MONTE CARLO METHODS

– PART 1

Francesco Pederiva

LISC, Interdisciplinary Laboratory for Computationsal Science, FBK and University of Trento
INFN, Gruppo Collegato di Trento

Outline

- Variational Monte Carlo
  - Many-dimensional integrals
  - Stochastic integration
  - Metropolis algorithm
  - Wave Functions (Bosons and Fermions)
- Diffusion Monte Carlo for Many Boson Systems
  - Propagation in imaginary time
  - Importance sampling
  - General treatment in terms of Green's Functions

Monte Carlo methods

Investigating the properties of nuclei and nuclear matter implies the solution of a *many-body* problem. Let us consider the non relativistic version. In general, we must solve a many-body Schroedinger equation:

$$\sum_{i=1}^{N} -\frac{\hbar^2}{2m_i}\nabla_i^2\Psi(r_1\ldots r_N,\alpha_1\ldots\alpha_N) + \hat{V}(r_1\ldots r_N,\sigma_1\ldots\sigma_N)\Psi(r_1\ldots r_N,\alpha_1\ldots\alpha_N) = E_n\Psi(r_1\ldots r_N,\alpha_1\ldots\alpha_N)$$

Where r_i are the space coordinates of the particles, and α_i are other possible degrees of freedom (e.g. the spin/isospin states of nucleons). The potential V is in general an operator that acts on both the space and/or spin/isospin degrees of freedom.
Many forms of approximate solutions exist.

IS IT POSSIBLE TO ACCURATELY SOLVE THE PROBLEM FOR AN ARBITRARY N?

MONTE CARLO METHODS

Let us assume that:

✓ We are at zero temperature (almost ok for most systems of interest);
✓ We are interested only in ground state properties.

Monte Carlo methods give us two options:

1. **Guess a reasonable form of the wavefunction, depending on a set of parameters, and apply the variational principle computing the expectation of the Hamiltonian by means of Monte Carlo methods. This is called *VARIATIONAL MONTE CARLO (VMC)*.**

2. **Solve the problem by projecting the ground state from an arbitrary initial guess of the wavefunction by means of a propagation in imaginary time. There is a family of such algorithms, the most celebrated of which is the so-called *DIFFUSION MONTE CARLO (DMC)***

VARIATIONAL APPROACH

The method is based on assuming a given functional form of the wavefunction (necessarily approximate), depending on some parameters {p}. The variational theorem tells us that:

$$\frac{\left\langle \Psi(R,\{p\}) \middle| \hat{H}\Psi(R,\{p\}) \right\rangle}{\left\langle \Psi(R,\{p\}) \middle| \Psi(R,\{p\}) \right\rangle} = E_T(\{p\}) \geq E_0$$

The expectation of the Hamiltonian is a multi-dimensional integral in the degrees of freedom of the system (coordinates, spin and isospin).

MULTI-DIMENSIONAL INTEGRALS

STANDARD APPROACH

-Divide the integration domain in "small" hypercubes of side h.
-Compute the function in one representative point within each hypercube.
-Sum up the function values and multiply by the volume of the hypercube.
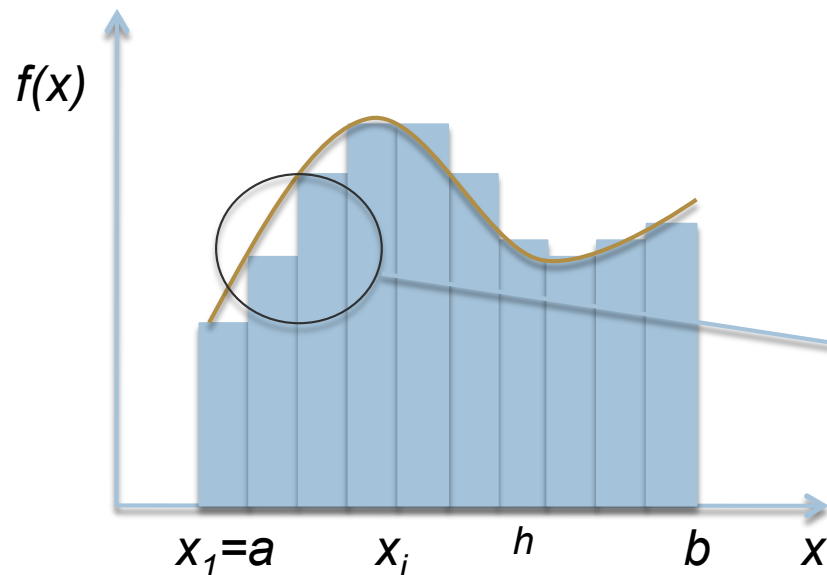
$$I = \int_a^b f(x)dx \cong \left[\sum_{i=1}^{(b-a)/h} f(x_i) \right] h$$

$$\sim f'(x)h$$

$$\Delta I_i \cong \frac{1}{2} f'(x) \cdot h^2$$

More clever methods can be used, but the error is always proportional to h^α

$f(x)$

$x_1 = a$ x_i h b x

MULTI-DIMENSIONAL INTEGRALS

In d dimensions the situation becomes even worse.
The important question is the following:

HOW MANY POINTS DO WE NEED TO REACH A GIVEN RELATIVE ERROR ON THE ESTIMATE OF THE INTEGRAL?

$$\frac{\Delta I_i}{I_i} = \varepsilon \qquad \frac{\nabla f(x_i) \cdot h^{d+1}}{f(x_i) \cdot h^d} \propto h \qquad but \qquad N = \left(\frac{L}{h}\right)^d \Rightarrow h \propto N^{\frac{1}{d}}$$

$$\varepsilon \propto N^{\frac{1}{d}} \Rightarrow N \propto \frac{1}{\varepsilon^d}$$

e.g.: ^{16}O, central potential, d=48
Require ε=0.1

N=10^{48}
With a Pflop machine $T_{CPU}>10^{33}s$!!!!!

CENTRAL LIMIT THEOREM

Fortunately we can rely on stochastic processes to speed up the convergence….

The key ingredient is the *Central Limit Theorem.*

Consider a set of N continuum random variables, each one described by the same *probability density P(x),* and a function *f(x)*. We define a new random variable S_N as:

$$S_N = \frac{1}{N} \sum_{i=1}^{N} f(x_i)$$

where each of the x_i is *sampled* from the probability density *P(x).*
ALL SAMPLES MUST BE STATISTICALLY INDEPENDENT

CENTRAL LIMIT THEOREM

Under the hypothesis that the samples are independent, the Central Limit Theorem proves that:

$$P(S_N) = \frac{1}{\sqrt{2\pi\sigma_N^2}} e^{\frac{(S_N - \langle f \rangle)^2}{2\sigma_N^2}}$$

Estimate of the error on the estimate of <f>…

where:

$$\langle f \rangle = \frac{\int P(x) f(x) dx}{\int P(x) dx}$$

should P not be normalized….

$$\sigma_N = \sqrt{\frac{1}{N}\left(\langle f^2 \rangle - \langle f \rangle^2\right)}$$

THIS RESULT HOLDS FOR ANY DIMENSIONALITY OF THE SPACE IN WHICH x IS DEFINED!

STOCHASTIC INTEGRATION

The CLT prescribes a very simple way to integrate a multi-dimensional function *F(x)*.

