Physics 514 – Quantum Monte Carlo

Emanuel Gull

December 1, 2016

Draft – Please report mistakes!

Draft – Please report mistakes!

Chapter 1

Quantum Monte Carlo (QMC)

1.1 Introduction

The term 'quantum Monte Carlo' has been applied to a range of related methods to simulate quantum many-body systems. So far, our efforts have been focused on obtaining a wave function $\psi(r_1, \ldots, r_N)$ to describe a system. Finding the exact wave function is not practically feasible unless the number of basis states is very small. Therefore, DMRG (and other variational approaches) proceed by carefully choosing a small but accurate basis set and finding the best wave function in that basis: they are *spectral* approaches. Quantum Monte Carlo methods (of the type discussed here) instead obtain a *noisy* but *unbiased* estimator of the wave function in the original, large basis. The two main variants of this type are the 'Diffusion' Monte Carlo (or Green's function Monte Carlo) and the Path integral Monte Carlo. Diffusion Monte Carlo has been developed into a precise tool to compute the ground state properties of systems with up

to about a thousand particles. Path integral Monte Carlo is technically similar, but gives finite-temperature properties. The path integral formalism carries over to the relativistic case, and forms an essential basis for the numerical study of quantum field theories.

1.2   Notation and general idea

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

$|\psi(t)\rangle$ solves the time-dependent Schrödinger equation $i\hbar\partial_t|\psi(t)\rangle = H|\psi(t)\rangle$

$$|\psi(t)\rangle = \exp(-\frac{i}{\hbar}Ht)|\psi(0)\rangle \tag{1.1}$$

We now perform a 'Wick rotation': an operation that rotates 'real' time t to 'imaginary time' τ. We define an imaginary time $\tau = it$, so that

$$|\psi(\tau)\rangle = \exp(\frac{-\tau}{\hbar}H)|\psi(0)\rangle \tag{1.2}$$

In the eigenbasis of H:

$$H|\psi_k\rangle = E_k|\psi_k\rangle, \qquad k = 0, 1, \ldots; \quad E_0 \leq E_1 \leq \ldots \tag{1.3}$$

the state $|\psi(\tau)\rangle$ is

$$|\psi(\tau)\rangle = \sum_k \exp(-\frac{\tau}{\hbar}E_k)\langle\psi_k|\psi(0)\rangle|\psi_k\rangle \tag{1.4}$$

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

In both approaches, the time evolution is preformed as a sum over histories, each history having a different probability. Note that the same set of equations is used for financial pricing, so this is useful. . .

1.3 Expression for the Green's function

We define the Green's function as

$$G(x, t; x_0, 0) = \langle x | \exp(-\frac{i}{\hbar} H t) | x_0 \rangle \qquad (1.5)$$

so that

$$\psi(x, t) = \int dx_0 G(x, t; x_0, 0) \psi(x_0, 0). \qquad (1.6)$$

G is a Green's function, in the sense that it is a solution for the Schrödinger equation for $\psi(x, 0) = \delta(x - x_0)$.

$$|\psi(t)\rangle = \int dx \psi(x, t) | x \rangle = \int dx | x \rangle \int dx_0 G(x, t; x_0, 0) \psi(x_0, 0) \qquad (1.7)$$

$$= \int dx \int dx_0 | x \rangle \langle x | \exp(-\frac{i}{\hbar} H t) | x_0 \rangle \psi(x_0, 0) \qquad (1.8)$$

$$= \exp(-\frac{i}{\hbar} H t) | \psi(0) \rangle \qquad (1.9)$$

An important property of the Green's function is

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

(1.10)

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

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

(1.11)

We proceed by dividing t into N subintervals $\delta t = t/N$. (Later we will take the limit $N \to \infty$):

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

(1.12)

with $x_N \equiv x$. The task is then to evaluate an elementary matrix element $\langle x_k| \exp(-\frac{i\delta t}{\hbar}H)|x_{k-1}\rangle$. Here we encounter a problem: the Hamiltonian $H = \frac{p^2}{2m} + V(x)$ is an operator made of two pieces which do not commute:

- The potential energy operator $V(x)$ is diagonal in position space $|x\rangle$

- The kinetic energy operator $\frac{p^2}{2m}$ is diagonal in momentum space $|p\rangle$.

