Introduction to Variational and Projector Monte Carlo

Cyrus Umrigar

Physics Department, Cornell University, Ithaca.

Email: CyrusUmrigar@cornell.edu

Solving the Many-Body Schrödinger Equation

Straightforward approach:

1. Expand the many-body wavefunction as a linear combination of (possibly nonorthogonal) basis states (determinants for Fermions).

2. Compute Hamiltonian and overlap matrices, H and S in this basis

3. Solve the generalized eigenvalue problem $Hc = ESc$

Problem:

The number of many-body states grows combinatorially in the number of single particle basis states and the number of particles, $\binom{N_{\rm orb}}{N_\uparrow} \times \binom{N_{\rm orb}}{N_\downarrow}$, e.g.

Molecules with 20 electrons in 200 orbitals: $\binom{200}{10}^2 = 5.0 \times 10^{32}$

(Partial) Solutions:

1. **DMRG:** Very efficient for low-dimensional problems. Steve White, Garnet Chan

2. **Selected CI:** If only a small fraction, say 10^{12} of these states are important, then one can use smart methods for finding the most important say 10^9 states, diagonalizing and then include rest of 10^{12} states using perturbation theory.

3. **Quantum Monte Carlo:** Applicable to large finite Hilbert spaces as well as infinite Hilbert spaces!

Cyrus J. Umrigar

What is Quantum Monte Carlo?

Stochastic implementation of the power method for projecting out the dominant eigenvector of a matrix or integral kernel.

"Dominant state" means state with largest absolute eigenvalue.

If we repeatedly multiply an arbitrary vector, not orthogonal to the dominant state, by the matrix, we will eventually project out the dominant state.
Power method is an iterative method for eigenvalue problems (less efficient than Lanczos or Davidson). However, *stochastic* power method, QMC, is powerful.

QMC methods are used only when the number of states is so large ($> 10^{10}$) that it is not practical to store even a single vector in memory. Otherwise use exact diagonalization method, e.g., Lanczos or Davidson. At each MC generation, only a sample of the states are stored, and expectation values are accumulated.

QMC methods are used not only in a large discrete space but also in a continuously infinite space. Hence "matrix or integral kernel" above. In the interest of brevity I will use either discrete or continuous language (sums and matrices or integrals and integral kernels), but much of what is said will apply to both situations.

Cyrus J. Umrigar

When to use Monte Carlo Methods

Monte Carlo methods: A class of computational algorithms that rely on repeated random sampling to compute results.

A few broad areas of applications are:

1. physics
2. chemistry
3. engineering
4. finance and risk analysis

When are MC methods likely to be the methods of choice?

1. When the problem is many-dimensional and approximations that factor the problem into products of lower dimensional problems are inaccurate.
2. A less important reason is that if one has a complicated geometry, a MC algorithm may be simpler than other choices.

Obvious drawback of MC methods: There is a statistical error.

Frequently there is a tradeoff between statistical error and systematic error (needed to overcome *sign problem*), so need to find the best compromise.

MC Simulations versus MC calculations

One can distinguish between two kinds of algorithms:

1. The system being studied is stochastic and the stochasticity of the algorithm mimics the stochasticity of the actual system. e.g. study of neutron transport and decay in nuclear reactor by following the trajectories of a large number of neutrons. Such problems are suitable for MC algorithms in a very obvious way.

2. Much more interesting are applications where the system being studied is not stochastic, but nevertheless a stochastic algorithm is the most efficient, or the most accurate, or the only feasible method for studying the system. e.g. the solution of a PDE in a large number of variables, e.g., the solution of the Schrödinger equation for an N-electron system, with say $N = 100$ or 1000. (Note: The fact that the wavefunction has a probabilistic interpretation has *nothing* to do with the stochasticity of the algorithm. The wavefunction itself is perfectly deterministic.)

I prefer to use the terminology that the former are MC simulations whereas the latter are MC calculations but not everyone abides by that terminology.

Cyrus J. Umrigar

Early Recorded History of Monte Carlo

1777   Comte de Buffon: If a needle of length L is thrown at random onto a plane ruled with straight lines a distance $d(d > L)$ apart, then the probability P of the needle intersecting one of those lines is $P = \frac{2L}{\pi d}$.
Laplace: This could be used to compute π (inefficiently).

1930s   First significant scientific application of MC: Enrico Fermi used it for neutron transport in fissile material.
Segre: "Fermi took great delight in astonishing his Roman colleagues with his "too-good-to-believe" predictions of experimental results."

1940s   Monte Carlo named by Nicholas Metropolis and Stanislaw Ulam

1953   Algorithm for sampling any probability density
Metropolis, Rosenbluth, Rosenbluth, Teller and Teller
(generalized by Hastings in 1970)

1962,1974   First PMC calculations, Kalos, and, Kalos, Levesque, Verlet.
1965   First VMC calculations (of liquid He), Bill McMillan.

Compte de Buffon

I gave a series of lectures at the
University of Paris.
After my first lecture, my host,
Julien Toulouse, took me for a
short walk to the Jardin de Plantes

Compte de Buffon

I gave a series of lectures at the University of Paris.

After my first lecture, my host, Julien Toulouse, took me for a short walk to the Jardin de Plantes to meet Buffon!

Here he is:

Among other things, he wrote a 36 volume set of books on the Natural History of the Earth!
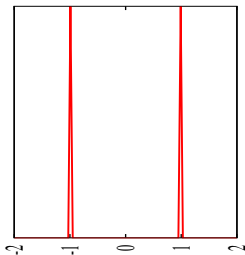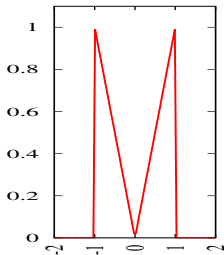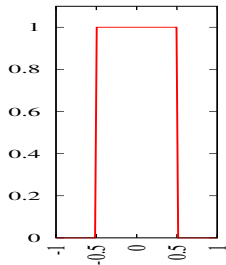
Central Limit Theorem

de Moivre (1733), Laplace (1812), Lyapunov (1901), Pólya (1920)

Let $X_1, X_2, X_3, \cdots, X_N$ be a sequence of N independent random variables sampled from a probability density function with a finite expectation value, μ, and variance σ^2. The central limit theorem states that as the sample size N increases, the probability density of the sample average, $\bar{X}$, of these random variables approaches the normal distribution,

$\sqrt{\frac{N}{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2/N)}$, with mean μ, and variance σ^2/N, irrespective of the original probability density function, e.g.:

The rate at which they converge will however depend on the original PDF.