1) Choose a probability density $P(x)$ that **can be sampled**, defined on a domain including the domain of $F(x)$. Because P(x) has to be positive definite, we can always rewrite the integral of F as:

$$I = \int P(x) f(x) dx = \int P(x) \frac{F(x)}{P(x)} dx$$

2) Sample N (with N "large") points from the probability density P(x)

3) Average the N values of the function $f(x_i)$ and the N values of $f(x_i)^2$. The estimate of the integral will be:

$$I \cong \frac{1}{N} \sum_{i=1}^{N} f(x_i) \pm \sqrt{\frac{1}{N}\left[\left(\frac{1}{N}\sum_{i=1}^{N} f(x_i)^2\right) - \left(\frac{1}{N}\sum_{i=1}^{N} f(x_i)\right)^2\right]}$$

STOCHASTIC INTEGRATION

The previous expression gives us a very important indication:

> **THE ERROR ON THE ESTIMATE OF THE INTEGRAL DECREASES AS $1/N^{1/2}$ REGARDLESS OF THE DIMENSIONALITY!**

This means that for reaching a given accuracy ε one needs a number of points N growing as $1/\varepsilon^2$. For d larger than a few units this is an enormous gain in computational time.
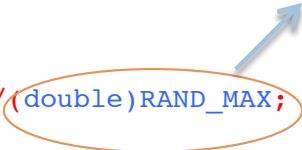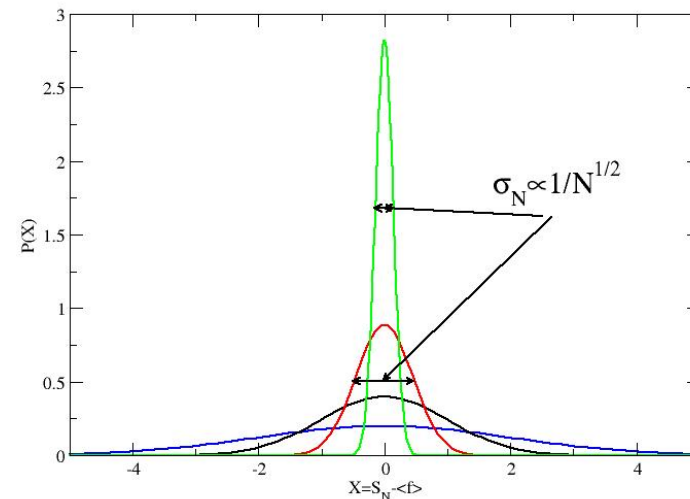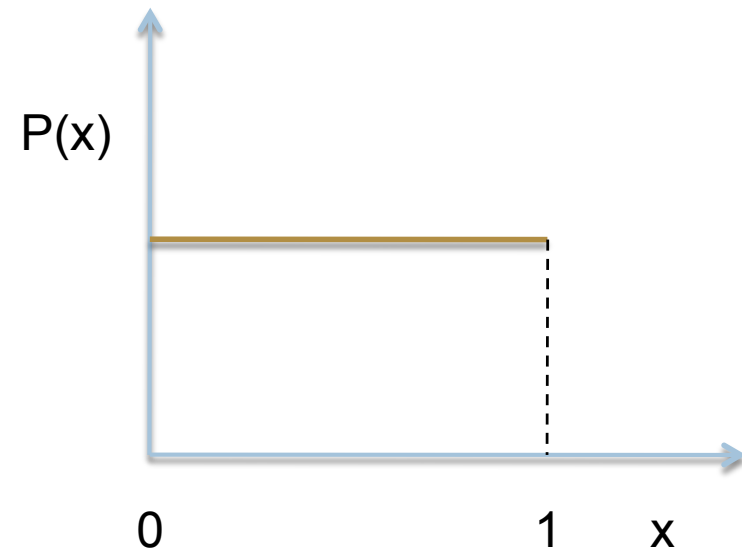
SAMPLING

How to sample a generic probability density?

Computers provide us with a generator of uniformly distributed random numbers between 0 and 1

```
#include <math.h>
#include <stdio.h>
#include <stdlib.h> /* necessary for the rand() function*/
int main()
{

……

srand(3);

……

y=y+(double)rand()/(double)RAND_MAX;


}
```

Normalization. In C rand() produces integer numbers uniformly distributed between 0 and RAND_MAX

P(x)

0   1   x

SAMPLING

In general we need to sample more complex P(X), such as the square modulus of a variational wavefunction.

The simplest and most reliable method is based on the concept of "random walk". The game consists of generating a random variable applying a transformation to another. This "moving" point is called **walker.**

$$x_{i+1} = t(x_i)$$

$$P_{i+1}(x_{i+1}) = \int T(x_{i+1} \leftarrow x_i) P_i(x_i) dx_i$$

If we apply the same transformation k times to some initial value we obtain:

$$P_k(x_k) = \int T(x_k \leftarrow x_{k-1}) \cdots T(x_3 \leftarrow x_2) T(x_2 \leftarrow x_1) P_1(x_1) dx_1 dx_2 \cdots dx_{k-1}$$

SAMPLING

It can be proved that under some very general conditions the probability density P_k converges to a limiting distribution, which depends only on T

$$\lim_{k \to \infty} P_k(x_k) = P_\infty(x)$$

Is it possible to build a T such that we eventually sample the P that we are interested in? Let us assume a further condition, i.e. that the asymptotic distribution is an "equilibrium" state (meaning that point by point there is no net flux of probability):

$$T(y \leftarrow x)P(x) = T(x \leftarrow y)P(y)$$

This is called **DETAILED BALANCE CONDITION,** and any T satisfying this condition will eventually sample $P(x)$. It is called "detailed" because it does not hold only on average, but **POINT BY POINT!**

SAMPLING

Now, quite arbitrarily, we split T in two factors: one factor G that we can sample, and another unknown part A. Then, the detailed balance condition reads:

$$G(y \leftarrow x)A(y \leftarrow x)P(x) = G(x \leftarrow y)A(x \leftarrow y)P(y)$$

that we can rewrite in the following form:

$$\frac{A(y \leftarrow x)}{A(x \leftarrow y)} = \frac{G(x \leftarrow y)P(y)}{G(y \leftarrow x)P(x)}$$

The quotient on the l.h.s. of the equation can be read as a *probability* of accepting the **new proposed value y as the next member of the chain** versus its complementary, which means "coming back" to x from y , and therefore **keeping x as the next element.**

METROPOLIS ALGORITHM

There are some very simple and obvious choices for *G*. The easier thing to do is to shift a point in the configuration space of a random vector having components uniformly distributed in some interval *[-Δ/2,Δ/2]*

$$\vec{x}_{i+1} = \vec{x}_i + \vec{\xi}$$

```
for(j=0;j<nparticles;j++)
   {
     ……
     dx=delta*(0.5-(double)rand()/(double)RAND_MAX);
     new.x=walker[j].x+dx;
     dy=delta*(0.5-(double)rand()/(double)RAND_MAX);
     new.y=walker[j].y+dy;
     dz=delta*(0.5-(double)rand()/(double)RAND_MAX);
     new.z=walker[j].z+dz;

     ……
```

$-\Delta/2$ $\Delta/2$
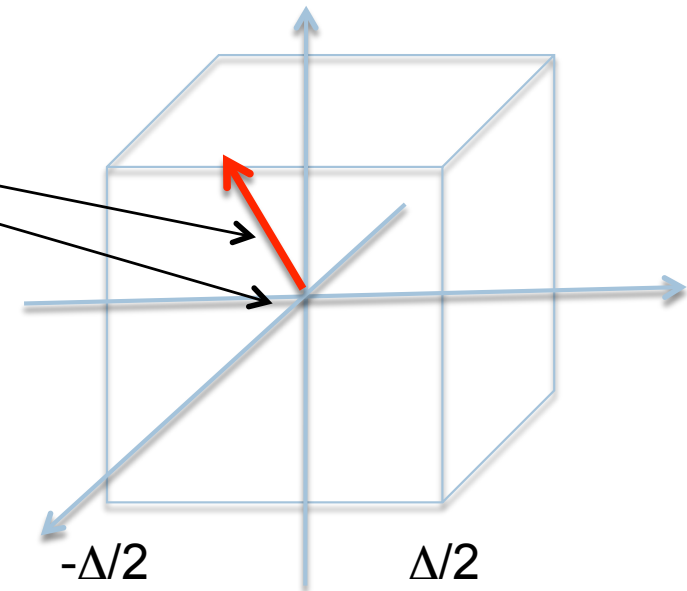
METROPOLIS ALGORITHM

At this point we have to decide which point (old or new) is to be kept as next point in the chain.
First, notice that the proposed move is symmetric in the old and new positions. This means that:

$$\frac{A(x_{i+1} \leftarrow x_i)}{A(x_i \leftarrow x_{i+1})} = \frac{P(x_{i+1})}{P(x_i)}$$

So the only thing we need to do is to compute the probability density we want to sample in the new and old position and take the ratio.