Baker-Campbell-Hausdorff states that

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]+\cdots}$$

(1.13)

We can neglect these higher order terms, which are of order $O(\delta t^2)$, since we are interested in $\delta t \to 0$, and

write

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

$$(1.14)$$

where we chose a symmetrized version of the interaction. We then insert **1** in the momentum basis:

$$\int \frac{dp}{2\pi}|p\rangle\langle p| = \mathbf{1}, \qquad (1.15)$$

in which the kinetic part is diagonal. $\langle x_k|\exp(-\frac{i}{\hbar}\delta t H)|x_{k-1}\rangle$ becomes

$$\exp\left[-\frac{i}{\hbar}\delta t\frac{V(x_k)+V(x_{k-1})}{2}\right]\int\frac{dp_k}{2\pi}\frac{dp_{k-1}}{2\pi} \quad (1.16)$$

$$\times\langle x_k|p_k\rangle\langle p_k|\exp\left(-\frac{i}{\hbar}\delta t\frac{p^2}{2m}\right)|p_{k-1}\rangle\langle p_{k-1}|x_{k-1}\rangle$$

The change of basis from position to momentum basis is encoded in the matrix elements $\langle x|p\rangle = \exp(\frac{ipx}{\hbar})$, thus (without the local part)

$$= \cdots\exp\left(\frac{i}{\hbar}p_k x_k\right)\delta(p_k - p_{k-1})\exp\left(-\frac{i}{\hbar}\delta t\frac{p_k^2}{2m}\right)\exp\left(-\frac{i}{\hbar}\delta t p_{k-1}x_{k-1}\right)$$

$$(1.17)$$

$$= \cdots\int\frac{dp_k}{2\pi}\exp\left(\frac{i}{\hbar}p_k(x_k - x_{k-1})\right)\exp\left(-\frac{i}{\hbar}\delta t\frac{p_k^2}{2m}\right)$$

$$(1.18)$$

This is a Gaussian integral,[1] which we can evaluate by completing the square. Putting everything together, one

---

[1] Use $\int e^{-\frac{A}{2}p^2 + Bp}dp = \sqrt{\frac{2\pi}{A}}e^{\frac{1}{2}\frac{B^2}{A}}$ with $A = \frac{i}{\hbar m}$ and $B = \frac{i}{\hbar}(x_k - x_{k-1})$.

obtains

$$\langle x_k| \exp\left(-i\frac{\delta t}{\hbar}H\right)|x_{k-1}\rangle \tag{1.19}$$

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

Finally we perform the Wick's rotation to imaginary time $\tau = it$

$$\langle x_k| \exp\left(-\frac{\delta\tau}{\hbar}H\right)|x_{k-1}\rangle \tag{1.20}$$

$$\approx C\exp\left[-\frac{\delta\tau}{\hbar}\left(\frac{m}{2}\left(\frac{x_k-x_{k-1}}{\delta\tau}\right)^2+\frac{V(x_k)+V(x_{k+1})}{2}\right)\right]$$

This exponent is just what one would expect from a Hamiltonian: $H = \frac{1}{2}mv^2 + V$. The constant C is independent of x_k, x_{k+1} and drops out of all observables. Eq. 1.20 is at the core of all quantum Monte Carlo wave function simulation algorithms and becomes exact as $\delta\tau \to 0$.

1.4   Diffusion Monte Carlo

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

For this to be possible, the ground state wave-function must be real and positive everywhere (up to a global phase, which is arbitrary). For many unfrustrated bosonic

systems such a choice is possible. Excited state wave functions have nodes; fermionic ground-state wave functions are totally antisymmetrized objects and therefore also have nodes. We will discuss these difficulties later.

The simplest form of DMC starts with a set of walkers, labeled by j, at time zero. Walkers have a position x_j and a weight w_j. A convenient starting point is $\psi(x, \tau = 0) = \delta(x - x_0)$, so that all the walkers are in the same state: $x_0^j = x_0, w_0^j = w_0$. The time evolution of the wave-function

$$\psi(x, \tau) = G(x, \tau; x_0, 0) \qquad (1.21)$$

where G is a product of elementary factors of Eq. 1.20. Each elementary factor is factorized into its kinetic part and its potential part, which are applied in succession to each walker. Given a walker j and position x_{k-1} with weight w_{k-1}, the new position x_k and the new weight w_k are obtained as follows:

- For x_k the kinetic part gives a Gaussian distribution centered around x_{k-1}. This distribution can be sampled stochastically, by drawing $(x_k - x_{k1})$ from a Gaussian distribution with variance $\hbar\delta\tau/m$. This step corresponds to a *diffusion* around x_{k-1} and gives its name to the algorithm.