Cyrus J. Umrigar

(Weak) Law of Large Numbers

Cardano, Bernouli, Borel, Cantelli, Kolmogorov, Khinchin

Let $X_1, X_2, X_3, \cdots, X_N$ be a sequence of N independent random variables sampled from a probability density function with a finite expectation value, μ, but not necessarily a finite variance σ^2. Then for any $\epsilon > 0$,

$$\lim_{N \to \infty} P(|\bar{X} - \mu| \geq \epsilon) = 0$$

However, the rate at which it converges may be very slow.
So, employ distributions with a finite variance whenever possible.

Lorentzian

Does the Central Limit Theorem or the Law of Large Numbers apply to a Lorentzian (also known as Cauchy) probability density function

$$L(x) \;=\; \frac{1}{\pi} \frac{1}{1 + x^2}?$$

Cyrus J. Umrigar

Lorentzian

A Lorentzian (also known as Cauchy) probability density function

$$L(x) \;=\; \frac{1}{\pi}\frac{1}{1+x^2}$$

not only violates the conditions for the Central Limit Theorem but also the conditions for the Law of Large Numbers, since not only the variance but even the mean is undefined.

$$\int_{-\infty}^{\infty} xL(x)dx \;=\; \left(\int_{-\infty}^{a} + \int_{a}^{\infty}\right) xL(x)dx$$
$$= \; -\infty + \infty$$

Averages over a Lorentzian have the same spread of values as the original values!

So, although the Lorentzian looks much "nicer" than the other 3 functions we showed, it is a problem!

Cyrus J. Umrigar

Chebychev Inequality

The Central Limit Theorem by itself does not tell you how quickly the averages converge to a Gaussian distribution.

If we have not averaged enough, for an arbitrary distribution with finite mean μ and finite variance σ^2, we have much weaker bounds given by Chebychev's inequality:

The probability of a variable lying between $\mu - n\sigma$ and $\mu + n\sigma$ is $> 1 - 1/n^2$, as compared to $\text{erf}(n/\sqrt{2})$ for a Gaussian.

Prob. of being within 1σ of μ is $\geq$ 0%   versus 68.3%   for Gaussian
Prob. of being within 2σ of μ is $\geq$ 75%   versus 95.4%   for Gaussian
Prob. of being within 3σ of μ is $\geq$ 89%   versus 99.7%   for Gaussian
Prob. of being within 4σ of μ is $\geq$ 94%   versus 99.994% for Gaussian

The worst case occurs for a distribution with probability $1 - 1/n^2$ at μ and probability $1/2n^2$ at $\mu - n\sigma$ and $\mu + n\sigma$.

What if the population variance $\sigma^2 = \infty$ but we do not know that beforehand? The computed sample variance will ofcourse always be finite. The practical signature of an infinite variance estimator is that the estimated σ increases with sample size, N and tends to have upward jumps. So the estimated error of the sample mean, $\sigma_N = \sigma/\sqrt{N}$, goes down more slowly than $\frac{1}{\sqrt{N}}$, or even does not go down at all.

Cyrus J. Umrigar

Monte Carlo versus Deterministic Integration methods

Deterministic Integration Methods:

Integration Error, ϵ, using N_{int} integration points:

1-dim Simpson rule: $\epsilon \leq cN_{\text{int}}^{-4}$,   (provided derivatives up to 4th exist)

d-dim Simpson rule: $\epsilon \leq cN_{\text{int}}^{-4/d}$,   (provided derivatives up to 4th exist)

This argument is correct for functions that are approximately separable.

Monte Carlo:

$\epsilon \sim \sigma(T_{\text{corr}}/N_{\text{int}})^{1/2}$, independent of dimension!, according to the central limit theorem since width of gaussian decreases as $(T_{\text{corr}}/N_{\text{int}})^{1/2}$ provided that the variance of the integrand is finite. (T_{corr} is the autocorrelation time.)

Very roughly, Monte Carlo becomes advantageous for $d > 8$.

For $d = 50$, even 2 grid points per dimensions gives $N_{\text{int}} \approx 10^{15}$, so deterministic integration not possible.

For a many-body wavefunction $d = 3N_{\text{elec}}$ and can be a few thousand!

Cyrus J. Umrigar

Monte Carlo Integration

$$I \;=\; \int_V f(x)dx \;=\; V\,\overline{f} \pm V\sqrt{\frac{\overline{f^2} - \overline{f}^2}{N-1}}$$

$$\text{where} \quad \overline{f} \;=\; \frac{1}{N}\sum_i^N f(x_i), \quad \overline{f^2} = \frac{1}{N}\sum_i^N f^2(x_i)$$

and the points x_i are sampled uniformly in V. Many points may contribute very little.

Monte Carlo Integration

$$I = \int_V f(x)dx = V\overline{f} \pm V\sqrt{\frac{\overline{f^2} - \overline{f}^2}{N-1}}$$

where $\quad \overline{f} = \frac{1}{N}\sum_i^N f(x_i), \quad \overline{f^2} = \frac{1}{N}\sum_i^N f^2(x_i)$

and the points x_i are sampled uniformly in V. Many points may contribute very little.

Importance sampling

$$I = \int_V g(x)\frac{f(x)}{g(x)}dx = \overline{\left(\frac{f}{g}\right)} \pm \sqrt{\frac{\overline{\left(\frac{f}{g}\right)^2} - \overline{\left(\frac{f}{g}\right)}^2}{N-1}}$$

where the probability density function $g(x) \geq 0$ and $\int_V g(x)dx = 1$.
If $g(x) = 1/V$ in V then we recover original fluctuations but if $g(x)$ mimics $f(x)$ then the fluctuations are much reduced. Optimal g is $|f|$. Need: a) $g(x) \geq 0$, b) know integral of $g(x)$, and, c) be able to sample it.
Importance sampling can turn an $\infty-$variance estimator into a finite variance one!

