{\mathcal{O}_T(-1)}{\mathcal{O}_T}, \tag{11.14}$$

with

$$\mathcal{O}_T(1) = \det[\Psi_T^T b_V(1)\Phi], \quad \mathcal{O}_T(-1) = \det[\Psi_T^T b_V(-1)\Phi] \tag{11.15}$$

and

$$N = P(1)\frac{\mathcal{O}_T(1)}{\mathcal{O}_T} + P(-1)\frac{\mathcal{O}_T(-1)}{\mathcal{O}_T}. \tag{11.16}$$

If either p_1 or p_2 is not positive, we set it equal to zero. If both are not positive, we set $\mathcal{O}_T < 0$ and return from the procedure. Otherwise, we choose x' equal to $+1$ or -1 with probability p_1 or p_2. With this x' we then update the walker

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

$$w \leftarrow wN, \quad O_T \leftarrow O_T(x'), \quad \Phi \leftarrow b_V(x')\Phi.$$

We also update $[\Psi_T^T\Phi]^{-1} \leftarrow [\Psi_T^T\Phi']^{-1}$.

We could have simplified the expressions for p_1 and p_2, but we deliberately wrote them with explicit ratios of determinants. We recall from the chapter on the determinant methods (Chapter 7) that when updating a change in the Hubbard-Stratonovich fields x_i, we needed to compute the ratio of two determinants, a computation that simplifies significantly if we know the i-th diagonal element of the zero-temperature Green's function $G = [I - \Phi\left(\Psi_T^T\Phi\right)^{-1}\Psi_T]^T$. Knowing this element requires knowing the inverse of the overlap matrix $\Psi_T^T\Phi$. The important point is that in the power method, we do not need to know all spatial elements of the Green's function, just the one for the site whose auxiliary field we are updating. By inputting the determinant and inverse of the overlap matrix, plus the state of the walker, we can easily perform the updating of this first field:

$$G_{ii} = 1 - \sum_{jk} \Phi_{ij}\left[\left(\Psi_T^T\Phi\right)^{-1}\right]_{jk}[\Psi_T]_{jk}.$$

To cycle to the next field, we need to update those quantities necessary to compute the diagonal element of the Green's function corresponding to this field. These quantities are the value of the new determinant, the new state of the walker, and the inverse of the new overlap matrix. We can get the new overlap determinant from the ratio $\mathcal{O}'_T/\mathcal{O}_T$ and the old overlap $\mathcal{O}_T$ by simple multiplication. To get the new overlap matrix, we first need to update the state of the walker, which we do by the multiplication of the matrix representing the old state by the matrix $b_V(x')$ that propagates it. To get the new inverse, we do not form $\Psi_T^T\Phi'$ and invert it, but as we did in the determinant method and in the continuous-time impurity algorithms (Chapters 7 and 8), we exploit the fact that Φ changes by only a few elements (here, just one row) and that we know the original inverse. If we call A^{-1} the known inverse, u the i-th row of Ψ_T, and v the i-th row of Φ, then the new inverse is (Sherman-Morrison formula; Meyer, 2000; Press et al., 2007)

$$\left(A + uv^T\right)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1 + uA^{-1}v^T}.$$

Thus, we can update the inverse of the overlap matrix by using (7.45) (see also Section 8.4.2) after notationally substituting G_{ij} with $A_{ij} = [\Psi_T^T\Phi']_{ij}$. This more efficient way of computing is the reason we compute the inverse of the overlap matrix at the end of the procedure that computes $B_{K/2}$ and B_K, and also while propagating by B_V, the reason we update this inverse after each selection of the new Hubbard-Stratonovich fields. At the end of a sweep we do not have the values of the Green's function matrix needed for doing measurements. Generally, we measure

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

every few sweeps or so and simply compute the correct Green's function when it is time to do measurements.

As a power method the constrained-path method benefits from stochastic reconfiguration (Section 10.3). One advantage is the standard variance reduction. A second is a replenishment of the walker population. As the constraint removes negatively signed walkers, the population would drop to zero if not replenished. Here, the comb is a useful procedure (Algorithm 38).

The repeated matrix-matrix multiplications inherent to the method accumulate numerical errors. The matrix-product stabilization procedures are simpler than the one for the finite-temperature determinant method (Chapter 7), but the same as for the zero-temperature method (Appendix I). The stabilization reduces to periodically orthonormalizing the columns of the propagation matrix Φ. Again, the modified Gram-Schmidt method works well (Section 7.4). In Appendix I, we give a pseudocode (Algorithm 50) for performing this orthonormalization. The justification for just needing this part of the matrix factorization follows the same arguments as given there for the zero-temperature determinant method.

