Quantum Monte Carlo Methods

Basic principles and algorithm

Yu Zhai

Institute of Theoretical Chemistry, Jilin University
yuzhai@mail.huiligroup.org

September 27, 2016

Outline

Outline

Local Energy

Energy can be expressed as

$$\begin{aligned} E &= \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \\ &= \frac{\int \Psi^*(\mathbf{R}) \hat{H} \Psi(\mathbf{R}) \, \mathrm{d}\, \mathbf{R}}{\int \Psi^*(\mathbf{R}) \Psi(\mathbf{R}) \, \mathrm{d}\, \mathbf{R}}, \end{aligned} \tag{1}$$

assuming that $\Psi(\mathbf{R})$ is a real function, we got

$$E = \frac{\int \Psi(\mathbf{R}) \hat{H} \Psi(\mathbf{R}) \, \mathrm{d}\, \mathbf{R}}{\int \Psi(\mathbf{R})^2 \, \mathrm{d}\, \mathbf{R}}. \tag{2}$$

Local Energy

Let us do some transformation

$$\begin{aligned} E &= \frac{\int \Psi(\mathbf{R})\Psi(\mathbf{R})[\Psi(\mathbf{R})^{-1}\hat{H}\Psi(\mathbf{R})]\,\mathrm{d}\,\mathbf{R}}{\int \Psi(\mathbf{R})^2\,\mathrm{d}\,\mathbf{R}} \\ &= \frac{\int \Psi(\mathbf{R})^2 E_{\mathrm{L}}(\mathbf{R})\,\mathrm{d}\,\mathbf{R}}{\int \Psi(\mathbf{R})^2\,\mathrm{d}\,\mathbf{R}}, \end{aligned} \tag{3}$$

we define $E_{\mathrm{L}}(\mathbf{R})$ as *local energy*, which is a constant when Ψ is an eigenfunction of $\hat{H}$.

Variational Monte Carlo

Actually a method using Variational Method, Monte Carlo Method is just a way to do the integral.

$$\begin{aligned} E &= \frac{\int \Psi(\mathbf{R})^2 E_{\mathrm{L}}(\mathbf{R})\,\mathrm{d}\,\mathbf{R}}{\int \Psi(\mathbf{R})^2\,\mathrm{d}\,\mathbf{R}} \\ &= \int \rho(\mathbf{R}) E_{\mathrm{L}}(\mathbf{R})\,\mathrm{d}\,\mathbf{R} \\ &= \sum_{m=1}^{M} E_{\mathrm{L}}(\mathbf{R}_m)/M, \end{aligned} \tag{4}$$

where the sampling is controlled by Metropolis algorithm

$$A(\mathbf{R} \leftarrow \mathbf{R}') = \min\left(1, \frac{\rho(\mathbf{R})}{\rho(\mathbf{R}')}\right). \tag{5}$$

VMC is actually a kind of variational methods.

Outline

Diffusion

Let the density of a kind of Brownian particles be $\rho(\mathbf{R}, t)$. The *diffusion equation* of them reads

$$\frac{\partial \rho(\mathbf{R})}{\partial t} = \gamma \nabla^2 \rho(\mathbf{R}). \tag{6}$$

Here I do not give the prrof of this equation, but describe how to perform the simulation of diffusion.

Green's function for the diffusion equation

Green's function is used to describe the evolution of the system.

$$\rho(\mathbf{R},t)=\int \mathrm{d}\,\mathbf{R}'\,G(\mathbf{R}\leftarrow\mathbf{R}';t)\rho(\mathbf{R}',t=0), \tag{7}$$

Total number of the Brownian particles is a constant.

$$G(\mathbf{R}\leftarrow\mathbf{R}';0)=\delta(|\,\mathbf{R}-\mathbf{R}'\,|), \tag{8}$$

$$\int \mathrm{d}\,\mathbf{R}'\,G(\mathbf{R}\leftarrow\mathbf{R}';0)=1. \tag{9}$$

We define the Green's function as

$$G(\mathbf{R}\leftarrow\mathbf{R}';0)=\frac{1}{\sqrt{4\pi\gamma t}}\exp[-(\mathbf{R}-\mathbf{R}')^2/(4\gamma t)]. \tag{10}$$

Schrödinger Equation

Free particle, which means there is no potential energy in the system, can be describe by the following equation

$$i\hbar\frac{\partial\Psi(\mathbf{R},t)}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\Psi(\mathbf{R},t), \tag{11}$$

and let $t = -i\tau$ to get

$$\frac{\partial\Psi(\mathbf{R},\tau)}{\partial \tau} = \frac{\hbar^2}{2m}\nabla^2\Psi(\mathbf{R},\tau). \tag{12}$$

The imaginary-time Schrödinger equation is equivalent to diffusion equation [Eq.(6)].