Cyrus J. Umrigar

Illustration of Importance Sampling

$f(x)$ is the function to be integrated. $g(x)$ is a function that is "similar" to $f(x)$ and has the required properties: a) $g(x) \geq 0$, b) $\int dx \, g(x) = 1$, and, c) we know how to sample it. $\int f(x)dx$ can be evaluated efficiently by sampling $g(x)$ and averaging $f(x)/g(x)$.

Cyrus J. Umrigar
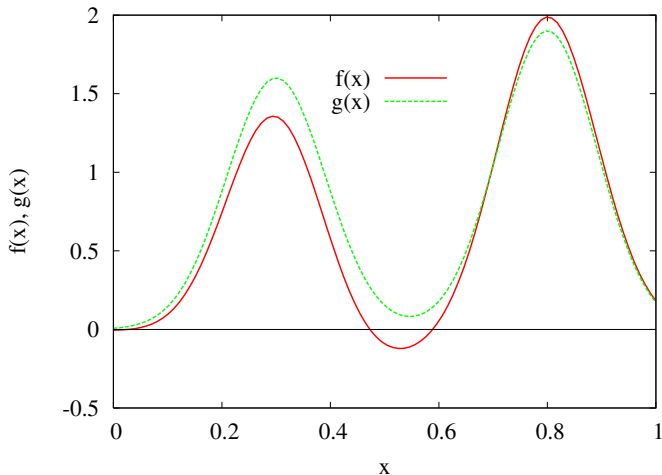
Quantum Monte Carlo in a Nutshell

Cyrus J. Umrigar

Definitions

Given a complete or incomplete basis: $\{|\phi_i\rangle\}$, either discrete or continuous

$$\text{Exact} \quad |\Psi_0\rangle = \sum_i e_i|\phi_i\rangle, \quad \text{where,} \quad e_i = \langle\phi_i|\Psi_0\rangle$$

$$\text{Trial} \quad |\Psi_T\rangle = \sum_i t_i|\phi_i\rangle, \quad \text{where,} \quad t_i = \langle\phi_i|\Psi_T\rangle$$

$$\text{Guiding} \quad |\Psi_G\rangle = \sum_i g_i|\phi_i\rangle, \quad \text{where,} \quad g_i = \langle\phi_i|\Psi_G\rangle$$

(If basis incomplete then "exact" means "exact in that basis".)

Ψ_T used to calculate variational and mixed estimators of operators $\hat{A}$, i.e.,
$$\langle\Psi_T|\hat{A}|\Psi_T\rangle/\langle\Psi_T|\Psi_T\rangle, \quad \langle\Psi_T|\hat{A}|\Psi_0\rangle/\langle\Psi_T|\Psi_0\rangle$$

Ψ_G used to alter the probability density sampled, i.e., Ψ_G^2 in VMC, $\Psi_G\Psi_0$ in PMC.

Ψ_G must be such that $g_i \neq 0$ if $e_i \neq 0$. If Ψ_T also satisfies this condition then Ψ_G can be chosen to be Ψ_T. Reasons to have $\Psi_G \neq \Psi_T$ are: a) rapid evaluation of "local energy", b) have finite-variance estimators. To simplify expressions, we sometimes use $\Psi_G = \Psi_T$ or $\Psi_G = 1$ in what follows.

Cyrus J. Umrigar

Variational MC

$$E_V = \frac{\langle \Psi_T | \hat{H} | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} = \frac{\sum_{ij}^{N_{st}} \langle \Psi_T | \phi_i \rangle \; \langle \phi_i | \hat{H} | \phi_j \rangle \; \langle \phi_j | \Psi_T \rangle}{\sum_i^{N_{st}} \langle \Psi_T | \phi_k \rangle \; \langle \phi_k | \Psi_T \rangle}$$

$$= \frac{\sum_{ij}^{N_{st}} t_i H_{ij} t_j}{\sum_k^{N_{st}} t_k^2} = \sum_i^{N_{st}} \frac{t_i^2}{\sum_k^{N_{st}} t_k^2} \; \frac{\sum_j^{N_{st}} H_{ij} t_j}{t_i}$$

$$= \sum_i^{N_{st}} \frac{t_i^2}{\sum_k^{N_{st}} t_k^2} \; E_L(i) \approx \frac{\left[\sum_i^{N_{MC}} E_L(i) \right]_{\Psi_T^2}}{N_{MC}} \xrightarrow{\Psi_G \neq \Psi_T} \frac{\left[\sum_i^{N_{MC}} \left(\frac{t_i}{g_i} \right)^2 E_L(i) \right]_{\Psi_G^2}}{\left[\sum_k^{N_{MC}} \left(\frac{t_k}{g_k} \right)^2 \right]_{\Psi_G^2}}$$

Sample probability density function $\frac{g_i^2}{\sum_k^{N_{st}} g_k^2}$ using Metropolis-Hastings, if Ψ_G complicated.

Value depends only on Ψ_T. Statistical error depend on Ψ_T and Ψ_G.

Energy bias and statistical error vanish as $\Psi_T \to \Psi_0$.

For fixed Ψ_T, $\Psi_G = \Psi_T$ does not minimize statistical fluctuations!

In fact $\Psi_G \neq \Psi_T$ needed when optim. to get finite variance.

$\Psi_G = \Psi_T$ allows simple unbiased estimator. Ratio of expec. val. $\neq$ expec. val. of ratios.