As with the finite- and zero-temperature determinant methods, with the constrained-path method we can compute the expectation value of any observable from sums of products of Green's functions. Almost all quantities of interest require the use of back-propagation. As discussed in the previous sections, back-propagation replaces $\langle\psi_T|$ in the mixed estimate with $\langle\psi_B| = \langle\psi_T|B_L B_{L-1}\cdots B_1$ where $B_i = B_{K/2}B_V(i)B_{K/2}$ and $B_V(i)$ is the B_V matrix for the i-th term in the product. The Hubbard-Stratonovich fields used in the left-to-right multiplication are the ones used in the right-to-left multiplication $B_L B_{L-1}\cdots B_1|\phi\rangle$. As we also noted the walkers now have to carry the identification of their parent over these L steps, and to execute the back-propagation we have to store the Hubbard-Stratonovich fields generated for each walker in the propagation $B_V = \prod b_V(x)$. This bookkeeping is most conveniently done as the last step in the B_V procedure. When we stochastically reconfigure, we need to maintain ancestry.

Recently, Shi and Zhang (2013) have illustrated how using symmetries can increase the accuracy and efficiency for ground state simulations that use auxiliary fields.

11.6 Constrained-phase method

The *constrained-phase method*, sometimes called the phaseless method, extends the constrained-path method to zero-temperature Fermion simulations that have complex weights (Ortiz et al., 1997; Schmidt and Fantoni, 1999; Schmidt et al., 2001, 2003; Sarsa et al., 2003; Zhang and Krakauer, 2003; Zhang, 2013). Complex weights arise for multiple reasons. The prime one is the system being simulated has

broken time-reversal symmetry. Typically, this symmetry is broken by an external magnetic field, making the wave function complex. In lattice models, the magnetic field enters via the hopping matrix element t_{ij}, which becomes $t_{ij} \exp[i \int A \cdot d\ell]$ because of a vector potential A acting along the path ℓ connecting lattice positions i and j. A net phase through a closed path is a measure of the magnetic flux threading the loop. The Hamiltonian matrix is now Hermitian as opposed to being simply symmetric. Strictly Hermitian Hamiltonians also result when they include a spin-orbit interaction, a situation of great current interest in many-electron physics. If the Hubbard-Stratonovich transformation is not chosen to be real, complex weights can develop (Appendix G). Such a choice is convenient for electron-electron interactions whose form is more general than the simple on-site Hubbard interaction.

An artificial but useful way to obtain a complex wave function is to add a phase shift to the periodic boundary conditions, for example, in one dimension by using

$$\psi(x + L) = \exp(i\varphi)\, \psi(x). \tag{11.17}$$

This boundary condition breaks the symmetries that were creating the degeneracies among the eigenvalues and is the basis for the *twisted-phase boundary condition method* (Gammel et al., 1993; Lin et al., 2001). This method averages the properties of the system over many different values of φ uniform randomly chosen over the interval $[0, 2\pi]$. The degeneracy splitting and the averaging remove closed-shell effects in the total energy as a function of electron density and create a density dependence closer to that of an infinite system. For small finite systems, *closed-shell effects* appear as discontinuities in the slope of the total energy at densities where the up and down electrons each have a gap between the energy level of the highest occupied electron state and the next unoccupied state. Typically, the highest occupied energy level of each spin is degenerate because of a point-group symmetry of the Brillouin zone for the bands of the noninteracting electrons. Interactions in small systems often only sparsely fill the gaps that were present in the noninteracting problem.

The constrained-phase method is the analog of the fixed-phase method (Ortiz et al., 1993) that was developed for continuum problems in the presence of a magnetic field. The latter method substitutes the wave function expressed as

$$\psi(C) = |\psi(C)| \exp(i\varphi(C))$$

into Schrödinger's equation to generate two coupled partial differential equations, one for the amplitude $|\psi(C)|$ and the other for the phase $\varphi(C)$. Then, in the equation for $|\psi(C)|$, it approximates $\varphi(C)$ with the phase $\varphi_T(C)$ of the trial state. The diffusion Monte Carlo method now solves the equation for $|\psi(C)|$. No sign problem exists as this function is by construction everywhere nonnegative.

For lattice Fermion problems, where we propagate one Slater determinant into another (Appendix F), we generate a constrained-phase approximation in a manner similar to the way we created the constrained-path approximation. There, for real $\langle\psi_T|\psi(\tau)\rangle$, we achieve positivity by requiring $\langle\psi_T|\psi(\tau)\rangle > 0$ and by applying the constraint to each walker at each step of the imaginary-time propagation. Here, for complex $\langle\psi_T|\psi(\tau)\rangle$, we achieve reality and positivity by requiring either $\langle\psi_T|\psi(\tau)\rangle = \mathrm{Re}\{\langle\psi_T|\psi(\tau)\rangle\} > 0$ or $\langle\psi_T|\psi(\tau)\rangle = |\langle\psi_T|\psi(\tau)\rangle| > 0$. As before, we apply the constraint to one walker at a time. For either option we define

$$\langle\phi|\psi(\tau)\rangle = e^{i\varphi_\tau}|\langle\phi|\psi(\tau)\rangle|$$

and fix the phase by taking $\varphi_\tau(\phi) = \varphi_T(\phi) = \mathrm{Im}\{\ln\langle\phi|\psi_T\rangle\}$. Both options produce the exact result if the trial state is the exact ground state.