Schrödinger Equation

For practical sysytem, potential energy is included in the Hamiltonian, of which the imaginary-time Schrödinger equation is

$$\frac{\partial \Psi(\mathbf{R},\tau)}{\partial \tau} = [\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{R})]\Psi(\mathbf{R},\tau). \tag{13}$$

The Green's function of this system is

$$\begin{aligned} G(\mathbf{R} \leftarrow \mathbf{R}';\tau) =& \langle \mathbf{R} \,|\, \exp[-\tau(\hat{H} - E_{\mathrm{T}})] \,|\, \mathbf{R}' \rangle \\ =& \langle \mathbf{R} \,|\, \exp[-\tau(\hat{T} + \hat{V} - E_{\mathrm{T}})] \,|\, \mathbf{R}' \rangle \end{aligned}$$

Schrödinger Equation

For practical sysytem, potential energy is included in the Hamiltonian, of which the imaginary-time Schrödinger equation is

$$\frac{\partial \Psi(\mathbf{R},\tau)}{\partial \tau} = [\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{R})]\Psi(\mathbf{R},\tau). \tag{13}$$

The Green's function of this system is

$$\begin{aligned} G(\mathbf{R} \leftarrow \mathbf{R}';\tau) =& \langle \mathbf{R} \,|\, \exp[-\tau(\hat{H} - E_{\mathrm{T}})] \,|\, \mathbf{R}' \rangle \\ =& \langle \mathbf{R} \,|\, \exp[-\tau(\hat{T} + \hat{V} - E_{\mathrm{T}})] \,|\, \mathbf{R}' \rangle \end{aligned}$$

with τ small enough

Schrödinger Equation

For practical sysytem, potential energy is included in the Hamiltonian, of which the imaginary-time Schrödinger equation is

$$\frac{\partial \Psi(\mathbf{R},\tau)}{\partial \tau} = [\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{R})]\Psi(\mathbf{R},\tau). \tag{13}$$

The Green's function of this system is

$$\begin{aligned} G(\mathbf{R} \leftarrow \mathbf{R}';\tau) =& \langle \mathbf{R} \,|\, \exp[-\tau(\hat{H} - E_{\mathrm{T}})] \,|\, \mathbf{R}' \rangle \\ =& \langle \mathbf{R} \,|\, \exp[-\tau(\hat{T} + \hat{V} - E_{\mathrm{T}})] \,|\, \mathbf{R}' \rangle \end{aligned}$$

with τ small enough

$$\begin{aligned} \approx& \exp[-\tau(V(\mathbf{R}) - E_{\mathrm{T}})/2] \langle \mathbf{R} \,|\, \exp[-\tau(\hat{T})] \,|\, \mathbf{R}' \rangle \\ & \times \exp[-\tau(V(\mathbf{R}') - E_{\mathrm{T}})/2] \\ \approx& (2\pi\tau)^{-3N/2} \exp[-(\mathbf{R}-\mathbf{R}')^2/(2\tau)] \\ & \times \exp[-\tau[V(\mathbf{R}) + V(\mathbf{R}') - 2E_{\mathrm{T}}]/2]. \end{aligned} \tag{14}$$

DMC leads to ground state

Any initial wave function can be expended in the eigenfunction of Hamiltonian

$$\Psi(\mathbf{R},\tau) = \sum_n c_n \exp(-E_n\tau/\hbar)\psi_n(\mathbf{R}). \tag{15}$$

The wave function damp as $\tau \to \infty$, but the higher energylevels damp faster. Let us add a bias energy to the system which is a constant everywhere to make the energy of ground state zero and we get

$$\Psi_0(R) = \lim_{\tau\to\infty} \Psi(\mathbf{R},\tau) = \psi_0(\mathbf{R}), \tag{16}$$

and the bias energy is the ground state energy.

