User's guide to Monte Carlo methods for evaluating path integrals

Marise J. E. Westbroek and Peter R. King
Department of Earth Science and Engineering,
Imperial College London, London SW7 2BP, United Kingdom

Stephan Dürr
University of Wuppertal, Gaußstraße 20, D-42119 Wuppertal, Germany and
Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52425 Jülich, Germany

Dimitri D. Vvedensky
The Blackett Laboratory, Imperial College London, London SW7 2AZ, United Kingdom

We give an introduction to the calculation of path integrals on a lattice, with the quantum harmonic oscillator as an example. In addition to providing an explicit computational setup and corresponding pseudocode, we pay particular attention to the existence of autocorrelations and the calculation of reliable errors. The over-relaxation technique is presented as a way to counter strong autocorrelations. The simulation methods can be extended to compute observables for path integrals in other settings.

I. INTRODUCTION

Undergraduate physics students are familiar with quantum mechanics as formulated by Schrödinger, Heisenberg, and others in the 1920s.[1,2] Schrödinger's equation and Heisenberg's matrix mechanics are based on Hamiltonian classical mechanics, which provides a direct connection between (classical) Poisson brackets and (quantum) commutators of observables. In 1933, Dirac[3,4] proposed an approach to quantum mechanics based on the Lagrangian, which he regarded as more fundamental than the Hamiltonian. Dirac suggested that the transition amplitude in quantum mechanics, also called the propagator, corresponds to the quantity $\exp(iS/\hbar)$, in which S is the classical action evaluated along the path the particle takes.

In 1948, Feynman[5] extended Dirac's ideas and formulated quantum mechanics based on the sum over *all* paths between fixed initial and final states. Each path contributes a pure phase $\exp(iS/\hbar)$ to the propagator, as Dirac suggested, with the amplitudes of the paths combined according to the usual quantum mechanical procedure for the superposition of amplitudes. Because the sum over paths is typically an integral over a continuum of paths, this procedure is now known as the *path integral method*.

Feynman derived his path integral method in a seminal paper[5] that laid the foundation for many formal developments and applications of path integrals in other areas of physics,[6,7] most notably, in quantum field theory,[8] statistical mechanics,[9] and stochastic dynamics.[10] In fact, although largely unknown to the physics community at that time, the notion of the integral over paths had been introduced in the 1920s by the mathematician Norbert Wiener[11] for diffusion and Brownian motion. Wiener's presentation had a similar formal structure to the Feynman path integral, though in a purely classical context.[6,12]

Path integrals provide an intuitively appealing framework for interpreting many aspects of quantum mechanics. A fundamental property of path integrals is the emergence of the classical limit as $\hbar \to 0$. In quantum mechanics ($\hbar \neq 0$), the classical path and nearby paths contribute constructively to the path integral, and others oscillate rapidly and cancel. As $\hbar \to 0$, the nearby paths oscillate rapidly and also cancel, leaving only the contribution from the classical path, which minimizes the action.[13] Quantum mechanical paths can explore regions unavailable to the classical path, leading to phenomena such as tunneling. The double-slit experiment, which is a conceptually simple demonstration of a fundamental difference between classical and quantum physics[14,15] is an example where the path integral provides a compelling basis for interpreting this experiment.

The path integral is an explicit expression for the probability amplitude. The actual calculation of these amplitudes depends on the problem of interest. In rare cases, such as the harmonic oscillator, the path integral can be evaluated exactly, but typically either an approximate solution is found, or a perturbative expansion is done. Mean-field theory, steepest descent, and the renormalization group are established methods for obtaining exact or approximate solutions from path integrals.[16,17]

However, there are situations when approximate solutions are ineffective. Among the best-known such example is quantum chromodynamics (QCD), the prevailing theory of hadronic matter and a component of the standard model of particle physics. In a regime where the coupling constant is small, calculations based on perturbation theory have been successful. In the strong coupling regime, however, such calculations fail, and an alternative approach is required. In this case, lattice QCD, in which the original theory is discretized on a space-time lattice, provides a framework for the non-perturbative numerical evaluation of amplitudes and matrix elements. The methodology is based on Markov chain Monte Carlo methods,[18–20] the subject of this paper.

In the following we provide a guide to the numerical evaluation of path integrals, using the harmonic oscillator as an example. We will focus on trajectories $x(t)$ in one spatial dimension. The time t is described by a lattice and takes discrete values. In addition to conceptual simplicity, this toy model has the advantage of having an exact solution, which enables the verification of the methodology. Although general descriptions of the computational procedures are available,[18] we provide a pedagogical description of the implementation, methods for error analysis, and suggestions for improving the computational procedures.

The organization of our paper is as follows. The theoretical framework of our calculations is set out in Sec. II, including the derivation of the path integral and the correlation functions we will use. Our computational procedure is summarized in Sec. III, including the definition of observables, the updating algorithm, the notion of thermalization, and correlations within the sampled paths. The jackknife

analysis of the variance of correlated variables is the subject of Sec. IV, and the autocorrelation time of a sequence of configurations is discussed in Sec. V. The technique of over-relaxation for reducing autocorrelation times is introduced in Sec. VI. Advanced topics based on the harmonic oscillator are discussed in Sec. VII.

II. THEORETICAL BACKGROUND

The solution to the initial-value problem of the Schrödinger equation,

$$i\hbar\frac{\partial\psi}{\partial t} = \hat{H}\psi\,, \tag{1}$$

can be written as

$$\psi(x,t) = e^{-i\hat{H}t/\hbar}\psi(x,0)\,, \tag{2}$$

where the exponential factor is known as the *evolution operator*. The exponential of an operator $\hat{O}$ is defined by the Taylor series of the exponential function:

$$e^{\hat{O}} = \sum_{n=0}^{\infty}\frac{\hat{O}^n}{n!}\,. \tag{3}$$

Equation (2) is only a formal solution to Eq. (1) because obtaining an explicit solution from the evolution operator is no simpler than solving the original equation.

The connection between the evolution operator and Feynman's path integral can be made by considering the matrix elements of the evolution operator between any two initial and final position eigenstates. In Dirac's bra-ket notation[27]

$$\langle x_f|e^{-i\hat{H}(t_f-t_i)/\hbar}|x_i\rangle = \langle x_f,t_f|x_i,t_i\rangle\,. \tag{4}$$

These matrix elements embody all the information about how a system with the Hamiltonian $\hat{H}$ evolves, or propagates, in time, and is known as the *propagator*. In particular, the evolution of the wave function is given by

$$\begin{aligned}\psi(x_f,t_f) &= \langle x_f,t_f|\psi\rangle &(5)\\ &= \int\langle x_f,t_f|x_i,t_i\rangle\langle x_i,t_i|\psi\rangle\,dx_i &(6)\\ &= \int\langle x_f,t_f|x_i,t_i\rangle\psi(x_i,t_i)\,dx_i\,, &(7)\end{aligned}$$

which shows that the propagator (4) is a type of Green function known as the fundamental solution of Eq. (1).

A. Derivation of the path integral

The standard derivation of the path integral from the evolution operator considers the evolution of a system over a short time δt. The method can be demonstrated for the Hamiltonian

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{x})\,, \tag{8}$$

of a particle of mass m moving in a potential V, where $\hat{p}$ and $\hat{x}$ signify momentum and position operators.

The propagator to be evaluated is

$$\langle x_f, t_i + \delta t | x_i, t_i \rangle = \langle x_f | e^{-i\hat{H}\delta t/\hbar} | x_i \rangle \tag{9}$$

$$= \int \langle x_f | p \rangle \langle p | e^{-i\hat{H}\delta t/\hbar} | x_i \rangle \, dp \, . \tag{10}$$

We expand the exponential to first order in δt:

$$\langle p | e^{-i\hat{H}\delta t/\hbar} | x_i \rangle = \left\langle p \left| 1 - \frac{i\hat{H}\delta t}{\hbar} + \mathcal{O}(\delta t)^2 \right| x_i \right\rangle . \tag{11}$$

The explicit mention of $\mathcal{O}(\delta t)^2$ corrections will be henceforth omitted.