Cyrus J. Umrigar

Projector MC

$E_0 = \dfrac{\langle \Psi_0 | \hat{H} | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle} = \dfrac{\langle \Psi_0 | \hat{H} | \Psi_T \rangle}{\langle \Psi_0 | \Psi_T \rangle}$

Projector: $\quad |\Psi_0\rangle = \hat{P}(\infty) \, |\Psi_T\rangle = \lim_{n \to \infty} \hat{P}^n(\tau) \, |\Psi_T\rangle$

$$E_0 = \frac{\langle \Psi_0 | \hat{H} | \Psi_T \rangle}{\langle \Psi_0 | \Psi_T \rangle} = \frac{\sum_{ij}^{N_{\mathrm{st}}} \langle \Psi_0 | \phi_i \rangle \, \langle \phi_i | \hat{H} | \phi_j \rangle \, \langle \phi_j | \Psi_T \rangle}{\sum_k^{N_{\mathrm{st}}} \langle \Psi_0 | \phi_k \rangle \, \langle \phi_k | \Psi_T \rangle}$$

$$= \frac{\sum_{ij}^{N_{\mathrm{st}}} e_i H_{ij} t_j}{\sum_k^{N_{\mathrm{st}}} e_k \, t_k} = \sum_i^{N_{\mathrm{st}}} \frac{e_i t_i}{\sum_k^{N_{\mathrm{st}}} e_k \, t_k} \, \frac{\sum_j^{N_{\mathrm{st}}} H_{ij} t_j}{t_i}$$

$$= \sum_i^{N_{\mathrm{st}}} \frac{e_i t_i}{\sum_k^{N_{\mathrm{st}}} e_k \, t_k} \, E_{\mathrm{L}}(i) \approx \frac{\left[\sum_i^{N_{\mathrm{MC}}} E_{\mathrm{L}}(i) \right]_{\Psi_T \Psi_0}}{N_{\mathrm{MC}}} \xrightarrow{\Psi_G \neq \Psi_T} \frac{\left[\sum_i^{N_{\mathrm{MC}}} \left(\frac{t_i}{g_i} \right) E_{\mathrm{L}}(i) \right]_{\Psi_G \Psi_0}}{\left[\sum_k^{N_{\mathrm{MC}}} \left(\frac{t_k}{g_k} \right) \right]_{\Psi_G \Psi_0}}$$

Sample $e_i g_i / \sum_k^{N_{\mathrm{st}}} e_k g_k$ using *importance-sampled* projector.

For exact PMC, value indep. of Ψ_T, Ψ_G, statistical error depends on Ψ_T, Ψ_G.

Statistical error vanishes as $\Psi_T \to \Psi_0$.

For fixed Ψ_T, $\Psi_G = \Psi_T$ does not minimize statistical fluctuations!

Cyrus J. Umrigar

Variational and Projector MC

$$E_V = \frac{\left[\sum_i^{N_{\mathrm{MC}}} \left(\frac{t_i}{g_i}\right)^2 E_{\mathrm{L}}(i)\right]_{\Psi_{\mathrm{G}}^2}}{\left[\sum_k^{N_{\mathrm{MC}}} \left(\frac{t_k}{g_k}\right)^2\right]_{\Psi_{\mathrm{G}}^2}} \qquad \text{(Value depends on } \Psi_{\mathrm{T}}, \text{ error } \Psi_{\mathrm{T}}, \Psi_{\mathrm{G}})$$

$$E_0 = \frac{\left[\sum_i^{N_{\mathrm{MC}}} \left(\frac{t_i}{g_i}\right) E_{\mathrm{L}}(i)\right]_{\Psi_{\mathrm{G}}\Psi_0}}{\left[\sum_k^{N_{\mathrm{MC}}} \left(\frac{t_k}{g_k}\right)\right]_{\Psi_{\mathrm{G}}\Psi_0}} \qquad \text{(Value exact}^{\dagger}. \text{ Error depends on } \Psi_{\mathrm{T}}, \Psi_{\mathrm{G}}.)$$

$$E_{\mathrm{L}}(i) = \frac{\sum_j^{N_{\mathrm{st}}} H_{ij} t_j}{t_i}$$

In both VMC and PMC weighted average of the *configuration value of $\hat{H}$* aka *local energy, $E_{\mathrm{L}}(i)$,* but from points sampled from different distributions.

This is practical for systems that are large enough to be interesting if

1. $t_i = \langle \phi_i | \Psi_{\mathrm{T}} \rangle$, $g_i = \langle \phi_i | \Psi_{\mathrm{G}} \rangle$ can be evaluated in polynomial time, say N^3
2. the sum in $E_{\mathrm{L}}(i)$ can be done quickly, i.e., $\hat{H}$ is sparse (if space discrete) or semi-diagonal, i.e. $V(\mathbf{R})$ is local (if space continuous).

† In practice, usually necessary to make approximation (e.g. FN) and value depends on Ψ_{G}.

Cyrus J. Umrigar

Variational Monte Carlo in Real Space

Now we make things more concrete by considering the example of variational MC in real space, i.e. the MC walk is in the space of position eigenstates.

Variational Monte Carlo in Real Space

W. L. McMillan, Phys. Rev. **138**, A442 (1965)

Monte Carlo is used to perform the many-dimensional integrals needed to calculate quantum mechanical expectation values. e.g.

$$
\begin{aligned}
E_T &= \frac{\int d\mathbf{R}\,\Psi_{\mathrm{T}}^*(\mathbf{R})\,\mathcal{H}\,\psi_{\mathrm{T}}(\mathbf{R})}{\int d\mathbf{R}\,\psi_{\mathrm{T}}^2(\mathbf{R})} \\[2mm]
&= \int d\mathbf{R}\;\frac{\psi_{\mathrm{T}}^2(\mathbf{R})}{\int d\mathbf{R}\,\psi_{\mathrm{T}}^2(\mathbf{R})}\;\frac{\mathcal{H}\psi_{\mathrm{T}}(\mathbf{R})}{\psi_{\mathrm{T}}(\mathbf{R})} \\[2mm]
&= \frac{1}{N}\sum_i \frac{\mathcal{H}\Psi_{\mathrm{T}}(\mathbf{R}_i)}{\Psi_{\mathrm{T}}(\mathbf{R}_i)} = \frac{1}{N}\sum_i E_L(\mathbf{R}_i)
\end{aligned}
$$

Energy is obtained as an arithmetic sum of the *local energies* $E_L(\mathbf{R}_i)$ evaluated for configurations sampled from $\psi_{\mathrm{T}}^2(\mathbf{R})$ using a generalization of the Metropolis method. If ψ_{T} is an eigenfunction the $E_L(\mathbf{R}_i)$ do not fluctuate. Accuracy of VMC depends crucially on the quality of $\psi_{\mathrm{T}}(\mathbf{R})$. Diffusion MC does better by projecting onto ground state.