The basic step is to propagate $|\phi\rangle$ to $|\phi'\rangle$, that is, $\exp[-\Delta\tau H]|\phi\rangle = |\phi'\rangle$, and replace $|\phi'\rangle$ with $|\tilde{\phi}\rangle$ in one of two ways. For $\mathrm{Re}\{\langle\psi_T|\psi(\tau)\rangle\} > 0$,

$$|\tilde{\phi}\rangle = \cos\varphi_T(\phi')e^{-i\varphi_T(\phi')}|\phi'\rangle,$$

while for $|\langle\psi_T|\psi(\tau)\rangle| > 0$,

$$|\tilde{\phi}\rangle = e^{-i\varphi_T(\phi')}|\phi'\rangle.$$

These replacements differ only in subtle ways, but the first replacement sometimes leads to energy estimates with smaller variance. However, testing of the fixed-phase method has been limited to date.

With either option, $\mathcal{O}_T(\phi)$ is real but $\mathcal{O}_T(\phi')$ is not. The replacement of $|\phi'\rangle$ with $|\tilde{\phi}\rangle$ creates a $\mathcal{O}_T(\tilde{\phi})$ that is real. As defined above, $|\tilde{\phi}\rangle$ is of the form $a\det[\Phi']$, which equals $\det[a^{1/N_{\mathrm{elec}}}\Phi']$, which in turn defines $\tilde{\Phi}$ to be $a^{1/N_{\mathrm{elec}}}\Phi'$.

The conversion of the constrained-path algorithm (Algorithms 39 and 40) to the constrained-phase one involves just a few steps. The first step is switching to complex arithmetic and variables where necessary. The next steps are replacing in the procedures that propagate by $B_{K/2}$ and B_V the testing whether $\mathcal{O}_T(\phi') < 0$ and the updating of the walker weight with $w \leftarrow w\ |\mathcal{O}_T(\phi')|$. Then, in the procedure that executes B_V, we replace the $\mathcal{O}_T(i)$ in the computation of the p_i with $|\mathcal{O}_T(i)|/\mathcal{O}_T(\phi)$. Finally, after the execution of $B_{K/2}B_VB_{K/2}$, the phase is constrained. Here the first step is computing $\varphi_T(\phi')$ and testing whether its absolute value is less than $\pi/2$. If so, then $|\phi\rangle$ is updated with $|\tilde{\phi}\rangle$ and $\mathcal{O}_T(\phi)$ is updated with $\exp[-i\varphi_T(\phi')]\mathcal{O}_T(\tilde{\phi})$. If not, then the weight of the walker is set to zero.

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press

Suggested reading

S. Fahy and D. R. Hamann, "Diffusive behavior of states in the Hubbard-Stratonovich transformation," *Phys. Rev. B* **43**, 765 (1991).

B. L. Hammond, W. A. Lester, Jr., and P. J. Reynolds, *Monte Carlo Methods in ab initio Quantum Chemistry* (Singapore: World-Scientific, 1994), chapters 3 and 4.

Exercises

11.1 For the lattice fixed-node method, prove $H|\psi_T\rangle = H_{\text{eff}}|\psi_T\rangle$.

11.2 Derive (11.9).

11.3 For a nearest-neighbor tight-binding model on a square lattice with periodic boundary conditions, derive analytically the energy dispersion relation. For a succession of increasing lattice sizes, plot $E(k)$ versus k and document the scaling of the gap between the highest occupied state and the lowest unoccupied state at half-filling. Now derive the dispersion for the phased boundary condition, using (11.17):

$$\psi(x, y) = \psi(x + L, y)\, e^{i\phi} = \psi(x, y + L)\, e^{i\phi}.$$

For a succession of increasing lattice sizes and an increasing number of randomly chosen values of ϕ for each size, plot $E(k)$ versus k. Compare and contrast the behavior of the gaps, particularly near half-filling.

11.4 For a nearest-neighbor tight-binding model on a hexagonal lattice with periodic boundary conditions, derive analytically the energy dispersion relation. For a succession of increasing lattice sizes, plot $E(k)$ versus k and document the scaling of the gap between the highest occupied state and the zero mode and/or the lowest unoccupied state and zero mode at half-filling. Now derive the dispersion for the phased boundary condition, using (11.17) in each dimension. For a succession of increasing lattice sizes and an increasing number of randomly chosen values of ϕ for each size, plot $E(k)$ versus k. Compare and contrast the behavior of the gaps, particularly near half-filling.

11.5 If the Slater determinants $|\phi\rangle$ and $|\phi'\rangle$ are defined by the matrices Φ and Φ', show that the overlap integral equals $\det(\Phi^T\Phi')$.

https://doi.org/10.1017/CBO9780511902581.012 Published online by Cambridge University Press