- The potential part modifies the weight:

$$w_k = w_{k-1} \exp\left(-\frac{\delta\tau}{\hbar}\frac{V(x_k) - V(x_{k-1})}{2}\right) \qquad (1.22)$$

Both of these factors together allow a stochastic repre-

sentation of $\psi(x, k\delta\tau)$:

$$\psi(x, k\delta\tau) = \langle w_k \delta(x_k - x)\rangle, \qquad (1.23)$$

where $\langle \cdots \rangle$ denotes a Monte Carlo average over the m walkers.

One problem with this algorithm is that the weights w^j will vary considerably from one walker to the next, so that the contribution of many walkers to the average in Eq. 1.23 is negligible, the variance of the estimator large, and the numerical effort to simulate them will be wasted: Walkers that 'walk off' into regions with high potential energy will be exponentially suppressed. Ideally, all walkers should maintain about an equal weight for best efficiency. This can be achieved by 'cloning' the 'important' walkers and by 'killing off' those that are negligible, again stochastically:

Compute the nominal weight

$$w^* = \frac{1}{m}\sum_j^m w^j. \qquad (1.24)$$

Replace each walker j by a number of clones (all with weight w^*) equal to int$\left(\frac{w^j}{w^*} + r\right)$, where r is a random number distributed uniformly in $[0, 1[$: The average of this expression is $\frac{w^j}{w}$, so that each walker is replaced, on average, by its appropriate number of equal-weight clones. As a consequence of this, the average number m of walkers will fluctuate proportional to $\sqrt{m}$.

With this algorithm one can already obtain some interesting results. Two technical modifications are customary:

- Limit the maximum number of clones: The number
  of clones increases when the walkers reach a potential
  well. Since DMC is often used for Coulomb systems,
  where the potential is unbounded, this will avoid di-
  vergencies. Note that getting many new walkers im-
  plies having a large variation in potential energy over
  a small step and is therefore a sign for $\delta\tau$ being too
  large.

- As formulated above, the nominal weight w^* varies
  as $\exp(-\frac{E_0}{\hbar}\tau)$ for large τ. To avoid overflow or under-
  flow, one introduces a trial energy E_T and multiplies
  all weights w^j by $\exp(\frac{E_T}{\hbar}\delta\tau)$ after each step. Stabil-
  ity of the weights is achieved when $E_T = E_0$, whcih
  gives a simple means to compute the ground state
  energy.

1.4.1 Importance sampling

Fluctuations in the estimated wave-function mostly come
from the modification of the weight by the potential en-
ergy term. The variation of the weight of a single walker
signals a waste of computer time: a weight which tends
to zero indicates that the walker approaches a forbid-
den region (e.g. the edge of the box for a particle in a
box); a weight which diverges occurs at a potential sin-
gularity (e.g. at the location of an opposite charge in
the Coulomb potential). In both cases, it would be ad-
vantageous to incorporate prior knowledge of the wave-
function in the diffusion step, so that walkers would be
discouraged or encouraged to enter the respective re-
gions, and the weights would vary less.

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

$$-\hbar\frac{d}{d\tau}\psi = \left(-\frac{\hbar^2}{2m}\nabla^2 + V\right)\psi \qquad (1.25)$$

we obtain the equation $\Phi(x, \tau)$ by multiplication with ψ_T:

$$-\hbar\frac{d}{d\tau}\Phi = \psi_T\left(-\frac{\hbar^2}{2m}\nabla^2 + V\right)\psi \qquad (1.26)$$

$$= \left(-\frac{\hbar^2}{2m}\nabla^2 + V\right)\Phi + \frac{\hbar^2}{2m}\left(2\nabla\psi_T \cdot \nabla\psi + \psi\nabla^2\psi_T\right)$$
$$(1.27)$$

$$= \left[-\frac{\hbar^2}{2m}\nabla^2 + \frac{\hbar^2}{m}\left(\frac{\nabla\psi_T}{\psi_T}\right)\cdot\nabla + \left(V - \frac{\hbar^2}{2m}\frac{\nabla^2\psi_T}{\psi_T}\right)\right]\Phi$$
$$(1.28)$$

The first two terms define a new diffusion with an additional 'drift' term that can just be added to the diffusion process. The last term describes a modified potential, to be used in the update of the weights.