For the Hamiltonian in Eq. (8) the matrix elements of the operators on the right-hand side of Eq. (11) are evaluated using

$$\langle p | 1 | x_i \rangle = \langle p | x_i \rangle \tag{12}$$

$$\langle p | \hat{p}^2 | x_i \rangle = \langle p | \hat{p}^2 | p \rangle \langle p | x_i \rangle = p^2 \langle p | x_i \rangle \tag{13}$$

$$\langle p | V(\hat{x}) | x_i \rangle = \langle p | x_i \rangle \langle x_i | V(\hat{x}) | x_i \rangle = V(x_i) \langle p | x_i \rangle \, . \tag{14}$$

The short-time propagator in Eq. (11) can now be approximated as

$$\langle p | e^{-i\hat{H}\delta t/\hbar} | x_i \rangle \approx \left[1 - \frac{ip^2\delta t}{2m\hbar} - \frac{i\delta t}{\hbar} V(x_i) \right] \langle p | x_i \rangle \tag{15}$$

$$\approx \exp \left\{ - \frac{i}{\hbar} \left[\frac{p^2 \delta t}{2m} + V(x_i)\delta t \right] \right\} \langle p | x_i \rangle \, , \tag{16}$$

with the approximations becoming equalities for infinitesimal δt. We use

$$\langle p | x \rangle = \frac{e^{-ipx/\hbar}}{\sqrt{2\pi\hbar}} \, , \tag{17}$$

to obtain

$$\langle p | e^{-i\hat{H}\delta t/\hbar} | x_i \rangle = \frac{1}{\sqrt{2\pi\hbar}} \exp \left\{ - \frac{i}{\hbar} \left[p x_i + \frac{p^2 \delta t}{2m} + V(x_i)\delta t \right] \right\} . \tag{18}$$

We return to the right-hand side of Eq. (10) and invoke Eq. (17) to find,

$$\langle x_f, t_i + \delta t | x_i, t_i \rangle = \int \frac{dp}{2\pi\hbar} \exp \left\{ - \frac{i\delta t}{\hbar} \left[\frac{p(x_i - x_f)}{\delta t} + \frac{p^2}{2m} + V(x_i) \right] \right\} \tag{19}$$

$$= \sqrt{\frac{m}{2\pi i\hbar\delta t}} \exp \left\{ \frac{i}{\hbar} \left[\frac{m(x_f - x_i)^2}{2\delta t} - V(x_i)\delta t \right] \right\} . \tag{20}$$

The integral has been evaluated by completing the square in the argument of the exponential. If we make the identification

$$\left(\frac{dx}{dt} \right)^2 = \left(\frac{x_f - x_i}{\delta t} \right)^2 , \tag{21}$$

we see that the argument of the exponential on the right-hand side of Eq. (20) is the product of δt and the classical Lagrangian L:

$$L \, \delta t = \left[\frac{m}{2} \left(\frac{x_f - x_i}{\delta t} \right)^2 - V(x_i) \right] \delta t \, . \tag{22}$$

Hence, the short-time propagator reduces to

$$\langle x_f, t_i + \delta t | x_i, t_i \rangle = \sqrt{\frac{m}{2\pi i\hbar\delta t}} \, e^{iL\delta t/\hbar} \, . \tag{23}$$

We can now evaluate propagators over finite times by dividing the time interval into slices of duration δt,

$$\begin{aligned}\langle x_f, t_f | x_i, t_i \rangle = \iint \cdots \int & \langle x_f, t_f | x_{N-1}, t_{N-1} \rangle \\ & \times \langle x_{N-1}, t_{N-1} | x_{N-2}, t_{N-2} \rangle \cdots \langle x_2, t_2 | x_1, t_1 \rangle \\ & \times \langle x_1, t_1 | x_i, t_i \rangle \, dx_1 \, dx_2 \cdots dx_{N-1} \,, \end{aligned} \tag{24}$$

and applying Eq. (23) to each slice:

$$\langle x_f, t_f | x_i, t_i \rangle = \int \prod_{n=1}^{N-1} dx_n \exp\left[\frac{i\delta t}{\hbar} \sum_{n=1}^{N-1} L(t_n) \right]. \tag{25}$$

We have omitted the prefactors in Eq. (25) because they will not be needed in the following.

In the continuum limit ($N \to \infty$, $\delta t \to 0$, such that the product $N\delta t$ is fixed), the integral over positions at each time is the same as the integral over all paths between the initial and final positions:

$$\langle x_f, t_f | x_i, t_i \rangle = \int Dx(t) \, e^{-iS/\hbar} \,, \tag{26}$$

where $Dx(t) \equiv \prod_{n=1}^{N-1} dx_i$ and, as $N \to \infty$, the action S of the path $x(t)$ becomes

$$S = \int_{t_i}^{t_f} L(x(t)) \, dt = \int_{t_i}^{t_f} \left[\frac{m}{2} \left(\frac{dx}{dt} \right)^2 - V(x(t)) \right] dt \,. \tag{27}$$

B. Imaginary time path integrals

The path integral in Eqs. (26) and (27) yields transition amplitudes as the sum of the phases of all paths between the given initial and final positions. For our purposes *imaginary time* path integrals, where the time t is replaced by $-i\tau$, with τ real, are of primary interest.

There are two main applications of imaginary time path integrals. In statistical mechanics $\tau = \hbar/(k_B T)$, where k_B is Boltzmann's constant and T is the absolute temperature. Thus, for equal initial and final positions x, an integration over x produces the partition function Z:

$$Z = \int \langle x | e^{-\hat{H}\tau/\hbar} | x \rangle \, dx = \mathrm{Tr}\big(e^{-\hat{H}\tau/\hbar} \big) \,, \tag{28}$$

in which the trace Tr is the sum/integral of the diagonal elements of an operator.

Another application is the determination of the energy spectrum of a quantum system. This calculation utilizes the identity $1 = \sum_n |n\rangle\langle n|$ in terms of the eigenfunctions of the Hamiltonian, such that $\hat{H}|n\rangle = E_n |n\rangle$,

$$Z = \int \langle x | e^{-\hat{H}\tau/\hbar} | x \rangle \, dx = \int \sum_n \langle x | e^{-\hat{H}\tau/\hbar} | n \rangle \langle n | x \rangle \, dx \tag{29}$$

$$= \int \sum_n e^{-E_n \tau/\hbar} \psi_n(x) \bar{\psi}_n(x) dx = \sum_n e^{-E_n \tau/\hbar}, \tag{30}$$

where we have used the fact that $\psi_n(x)$ is normalized. Similarly, we can expand the propagator $\langle x_f, t_f | x_i, t_i \rangle$ in terms of the eigenfunctions $\{|n\rangle\}$:

$$\langle x_f, t_f | x_i, t_i \rangle = \sum_{n=0}^{\infty} e^{-E_n (t_f - t_i)/\hbar} \langle x_f | n \rangle \langle n | x_i \rangle. \tag{31}$$

The derivation of the imaginary-time path integral proceeds along the same lines as the real-time propagator, with the result corresponding to Eq. (20) given by

$$\langle x_f | e^{-\hat{H}\delta\tau/\hbar} | x_i \rangle = \sqrt{\frac{m}{2\pi\hbar\delta\tau}} \exp\left\{ -\frac{\delta\tau}{\hbar} \left[\frac{m}{2} \left(\frac{x_f - x_i}{\delta\tau} \right)^2 + V(x_i) \right] \right\}, \tag{32}$$

In the limit $N \to \infty$, Eq. (32) can be used to write the partition function in a form analogous to Eqs. (26) and (27):

$$Z = \mathrm{Tr}\big(e^{-\hat{H}(\tau_f - \tau_i)/\hbar}\big) = \int Dx(\tau)\, e^{-S/\hbar}\,, \tag{33}$$

where S is the (Euclidean) action over a path $x(\tau)$ with $\tau_f \geq \tau \geq \tau_i$, and $x(\tau_f) = x_f,\ x(\tau_i) = x_i$.

$$S = \int_{\tau_i}^{\tau_f} L(x(\tau))\, d\tau = \int_{\tau_i}^{\tau_f} \left[\frac{m}{2}\left(\frac{dx}{d\tau}\right)^2 + V(x(\tau))\right] d\tau\,. \tag{34}$$

The integrals in Eqs. (33) and (34) and their real-time counterparts in Eqs. (26) and (27) are over all paths weighted by Lagrangian-type quantities. However, in the imaginary-time formalism, quantities associated with the paths are real.

C. The quantum harmonic oscillator

The Hamiltonian for a particle of mass m bound by a harmonic potential with force constant k is

$$\hat{H} = \frac{\hat{p}^2}{2m} + \frac{k\hat{x}^2}{2} = \frac{\hat{p}^2}{2m} + \frac{m\omega^2\hat{x}^2}{2}\,, \tag{35}$$

where $\omega = \sqrt{k/m}$ is the natural frequency of the oscillator. The discretized Euclidean Lagrangian for this system is

$$L_i = \frac{m}{2}\left(\frac{x_{i+1} - x_i}{\delta\tau}\right)^2 + \frac{m\omega^2 x_i^2}{2}\,, \tag{36}$$

which allows us to express the Euclidean action and the partition function as

$$S = \sum_{n=1}^{N-1} L_i \tag{37}$$

$$Z = \int_{-\infty}^{\infty} \prod_{i=1}^{N-1} dx(\tau_i) \exp\left(-\frac{\delta\tau}{\hbar} S\right). \tag{38}$$

The energy eigenvalues of $\hat{H}$ are $E_n = \hbar\omega(n + \frac{1}{2})$ for $n = 0, 1, 2, \ldots$. The normalized ground state wave function is

$$\psi_0 = \left(\frac{m\omega}{\pi\hbar}\right)^{\frac{1}{4}} \exp\left(-\frac{m\omega x^2}{2\hbar}\right), \tag{39}$$

from which all other wave functions can be obtained through ladder operations. Expectations of observable quantities in the ground state are determined by $\psi_0(x)$.

III. COMPUTATIONAL METHOD

The formalism discussed in Sec. II will be applied to the harmonic oscillator. However, the range of applicability is much broader. The idea is that if the partition function can be constructed ("if the system can be simulated"), an arbitrary observable can be determined ("measured") with a statistical uncertainty that decreases as the simulation is extended. For the construction of such observables and their evaluation the complete tool set of statistical mechanics can be used. An overview of all parameters and their meanings is given in Table I.