Cyrus J. Umrigar

Three ingredients for accurate Variational Monte Carlo

1. A method for sampling an arbitrary wave function Metropolis-Hastings.
2. A functional form for the wave function that is capable of describing the correct physics/chemistry.
3. An efficient method for optimizing the parameters in the wave functions.

Metropolis-Hastings Monte Carlo

Metropolis method originally used to sample the Boltzmann distribution.
This is still one of its more common uses.

General method for sampling **any known** discrete or continuous density.
(Other quantum Monte Carlo methods, e.g., diffusion MC, enable one to
sample densities that are not explicitly known but are the eigenstates of
known matrices or integral kernels.)

Metropolis-Hastings has serial correlations. Hence, direct sampling methods
preferable, but rarely possible for complicated densities in many dimensions.

Cyrus J. Umrigar

Metropolis-Hastings Monte Carlo (cont)

A *Markov chain* is specified by two ingredients:
1) an initial state
2) a transition matrix $M(\mathbf{R}_f|\mathbf{R}_i)$ (probability of transition $\mathbf{R}_i \rightarrow \mathbf{R}_f$.)

$$M(\mathbf{R}_f|\mathbf{R}_i) \geq 0, \quad \sum_{\mathbf{R}_f} M(\mathbf{R}_f|\mathbf{R}_i) = 1. \quad \textit{Column-stochastic matrix}$$

To sample $\rho(\mathbf{R})$, start from an arbitrary $\mathbf{R}_i$ and evolve the system by repeated application of M that satisfies the *stationarity condition* (flux into state $\mathbf{R}_i$ equals flux out of $\mathbf{R}_i$):

$$\sum_{\mathbf{R}_f} M(\mathbf{R}_i|\mathbf{R}_f)\, \rho(\mathbf{R}_f) = \sum_{\mathbf{R}_f} M(\mathbf{R}_f|\mathbf{R}_i)\, \rho(\mathbf{R}_i) = \rho(\mathbf{R}_i) \quad \forall\, \mathbf{R}_i$$

i.e., $\rho(\mathbf{R})$ is a right eigenvector of M with eigenvalue 1.
Stationarity $\Rightarrow$ if we start with ρ, will continue to sample ρ.
Want more than that: *any* initial density should evolve to ρ.

$$\lim_{n \to \infty} M^n(\mathbf{R}_f|\mathbf{R}_i)\, \delta(\mathbf{R}_i) = \rho(\mathbf{R}_f), \quad \forall\, \mathbf{R}_i.$$

i.e., ρ should be the *dominant* right eigenvector.

Cyrus J. Umrigar

Want that *any* initial density should evolve to ρ.

$$\lim_{n \to \infty} M^n(\mathbf{R}_f | \mathbf{R}_i) \delta(\mathbf{R}_i) = \rho(\mathbf{R}_f), \quad \forall \, \mathbf{R}_i.$$

ρ should be the *dominant* right eigenvector. Additional conditions needed to guarantee this.

A nonnegative matrix M is said to be *primitive* if $\exists \, n$ such that M^n has all elements positive. (Can go from any state to any other in finite number of steps.)

(Special case of) Perron-Frobenius Theorem: A column-stochastic primitive matrix has a unique dominant eigenvalue of 1, with a positive right eigenvector and a left eigenvector with all components equal to 1 (by definition of column-stochastic matrix).

In practice, length of Monte Carlo should be long enough that there be a significant probability of the system making several transitions between the neighborhoods of any pair of representative states that make a significant contribution to the average. This ensures that states are visited with the correct probability with only small statistical fluctuations.

For example in a double-well system many transitions between the 2 wells should occur, but we can choose our proposal matrix to achieve this even if barrier between wells is high.

Cyrus J. Umrigar

Metropolis-Hastings Monte Carlo (cont)
Construction of M

Need a prescription to construct M, such that ρ is its stationary state. Impose *detailed balance* condition

$$M(\mathbf{R}_f|\mathbf{R}_i)\,\rho(\mathbf{R}_i) = M(\mathbf{R}_i|\mathbf{R}_f)\,\rho(\mathbf{R}_f)$$

Detailed balance more stringent than stationarity condition (removed the sums).
Detailed balance is not necessary but provides way to construct M.
Write elements of M as product of elements of a proposal matrix T and an acceptance Matrix A,

$$M(\mathbf{R}_f|\mathbf{R}_i) = A(\mathbf{R}_f|\mathbf{R}_i)\,T(\mathbf{R}_f|\mathbf{R}_i)$$

$M(\mathbf{R}_f|\mathbf{R}_i)$ and $T(\mathbf{R}_f|\mathbf{R}_i)$ are stochastic matrices, but $A(\mathbf{R}_f|\mathbf{R}_i)$ is not.
Detailed balance is now:

$$A(\mathbf{R}_f|\mathbf{R}_i)\,T(\mathbf{R}_f|\mathbf{R}_i)\,\rho(\mathbf{R}_i) = A(\mathbf{R}_i|\mathbf{R}_f)\,T(\mathbf{R}_i|\mathbf{R}_f)\,\rho(\mathbf{R}_f)$$

or
$$\frac{A(\mathbf{R}_f|\mathbf{R}_i)}{A(\mathbf{R}_i|\mathbf{R}_f)} = \frac{T(\mathbf{R}_i|\mathbf{R}_f)\,\rho(\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i)\,\rho(\mathbf{R}_i)}\,.$$

Metropolis-Hastings Monte Carlo (cont)
Choice of Acceptance Matrix A

$$\frac{A(\mathbf{R}_f|\mathbf{R}_i)}{A(\mathbf{R}_i|\mathbf{R}_f)} = \frac{T(\mathbf{R}_i|\mathbf{R}_f)\ \rho(\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i)\ \rho(\mathbf{R}_i)}\ .$$

Infinity of choices for A. Any function

$$F\left(\frac{T(\mathbf{R}_i|\mathbf{R}_f)\ \rho(\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i)\ \rho(\mathbf{R}_i)}\right) = A(\mathbf{R}_f|\mathbf{R}_i)$$

for which $F(x)/F(1/x) = x$ and $0 \leq F(x) \leq 1$ will do.
Choice of Metropolis *et al.* $F(x) = \min\{1, x\}$, maximizes the acceptance:

$$A(\mathbf{R}_f|\mathbf{R}_i) = \min\left\{1, \frac{T(\mathbf{R}_i|\mathbf{R}_f)\ \rho(\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i)\ \rho(\mathbf{R}_i)}\right\}\ .$$