Notice that the fact that only the ratio of the $P(x)$ appears makes the results **INDEPENDENT OF THE NORMALIZATION** of $P(x)$.
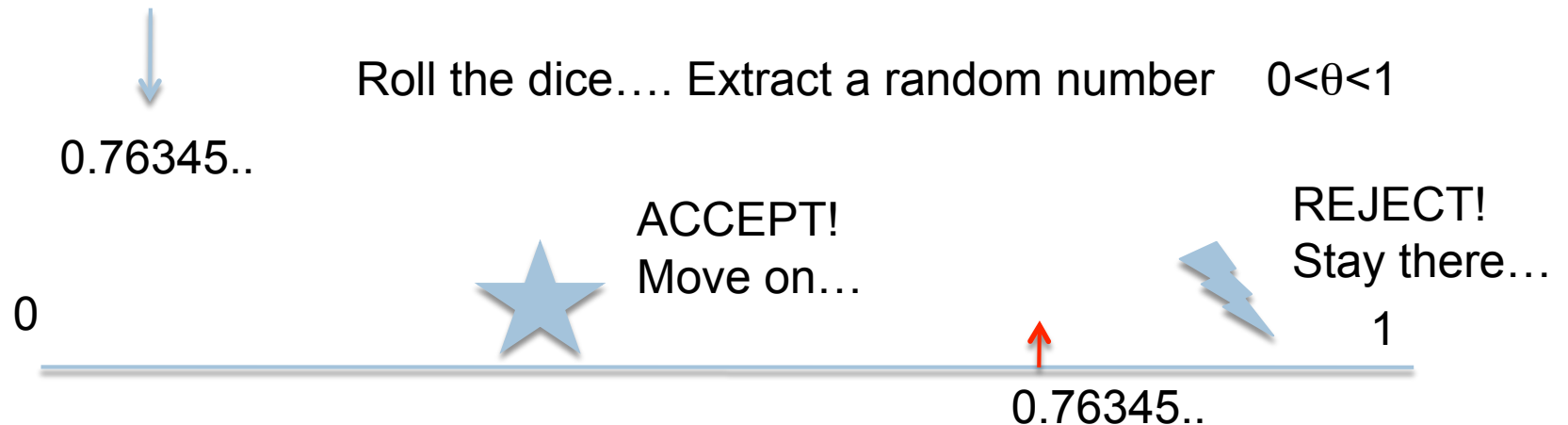
METROPOLIS ALGORITHM

Now, how do we exploit the computed ratio?

$$\frac{A(x_{i+1} \leftarrow x_i)}{A(x_i \leftarrow x_{i+1})} > 1$$

The probability of arriving at the new position is larger than that of going back. Let's stay there!

$$\frac{A(x_{i+1} \leftarrow x_i)}{A(x_i \leftarrow x_{i+1})} < 1$$

We have a finite probability of going back, and we need to *SAMPLE* this probability.

0.76345..

Roll the dice…. Extract a random number $0 < \theta < 1$

ACCEPT!
Move on…

REJECT!
Stay there…

0

1

0.76345..

AUTOCORRELATIONS

The Metropolis prescription *explicitly violates one of the requests of the central limit theorem*:

> **THE SAMPLED CONFIGURATIONS ARE IN GENERAL NOT STATISTICALLY INDEPENDENT!!!!**

$$\langle f(x_i)f(x_j)\rangle = \int P(x_i,x_j)f(x_i)f(x_j)dx_i dx_j \overset{???}{=}$$

$$= \int P(x_i)P(x_j)f(x_i)f(x_j)dx_i dx_j =$$

$$= \int P(x_i)f(x_i)dx_i \int P(x_j)f(x_j)dx_j =$$

$$= \langle f(x_i)\rangle\langle f(x_j)\rangle = \langle f(x_i)\rangle^2$$

Only if independent samples

AUTOCORRELATIONS

It is convenient to introduce an estimate of the autocorrelation normalized to the variance. We assume that it depends only on the difference t=j-i.

$$\overline{c}_t = \frac{\langle f(x_i)f(x_{i+t})\rangle - \langle f(x)\rangle^2}{\langle f^2(x)\rangle - \langle f(x)\rangle^2}$$

It is easy to prove that data correlated over a "time" *t* lead to an underestimate of the error. The error must be corrected in the following way:

$$\sigma_N^{true} = \sqrt{\sigma_N^2 \cdot t}$$
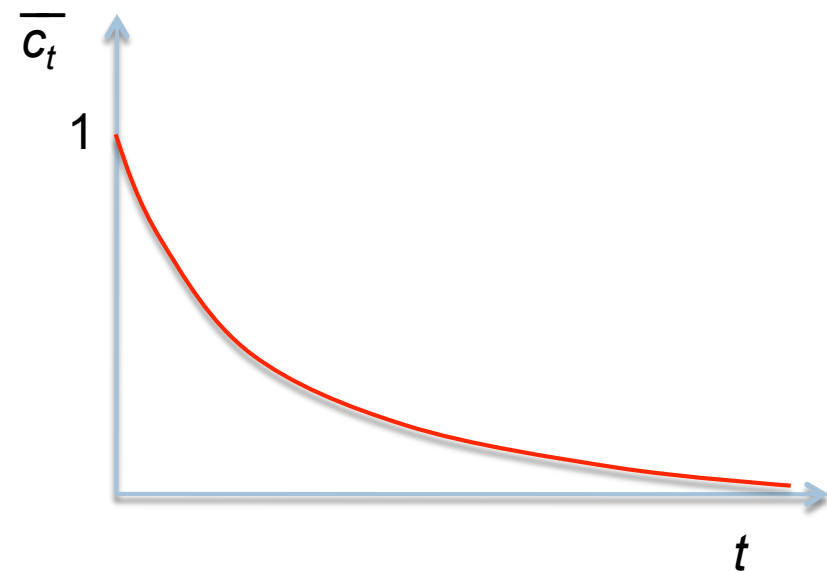
AVERAGES

As we saw, integrals are estimate by averaging the integrand function on the sampled configurations. Beware that **configurations repeated in the sequence because of rejection must be included in the averages each and every time they appear.**

IMPORTANT EXAMPLE:
EXPECTATION OF THE HAMILTONIAN

$P(x)$

"Local" energy

$$\frac{\left\langle \Psi(R,\{p\}) \middle| \hat{H}\Psi(R,\{p\}) \right\rangle}{\left\langle \Psi(R,\{p\}) \middle| \Psi(R,\{p\}) \right\rangle} = \frac{\int \left|\Psi(R,\{p\})\right|^2 \dfrac{\hat{H}\Psi(R,\{p\})}{\Psi(R,\{p\})} dR}{\int \left|\Psi(R,\{p\})\right|^2 dR} \cong$$

$$\cong \frac{1}{N}\sum_{i=1}^{N} \frac{\hat{H}\Psi(R_i,\{p\})}{\Psi(R_i,\{p\})} \pm \sqrt{\frac{1}{N}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{H}\Psi(R_i,\{p\})}{\Psi(R_i,\{p\})}\right)^2 - \left(\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{H}\Psi(R_i,\{p\})}{\Psi(R_i,\{p\})}\right)^2\right]}$$

EXERCISES

EASY:

Using the Metropolis algorithm compute:

$$\int_0^1 P(x)[e^x - 1]dx \qquad P(x) = 1, \ \ P(x) = 2x$$

Discuss how the variance changes by changing the Metropolis step.

ALMOST EASY:

Find the optimal variational parameters x_0 and σ for the following approximation of the first excited state of the 1D harmonic oscillator (h=m=1):

$$\psi(x) = e^{-\frac{(x-x_0)^2}{2\sigma^2}} - e^{-\frac{(x+x_0)^2}{2\sigma^2}}$$

Scuola "R. Anni" – Otranto, 30/05-04/06/2011
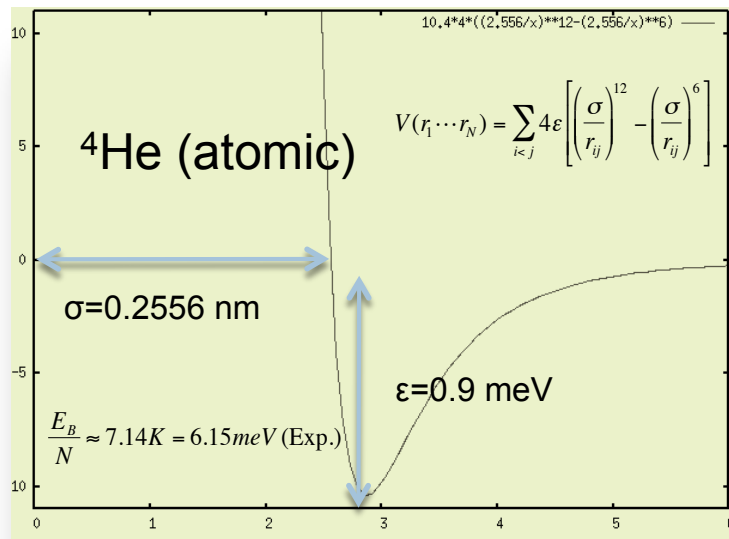
MONTE CARLO METHODS