Parameter	Meaning
N_τ	number of elements of the time lattice
$i\delta\tau$	Euclidean time, with $i \in \{1, \ldots, N_\tau\}$ the site index
t_{MC}	Monte Carlo time; refers to index of a path in the Markov chain
sweep	N_τ applications of the single-site Metropolis–Hastings algorithm
$N_{\mathrm{sep}} - 1$	number of discarded paths between successive paths used for measurement
$\delta\tau$	lattice spacing
$\tilde{m}$	dimensionless effective mass: $\tilde{m} = m\delta\tau$
$\tilde{\omega}$	dimensionless frequency: $\tilde{\omega} = \omega\delta\tau$
N	number of paths within an ensemble
ξ	"correlation time": Euclidean time for two-point correlations to diminish by a factor e
m_{eff}	effective mass: $m_{\mathrm{eff}} = 1/\xi$
N_B	number of bins in jackknife procedure
B	bin width $B = N/N_B$
$\tau_{O,\mathrm{exp}\,/\mathrm{int}}$	exponential and integrated autocorrelation time of observable O

Table I. An overview of all parameters and their meanings.

A. Monte Carlo methods

The simulation is done on a discrete time lattice with N_τ time slices with periodic boundary conditions so that the time slice $N_\tau + 1$ equals the time slice 1. To calculate the statistics of the observables, many particle trajectories of the form $(x_1, \ldots, x_{N_\tau})$ are needed, where each coordinate is a real number. Starting from an initial, thermalized configuration path$^{(0)}$ (see Sec. III D), the path is updated by the Metropolis–Hastings algorithm. The application of this elementary update to the variable x_i for each time slice i constitutes one "sweep" or one Monte Carlo step per site. One Metropolis sweep yields the next path in the sequence, e.g., path$^{(1)}$ from path$^{(0)}$. Because path$^{(\nu)}$ relies only on path$^{(\nu-1)}$, the trajectories constitute a Markov chain. The computational method is illustrated in Fig. 1. We distinguish between Euclidean time $\tau = i\delta\tau$, which indicates the index i of a lattice site, and Monte Carlo time, which refers to the index of a path ν in the Markov chain. Because each path in the chain is based on the previous path, the paths are correlated. We will discuss these correlations in more detail in Sec. V. To combat the autocorrelation, we discard a number $N_{\mathrm{sep}} - 1$ of paths between every two paths used for measurements. We will refer to the remaining paths used to calculate average quantities as "configurations."

B. Dimensionless variables and observables

Because computer code can handle only pure numbers, it is necessary to express the physics of the system in dimensionless form. A naive way of doing this is by expressing all quantities in metric units, e.g., meters. A disadvantage of this choice is that it leads to numbers that often span several orders of magnitude. To avoid this problem, we express all variables in terms of the appropriate power of the lattice spacing $\delta\tau$. To this end, we set $\hbar = 1 = c$, which implies that [time] = [length] = [mass^{-1}] = [energy^{-1}]. We introduce the dimensionless variables:

$$\tilde{m} = m\delta\tau, \quad \tilde{\omega} = \omega\delta\tau, \quad \tilde{x}_i = \frac{x_i}{\delta\tau}, \tag{40}$$

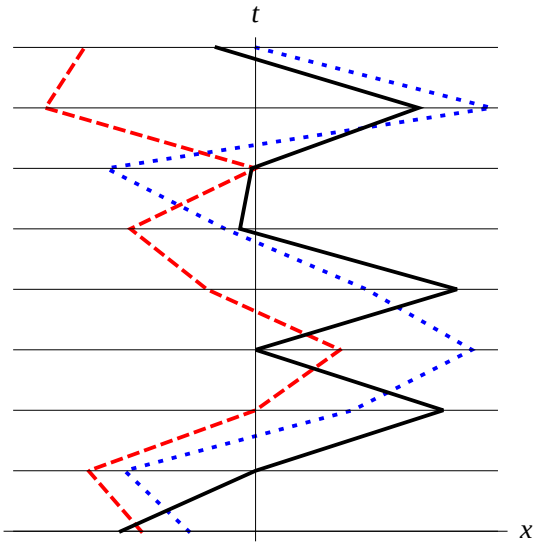

Figure 1. (Color online) Illustration of the computational method. The spatial location of the particle at time τ_i, where $i = 1, \ldots, N_\tau = 8$, can take any real value, but is constrained by the potential centered at the origin and its neighboring positions at times τ_{i-1} and τ_{i+1}. The solid line represents the thermalized path$^{(0)}$; the dotted line is the next path in the Markov chain, path$^{(1)}$; and the dashed line is the resulting trajectory after 19 further Metropolis sweeps.

The dimensionless action becomes

$$\tilde{S} = \sum_{i=1}^{N_\tau} \frac{1}{2}\tilde{m}(\tilde{x}_{i+1} - \tilde{x}_i)^2 + \frac{1}{2}\tilde{m}\tilde{\omega}^2\tilde{x}_i^2, \tag{41}$$

where $\tilde{m}$, $\tilde{\omega}$ and $\{\tilde{x}_i\}$ are dimensionless. Note that the summation range differs from the one in Eq. (38), due to periodic boundary conditions. We restrict ourselves to the subspace $\tilde{m} = \tilde{\omega}$ of the available parameter space. The parameter $\tilde{m}$ can thus be viewed as the effective lattice spacing of a harmonic oscillator with unit mass and unit natural frequency.

The continuum limit applies to observables and is not taken for an individual simulation. Rather, we run a series of simulations, each with a smaller effective lattice spacing (that is, a smaller value of $\tilde{m}$) and a greater value of N_τ than the previous, such that the product $N_\tau\tilde{m}$ is fixed.

For a quadratic action, all odd moments of $\hat{x}$ have zero expectation value. An analytic expression for $\langle\hat{x}^2\rangle$ is derived in Ref. 18:

$$\langle\hat{x}^2\rangle = \frac{1}{2\tilde{m}\tilde{\omega}\sqrt{1 + \frac{1}{4}\tilde{\omega}^2}}\left(\frac{1 + R^{N_\tau}}{1 - R^{N_\tau}}\right), \tag{42}$$

with the auxiliary variable

$$R = 1 + \frac{\tilde{\omega}^2}{2} - \tilde{\omega}\sqrt{1 + \frac{\tilde{\omega}^2}{4}}. \tag{43}$$

The other observable we need is

$$\langle\hat{x}^4\rangle = \frac{3}{(2\tilde{m}\tilde{\omega})^2(1 + \frac{1}{4}\tilde{\omega}^2)}\left(\frac{1 + R^{N_\tau}}{1 - R^{N_\tau}}\right)^2 = 3\langle\hat{x}^2\rangle^2. \tag{44}$$

Derivations of the expressions for Eqs. (42) and (44) are given in Ref. 27.

C. The Metropolis update

The core of our path-generating algorithm is an update of a single site based on the Metropolis–Hastings algorithm.[28,29] The output is a set of N paths $\{\tilde{x}_1, \ldots, \tilde{x}_{N_\tau}\}$ with Boltzmann weights $\rho[\{\tilde{x}_i\}] \sim \exp[-\tilde{S}[\{\tilde{x}_i\}]]$. The input of the Metropolis update is an array path with N_τ sites, a real number h, and the parameters $\tilde{m}$ and $\tilde{\omega}$. Periodic boundary conditions avoid the need to abandon (in the data) sites affected by the lattice edges. One sweep visits N_τ sites in random order. A site may be visited repeatedly or not at all, but the mean number of visits per sweep for each site is one. Each Metropolis update to a given site consists of four steps.

1. Generate a random number u from a uniform distribution in the interval $[-h, h]$.
2. Propose a change to the visited site, $\tilde{x}_i \to \tilde{x}'_i = \tilde{x}_i + u$.
3. Compute the change in the action $\delta\tilde{S}$ as a result of this trial modification.
4. Accept the change with probability $\min\{1, e^{-\delta\tilde{S}}\}$.

Pseudocode is provided in Ref. 27. The probability $\min\{1, e^{-\delta\tilde{S}}\}$ in step 4 implies that proposed modifications that lower the action are always accepted. A trial that would increase the action is accepted with probability $e^{-\delta\tilde{S}}$. This decision is made in an accept/reject step. The Metropolis update satisfies detailed balance:[19]

$$p(\tilde{x}_i \to \tilde{x}'_i) e^{-\tilde{S}(\tilde{x}_i)} = p(\tilde{x}'_i \to \tilde{x}_i) e^{-\tilde{S}(\tilde{x}'_i)} \,. \tag{45}$$

Because of this property, the estimated average of an observable $\hat{O}$ reduces to an arithmetic average. After one sweep, the acceptance ratio is computed. The value of h is adjusted to meet a predefined acceptance ratio. We chose the desired acceptance ratio to be 0.8 (which is a conventional choice in lattice QCD), even though we suspect the ideal value for the harmonic oscillator, with the choices of $\tilde{m} = \tilde{\omega}$ listed in Table II, to be smaller. Although algorithms with too low or too high an acceptance ratio are less efficient, the generated Boltzmann distribution of paths is unaffected by this choice.