Other less good choices for $A(\mathbf{R}_f|\mathbf{R}_i)$ have been made, *e.g.* $F(x) = \frac{x}{1+x}$

$$A(\mathbf{R}_f|\mathbf{R}_i) = \frac{T(\mathbf{R}_i|\mathbf{R}_f)\ \rho(\mathbf{R}_f)}{T(\mathbf{R}_i|\mathbf{R}_f)\ \rho(\mathbf{R}_f) + T(\mathbf{R}_f|\mathbf{R}_i)\ \rho(\mathbf{R}_i)}.$$

Metropolis: $T(\mathbf{R}_i|\mathbf{R}_f) = T(\mathbf{R}_f|\mathbf{R}_i)$,   Hastings: $T(\mathbf{R}_i|\mathbf{R}_f) \neq T(\mathbf{R}_f|\mathbf{R}_i)$

Cyrus J. Umrigar

Metropolis-Hastings Monte Carlo (cont)
Choice of Proposal Matrix T

So, the optimal choice for the acceptance matrix $A(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})$ is simple and known.

However, there is considerable scope for using one's ingenuity to come up with good proposal matrices, $T(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})$, that allow one to make large moves with large acceptances, in order to make the autocorrelation time small.

Cyrus J. Umrigar

Choice of Proposal Matrix T in Metropolis-Hastings (cont)

$$A(\mathbf{R}_f | \mathbf{R}_i) = \min \left\{ 1, \frac{T(\mathbf{R}_i | \mathbf{R}_f) \, \rho(\mathbf{R}_f)}{T(\mathbf{R}_f | \mathbf{R}_i) \, \rho(\mathbf{R}_i)} \right\}$$

Use freedom in T to make $\quad \dfrac{T(\mathbf{R}_i | \mathbf{R}_f) \, \rho(\mathbf{R}_f)}{T(\mathbf{R}_f | \mathbf{R}_i) \, \rho(\mathbf{R}_i)} \approx 1.$

$T(\mathbf{R}_f | \mathbf{R}_i) \propto \rho(\mathbf{R}_f)$ optimal if $T(\mathbf{R}_f | \mathbf{R}_i)$ can be sampled over all space – usually not the case. And if it is, then one would not use Metropolis-Hastings in the first place.

Otherwise, let $\quad T(\mathbf{R}_f | \mathbf{R}_i) = \dfrac{S(\mathbf{R}_f | \mathbf{R}_i)}{\int d\mathbf{R}_f \, S(\mathbf{R}_f | \mathbf{R}_i)} \approx \dfrac{S(\mathbf{R}_f | \mathbf{R}_i)}{S(\mathbf{R}_i | \mathbf{R}_i) \Omega(\mathbf{R}_i)}$

$S(\mathbf{R}_f | \mathbf{R}_i)$ is non-zero only in domain $D(\mathbf{R}_i)$ of volume $\Omega(\mathbf{R}_i)$ around $\mathbf{R}_i$).

$$\frac{A(\mathbf{R}_f, \mathbf{R}_i)}{A(\mathbf{R}_i, \mathbf{R}_f)} = \frac{T(\mathbf{R}_i | \mathbf{R}_f) \, \rho(\mathbf{R}_f)}{T(\mathbf{R}_f | \mathbf{R}_i) \, \rho(\mathbf{R}_i)} \approx \frac{\Omega(\mathbf{R}_i)}{\Omega(\mathbf{R}_f)} \frac{S(\mathbf{R}_i | \mathbf{R}_i)}{S(\mathbf{R}_f | \mathbf{R}_f)} \frac{S(\mathbf{R}_i | \mathbf{R}_f)}{S(\mathbf{R}_f | \mathbf{R}_i)} \frac{\rho(\mathbf{R}_f)}{\rho(\mathbf{R}_i)}$$

from which it is apparent that the choice

$$S(\mathbf{R}_f | \mathbf{R}_i) \overset{\propto}{\sim} \sqrt{\rho(\mathbf{R}_f)/\Omega(\mathbf{R}_f)} \quad \text{yields} \quad A(\mathbf{R}_f, \mathbf{R}_i)/A(\mathbf{R}_i, \mathbf{R}_f) \approx 1.$$

Cyrus J. Umrigar

Choice of Proposal Matrix T in Metropolis-Hastings (cont)

To be more precise, if the log-derivatives of $T(\mathbf{R}_f|\mathbf{R}_i)$ equal those of $\sqrt{\rho(\mathbf{R}_f)/\Omega(\mathbf{R}_f)}$ at $\mathbf{R}_f = \mathbf{R}_i$, the acceptance goes as $1 - \mathcal{O}((\mathbf{R}' - \mathbf{R})^3)$, i.e., the average acceptance goes as $1 - \mathcal{O}(\Delta^4)$, where Δ is the linear dimension of $D(\mathbf{R}_i)$.

Considerable improvement compared to using a symmetric $S(\mathbf{R}_f|\mathbf{R}_i)$ or choosing $S(\mathbf{R}_f|\mathbf{R}_i) \stackrel{\propto}{\sim} \rho(\mathbf{R}_f)$ for either of which we have acceptance $1 - \mathcal{O}((\mathbf{R}' - \mathbf{R})^1)$ and av. accep. $1 - \mathcal{O}(\Delta^2)$.

Another possible choice, motivated by (DMC) is

$$T(\mathbf{R}_f|\mathbf{R}_i) = \frac{1}{(2\pi\tau)^{3/2}} \exp\left[\frac{-(\mathbf{R}_f - \mathbf{R}_i - \mathbf{V}(\mathbf{R}_i)\tau)^2}{2\tau}\right], \quad \mathbf{V}(\mathbf{R}_i) = \frac{\nabla\Psi(\mathbf{R}_i)}{\Psi(\mathbf{R}_i)}$$

Advantage: allows Metropolis Monte Carlo and diffusion Monte Carlo programs to share almost all the code.
Such an algorithm is more efficient than one with a symmetric $S(\mathbf{R}_f|\mathbf{R}_i)$ or one for which $S(\mathbf{R}_f|\mathbf{R}_i) \stackrel{\propto}{\sim} \rho(\mathbf{R}_f)$, but less efficient than one for which $S(\mathbf{R}_f|\mathbf{R}_i) \stackrel{\propto}{\sim} \sqrt{\rho(\mathbf{R}_f)/\Omega(\mathbf{R}_f)}$.