The effect of a good trial wave function on the weights of the walkers is quite dramatic: assume that we know the exact ground state wave function ψ_0 and set $\psi_T = \psi_0$. Then the last term in Eq. 1.28 is $(V - (V - E_0))\Phi$, with E_0 the ground state energy. The new potential is therefore just the ground state energy, which is a constant independent of x. The weights of all walkers are multiplied with the same constant, and no cloning or killing of walker occurs.

1.4.2 Fermionic systems

A major application of Diffusion Monte Carlo is the calculation of the ground state energy of molecules or crystals, given the atomic/nuclear position. However, the electrons are indistinguishable fermions, and the wave function should be anti-symmetric under interchange of any two electrons. Thus, if it is positive when electrons 1 and 2 are at positions $(\mathbf{r}_1, \mathbf{r}_2)$, it is negative when they are at $(\mathbf{r}_2, \mathbf{r}_1)$.

This presents a problem to the algorithm above: our wave function is represented by a density of walkers, and densities cannot become negative: in its present formulation the method only works for many-body wave functions which have the same sign everywhere, *e.g.* (unfrustrated) bosons.

Configuration space has 'nodal surfaces' separating the regions where the wave functions are positive from regions where they are negative. If the location of these nodal surfaces was known, one could perform distinct simulations in each region of definite sign (as many as there are disconnected regions): If ϕ_0 is positive, then apply the algorithm above with a potential barrier preventing the walkers to cross the nodal surface. If ψ_0 is negative, apply the same algorithm with the substitution $\psi \leftarrow -\psi$. Usually, the location of the nodal surfaces is not known, and some ansatz is made. This strategy is called the *fixed node* approximation.

Fixing the nodal surface away from its ground state location can only increase energy, so that the fixed node approximation gives a variational upper bound to the

ground state energy.

One can try to relax the constraint of fixed nodal surfaces and move the nodal surfaces into the directions that most lower the energy. This will systematically lower the energy towards the accurate ground state energy. The cost of performing this relaxation is exponential in the number of particles – one of the many manifestations of the fermionic sign problem.

1.5   Path integral Monte Carlo

Diffusion Monte Carlo is a technique used to compute the ground state of many-particle quantum systems. A small modification allows to compute finite temperature averages of quantum systems.

We start by writing the propagation (time evolution) between two states $|\alpha\rangle$ and $|\beta\rangle$ separated by a time τ:

$$\langle\alpha| \exp\left(-\frac{\tau}{\hbar}H\right) |\beta\rangle = \int dx_1 dx_2 \ldots dx_{N-1} \langle x_k| \exp(-\frac{\delta\tau}{\hbar}H)|x_{k-1}\rangle$$

$$(1.29)$$

$$\sim C^N \int dx_1 \cdots dx_{N-1} \prod_{k=1}^{N} \exp\left[-\frac{\delta\tau}{\hbar}\left(\frac{1}{2}m\left(\frac{x_k - x_{k-1}}{\delta\tau}\right)^2 + \frac{V(x_k) + V(x_{k-1})}{2}\right.\right.$$

$$(1.30)$$

$$= C^N \int dx_1 \cdots dx_{N-1} \exp\left[-\frac{\delta\tau}{\hbar}\left(\frac{1}{2}m\sum_k\left(\frac{x_k - x_{k-1}}{\delta\tau}\right)^2 + \sum_k\frac{V(x_k) + V}{2}\right.\right.$$

$$(1.31)$$

This imaginary-time-discretized propagation from state $|\alpha\rangle$ to state $|\beta\rangle$ is a high-dimensional integral which we

can solve by Monte Carlo: we perform a stochastic random walk over all intermediate time steps. To converge to the desired equilibrium distribution:

- choose an arbitrary starting path with x_k distributed, *e.g.*, randomly.