– PART 2

Francesco Pederiva

LISC, Interdisciplinary Laboratory for Computationsal Science, FBK and University of Trento
INFN, Gruppo Collegato di Trento

Wave Functions

Example: the Lennard Jones potential seen in O. Benhar lectures:

^{4}He (atomic)

$\sigma=0.2556$ nm

$\varepsilon=0.9$ meV

$\dfrac{E_B}{N} \approx 7.14K = 6.15 meV$ (Exp.)

$$V(r_1 \cdots r_N) = \sum_{i<j} 4\varepsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right]$$

Let's consider a homogeneous liquid of ^{4}He atoms at T=0. What might be a reasonable wavefunction? Let's assume

$$\Psi(r_1 \cdots r_N) = \prod_{i<j}^{N} \exp[-\frac{1}{2} u(r_{ij})]$$

In general we would like that even when the potential diverges the local energy does not diverge! (cusp condition)

$$\frac{\hat{H}\Psi(r_1 \cdots r_N)}{\Psi(r_1 \cdots r_N)} = \frac{\hat{H}\prod_{i<j}^{N} \exp[-\frac{1}{2} u(r_{ij})]}{\prod_{i<j}^{N} \exp[-\frac{1}{2} u(r_{ij})]} \underset{\text{some } r_{ij} \to 0}{\longrightarrow} < \infty$$

$$u(r_{ij}) = \left(\frac{b}{r_{ij}} \right)^{5}$$

Exercise (some algebra)

METROPOLIS ALGORITHM

```
advance(struct particle walker[npmax])

{
    ......
uo=pair_wf(j,walker[j].x,walker[j].y,walker[j].z,walker);
    for(j=0;j<npart;j++)
    {
        dx=delta*(0.5-(double)rand()/(double)RAND_MAX);
        new.x=walker[j].x+dx;
        dy=delta*(0.5-(double)rand()/(double)RAND_MAX);
        new.y=walker[j].y+dy;
        dz=delta*(0.5-(double)rand()/(double)RAND_MAX);
        new.z=walker[j].z+dz;
    }

    un=pair_wf(j,new.x,new.y,new.z,walker);
    arg=un-uo;
    p=exp(-arg);
    csi=(double)rand()/(double)RAND_MAX;

    if(p>csi)
        {
            acc=acc+1.;
            walker[j].x=new.x;
            walker[j].y=new.y;
            walker[j].z=new.z;
        };
}
```

Move walkers

Acceptance/ rejection

$$\Psi(r_1 \cdots r_N) = \prod_{i<j}^{N} \exp[-\frac{1}{2} u(r_{ij})]$$

$$\sum_{i<j} u(r_{ij}) \Big|_{OLD}$$

$$\sum_{i<j} u(r_{ij}) \Big|_{NEW}$$

$$\frac{|\Psi(R_{new})|^2}{|\Psi(R_{old})|^2} = \exp\left[\sum_{i<j} u(r_{ij})\Big|_{OLD} - \sum_{i<j} u(r_{ij})\Big|_{NEW}\right]$$
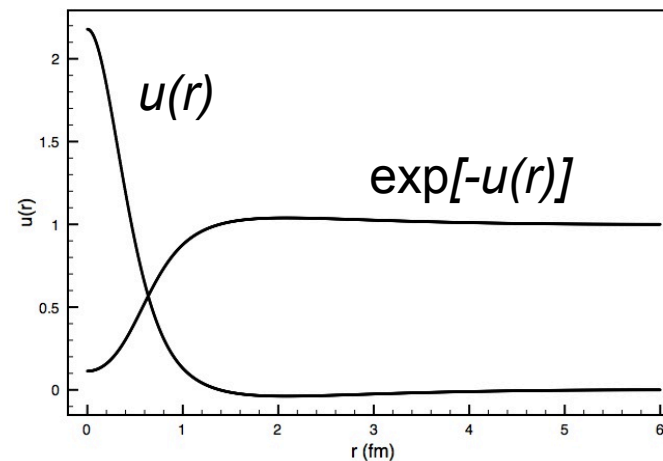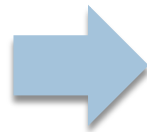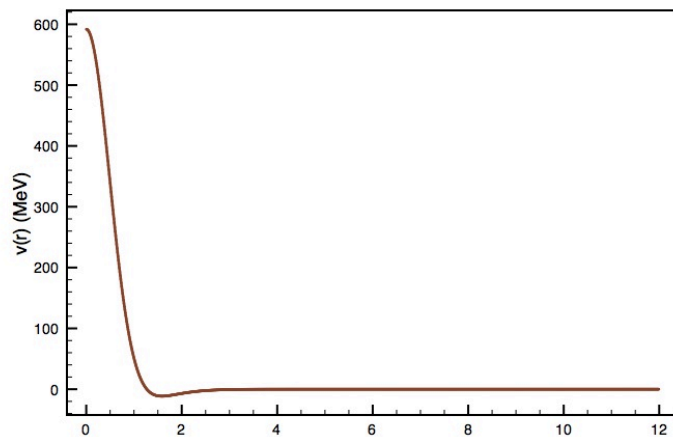
Wave Functions

An easy way to implement the cusp condition is to solve numerically the two-body problem:

$$-\frac{\hbar^2}{2m}\nabla^2 e^{-\frac{1}{2}u(r)} + [V(r) - E]e^{-\frac{1}{2}u(r)} = 0$$

in the interval *[0,h]*. *h* is a variational parameter (healing distance) and represents the distance at which *u(r)* becomes 0.

FERMIONS

When we deal with a many-Fermion system the trial wavefunction must be antisymmetric under particle permutations.

$$\Psi(PR) = (-1)^P \Psi(R)$$

This is usually achieved by using a Slater determinant of single particle wavefunctions

$$\Psi(R) = \exp\left[-\frac{1}{2}\sum_{i<j} u(r_{ij}) - \frac{1}{6}\sum_{i<j<k} u(\mathbf{r}_{ij}\cdot\mathbf{r}_{ik}) + \cdots\right] \times \begin{Vmatrix} f_1(\mathbf{r}_1) & f_1(\mathbf{r}_2) & \cdots & f_1(\mathbf{r}_N) \\ f_2(\mathbf{r}_1) & f_2(\mathbf{r}_2) & \cdots & f_2(\mathbf{r}_N) \\ \vdots & \vdots & \ddots & \vdots \\ f_N(\mathbf{r}_1) & f_N(\mathbf{r}_2) & \cdots & f_N(\mathbf{r}_N) \end{Vmatrix}$$

If the Hamiltonian commutes with σ_z it is convenient to write a product of two determinants, one for spin-up particles and one of spin-down particles.

FERMIONS

The computation of the determinant is expensive, scaling is with N^3. It is possible to save some time if we compute at first the *inverse* of the Slater matrix. Then we remember that:

$$\mathbf{A}^{-1} = \frac{1}{\|\mathbf{A}\|}\left(\mathbf{C}^T\right)_{ij} = \frac{1}{\|\mathbf{A}\|}\left(\mathbf{C}_{ji}\right)$$

where **C** is the matrix of the co-factors.
If we change the coordinates of *only one of the particles* in a Metropolis step, the ratio of the Slater determinants can be computed by multiplying just two rows (columns) of the Slater matrix and of its inverse.

FERMIONS

Example:
Let's move particle 2.

$$A = \begin{Vmatrix} f_1(\mathbf{r}_1) & f_1(\mathbf{r}_2) & \cdots & f_1(\mathbf{r}_N) \\ f_2(\mathbf{r}_1) & f_2(\mathbf{r}_2) & \cdots & f_2(\mathbf{r}_N) \\ \vdots & \vdots & \ddots & \vdots \\ f_N(\mathbf{r}_1) & f_N(\mathbf{r}_2) & \cdots & f_N(\mathbf{r}_N) \end{Vmatrix} \Rightarrow \begin{Vmatrix} f_1(\mathbf{r}_1) & f_1(\mathbf{r}_2^{new}) & \cdots & f_1(\mathbf{r}_N) \\ f_2(\mathbf{r}_1) & f_2(\mathbf{r}_2^{new}) & \cdots & f_2(\mathbf{r}_N) \\ \vdots & \vdots & \ddots & \vdots \\ f_N(\mathbf{r}_1) & f_N(\mathbf{r}_2^{new}) & \cdots & f_N(\mathbf{r}_N) \end{Vmatrix}$$

The ratio of the determinants simply becomes

$$\frac{\det(\mathbf{A}_{new})}{\det(\mathbf{A}_{old})} = f_1^{-1}(\mathbf{r}_2)f_1(\mathbf{r}_2) + f_2^{-1}(\mathbf{r}_2)f_2(\mathbf{r}_2) + \cdots + f_N^{-1}(\mathbf{r}_2)f_N(\mathbf{r}_2)$$

The square of this quantity enters the probability to sample..