D. Thermalization

The required thermalization process can start from an array of zeros (a "cold" start), random numbers (a "hot" start), or an initial path that is expected to be close to a thermalized path. The initial thermalization steps are not characteristic of the probability density $\rho[\{\tilde{x}_i\}] \sim \exp[-\tilde{S}\{\tilde{x}_i\}]$ and must be discarded lest they skew the simulation. A trial run is one way to choose the number of sweeps needed before the array qualifies as a thermalized path. An example of a thermalization process is given in Fig. 2. For each configuration, the observable $\langle \hat{x}^2 \rangle$ was measured to monitor its fluctuations around the expected values.[30] In our code we first average $\tilde{x}_i^2$ over the N_τ time slices in a given configuration, and then compute the ensemble average of that number over the configurations. For actions for which the exact answer is not known, independent "hot" and "cold" runs can help to establish the expectation value. In this case, the first 50–100 configurations should not be used.

E. Two-point correlation function

To make optimal use of CPU time, it is important to choose the number of sites N_τ as small as possible, but large enough to avoid finite-size effects. How is the lower bound on N_τ established? Correlations within a lattice are quantified by the connected two-point function:

$$G(\Delta\tau) = \langle x(\tau) x(\tau + \Delta\tau) \rangle - \langle x(\tau) \rangle \langle x(\tau + \Delta\tau) \rangle \,, \tag{46}$$

where we have written $x(\tau)$ instead of x_i to emphasize the dependence of G on the time difference $\Delta\tau$, where τ and $\Delta\tau$ can be any multiple of the lattice spacing. Because $\langle x(\tau) \rangle = 0$ for all τ for the

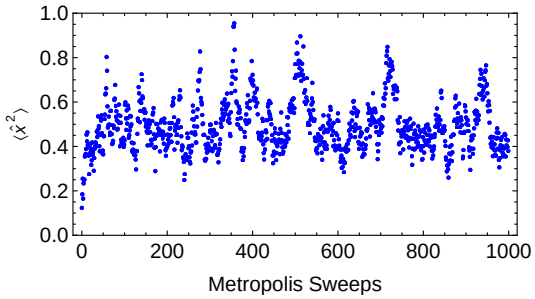

Figure 2. (Color online) A trial run for $\langle \hat{x}^2 \rangle$ to illustrate thermalization effects ($\tilde{m} = \tilde{\omega} = 0.1$, $N_\tau = 1200$). One thousand paths were discarded between every two configurations whose output is shown. The first 50–100 configurations should not be used for measurements.

$\tilde{m} = \tilde{\omega}$	N_τ	$\tilde{m} = \tilde{\omega}$	N_τ
1	120	0.1	1200
0.8	150	0.08	1500
0.6	200	0.06	2000
0.5	240	0.05	2400
0.3	400	0.03	4000
0.2	600	0.02	6000
		0.01	12000

Table II. Effective lattice spacings used for the results shown in Figs. 4–8.

harmonic oscillator, we work with the two-point function

$$G(\Delta\tau) = \langle x(\tau)x(\tau+\Delta\tau)\rangle \tag{47}$$

$$= \frac{1}{N_\tau}\sum_{i=1}^{N_\tau} \sum_{\substack{j \\ (j-i) \bmod N_\tau = \Delta\tau}} x(i)x(j). \tag{48}$$

An example of the exponential decay of $G(\Delta\tau)$ is shown in Fig. 3(a),[22]

$$G(\Delta\tau) = Ae^{-\Delta\tau/\xi} + Ae^{-(T-\Delta\tau)/\xi}, \tag{49}$$

where ξ is the correlation time and T is the final time. The second term in Eq. (49) is due to periodic boundary conditions. The total length of the lattice must be greater than ξ. We choose N_τ to be about $10\tilde{\xi}$ (where $\tilde{\xi}$ is the correlation time expressed in lattice units).

An estimate of $1/\tilde{\xi}$ can be obtained from the local logarithmic slope for suitable $\Delta\tau$:[22,31]

$$\frac{1}{\tilde{\xi}} = \frac{1}{2}\log\left[\frac{G(\Delta\tau-1)}{G(\Delta\tau+1)}\right]. \tag{50}$$

The quantity $1/\xi$ is known as the effective mass $m_{\rm eff}$. Figure 3(b) shows $\tilde{m}_{\rm eff}$ for the same set of paths used for Fig. 3(a).

Figure 4 suggests that there is a power-law dependence of $\tilde{\xi}$ on the effective lattice spacing. To construct Fig. 4, we repeated the procedure illustrated in Fig. 3 for 13 effective lattice spacings listed in Table II. The physical length, the product of $\tilde{m} = \tilde{\omega}$ and N_τ, was kept constant. With this choice of parameters, we were able to explore two orders of magnitude in the lattice spacing; the associated N_τ are round numbers.

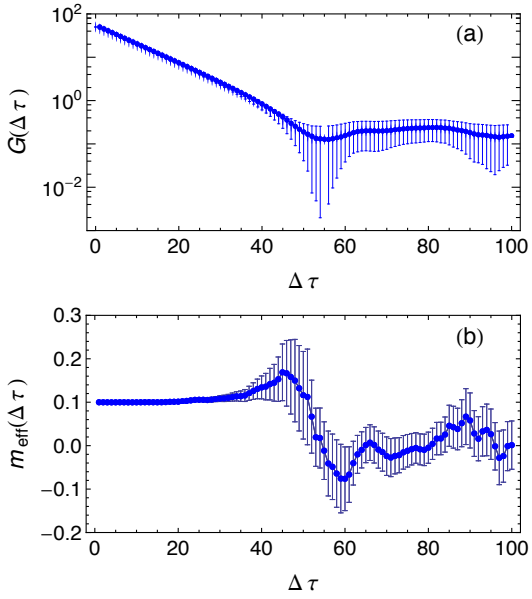

Figure 3. (a) The symmetrized two-point correlation function ($N_\tau = 1200$, $\tilde{m} = \tilde{\omega} = 0.1$, $N = 10^4$). The exponential decrease is swamped by noise after approximately 40 time slices and the magnitude of the error bars starts to increase significantly, and eventually the error bars become unreliable. (b) The effective mass $1/\tilde{\xi}$. The estimates of the errors are reliable until $\Delta\tau \approx 40$.[32]

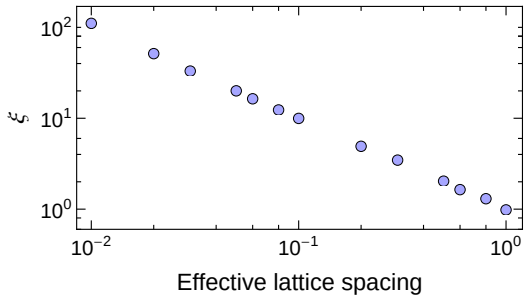

Figure 4. The correlation length ξ *versus* the effective lattice spacing ($N_{\rm sep} = 300$, $N = 10^4$). The points fall on a straight line, indicating a power-law dependence. The error bars were constructed using a jackknife analysis. Error bars are smaller than the symbol size.

IV. JACKKNIFE ANALYSIS

Suppose we compute values $O_1, \ldots, O_N$, of an observable $\hat{O}$ (typically a moment of $\hat{x}$), with the expectation value and variance:

$$\langle \hat{O} \rangle = \langle O \rangle \tag{51}$$

$$\langle (\hat{O} - \langle \hat{O} \rangle)^2 \rangle = \sigma_O^2 \,. \tag{52}$$

The quantity $E(O)$ provides an unbiased estimator of the mean:

$$E(O) = E_N(O) = \frac{1}{N} \sum_{i=1}^{N} O_i \,, \tag{53}$$

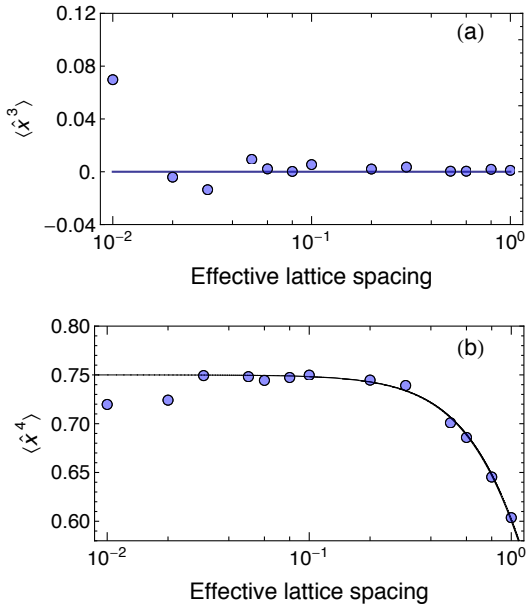

Figure 5. (a) Simulations of $\langle \hat{x}^3 \rangle$ as a function of the effective lattice spacing ($N_{\rm sep} = 300$). (b) Simulations of $\langle \hat{x}^4 \rangle$ as a function of the effective lattice spacing (with the same parameters). The solid line is the exact result Eq. (44). The error bars are smaller than the size of the symbols, because the naive error ignores correlations between measurements.

where N is the number of measurements. An unbiased estimator of the sample variance is given by

$$\sigma^2_{O,\,\rm std} = \frac{1}{N-1} \sum_{i=1}^{N} [O_i - E(O)]^2. \tag{54}$$

This estimator is applicable even if the measurements are somewhat correlated, that is, not entirely independent of each other. The square root of the variance of the estimator of the sample mean is the error in our estimate, not the square root of the variance of the distribution. For the former quantity it matters whether the measurements are independent or not. For uncorrelated (independent) measurements the relation is

$$\sigma^2_{E(O),\,\rm naive} = \frac{\sigma^2_{O,\,\rm std}}{N}. \tag{55}$$

The subscript "naive" refers to the assumption that the variables are not correlated. This assumption implies that the statistical error of the mean in Eq. (53) is given by

$$\mathrm{err}_O = \sqrt{\sigma^2_{E(O),\,\rm naive}} = \frac{\sigma_{O,\rm std}}{\sqrt{N}}. \tag{56}$$

For correlated data the factor of N in the denominator of Eq. (55) is replaced by $N_{\rm eff} < N$, and the statistical error of the mean differs from the naive estimate in Eqs. (55) and (56). That is, for correlated data the naive estimate underestimates the true statistical error of the sample mean as shown in Fig. 5 for $\langle \hat{x}^3 \rangle$ and $\langle \hat{x}^4 \rangle$. The statistical error is smaller than the size of the dots in Fig. 5, yet many of the dots do not lie on the exact curve. For a fixed number of sweeps between adjacent measurements ($N_{\rm sep} = 300$) the problem is seen to worsen for smaller effective lattice spacings.