Branching algorithm

1. Intialization Set the imaginary time step $\delta\tau$, the initial total number of walkers N_{T}, which is also the target, and the energy offset E_{T}.
2. Time step
   1. Diffusion Step Shift the walkers by a distance which is controlled by Gaussian distribution

$$p(x)\,\mathrm{d}\,x = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-x^2/2\sigma^2)\,\mathrm{d}\,x. \tag{17}$$

   Notice that the distrance $\delta\,\mathbf{R}$ should be proportional to $\delta\tau$.
   2. Branching Step *To be or not to be—that is the question.* The number of a single walker after branching is determined by

$$\lfloor e^{-\delta\tau[V(\mathbf{R})+V(\mathbf{R}')-2E_{\mathrm{T}}]/2} + \mathrm{rand}[0,1]\rfloor. \tag{18}$$

Branching algorithm

3. Adjusting E_{T} Higher E_{T} leads to more walkers[Notice that the handout of University of Buffalo is wrong in this part]. Hence

$$E_{\mathrm{T}} \leftarrow E_{\mathrm{T}} + \alpha \ln \frac{N_{\mathrm{T}}}{N}, \tag{19}$$

where α is a small positive number. Or

$$E_{\mathrm{T}} \leftarrow (E_{\mathrm{T}} + \langle E_{\mathrm{L}} \rangle)/2, \tag{20}$$

or

$$E_{\mathrm{T}} \leftarrow \langle E_{\mathrm{L}} \rangle + \alpha \ln \frac{N_{\mathrm{T}}}{N}, \tag{21}$$

Branching algorithm

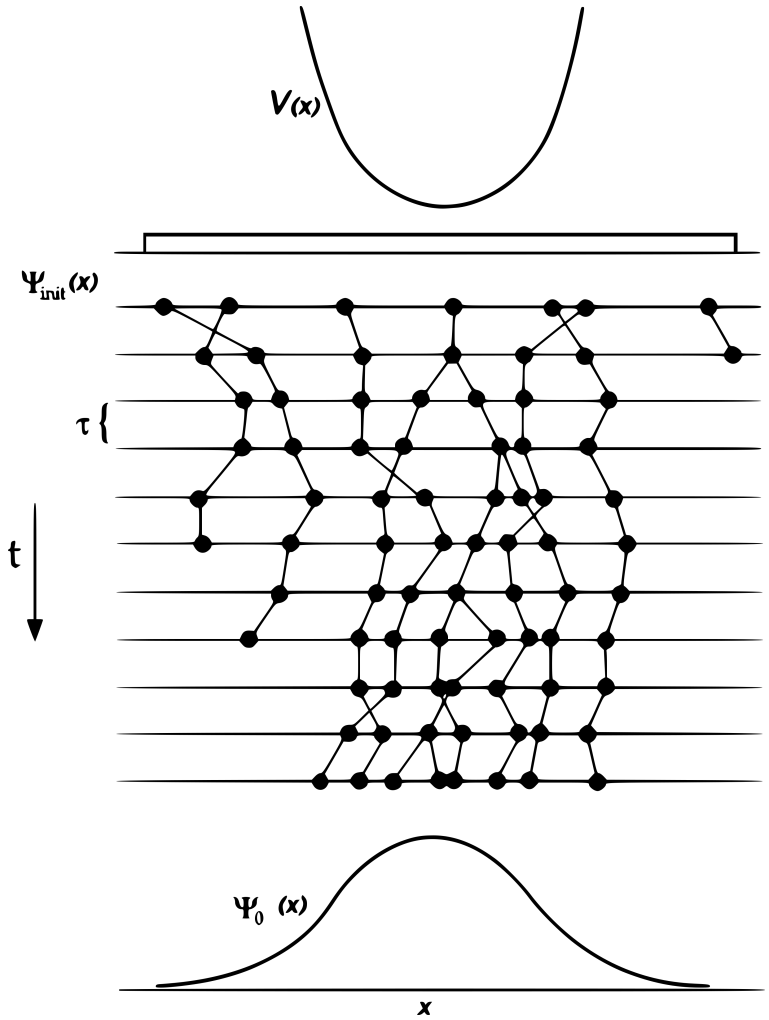

Figure: The branching algorithm.