FERMIONS

The same property is used to compute gradients and Laplacians of the wavefunction:

$$\partial_k \mathbf{A} = \begin{Vmatrix} f_1(\mathbf{r}_1) & \partial_k f_1(\mathbf{r}_2) & \cdots & f_1(\mathbf{r}_N) \\ f_2(\mathbf{r}_1) & \partial_k f_2(\mathbf{r}_2) & \cdots & f_2(\mathbf{r}_N) \\ \vdots & \vdots & \ddots & \vdots \\ f_N(\mathbf{r}_1) & \partial_k f_N(\mathbf{r}_2) & \cdots & f_N(\mathbf{r}_N) \end{Vmatrix}$$

That is:

$$\frac{\partial_k \det(\mathbf{A}_{new})}{\det(\mathbf{A}_{old})} = f_1^{-1}(\mathbf{r}_k)\partial_k f_1(\mathbf{r}_k) + f_2^{-1}(\mathbf{r}_k)\partial_k f_2(\mathbf{r}_k) + \cdots + f_N^{-1}(\mathbf{r}_k)\partial_k f_N(\mathbf{r}_k)$$

The same holds for second derivatives. These quantities enter the local energy.

FERMIONS

Are the zeros of the wavefunction a problem?
Not here. Near the nodes the wavefunction is linear in the coordinates.

$$\Psi(r)$$

$$\frac{\int |\Psi(\mathbf{r}_1 \cdots \mathbf{r}_N)|^2 \dfrac{\hat{H}\Psi(\mathbf{r}_1 \cdots \mathbf{r}_N)}{\Psi(\mathbf{r}_1 \cdots \mathbf{r}_N)} d\mathbf{r}_1 \cdots d\mathbf{r}_N}{\int |\Psi(\mathbf{r}_1 \cdots \mathbf{r}_N)|^2 d\mathbf{r}_1 \cdots d\mathbf{r}_N}$$

~ R²,
and always
positive

OK!

~ R

Wave Function Optimization

How do we optimize the parameters in a variational wavefunction?
We can certainly do it by hand… But if we want to be flexible in our
parametrization this is not certainly efficient!

There is a method allowing for a set of configurations sampled froma a
wavefunction with a set of parameters *{p}* to compute the local energy for a
different set of parameters *{p'}*

$$\frac{\left\langle \Psi(R,\{p'\}) \middle| \hat{H}\Psi(R,\{p'\}) \right\rangle}{\left\langle \Psi(R,\{p'\}) \middle| \Psi(R,\{p'\}) \right\rangle} = \frac{\int \left| \Psi(R,\{p'\}) \right|^2 \frac{\hat{H}\Psi(R,\{p'\})}{\Psi(R,\{p'\})} dR}{\int \left| \Psi(R,\{p'\}) \right|^2 dR} =$$

$$= \frac{\int \left| \Psi(R,\{p\}) \right|^2 \frac{\left| \Psi(R,\{p'\}) \right|^2}{\left| \Psi(R,\{p\}) \right|^2} \frac{\hat{H}\Psi(R,\{p'\})}{\Psi(R,\{p'\})} dR}{\int \left| \Psi(R,\{p\}) \right|^2 \frac{\left| \Psi(R,\{p'\}) \right|^2}{\left| \Psi(R,\{p\}) \right|^2} dR} \cong \frac{\sum_i w_i \frac{\hat{H}\Psi(R_i,\{p'\})}{\Psi(R_i,\{p'\})}}{\sum_i w_i}$$

Wave Function Optimization

The weight w_i is simply given by:

$$w_i = \frac{\left|\Psi(R_i,\{p'\})\right|^2}{\left|\Psi(R_i,\{p'\})\right|^2}$$

The method is often called *CORRELATED SAMPLING.*

Note that if w_i is not ~1 this method might become in turn very inefficient!! However, this method is extremely useful to compute **gradients and Laplacians of the local energy in parameters space**, which are the essential ingredient in any minimization or fitting method (steepest descent, Levemberg-Marquardt, Hessian based methods etc.)

DMC for central potentials

Important fact:
The Schroedinger equation in imaginary time is a diffusion equation (a transformed Fokker-Planck equation):

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(R) \right]\Psi(R,\tau) = -\hbar\frac{\partial}{\partial\tau}\Psi(R,\tau)$$

where R represent the coordinates of the particles, and τ = it is an imaginary time (would be real time in diffusion).

DMC for central potentials

Let us consider the formal solution:

$$\Psi(R,\tau) = e^{-\frac{H}{\hbar}\tau}\Psi(R,0)$$

We can expand the initial state on a basis of eigenfunction of the Hamiltonian, and get:

$$\Psi(R,\tau) = e^{-\frac{H}{\hbar}\tau}\Psi(R,0) =$$

$$= e^{-\frac{H}{\hbar}\tau}\sum_n c_n \Psi_n(R) = \sum_n e^{-\frac{E_n}{\hbar}\tau} c_n \Psi_n(R)$$

What about the normalization of the propagated wf?

DMC for central potentials

Let us multiply by a time-dependent factor $e^{\frac{E_0}{\hbar}\tau}$.
This yelds:

"Absolute" ground state of H. This is **ALWAYS** a nodeless function!

$$\Psi(R,\tau) = e^{-\frac{(H-E_0)}{\hbar}\tau}\Psi(R,0)$$

$$= e^{-\frac{(E_0-E_0)}{\hbar}\tau}c_0\Psi_0(R,0) + \sum_{n\neq0}e^{-\frac{(E_n-E_0)}{\hbar}\tau}c_n\Psi_n(R,0)$$

The expression now converges to the lowest energy eigenstate provided that the initial and final states are not orthogonal.

NB: because of the properties of the ground state we must restrict for the moment to **many-Boson systems**!

DMC for central potentials

For a generic Hamiltonian we do not know neither the explicit form of the propagator, nor the normalization (i.e. the lowest eigenvalue).
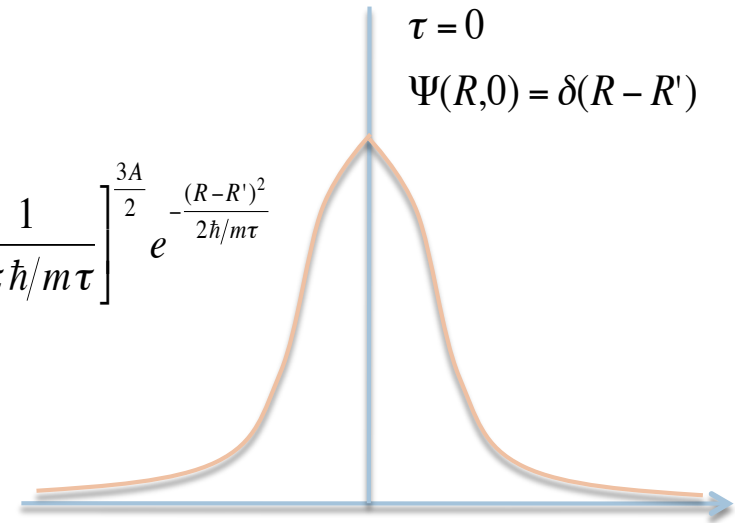*Looks useless!*

BUT...