The jackknife procedure[33] provides a more realistic estimate of the variance of the mean of a set of correlated variables. The jackknife variance of a parameter is found by systematically leaving out batches of observations from a dataset, calculating the variance each time a different batch is omitted, and finding the average of these variance calculations. A theoretical justification of this procedure is

given in Ref. 34. The N samples are divided into N_B blocks of bin width B. The block estimators are

$$o_k = \frac{1}{B}\sum_{i=1}^{B} O_{(k-1)B+i} \quad (k = 1, \dots N_B). \tag{57}$$

The bin width B should exceed the autocorrelation time of the observable to ensure that the N_B values can be treated as uncorrelated.[35] The bin-based variance of the mean is given by

$$\sigma^2_{E(O),\,\mathrm{bins}} = \frac{1}{N_B(N_B-1)}\sum_{k=1}^{N_B}[o_k - E(O)]^2. \tag{58}$$

The estimator o_k in Eq. (58) is an average over only the N_Bth fraction of all the measurements. This limited number of measurements may prevent the determination of o_k for some k (for instance a fit based on too few configurations may occasionally fail to converge).[36] This problem is overcome by using complementary bins:

$$\tilde{o}_k = \frac{1}{N-B}\left(\sum_{i=1}^{N} O_i - B o_k\right). \tag{59}$$

Rather than N_B estimators $o_1, \dots, o_{N_B}$, each containing B measurements, as in Eq. (57), we work with N_B jackknife estimators $\tilde{o}_1, \dots, \tilde{o}_{N_B}$, each based on $N - B \gg B$ measurements. The resulting complementary bin-based or jackknife variance of the mean is

$$\sigma^2_{E(O),\mathrm{jack}} = \frac{N_B - 1}{N_B}\sum_{k=1}^{N_B}[\tilde{o}_k - E(O)]^2. \tag{60}$$

Figure 6 features the same data set as Fig. 5, but this time the statistical errors are determined as $\sigma_{E(O),\,\mathrm{jack}}$ [see Eq. (60)]. The errors are assessed reliably; the $\sigma_{E(O),\,\mathrm{jack}}$ error bars miss the analytical curve at a frequency consistent with the expected 32%, i.e. one standard deviation away assuming a normal distribution.

Note that for $\langle \hat{x}^3 \rangle$ and $\langle \hat{x}^4 \rangle$ the bin-based estimator (58) would have served the same purpose. The jackknife formula (60) is built to reproduce the result of (58) whenever an observable can be determined from a single configuration. In general, this is not true (e.g. for "secondary" observables derived from "primary" observables through a fit). It is, therefore, common practice to use the jackknife procedure by default.

V. AUTOCORRELATION TIME

The correlation of a sequence of generated configurations arises naturally because one configuration differs from the next only by the result of a fixed number of sweeps. The autocorrelation time provides information on how strongly subsequent measurements are correlated. Because correlations lead to increased errors in measurements, its accurate assessment is important. The autocorrelation for an observable $\langle \hat{O} \rangle$, which takes values $\{O_i\}$, as a function of Monte Carlo time t_{MC} is defined as:

$$A_O(t_{\mathrm{MC}}) = E[(O_i - E(O_i))[O_{i+t_{\mathrm{MC}}} - E(O_{i+t_{\mathrm{MC}}}))] \tag{61}$$

$$= \frac{1}{N - t_{\mathrm{MC}} - 1}\sum_{i=1}^{N-t_{\mathrm{MC}}}[O_i - E(O_i)][O_{i+t_{\mathrm{MC}}} - E(O_{i+t_{\mathrm{MC}}})]\,, \tag{62}$$

where the average $E(O_i)$ is over the first $N - t_{\mathrm{MC}}$ measurements and the average $E(O_{i+t_{\mathrm{MC}}})$ over the last $N - t_{\mathrm{MC}}$ measurements. Note that $A_O(0) = \sigma^2_{O,\mathrm{std}}$.Comparisons between autocorrelation times for different observables are easier to make when the normalized $A_O(t_{\mathrm{MC}})/A_O(0)$ is considered instead.

Two parameters can be extracted from Eq. (61): the asymptotic (or exponential) and the integrated autocorrelation times. The autocorrelation function of O typically exhibits multi-exponential behavior:[22]

$$\frac{A_O(t_{\mathrm{MC}})}{A_O(0)} = a_0 e^{-t_{\mathrm{MC}}/\tau_0} + a_1 e^{-t_{\mathrm{MC}}/\tau_1} + \cdots\,, \tag{63}$$

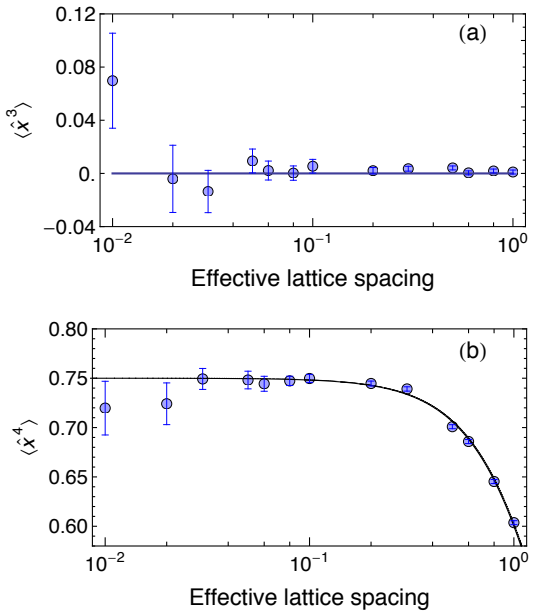

Figure 6. (a) Simulations of $\langle \hat{x}^3 \rangle$ as a function of the effective lattice spacing ($N_{\rm sep} = 300$) with jackknife error bars, which account for correlations within the data. (b) Simulations of $\langle \hat{x}^4 \rangle$ as a function of the effective lattice spacing ($N_{\rm sep} = 300$) with jackknife error bars. The data are always within two jackknife errors of the theoretical results in Eq. (44).

with $\tau_0 < \tau_1 < \tau_2 < \cdots$. Usually $a_0 \gg a_1 \gg a_2 \gg \cdots$, where $a_0 + a_1 + a_2 + \cdots = 1$. Determining the "true" exponential autocorrelation time, defined as $\max(\tau_0, \tau_1, \tau_2, \cdots)$, requires precise data at large $t_{\rm MC}$, which are not normally available. To obtain an estimate for the exponential autocorrelation time, we make a multi-exponential fit as in Eq. (63). In practice, such a fit is likely to contain one or two terms.

To determine the integrated autocorrelation time, we start from the variance of the unbiased estimator of the mean. Let $t_{\rm MC}$ be the absolute time difference between measurement i and j, such that

$$\left(\frac{1}{N}\sum_{i=1}^{N} O_i - E(O_i)\right)^2 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}[O_i - E(O_i)][O_j - E(O_j)] \tag{64}$$

$$= \frac{1}{N}\sigma^2_{O,\,\rm std} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[O_i - E(O_i)](O_j - E(O_j)] \tag{65}$$

$$= \frac{1}{N}\sigma^2_{O,\,\rm std} + \frac{2}{N^2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}[O_i - E(O_i)][O_j - E(O_j)] \tag{66}$$

$$= \frac{\sigma^2_{O,\,\rm std}}{N}\left\{1 + \frac{2}{N}\frac{1}{\sigma^2_{O,\,\rm std}}\sum_{t_{\rm MC}=1}^{N-1}\sum_{i=1}^{N-t_{\rm MC}}[O_i - E(O_i)][O_{i+t_{\rm MC}} - E(O_{i+t_{\rm MC}})]\right\}$$

$$= \frac{2\sigma^2_{O,\,\rm std}}{N}\left\{\frac{1}{2} + \frac{N - t_{\rm MC}}{N}\sum_{t_{\rm MC}=1}^{N-1}\frac{A_O(t_{\rm MC})}{A_O(0)}\right\}. \tag{67}$$