Outline

Importance sampling

If we already have got a trial wave function $\Psi_{\mathrm{T}}(\mathbf{R})$ which is the approximation of the accurate wave function, we can get a new function

$$f(\mathbf{R},\tau) = \Phi(\mathbf{R},\tau)\Psi_{\mathrm{T}}(\mathbf{R}). \tag{22}$$

Multiply the imainary-time Schrödinger equation by $\Psi_{\mathrm{T}}(\mathbf{R})$ and we can get

$$-\frac{\partial}{\partial\tau}f(\mathbf{R},\tau) = -\frac{1}{2}\nabla^2 f(\mathbf{R},\tau)+\nabla\cdot[\mathbf{v}_{\mathrm{D}}(\mathbf{R})f(\mathbf{R},\tau)]+[E_{\mathrm{L}}(\mathbf{R})-E_{\mathrm{T}}]f(\mathbf{R},\tau), \tag{23}$$

where the *drift velocity* $\mathbf{v}_{\mathrm{D}}(\mathbf{R})$ is defined as

$$\mathbf{v}_{\mathrm{D}}(\mathbf{R}) = \nabla \ln|\Psi_{\mathrm{T}}(\mathbf{R})| = \Psi_{\mathrm{T}}(\mathbf{R})^{-1}\nabla\Psi_{\mathrm{T}}(\mathbf{R}). \tag{24}$$

Green's function

To perform the sampling, again, we need the Green's function for the evolution of function $f(\mathbf{R}, \tau)$. We have

$$f(\mathbf{R}, \tau) = \int \tilde{G}(\mathbf{R} \leftarrow \mathbf{R}', \tau) f(\mathbf{R}, \tau) d\,\mathbf{R}', \tag{25}$$

and in *short-time approximation*

$$\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}', \tau) \approx G_{\mathrm{d}}(\mathbf{R} \leftarrow \mathbf{R}', \tau) G_{\mathrm{b}}(\mathbf{R} \leftarrow \mathbf{R}', \tau), \tag{26}$$

where

$$G_{\mathrm{d}}(\mathbf{R} \leftarrow \mathbf{R}', \tau) = (2\pi\tau)^{-3N/2} \exp[-(\mathbf{R} - \mathbf{R}' - \tau \mathbf{v}_{\mathrm{D}}(\mathbf{R}'))^2/(2\tau)], \tag{27}$$

and

$$G_{\mathrm{b}}(\mathbf{R} \leftarrow \mathbf{R}', \tau) = \exp[-\tau[E_{\mathrm{L}}(\mathbf{R}) + E_{\mathrm{L}}(\mathbf{R}') - 2E_{\mathrm{T}}]/2]. \tag{28}$$

Advantages

- More effcient. Walkers will concentrate in those where Ψ_{T}^2 is large.
- Using local energy instead of potential. The fluctuations are much diminished.
- Helpful in fixed-node constraint.

Improvement

In real simulation, the "short-time approximation" version of Green's function is used. It means the Freen's function is reliable only when $\delta\tau \to 0$. Therefore, time step far less than 1 a. u. is used.
The *exact* Green's function fulfills the *detailed balance condition*, which is the foundation of Monte Carlo Method

$$\tilde{G}(\mathbf{R} \leftarrow \mathbf{R}', \tau)\Psi_{\mathrm{T}}(\mathbf{R}')^2 = \tilde{G}(\mathbf{R}' \leftarrow \mathbf{R}, \tau)\Psi_{\mathrm{T}}(\mathbf{R})^2. \tag{29}$$

To meet this requirement, a rejection step is employed in diffusion step

$$\begin{aligned} A(\mathbf{R} \leftarrow \mathbf{R}') &= \min[1, \frac{G_{\mathrm{d}}(\mathbf{R} \leftarrow \mathbf{R}', \tau)G_{\mathrm{b}}(\mathbf{R} \leftarrow \mathbf{R}', \tau)\Psi_{\mathrm{T}}(\mathbf{R}')^2}{G_{\mathrm{d}}(\mathbf{R}' \leftarrow \mathbf{R}, \tau)G_{\mathrm{b}}(\mathbf{R}' \leftarrow \mathbf{R}, \tau)\Psi_{\mathrm{T}}(\mathbf{R})^2}] \\ &= \min[1, \frac{G_{\mathrm{d}}(\mathbf{R} \leftarrow \mathbf{R}', \tau)\Psi_{\mathrm{T}}(\mathbf{R}')^2}{G_{\mathrm{d}}(\mathbf{R}' \leftarrow \mathbf{R}, \tau)\Psi_{\mathrm{T}}(\mathbf{R})^2}]. \end{aligned} \tag{30}$$

Improvement

The imaginary time step τ is substituted by

$$\tau_{\text{eff}} = \tau \frac{\langle A \Delta \mathbf{R}^2 \rangle}{\langle \Delta \mathbf{R}^2 \rangle}. \tag{31}$$

Branching algorithm

So we should adjust the algorithm.