• Maybe we could use an approximate propagator…

• We might try to control the normalization "along the way"

Propagator

Case m $\longrightarrow$ 0

The Hamiltonian reduces to:

$$\left[-\frac{\hbar^2}{2m} \sum_{i=1}^{A} \nabla_i^2 \right] \Psi(R,\tau) = -\hbar \frac{\partial}{\partial \tau} \Psi(R,\tau)$$

$\tau = 0$

$\Psi(R,0) = \delta(R - R')$

$$\Psi(R,\tau) = \left[\frac{1}{2\pi\hbar/m\tau} \right]^{\frac{3A}{2}} e^{-\frac{(R-R')^2}{2\hbar/m\tau}}$$

The lowest eigenstate is a constant, and the lowest eigenvalue is 0:

no extra normalization needed!

$$\left\langle R \left| e^{-\frac{H}{\hbar}\tau} \Psi \right. \right\rangle = \left\langle R \left| e^{-\frac{H}{\hbar}\tau} \int dR' \left| R' \right\rangle \left\langle R' | \Psi \right\rangle = \int G_0(R,R',\tau) \Psi(R') dR' \right.$$

where $\quad G_0(R,R',\tau) = \left[\frac{1}{2\pi\sigma^2} \right]^{\frac{3A}{2}} e^{-\frac{(R-R')^2}{2\sigma^2}} \quad$ *and* $\quad \sigma^2 = \frac{\hbar^2}{m} \boxed{\frac{\tau}{\hbar}}$
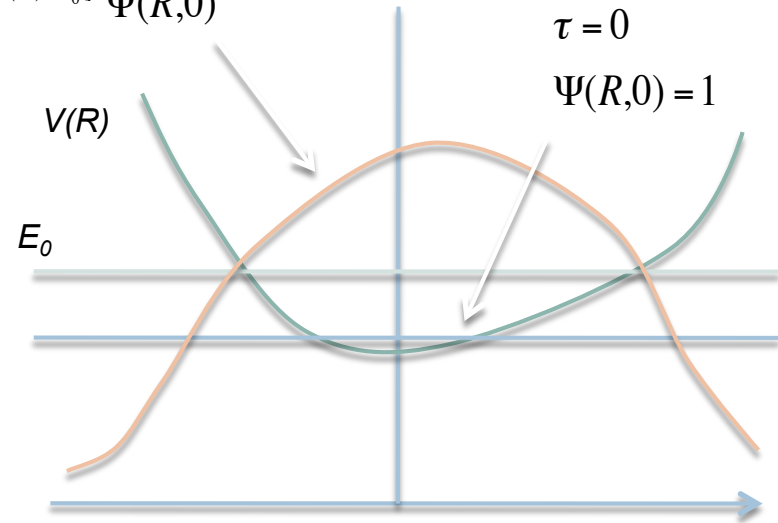
Inverse of an energy

Free particle propagator

Propagator

$$\Psi(R,\tau) = e^{-[V(R)-E_0]\tau}\Psi(R,0)$$

$\tau = 0$

$\Psi(R,0) = 1$

Case m $\longrightarrow$ ∞

$$\hbar = 1$$

V(R)

E_0

The Hamiltonian now reads:

$$[V(R) - E_0]\Psi(R,\tau) = -\frac{\partial}{\partial\tau}\Psi(R,\tau)$$

Here we include the normalization. The solution, once more written in terms of eigenstates of the position, is trivial:

$$\langle R|e^{-[V(R)-E_0]\tau}\Psi\rangle =$$

Probability for the walker to generate copies of itself in a time τ (source/sink probability)

$$= \langle R|e^{-[V(R)-E_0]\tau}\int dR'|R'\rangle\langle R'|\Psi\rangle = \int e^{-[V(R')-E_0]\tau}\Psi(R')\delta(R-R')dR'$$

The latter expression does not imply any time evolution of the position, but simply a "reshaping" of the wavefunction in the imaginary time interval τ

Trotter-Suzuki break-up

In real life we are never in either of these two limts. However, we can invoke the Trotter-Suzuki formula in order to get closer to this situation, provided that the propagation occurs for a short time $\Delta\tau$:

$$\langle R | e^{-\frac{(H-E_0)}{\hbar}\tau} | R' \rangle = \langle R | e^{-\frac{[V(R)-E_0]\Delta\tau}{\hbar}\frac{1}{2}} e^{-\frac{T}{\hbar}\tau} e^{-\frac{[V(R')-E_0]\Delta\tau}{\hbar}\frac{1}{2}} | R' \rangle + o(\Delta\tau^3)$$

$$= \left[\frac{1}{2\pi\hbar/m\,\Delta\tau} \right]^{\frac{3A}{2}} e^{-\frac{(R-R')^2}{2\hbar/m\Delta\tau}} e^{-\left[\frac{V(R)+V(R')}{2}-E_T\right]\frac{\Delta\tau}{\hbar}}$$

In order to project out the ground state of the Hamiltonian, we need a large τ. So the short-time propagator must be repeatedly applied.

How does it work?

There are many different ways to achieve a propagation in imaginary time (see e.g. PIGS or reptation methods).

*Diffusion Monte Carlo is based on the approximation of the wavefunction in terms of an **expansion over a limited number of eigenstates of the position (walkers**).*

$$\Psi(R) = \langle R | \Psi \rangle = \langle R | \int dR' | R' \rangle \langle R' | \Psi \rangle \approx$$

$$\approx \langle R | \sum_{k=1}^{M} | R'_k \rangle \langle R'_k | \Psi \rangle = \sum_{k=1}^{M} \delta(R - R'_k) \Psi(R'_k)$$

How it works…

*Once more, **expectation values of estimators** can be easily computed starting from the walkers positions. Let us assume that the propagation has been carried out for a sufficiently long imaginary time. Then the walkers will be distributed proportionnaly to the **ground state wavefunction**. As for the VMC case, the expectation of the Hamiltonian will become:*

$$E_0 = \frac{\left\langle \Psi_T \middle| \hat{H}\Psi_0 \right\rangle}{\left\langle \Psi_T \middle| \Psi_0 \right\rangle} = \frac{\left\langle \Psi_0 \middle| \hat{H}\Psi_T \right\rangle}{\left\langle \Psi_0 \middle| \Psi_T \right\rangle} =$$

$$\frac{\int dR \Psi_0(R) \hat{H}\Psi_T(R)}{\int dR \Psi_0(R) \Psi_T(R)} \approx \frac{\sum_{k=1}^{M} \hat{H}\Psi_T(R'_k)}{\sum_{k=1}^{M} \Psi_T(R'_k)}$$

If M is sufficiently large this is a Monte Carlo average

Propagation of walkers

The propagation of the single walker is easily implemented for the Trotter-Suzuki propagator. In fact the functions which are actually propagated are just Delta functions.

KINETIC PROPAGATOR

Displace the walker sampling from the Gaussian
$$\frac{1}{\sqrt{2\pi\hbar/m\,\Delta\tau}}\,e^{-\frac{(R-R')^2}{2\hbar/m\Delta\tau}}$$

(this can be easily achieved by using e.g. the Box-Muller formula)

$$R_1 - R_1' = \sqrt{\frac{\Delta\tau}{\hbar}\left(\frac{\hbar^2}{m}\right)} \times \text{Normally distributed random number}$$
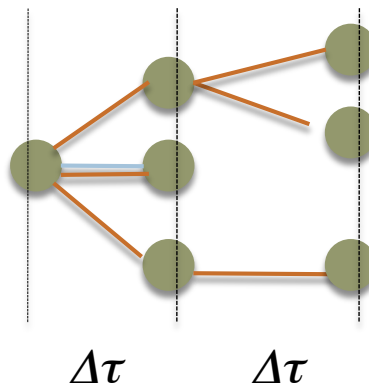
Propagation of walkers