For $N \gg 1$ the right-hand side of Eq. (64) approaches$(2\sigma^2_{O,\rm std}/N)\tau_{O,\,\rm int}$, where

$$\tau_{O,\,\rm int} = \frac{1}{2} + \sum_{t_{\rm MC}=1}^{N-1}\frac{A_O(t_{\rm MC})}{A_O(0)}. \tag{68}$$

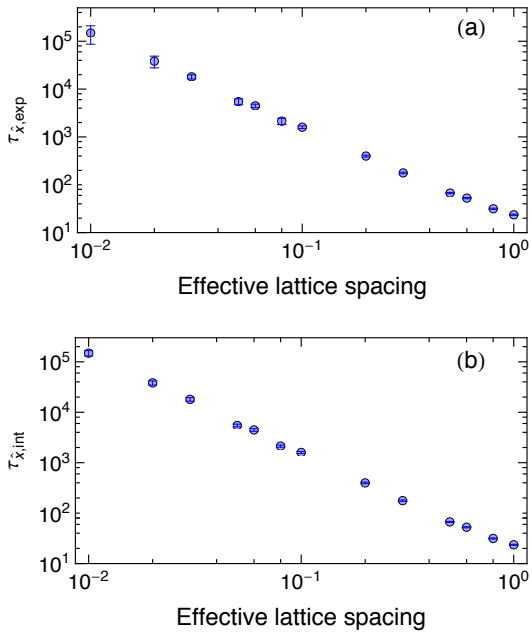

Figure 7. (a) Number of lattice updates required to generate a path independent of its predecessor as a function of the lattice spacing $\tilde{m}$. The exponential autocorrelation time[37] of the observable $\hat{x}$ exhibits power-law behavior as a function of $\tilde{m}$, $\tau_{\hat{x},\mathrm{exp}} \sim \tilde{m}^{-1.86}$. (b) The integrated autocorrelation time[38] as a function of $\tilde{m}$. As for the exponential autocorrelation time, the time required for two independent paths exhibits power-law behavior, $\tau_{\hat{x},\mathrm{int}} \sim \tilde{m}^{-1.85}$.

Comparison of Eq. (64) with the naive variance of the mean in Eq. (55) shows that the effective number of independent measurements is $N_{\mathrm{eff}} = N/(2\tau_{O,\,\mathrm{int}})$.

To compute the integrated correlation time, we have to cut off the sum where the exponential relation for $A_O(t_{\mathrm{MC}})$ breaks down. If the cutoff is clearly defined (for example, at the first value where the autocorrelation becomes negative), the integrated autocorrelation time takes a unique value for a given data set, and is thus less subjective than the exponential correlation time. However, the latter clearly captures the exponential behavior of the autocorrelations.

Figure 7 shows the dependence of the autocorrelation time of an observable $\hat{X}$ on the lattice spacing. For a given operator $\hat{O}$, we expect[22,39] $\tau_O \sim \tilde{\xi}^z$ and $z \simeq 2$ for local updating algorithms.[24] Similar power-law behavior can be observed for the integrated autocorrelation time. We have chosen the number of lattice sites N_τ to be inversely related to the lattice spacing to keep the time T in physical units constant. Hence, when moving toward the continuum, the time increases as $\tau_O \tilde{\xi}^d$ in d dimensions (in our case $d = 1$).

Another way to express the computational difficulties in taking the continuum limit is that for fixed computational time, the exponential increase in the autocorrelation time as a function of the inverse lattice spacing causes an increase of the statistical errors, meaning that the reliability of results is limited for very small lattice spacings. A possible remedy is to change the updating strategy. The multigrid method, which employs intrinsically nonlocal updates, is an effective way to address large correlation lengths.

VI. OVER-RELAXATION

The goal of the over-relaxation method is to reduce autocorrelation times. To this end, a trial change $\tilde{x}_{i,\,\mathrm{trial}}$ far from the old value $\tilde{x}_i$, but involving small changes in the action, is proposed. Creutz[41] and Brown and Woch[42] suggest:

$$\tilde{x}_{i,\,\mathrm{trial}} = \frac{\tilde{x}_{i-1} + \tilde{x}_{i+1}}{1 + \frac{1}{2}\tilde{\omega}^2} - \tilde{x}_i \,. \tag{69}$$

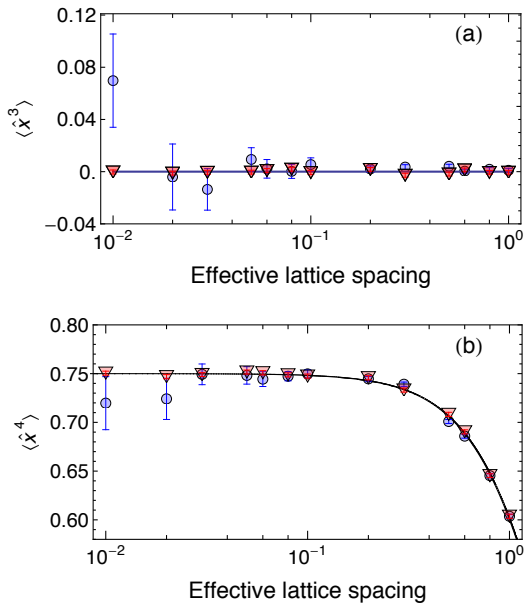

Figure 8. Over-relaxed simulations for the observables $\langle \hat{x}^3 \rangle$ and $\langle \hat{x}^4 \rangle$ are shown by the triangles ($N_\tau = 120/\tilde{m}$; $N_{\rm sep} = 300$). The reliability of the results is improved with no additional computational expense.

The variable

$$\tilde{x}_{i,\,\rm mid} \equiv \frac{\tilde{x}_{i-1} + \tilde{x}_{i+1}}{2 + \tilde{\omega}^2} \tag{70}$$

minimizes the part of the action that depends on $\tilde{x}_i$. Thus, $\tilde{x}_{i,\,\rm trial}$ lies "on the other side" of this minimum for fixed x_{i-1} and x_{i+1}. The update $\tilde{x}_i \to \tilde{x}_{i,\,\rm trial}$ is microcanonical, meaning that the action is constant under this change and no Metropolis accept/reject step is needed. The disadvantage of Eq. (69) is that the procedure is not applicable to actions for which $\tilde{x}_{i,\,\rm mid}$ cannot be found exactly, such as an anharmonic oscillator. Because the ratio of the kinetic to potential energy increases strongly in the continuum limit, we propose a trial change that preserves the kinetic part of the action:

$$\tilde{x}_{i,\,\rm kin} = (\tilde{x}_{i-1} + \tilde{x}_{i+1}) - \tilde{x}_i, \tag{71}$$

followed by a standard accept/reject procedure.[43] Equation (71) can be used for any potential term. The only change in the pseudocode given in Ref. 27 is the value of $x_{\rm new}$. Usually all sweeps are Metropolis sweeps; for the over-relaxation routine used to create Fig. 8, four in five Metropolis sweeps were exchanged for over-relaxed sweeps. The error bars are visibly smaller with over-relaxation. Because the CPU time needed for an over-relaxed sweep is comparable to that required for an ordinary Metropolis sweep, the accuracy of the measurements is significantly improved at constant computational cost.

VII. ADVANCED TOPICS

A. The ground state wave function

Section III discussed how to determine any moment $\langle x^n \rangle$ of the ground state wave function of the harmonic oscillator. Here we point out that information about the complete ground state wave function (more precisely $|\psi_0(x)|^2$, because only the modulus squared of a wave amplitude can be measured in quantum mechanics) is directly accessible from the path integral. The squared modulus $|\psi_0(x)|^2$ is approximated by a histogram of the data for $x(\tau)$, which are available from the simulation. A histogram (over all configurations and all time slices within each configuration) is shown in Fig. 9. The data and full curve deviate from the dashed curve, which represents the continuum result (39).

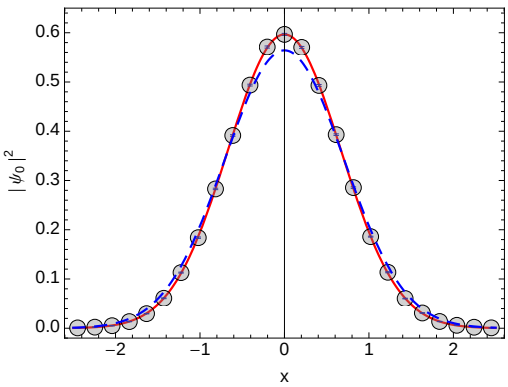

Figure 9. (color online) Modulus squared of the ground state wave function for the harmonic oscillator. Measurements were based on $N = 10^4$ paths with $\tilde{m} = \tilde{\omega} = 1$, $N_\tau = 120$ and bin size $\Delta x = 0.1$. The error bars are smaller than the symbol size. The solid curve (red) represents the expression for $|\psi_0|^2$ on a discrete lattice for this choice of parameters, and the dashed curve (blue) is its continuum counterpart.

Here the correction factor $(1 + \frac{1}{4}\tilde{\omega}^2)^{1/2}$ is apparent. This factor, seen in Eqs. (42) and (44) and in Eq. (72), accounts for the discreteness of Euclidean time. In the naive continuum limit $\tilde{\omega} \to 0$, and the correction factor approaches one.