- propose to move the path at x_k to x'_k. The probability of the new state x_k is, up to an irrelevant proportionality constant,

$$p(x_k) \propto \exp\left[-\frac{\delta\tau}{\hbar}\left(\frac{1}{2}m\left(\left(\frac{x_{k+1}-x_k}{\delta\tau}\right)^2 + \left(\frac{x_k - x_{k-1}}{\delta\tau}\right)^2\right) + V(x_k)\right)\right],$$
(1.32)

and we can implement a Metropolis random walk: draw the new $x'_k = x_k + \delta x$, where δx is drawn from an even distribution, e.g. uniform in $[-\Delta, \Delta]$ with a constant Δ

- accept the transition to the new path with probability $p(x_k \to x'_k) = \min(1, \frac{p(x'_k)}{p(x_k)})$.

Note that the the step size, governed by Δ, is the result of a trade-off: a large step causes large variations in Eq. 1.32, and therefore a small average acceptance probability; a small step causes a slow evolution in Monte Carlo time, resulting in a large autocorrelation time.

1.5.1   Finite Temperature

The canonical partition function in quantum statistical mechanics at inverse temperature β is defined as

$$Z = \text{Tr} \exp(-\beta H)$$
(1.33)

We can expand Z in the eigenbasis of H, to obtain an expression that resembles the one from classical statistical mechanics:

$$Z = \operatorname{Tr} \exp(-\beta H) = \sum_k \langle \psi_k | \exp(-\beta H) | \psi_k \rangle = \sum_k \exp(-\beta E_k)$$

(1.34)

We can, similarly, write Z in a position basis:

$$Z = \sum_{|x\rangle} \langle x | \exp(-\beta H) | x \rangle.$$

(1.35)

This expression looks like the one for the imaginary time propagator, with the time τ replaced by the inverse temperature β, the first state identified with the last state, and a summation over all initial states.

The paths are then loops on a cylinder with extent β, the imaginary time direction is compact. The ground state $(T \to 0)$ limit is obtained by making the imaginary time extent infinite.

This partition can be interpreted as the partition function of a 1d classical gas of loops, distributed according to

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

(1.36)

which is the $\delta\tau \to 0$ limit of Eq. 1.29 for periodic paths, and

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

(1.37)

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

This path integral formulation builds the foundation of all quantum field theory algorithms.

1.5.2 Observables

The expectation value of an observable W is given by

$$\langle W \rangle = \frac{1}{Z}\text{Tr}\, W \exp(-\beta H). \tag{1.38}$$

When expanded in the eigenbasis of H we obtain

$$\langle W \rangle = \frac{\sum_k \langle \psi_k|W|\psi_k\rangle \exp(-\beta E_k)}{\sum_k \langle \psi_k|\psi_k\rangle \exp(-\beta E_k)} \tag{1.39}$$

$$=_{\beta \to \infty} \frac{\langle \psi_0|W|\psi_0\rangle}{\langle \psi_0|\psi_0\rangle} \tag{1.40}$$

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

$$\langle \delta(x_0) \rangle =_{\beta \to \infty} |\langle \psi_0|x_0\rangle|^2 = |\psi_0(x_0)|^2. \tag{1.41}$$

This implies that the number of paths going through x_0 (at any Euclidian time) is proportional to the square of the wave function.

Compare this with Diffusion Monte Carlo, where $\langle \delta(x_0) \rangle \propto \psi_0(x_0)$.

One can also measure correlation functions at two times ('Green's functions'). These give information about excited states. In particular, it is possible to extract the

lowest excited state from deviations to the exponential decay to the ground state.

The ground state energy itself is more delicate to obtain, for the following reason. Consider the kinetic energy part of Eq. 1.29. It defines a Gaussian distribution for $(x_k - x_{k-1})$, with variance $\langle (x_k - x_{k-1})^2 \rangle \sim O(\delta\tau)$. As appropriate for Brownian motion, the mean-square distance grows linearly with time. However, this implies that the mean square velocity, i.e. the kinetic energy, grows as $1/\delta\tau$. This is, again, expected from Brownian motion: a Brownian path is continuous, but not differentiable.

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