POTENTIAL PROPAGATOR

The propagator is interpreted as the "weight" the walker will have after an imaginary time interval $\Delta\tau$. This can be naively interpreted as the probability of the walkers of "surviving" or "procreating" at that position.
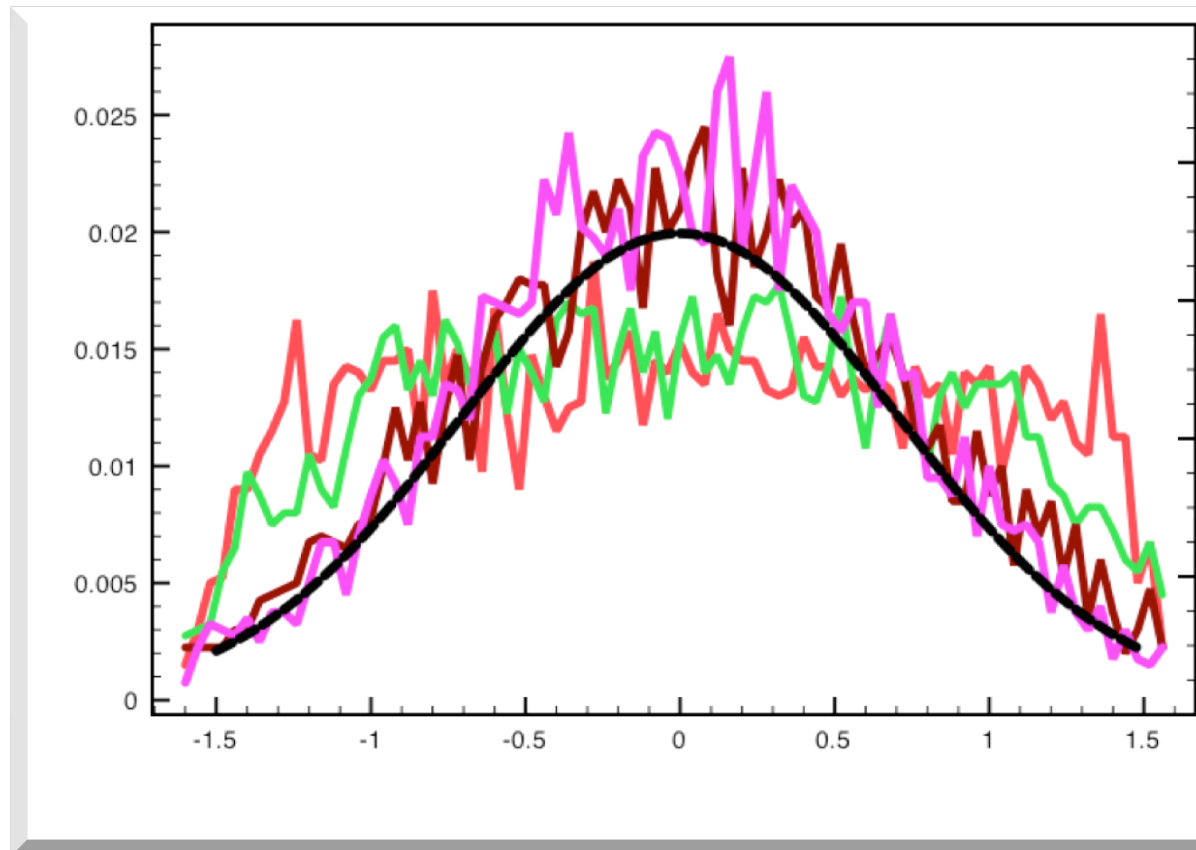
$$w = e^{-\left[\frac{V(R)+V(R')}{2}-E_0\right]\frac{\Delta\tau}{\hbar}}$$

Number of walkers $=$
$\text{INT}(w + \xi), \; \xi \in (0,1]$ random

"BRANCHING" PROCESS

Each walker generates descendants that can multiply in turn or die off. Obviously each walker is propagated with the Gaussian kernel at each time step.

$\Delta\tau$ $\Delta\tau$

Example

$$\hat{H} = -\frac{1}{2}\frac{\partial^2}{\partial x^2} + \frac{1}{2}x^2$$

Histograms of the distribution of walkers after 1, 10, 100, 1000, 5000 steps of 0.01 units

Normalization

What about the normalization of the state?
A possible way of fixing the energy shift E_0 is to attempt to maintain the fluctuations of the number of walkers under control.

This can be achieved by artificially modifying the weight of the walkers in order to react to population fluctuations.

$$\hbar = 1 \qquad w = \frac{N_T}{N} e^{-\left[\frac{V(R)+V(R')}{2} - E_0\right]\Delta\tau} = e^{-\left[\frac{V(R)+V(R')}{2} - \tilde{E}\right]\Delta\tau}$$

where

$$\tilde{E} = E_0 + \frac{1}{\Delta\tau}\ln\left(\frac{N_T}{N}\right)$$

The trial eigenvalue is continuosly updated.

Importance sampling

The algorithm as it was shown so far, does not work for potentials presenting a divergent behavior.

• A strongly repulsive potential will result in a very fast absorption of walkers, eventually killing the whole population. (e.g. Repulsive Coulomb potential, Lennard-Jones, N-N hard core…)

• An attractive potential will generate an exponentially growing population (e.g. Coulomb attraction between the nucleus and electrons in an atom)

Importance sampling

The problem can be addressed by changing the target distribution of walker into the product of the propagated state with a trial wave function determined, for instance, by means of a variational calculation.

Let us define a pseudoforce

$$F(R) = 2\nabla \ln \Psi_T(R) = -2\frac{\nabla \Psi_T(R)}{\Psi_T(R)}$$

and the density of walkers

$$f(R,\tau) = \Psi_T(R)\Psi(R,\tau)$$

Importance sampling

We can notice that:

$$\nabla\left[f(R,\tau)F(R)\right] = 2\nabla\Psi_T(R)\nabla\Psi(R,\tau) + 2\Psi(R,\tau)\nabla^2\Psi_T(R)$$

$$\nabla^2 f(R,\tau) = \Psi_T(R)\nabla^2\Psi(R,\tau) + 2\nabla\Psi_T(R)\nabla\Psi(R,\tau) + \Psi(R,\tau)\nabla^2\Psi_T(R)$$

By combining the two expressions we obtain:

$$\Psi_T(R)\nabla^2\Psi(R,\tau) = \nabla^2 f(R,\tau) - \nabla\left[f(R,\tau)F(R)\right] + f(R,\tau)\frac{\nabla^2\Psi_T(R)}{\Psi_T(R)}$$

Importance sampling

Let us multiply the Schroedinger equation in imaginary time by the importance function

$$\Psi_T(R)\left[-\frac{\hbar^2}{2m}\nabla^2 + V(R)\right]\Psi(R,\tau) = -\Psi_T(R)\hbar\frac{\partial}{\partial\tau}\Psi(R,\tau)$$

By using the previous result we obtain

Fokker Planck equation

$$\hbar\frac{\partial f(R,\tau)}{\partial\tau} = \frac{\hbar^2}{2m}\sum_i \nabla^2 f(R_i,\tau) - \frac{\hbar^2}{2m}\sum_i \nabla\left[f(R_i,\tau)F(R_i)\right] - E_L f(R,\tau)$$

where $\quad E_L(R) = -\dfrac{\hbar^2}{2m}\sum_i\dfrac{\nabla^2\Psi_T(R_i)}{\Psi_T(R_i)} + V(R) \quad$ *is the local energy.*

Importance sampling

We can repeat the same analysis (zero and infinite mass limits). The outcome are the following propagators:

Kinetic propagator

$$\tilde{G}_0(R,R',\Delta\tau) \propto e^{-\left[\dfrac{\left[R-R'-F(R')\frac{\Delta\tau}{\hbar}\left(\frac{\hbar^2}{m}\right)\right]^2}{2\frac{\Delta\tau}{\hbar}\left(\frac{\hbar^2}{m}\right)}\right]}$$

$E_L(R)$

Drives the walkers towards regions o high probability

Potential propagator

$$w = e^{-\left[\frac{E_L(R)+E_L(R')}{2}-E_0\right]\frac{\Delta\tau}{\hbar}}$$

The presence of the local energy strongly reduces fluctuations. If $\Psi_T(R)$ is exact, and E_0 is exact, then $w=1$

THE RESULTING ALGORITHM IS EXTREMELY EFFICIENT

Importance sampling

It is possible to overcame these problems by modifying the propagator.
We start from the expression:

$$\psi(R, \Delta\tau) = \int G(R, R', \Delta\tau)\psi(R', 0)\, dR'$$

where G is approximated by

From now on : $\hbar = m = 1$

$$G(R, R', \Delta\tau) \approx \frac{1}{\sqrt{(2\pi\Delta\tau)^{3N}}} e^{-\frac{(R-R')^2}{2\Delta\tau}} e^{-[V(R')-E_T]\Delta\tau}$$

τ = Inverse of an energy

Importance sampled GF

Let us multiply the propagated wavefunction by an (variational) approximation of the ground state wavefunction $\psi_T(R)$:

$$\psi_T(R)\psi(R,\Delta\tau) = \int G(R,R',\Delta\tau)\psi_T(R)\psi(R',0)dR'$$

which can in turn be rewritten as:

$$\psi_T(R)\psi(R,\Delta\tau) = \int G(R,R',\Delta\tau)\frac{\psi_T(R)}{\psi_T(R')}\psi_T(R')\psi(R',0)dR'$$