1. Initialization + Trial function.
2. Time step
   1. Diffusion step $\mathbf{R} = \mathbf{R}' + \chi + \tau \mathbf{v}_{\mathrm{D}}(\mathbf{R}')$
   2. Branching step $\lfloor e^{-\delta\tau[E_{\mathrm{L}}(\mathbf{R})+E_{\mathrm{L}}(\mathbf{R}')-2E_{\mathrm{T}}]/2} + \mathrm{rand}[0,1] \rfloor$

E_{T} as an energy estimator

- Even in DMC with importance sampling, the variable is still the energy of Ψ_0. Hence, it can be used as a good energy estimator.
- However, this estimator has a serious problem that it is not *stable* enough. When the number of walker is more than the desired number, E_{T} get a little bit lower, but not low enengh to bring the walkers' number down.
- An improved method came to me and I think I can change the fluctuaction of E_{T} from low frency to high frequace, which reads

$$E_{\mathrm{T}} \leftarrow E_{\mathrm{T}} + \ln \frac{N_{\mathrm{T}}}{N} / \tau. \tag{32}$$

Why?

Just consider the way perform branching step. A not-so-good α in Eq.19 will lead to the result that the evolution of E_{T} too fast or slow for the change of number of walkers. Hence $1/\tau$ is used here to get the right α.