B. Measuring energy differences

The energy spectrum of the harmonic oscillator is $E_n = (n + \frac{1}{2})\omega$ for $n = 0, 1, 2, \ldots$ ($\hbar = 1$). It is possible to verify this energy spectrum by lattice simulations, up to the zero-point energy. In other words, we can measure the energy differences $\tilde{E}_m - \tilde{E}_n$ for any m and n, but not the energies themselves.

To understand how, we revisit the two-point correlator $G(\Delta\tau)$ in Eq. (47). It is instructive to rewrite it in terms of the ladder operators $\hat{a}$ and $\hat{a}^\dagger$ by substituting

$$\hat{x} = i\sqrt{\frac{1}{2\tilde{m}\tilde{\omega}(1 + \frac{1}{4}\tilde{\omega}^2)^{1/2}}}(\hat{a} - \hat{a}^\dagger), \tag{72}$$

which follows from the definition of $\hat{a}$ and $\hat{a}^\dagger$ in terms of $\hat{x}$ and $\hat{p}$ (see Sec. VII and Ref. 27 for details and the convention used). The operators $\hat{a}$ and $\hat{a}^\dagger$ act on the normalized eigenstates $|n\rangle$ of the number operator $\hat{a}^\dagger\hat{a}$, $\hat{a}|0\rangle = 0$ and $\left(\hat{a}^\dagger\right)^n |0\rangle = \sqrt{n!}|n\rangle$. With this substitution the two-point correlator takes the form

$$G(\Delta\tau) = \frac{1}{2\tilde{m}\tilde{\omega}(1 + \frac{1}{4}\tilde{\omega}^2)^{1/2}}\langle 0|\hat{a}(\tau)\hat{a}^\dagger(\tau + \Delta\tau)|0\rangle, \tag{73}$$

because each one of the remaining three terms in the two-point correlator $G(\Delta\tau)$ with the substitution given in Eq. (72) has an $\hat{a}^\dagger(\tau)$ operator acting on $\langle 0|$ to the left, and/or an $\hat{a}(\tau + \Delta\tau)$ acing on $|0\rangle$ to the right. After inserting a complete set of states $I = \sum |n\rangle\langle n|$ between the two ladder operators, we find that only the state $|1\rangle\langle 1|$ contributes. Finally, we write $\hat{A}(t) = \exp(\hat{H}t)\hat{A}\exp(-\hat{H}t)$ to find

$$\tilde{m}G(\Delta\tau) = \frac{1}{2\tilde{m}\tilde{\omega}(1 + \frac{1}{4}\tilde{\omega}^2)^{1/2}}\langle 0|\hat{a}(\tau)|1\rangle\langle 1|e^{-E_1\Delta\tau}\hat{a}^\dagger(\tau)e^{E_0\Delta\tau}|0\rangle \tag{74}$$

$$= \frac{1}{2\tilde{m}\tilde{\omega}(1 + \frac{1}{4}\tilde{\omega}^2)^{1/2}}e^{-(E_1 - E_0)\Delta\tau}\langle 0|\hat{a}(\tau)|1\rangle\langle 1|\hat{a}^\dagger(\tau)|0\rangle \tag{75}$$

$$= e^{-(E_1 - E_0)\Delta\tau}G(0), \tag{76}$$

which implies that the slope of $\log(G(\Delta\tau))$ *versus* $\Delta\tau$ measures the energy difference $E_1 - E_0$ (see Fig. 3).

To determine the difference $E_n - E_0$ for arbitrary n, we need to consider higher order multi-point functions. Equation (72) implies that any function $\langle 0|\hat{x}(\tau_1)\dots\hat{x}(\tau_{2n+1})|0\rangle$ with an odd number of $\hat{x}$ is zero. Accordingly, the first viable option is the 4-point function $\langle 0|\hat{x}(\tau_1)\hat{x}(\tau_2)\hat{x}(\tau_3)\hat{x}(\tau_4)|0\rangle$. For simplicity, we consider the restricted 4-point function with $\tau_1 = \tau_2$ and $\tau_3 = \tau_4$:

$$F(\Delta\tau) = \langle 0|\hat{x}^2(\tau)\hat{x}^2(\tau+\Delta\tau)|0\rangle. \tag{77}$$

Squaring Eq. (72) yields

$$\hat{x}^2 = -\frac{1}{2\tilde{m}\tilde{\omega}(1+\frac{1}{4}\tilde{\omega}^2)^{1/2}}(\hat{a}^2 - \hat{a}\hat{a}^\dagger - \hat{a}^\dagger\hat{a} + \hat{a}^{\dagger 2}) \tag{78}$$

and substituting this expression into Eq. (77) gives

$$\begin{aligned} F(\Delta\tau) = {} & \frac{1}{4\tilde{m}^2\tilde{\omega}^2(1+\frac{1}{4}\tilde{\omega}^2)} \\ & \times \langle 0|[\hat{a}^2 - \hat{a}\hat{a}^\dagger](\tau)[-\hat{a}\hat{a}^\dagger + \hat{a}^{\dagger 2}](\tau+\Delta\tau)|0\rangle, \end{aligned} \tag{79}$$

because all other terms are zero. Only two of the four terms survive, and we find

$$\begin{aligned} F(\Delta\tau) = {} & \frac{1}{4\tilde{m}^2\tilde{\omega}^2(1+\frac{1}{4}\tilde{\omega}^2)} \\ & \times [\langle 0|\hat{a}^2(\tau)\hat{a}^{\dagger 2}(\tau+\Delta\tau)|0\rangle \\ & + \langle 0|\hat{a}\hat{a}^\dagger(\tau)\hat{a}\hat{a}^\dagger(\tau+\Delta\tau)|0\rangle]. \end{aligned} \tag{80}$$

As before, we insert a complete set of states $|n\rangle\langle n|$ between τ and $\tau+\Delta\tau$. It follows that the first term receives only a contribution from the $n = 2$ state, and the second term only from the $n = 0$ ground state. We again use the Heisenberg picture and find that the first term carries a time-dependence $\exp[-(E_2 - E_0)\Delta\tau]$, and the second term carries no time-dependence. We thus define the connected correlator

$$F_{\rm conn}(\Delta\tau) = \langle 0|\hat{x}^2(\tau)\hat{x}^2(\tau+\Delta\tau)|0\rangle - |\langle 0|\hat{x}^2(\tau)|0\rangle|^2, \tag{81}$$

which has the time-dependence

$$F_{\rm conn}(\Delta\tau) = e^{-(E_2-E_0)\Delta\tau}F_{\rm conn}(0). \tag{82}$$

We see that this correlator decreases twice as fast (for the same parameters) as the 2-point function shown in Fig. 3. It is possible to generalize these considerations and to construct suitable combinations of $2n$-point functions that decrease as $\exp(-(E_n - E_0)\Delta\tau)$. Hence, we can determine the energy spectrum of the harmonic oscillator through simulations. However, the higher multi-point functions will be noisier than the 2-point function.

C. The anharmonic oscillator

The anharmonic oscillator is interesting, not just as an application of the Monte Carlo Markov chain method developed here, but as a system where exact solutions are not available. The Rayleigh–Schrödinger perturbation is known[44] to diverge, which has led to the development of approximate methods to estimate and place bounds on the energy levels of this system.[45]

The action of the quantum anharmonic oscillator is

$$\tilde{S} = \sum_{i=1}^{N_\tau} \frac{1}{2}\tilde{m}(\tilde{x}_{i+1} - \tilde{x}_i)^2 + \frac{1}{2}\tilde{m}\tilde{\omega}^2\tilde{x}_i^2 + \frac{1}{4}\lambda\tilde{x}_i^4. \tag{83}$$

All the techniques we have discussed can be applied to the anharmonic oscillator. We have chosen the over-relaxation update $x_i \to \tilde{x}_i''$ in Eq. (71) such that it carries over to actions with an anharmonic term. That being said, the anharmonic oscillator is a much harder problem than the harmonic oscillator. We hope to come back to it in a future publication.

VIII. SUGGESTED PROBLEMS

As an exercise for the reader, we suggest developing the machinery in this paper step by step.

(a) Write code for a Metropolis sweep in a programming language of choice. Study the pseudocode in Ref. 27 to get started.

(b) Take $\tilde{m} = \tilde{\omega} = 1$ and $N_\tau = 120$, which is our coarsest lattice. To thermalize the path, perform 100 Metropolis sweeps before saving it. Carry out 12 Metropolis sweeps (written in a do/for loop) before saving the next path. Repeat this procedure 10,000 times, always with $N_{\rm sep} = 12$ (i.e., running the Metropolis sweep 12 times before saving the next path). This makes for a grand total of 120,000 Metropolis sweeps in order to generate 10,000 paths. Save each path externally. As a sanity check, make a probability density diagram of all simulated positions (i.e., of $120 \times 10,000$ numbers) with bin size $\Delta x = 0.1$. A plot of these measurements should reproduce Fig. 9 (data).

(c) Measure a few observables based on the stored 10,000 configurations. We suggest starting with $\langle \hat{x} \rangle$ and $\langle \hat{x}^2 \rangle$; the former expectation value should be consistent with zero, the latter is given in Eqs. (42) and (43). Modify your program so that it determines, after $N_{\rm sep}$ Metropolis sweeps, the average $\langle \hat{x} \rangle$ and the average $\langle \hat{x}^2 \rangle$ of that path, and writes it to disk, rather than the path itself. Extend your program so that it determines, from the same configurations, the intra-path averaged quantities $\langle \hat{x}^3 \rangle$ and $\langle \hat{x}^4 \rangle$, and writes these two numbers to disk, too.