This defined a new "dressed" propagator:

$$\tilde{G}(R,R',\Delta\tau) = G_0(R,R',\Delta\tau)\frac{\Psi_T(R)}{\Psi_T(R')}$$

Importance sampled GF

The smearing of the initial density (which is a sum of delta functions) made by the propagator does not conserve in general the normalization of the wavefunction. We should then compute it as the integral over all the possible arrival points starting from a given point R'.

$$N(R') = \int G(R,R',\Delta\tau)\frac{\psi_T(R)}{\psi_T(R')}dR$$

Expanding the ratio of the importance functions (remembering that R-R' is order $\Delta\tau$)

$$N(R') \cong \int G(R,R'\Delta\tau)\left[1+\frac{\nabla\psi_T(R')}{\psi_T(R')}(R-R')+\right.$$

$$\left.+\frac{1}{\psi_T(R')}\sum_{i,j,\alpha,\beta}\frac{\partial^2\psi_T(R')}{\partial x_{i\alpha}\partial x_{j\beta}}(x_{i\alpha}-x'_{i\alpha})(x_{j\beta}-x'_{j\beta})\right]dR$$

Importance sampled GF

Integrating one obtains:

$$N(R') \cong \left[1 + \frac{\Delta\tau}{2} \frac{1}{\psi_T(R')} \sum_i \nabla^2 \psi_T(R') \right] e^{-\Delta\tau[V(R') - E_T]} \cong$$

$$\cong \exp\left[\left(\frac{1}{2} \frac{\nabla^2 \psi_T(R')}{\psi_T(R')} + E_T - V(R') \right) \Delta\tau \right] \cong$$

$$\cong \exp\left[-\Delta\tau \left(\frac{H\psi_T(R')}{\psi_T(R')} - E_T \right) \right]$$

If the trial wavefunction *satisfies the cusp condition* this **is always bound**!!!

Note that we could sample the dressed propagator by using the simple Gaussian propagator and using the ratio of the importance functions as a probability of acceptance rejection. *In some cases (e.g. in nuclear physics calculations) this might be a better procedure!*
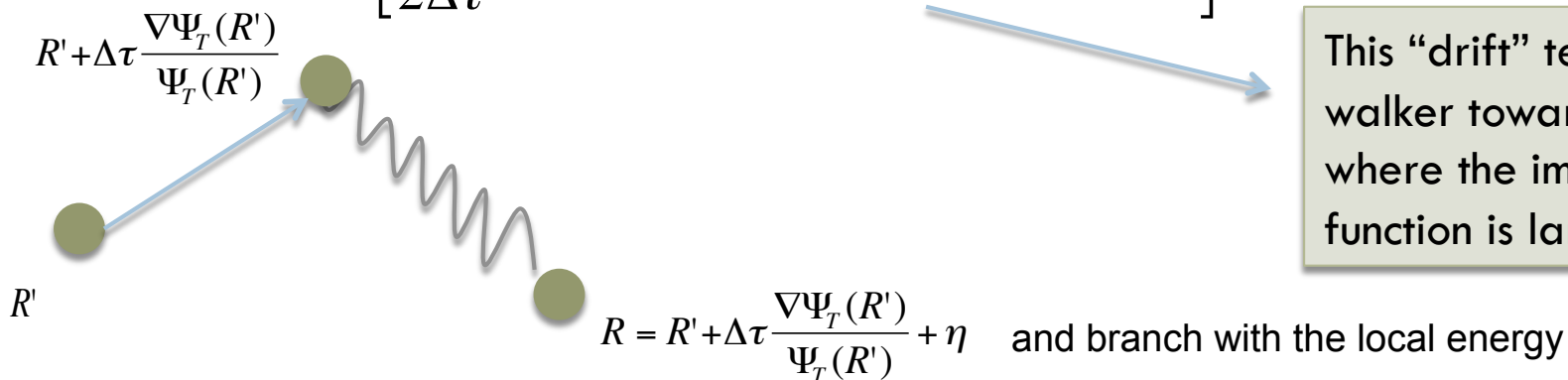
Importance sampled GF

Expanding the ratio inside the integral we can see that importance sampling naturally leads to the appearance of a "drift" term:

$$G(R,R',\Delta\tau) \cong \left[1 + \frac{\nabla\psi_T(R')}{\psi_T(R')}(R-R')\right]\exp\left[-\frac{(R-R')^2}{2\Delta\tau}\right] \cong$$

$$\cong \exp\left[\nabla\log\psi_T(R')(R-R') - \frac{(R-R')^2}{2\Delta\tau}\right] \cong$$

$$\cong \exp\left[\frac{1}{2\Delta\tau}\left(R-R'-\nabla\log\psi_T(R')\Delta\tau\right)^2 + o(\Delta\tau)\right]$$

Same order as the error made when applying the Trotter-Suzuki breakup

This "drift" term drives a walker towards points where the importance function is large.

$$R'+\Delta\tau\frac{\nabla\Psi_T(R')}{\Psi_T(R')}$$

$$R'$$

$$R = R'+\Delta\tau\frac{\nabla\Psi_T(R')}{\Psi_T(R')} + \eta \quad \text{and branch with the local energy}$$

Estimators

Importance sampling does not affect the capability of computing the exact ground state eigenvalue of the Schroedinger equation.

$$\frac{\int \Psi(R,\tau)\hat{H}\Psi_T(R)dR}{\int \Psi(R,\tau)\Psi_T(R)dR} = \frac{\int \Psi_T(R)\Psi(R,\tau)\dfrac{\hat{H}\Psi_T(R)}{\Psi_T(R)}dR}{\int \Psi(R,\tau)\Psi_T(R)dR} =$$

$$= \frac{\int f(R,\tau)E_L(R)dR}{\int f(R,\tau)dR} = \frac{1}{M}\sum_{k=1}^{M}E_L(R_k)$$

On the other hand, the above integral converges to:

$$\frac{\int \Psi(R,\tau)\hat{H}\Psi_T(R)dR}{\int \Psi(R,\tau)\Psi_T(R)dR} \xrightarrow{\tau\to\infty} \frac{\left\langle \Psi_0(R)\middle|\hat{H}\Psi_T(R)\right\rangle}{\left\langle \Psi_0(R)\middle|\Psi_T(R)\right\rangle} = \frac{\left\langle \Psi_T(R)\middle|\hat{H}\Psi_0(R)\right\rangle}{\left\langle \Psi_T(R)\middle|\Psi_0(R)\right\rangle} = E_0$$

Estimators

All estimators other than the energy will be evaluated as matri elements between the exact and the importance function. There are many techniques to correct this problem.

However, let us suppose that the importance function is a good approximation of the ground state, i.e.

$$\Psi_T(R) = \Psi_0(R) + \varepsilon\Psi_R(R)$$

then:

$$\left\langle \Psi_T(R) \middle| O(R)\Psi_0(R) \right\rangle = \left\langle \Psi_0(R) + \varepsilon\Psi_R(R) \middle| O(R)\Psi_0(R) \right\rangle$$

$$\left\langle \Psi_T(R) \middle| O(R)\Psi_T(R) \right\rangle = \left\langle \Psi_0(R) + \varepsilon\Psi_R(R) \middle| O(R)\left[\Psi_0(R) + \varepsilon\Psi_R(R)\right] \right\rangle =$$

$$= \left\langle \Psi_0(R) \middle| O(R)\Psi_0(R) \right\rangle + 2\varepsilon\left\langle \Psi_T(R) \middle| O(R)\Psi_R(R) \right\rangle$$

Estimators

Combining the previous two expressions we get:

$$2\langle \Psi_T(R) | O(R) \Psi_0(R) \rangle - \langle \Psi_T(R) | O(R) \Psi_T(R) \rangle = \langle \Psi_0(R) | O(R) \Psi_0(R) \rangle + o(\varepsilon^2)$$

Therefore, if we combine the result of a DMC calculation with the result of the variational calculation, we can obtain a better estimate of quantities other than the energy.

Biases

There are additional issues to be considered in calculations. In particular, the results are biased by:

• The Trotter-Suzuki breakup is not exact for any finite imaginary time step. It is therefore necessary to extrapolate to $\Delta\tau \rightarrow 0$.

• The population control quenches necessary fluctuations in the number of walkers. It is therefore necessary to extrapolate for $N_{walkers} \rightarrow \infty$