These arguments are rigorous only in the small-step limit and are applicable only to functions with sufficiently many derivatives within $D(\mathbf{R}_i)$. In practice these ideas yield large reduction in the autocorrelation time provided that we employ a coordinate system such that ρ has continuous derivatives within $D(\mathbf{R}_i)$.

Cyrus J. Umrigar

Some examples

We want to sample from $|\Psi(\mathbf{R})|^2$.

We propose moves with probability density

$$T(\mathbf{R}_f|\mathbf{R}_i) = \frac{S(\mathbf{R}_f|\mathbf{R}_i)}{\int d\mathbf{R}_f\, S(\mathbf{R}_f|\mathbf{R}_i)} \approx \frac{S(\mathbf{R}_f|\mathbf{R}_i)}{S(\mathbf{R}_i|\mathbf{R}_i)\Omega(\mathbf{R}_i)}$$
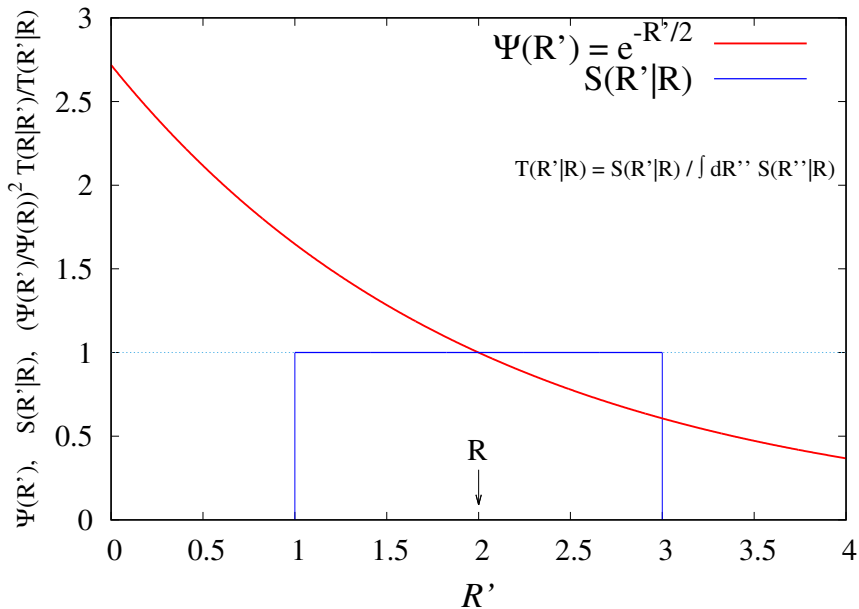
and since the acceptance is

$$A(\mathbf{R}_f|\mathbf{R}_i) = \min\left\{1, \frac{|\Psi(\mathbf{R}_f)|^2\ T(\mathbf{R}_i|\mathbf{R}_f)}{|\Psi(\mathbf{R}_i)|^2\ T(\mathbf{R}_f|\mathbf{R}_i)}\right\}$$
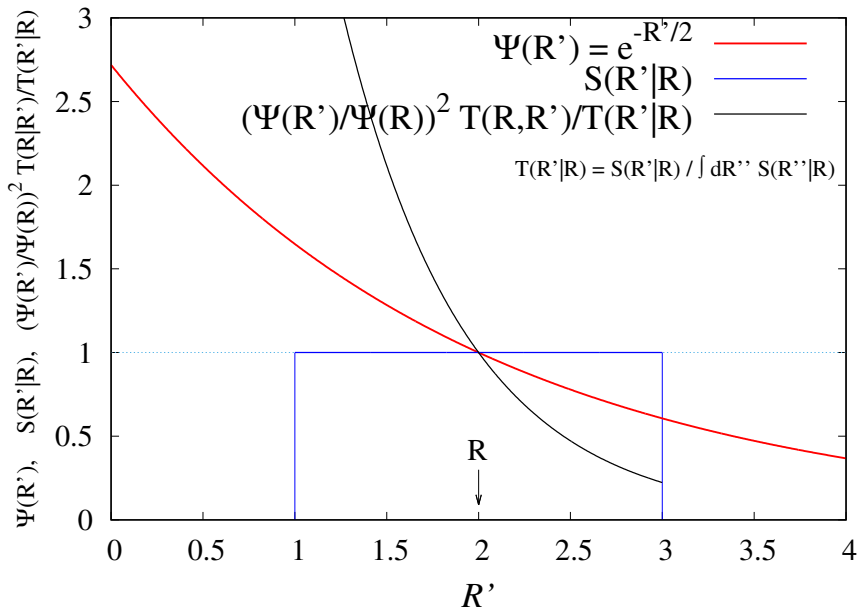
we want

$$\frac{|\Psi(\mathbf{R}_f)|^2\ T(\mathbf{R}_i|\mathbf{R}_f)}{|\Psi(\mathbf{R}_i)|^2\ T(\mathbf{R}_f|\mathbf{R}_i)}$$

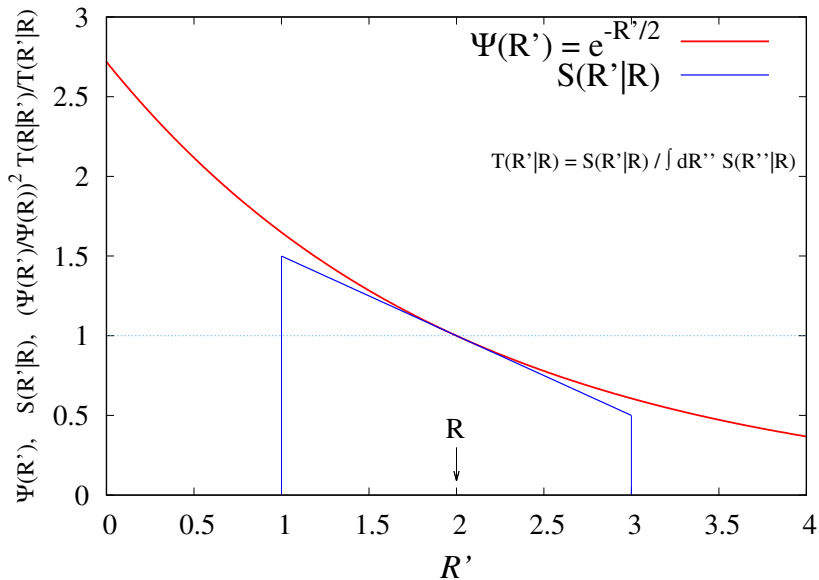to be as close to 1 as possible. Let's see how it changes with $T(\mathbf{R}_f|\mathbf{R}_i)$.