Fluctruation of E_{T}

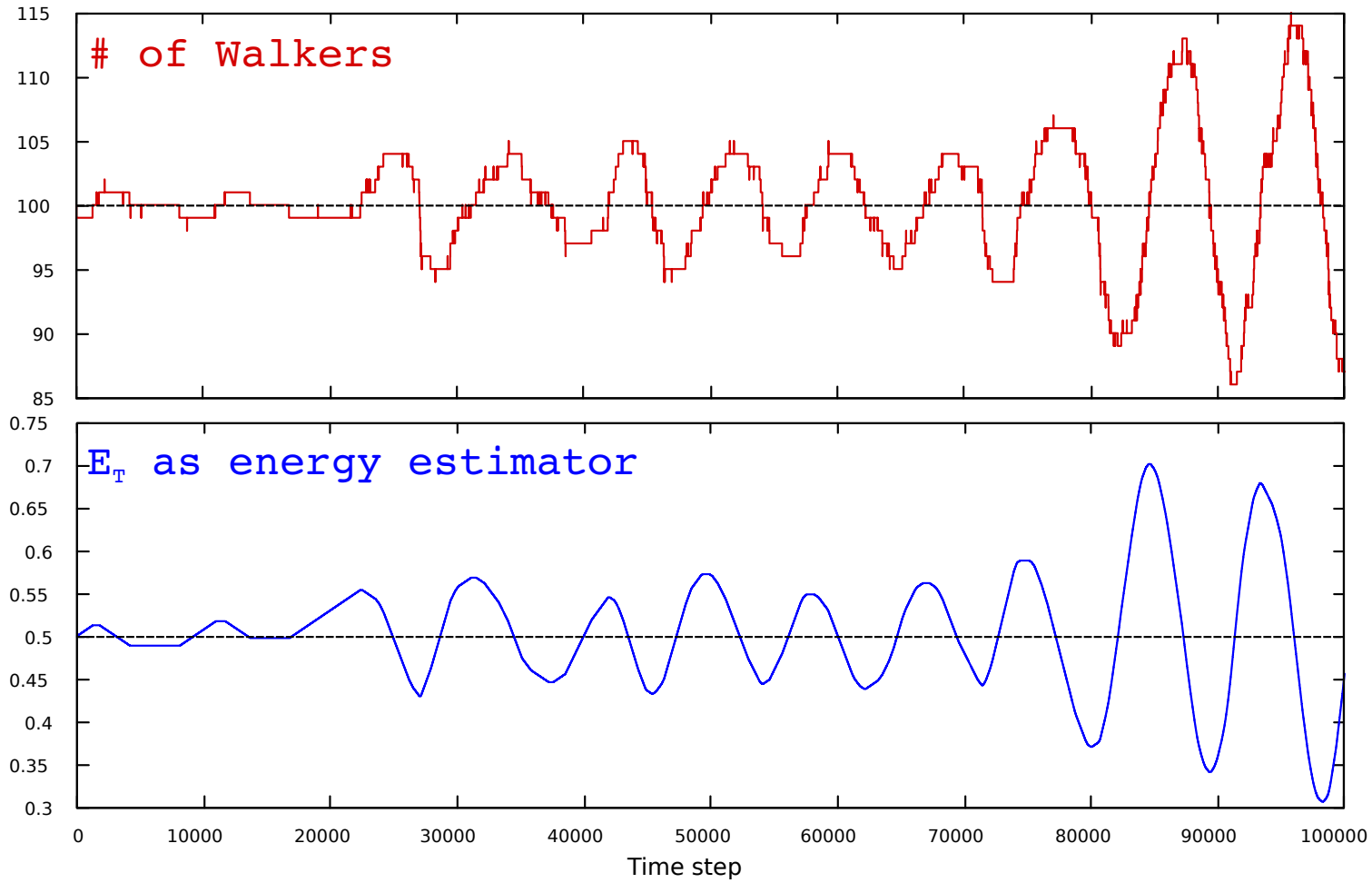

Figure: The fluctruation of E_{T}.

Distribution of E_{T}

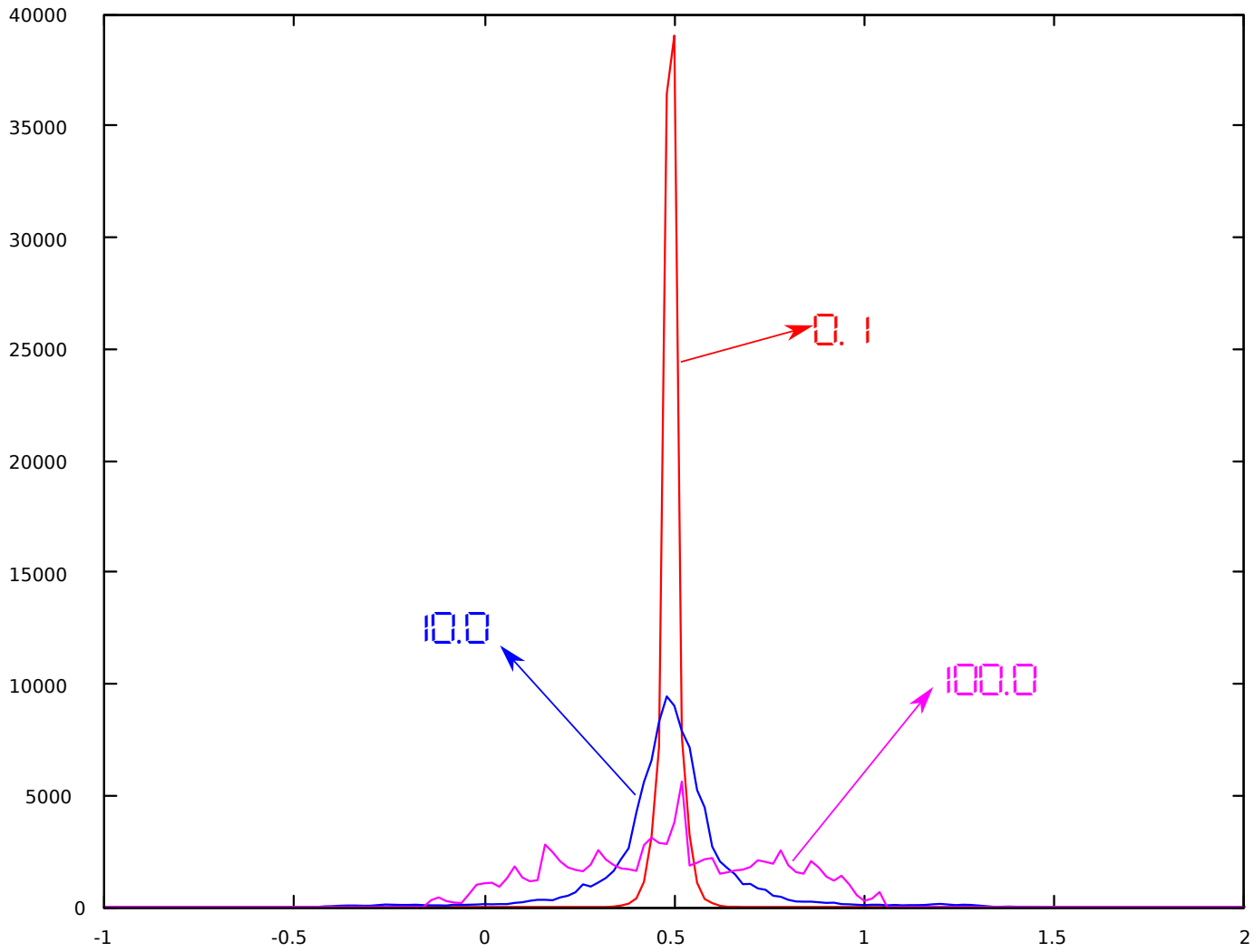

Figure: The histogram of E_{T}.