(d) Repeat step (c) on progressively finer lattices, and combine the 10,000 numbers of $\langle \hat{x}^n \rangle$ into an ensemble average $\langle \hat{x}^n \rangle$ for every set of parameters. The formulas for $\langle \hat{x}^2 \rangle$ and $\langle \hat{x}^4 \rangle$ as a function of the lattice spacing are given in Eqs. (42)-(44). Our values for $\tilde{m} = \tilde{\omega}$ and N_τ are stated in Table II.

(e) For each observable and choice of parameters, obtain an estimate of the asymptotic autocorrelation time A_O. While A_O can be determined through a fit, as in Eq. (63), a good way to get a feel for it is by using the matlab routine "UWerr.m"[36], which is freely available online. For a given dataset, it calculates an estimate of the asymptotic autocorrelation time and the associated statistical error. Note that $N_{\rm sep}$ must increase for progressively finer lattices. Is $N_{\rm sep} = N_\tau/10$ sufficient to keep the autocorrelation time to a reasonable magnitude (i.e., small enough to have a sufficient number of independent "bins" on which to do a jackknife analysis)?

(f) Redo the calculations in (e) with a jackknife analysis (pseudocode is provided in Ref. 27). The bin width should always be greater than the autocorrelation time of the dataset under consideration.

(g) Plot the results under (e) and (f) to visualize the reliability of the jackknife errors.

IX. DISCUSSION

How can we be confident in simulation data in the absence of predictions? The theory and data we have presented highlight two effects as the continuum limit is approached. The lattice-dependent theoretical answer converges to a limiting value; in the same limit, the noise-to-signal ratio blows up. Two main themes are of interest for the computation of path integrals on a lattice: the limiting answer and the trade-off between accuracy and required computer time. Because computer time requirements depend strongly on the autocorrelations present in the Markov chain, we should start with a systematic determination of the autocorrelation time τ_O as a function of the effective lattice spacing, keeping the product $N_\tau \Delta\tau$ fixed. For reliable estimates it is best to start with coarse lattices and high statistics and to determine the behavior of the autocorrelation function for progressively finer lattices. Once τ_O is established as a function of the effective lattice spacing, we can calculate the desired observable with an appropriate number of intermediate paths. The error bars should be determined with the jackknife procedure using an appropriate bin width.

The answer in the continuum limit should reveal itself through a stable fit to the data obtained at the finest few lattice spacings, similar to our Fig. 8. If this stable fit is possible, we can have confidence in the continuum answer, even if there is no analytic answer available for comparison.

ACKNOWLEDGMENTS

MJEW was supported through a Janet Watson scholarship from the Department of Earth Science and Engineering and a studentship in the Centre for Doctoral Training on Theory and Simulation of

Materials funded by the EPSRC (EP/L015579/1), both at Imperial College London. SD acknowledges partial support by DFG through SFB-TR-55.

---

[1] B. L. van der Waerden, *Sources of Quantum Mechanics* (Dover, New York, 1968).
[2] S. G. Brush, "Resource letter HP-1: History of physics," Am. J. Phys. **55**, 683–690 (1987).
[3] P. A. M. Dirac, "The Lagrangian in quantum mechanics," Physik. Z. Sowjetunion **3**, 64–72 (1933).
[4] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, London, 1947).
[5] R. P. Feynman, "Space-time approach to non-relativistic quantum mechanics," Rev. Mod. Phys. **20**, 367–387 (1948).
[6] S. G. Brush, "Functional integrals and statistical physics," Rev. Mod. Phys. **33**, 79–92 (1961).
[7] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw–Hill, New York, 1965).
[8] K. Huang, *Quantum Field Theory: From Operators to Path Integrals* 2nd ed. (John Wiley & Sons, Weinheim, 2010).
[9] F. W. Weigel, "Path integral methods in statistical mechanics," Phys. Rep. **16**, 57–114 (1975).
[10] H. Wio, *Path Integrals for Stochastic Processes: An Introduction* (World Scientific, Singapore, 2013).
[11] N. Wiener, "The average value of a functional," Proc. London Math. Soc. **22**, 454–467 (1921).
[12] M. Kac, "Wiener and integration in function spaces," Bull. Amer. Math. Soc. **72**,, 52–68 (1966).
[13] L. D. Landau and E. M. Lifschitz, *Classical Mechanics* (Pergamon Press, Oxford, 1960).
[14] R. P. Feynman, *The Feynman Lectures in Physics*, Vol. 3 (Addison–Wesley, Reading, MA, 1965), pp. 3-1–3-7.
[15] R. Sawant, J. Samuel, A. Sinha, S. Sinha, and U. Sinha, "Nonclassical paths in quantum interference experiments," Phys. Rev. Lett. **113**, 120406-1–5 (2014).
[16] D. J. Amit, *Field Theory, The Renormalization Group, and Critical Phenomena* (World Scientific, Singapore, 1984).
[17] J. Zinn-Justin, *Path Integrals in Quantum Mechanics* (Oxford University Press, Oxford, 2006).
[18] M. Creutz and B. Freedman, "A statistical approach to auantum mechanics," Ann. Phys. **132**, 427–462 (1981).
[19] C. Morningstar, "The Monte Carlo method in quantum field theory," arXiv:hep-lat/0702020.
[20] The application of Monte Carlo methods to quantum field theory is developed at length in Refs. 21–26.
[21] M. Creutz, *Quarks, Gluons and Lattices* (Cambridge University Press, 1983).
[22] C. Gattringer and C. B. Lang, *Quantum Chromodynamics on the Lattice: An Introductory Presentation* (Springer, 2010).
[23] T. DeGrand and C. DeTar, *Lattice Methods for Quantum Chromodynamics* (World Scientific, 2006).
[24] I. Montvay and G. Münster, *Quantum Fields on a Lattice* (Cambridge University Press, 1994).
[25] H. J. Rothe, *Lattice Gauge Theories: An Introduction*, 4th ed. (World Scientific, 2012).
[26] J. Smit, *Introduction to Quantum Fields on a Lattice* (Cambridge University Press, Cambridge, 2002).
[27] Reference for Supplementary Information.
[28] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, "Equation of state calculations by fast computing machines," J. Chem. Phys. **21**, 1087–1092 (1953).
[29] W. A. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," Biometrica **57**, 97–190 (1970).
[30] Different observables can approach their equilibrium values at different rates.
[31] The local logarithmic slope can also be determined less locally, as long as one stays within the range for which Eq. (49) holds.
[32] If the error on the error had started to increase at some $\Delta t^* < 40$, Δt^* would have defined the cutoff for this parameter set.
[33] M. H. Quenouille, "Problems in plane sampling," Ann. Math. Statist. **20**, 355–375 (1949).
[34] J. Shao and D. Tu, *The Jackknife and Bootstrap* (Springer, 1996).
[35] Initially the jackknife variance will increase monotonically with bin size. After a while, it will reach a plateau *versus* the bin size, and eventually the jackknife variance will fluctuate; that is, vary non-monotonically as a function of the bin size. We should choose the bin size in the plateau region.
[36] U. Wolff, "Monte Carlo errors with less errors," Comput. Phys. Commun. **156**, 143–153 (2004).
[37] The Matlab function UWerr.m, Version 6, described in Ref. 36, was used to determine the statistical errors of the asymptotic autocorrelation time.
[38] If the asymptotic autocorrelation time is fully determined by the term e^{-t/τ_0}, there is no fundamental distinction between the asymptotic and the integrated autocorrelation time, because $\tau_{O,\mathrm{int}} \simeq \frac{1}{2} + \int_1^\infty e^{-\tau/\tau_0} \simeq \int_0^\infty e^{-t\tau_0} = \tau_0 = \tau_{O,\mathrm{asy}}$.
[39] W. Janke and T. Sauer, "Path integral Monte Carlo using multigrid techniques," Chem. Phys. Lett. **201**,

499–505 (1993).
[40] A. Goodman and D. Sokal, "Multigrid Monte Carlo method for lattice field theories," Phys. Rev. Lett. **56**, 1015–1018 (1986).
[41] M. Creutz,"Overrelaxation and Monte Carlo simulation," Phys. Rev. D **36**, 515–519 (1987).
[42] F. R. Brown and T. J. Woch, "Overrelaxed heat-bath and Metropolis algorithms for accelerating pure gauge Monte Carlo simulations," Phys. Rev. Lett. **58**, 2394–2396 (1987).
[43] In the continuum limit, the term $1+\tilde{\omega}^2/2$ in Eq. (69) approaches 1. Thus, in this limit, the trial change $\tilde{x}_i''$ approximates the choice of $\tilde{x}_i'$ by Creutz,[41] and Brown and Woch.[42] We note that the acceptance rate for $\tilde{x}_i''$ appraches one in the continuum limit, where the kinetic term dominates.
[44] C. M. Bender and T. T. Wu, "Anharmonic oscillator," Phys. Rev. **184**, 1231–1260 (1969).
[45] F. T. Hioe, and E. W. Montroll, "Quantum theory of anharmonic oscillators. I. Energy levels of oscillators with positive quartic anharmonicity," J. Math. Phys. **16**, 1945–1955 (1975).