Mixed estimator

If $\hat{A}$ commutes with $\hat{H}$, we have

$$\frac{\langle \Psi_{\mathrm{T}}|\hat{A}|\Psi_0\rangle}{\langle \Psi_{\mathrm{T}}|\Psi_0\rangle} = \frac{\langle \Psi_0|\hat{A}|\Psi_0\rangle}{\langle \Psi_0|\Psi_0\rangle}. \tag{33}$$

Let $\hat{A} = \hat{H}\Psi_{\mathrm{T}}/\Psi_{\mathrm{T}}$, and we have

$$\begin{aligned} E_0 = \frac{\langle \Psi_0|\hat{H}|\Psi_0\rangle}{\langle \Psi_0|\Psi_0\rangle} &= \frac{\int \mathrm{d}\,\mathbf{R}\,\Psi_{\mathrm{T}}\hat{H}\Psi_{\mathrm{T}}/\Psi_{\mathrm{T}}\Psi_0}{\int \mathrm{d}\,\mathbf{R}\,\Psi_{\mathrm{T}}\Psi_0} = \frac{\int \mathrm{d}\,\mathbf{R}\,E_{\mathrm{L}}(\mathbf{R})f(\mathbf{R})}{\int \mathrm{d}\,\mathbf{R}\,f(\mathbf{R})} \\ &\approx \frac{1}{M}\sum_m E_{\mathrm{L}}(\mathbf{R}_m). \end{aligned} \tag{34}$$

Y.'s note

As I tested, the mixed estimator is highly dependent on the trial (guide) function. It means that the DMC simulation has to follow a VMC simulation or something to make sure that the trial function is as accurate as possible.

Extrapolated estimator

For observables not commuting with $\hat{H}$, *extrapolated estimator* is often used, which reads

$$\langle \Psi_0 | \hat{A} | \Psi_0 \rangle \approx 2 \langle \Psi_T | \hat{A} | \Psi_0 \rangle - \langle \Psi_T | \hat{A} | \Psi_T \rangle. \tag{35}$$

For non-negative quantities, it can also be

$$\langle \Psi_0 | \hat{A} | \Psi_0 \rangle \approx \langle \Psi_T | \hat{A} | \Psi_0 \rangle^2 / \langle \Psi_T | \hat{A} | \Psi_T \rangle. \tag{36}$$

Is this method robust?

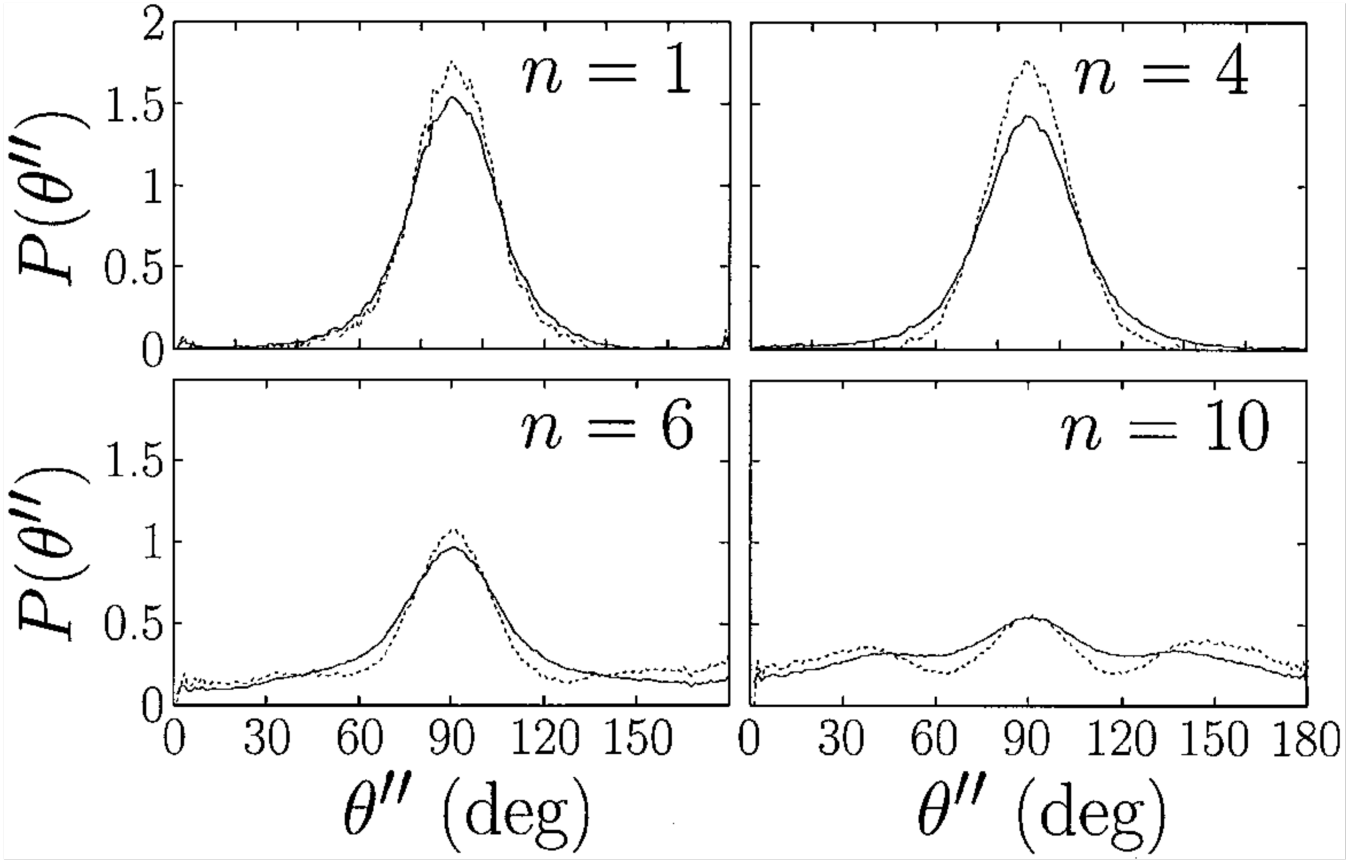

FIG. The mixed (full line) and extrapolated (dashed line) probability distribution functions $P(\theta'')$ for $He_n(HF)_2$, $n = 1, 4, 6, 10$.

Figure: Distribution of the He's, from *Phys.Rev.Lett.*, **88**, 123401 (2002).

Drawbacks?

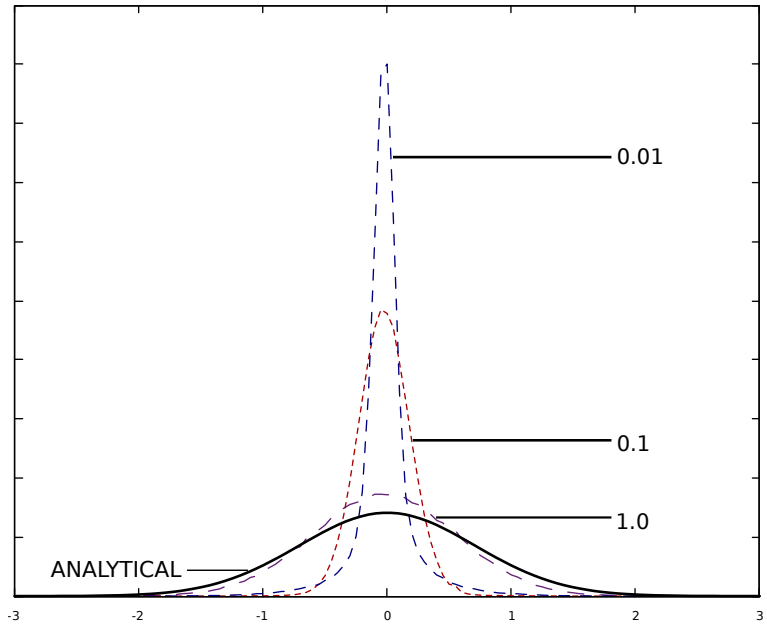

Figure: Harmonic Oscillator, $\Psi(x) = \pi^{-1/4} \cdot \exp(-x^2/2)$.

Y.'s note

As I tested, the distribution of the system fully depends on the size of the time step. If the time step is too small, something terrible will happen because of the shortcoming of the algorithm.

The end.

Thank you for your attention.