Symmetrical T in Metropolis

$$\Psi(R') = e^{-R'/2}$$
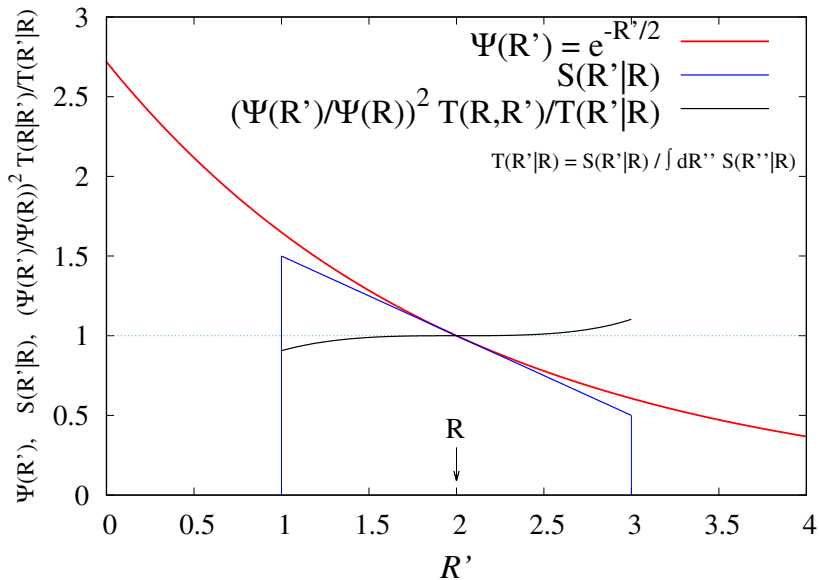$$S(R'|R)$$

$$T(R'|R) = S(R'|R) / \int dR'' \, S(R''|R)$$

Cyrus J. Umrigar

Symmetrical T in Metropolis

$\Psi(R') = e^{-R'/2}$
$S(R'|R)$
$(\Psi(R')/\Psi(R))^2 \, T(R,R')/T(R'|R)$

$T(R'|R) = S(R'|R) / \int dR'' \, S(R''|R)$

Cyrus J. Umrigar

Non-symmetrical linear T in Metropolis-Hastings

$\Psi(R') = e^{-R'/2}$

$S(R'|R)$

$T(R'|R) = S(R'|R) / \int dR'' \ S(R''|R)$

Y-axis: $\Psi(R')$, $S(R'|R)$, $(\Psi(R')/\Psi(R))^2 \ T(R|R')/T(R'|R)$

X-axis: R'

R

Cyrus J. Umrigar

Non-symmetrical linear T in Metropolis-Hastings

Cyrus J. Umrigar

When will the above not work so well?

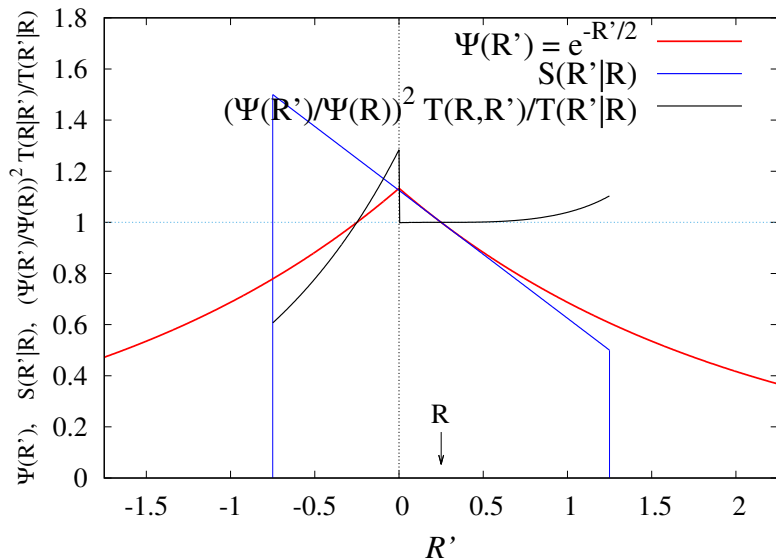What assumptions have we made in both of the non-symmetric choices above?

When will the above not work so well?

What assumptions have we made in both of the non-symmetric choices above?

Answer: In both cases we are utilizing the gradient of the function to be sampled and are implicitly assuming that it is smooth.

Let's see what happens when it is not.

Non-symmetrical linear T in Metropolis-Hastings

Plot legend:
- $\Psi(R') = e^{-R'/2}$
- $S(R'|R)$
- $(\Psi(R')/\Psi(R))^2 \, T(R,R')/T(R'|R)$

y-axis: $\Psi(R')$, $S(R'|R)$, $(\Psi(R')/\Psi(R))^2 \, T(R|R)/T(R'|R)$

x-axis: R'

When the gradient has a discontinuity the acceptance goes down.

Cyrus J. Umrigar

Choice of Proposal Matrix T in Metropolis-Hastings (cont)

How to make large moves with high acceptance in spite of wavefunctions that have cusps at nuclei?

1. Make moves in spherical polar coordinates, centered on the nearest nucleus.
2. Radial move is proportional to distance to nucleus, say in interval $[\frac{r}{5}, 5r]$.
3. Angular move gets *larger* as electron approaches nucleus.

Using these ideas an autocorrelation time $T_{\text{corr}} \approx 1$ can be achieved!

Details are in: Accelerated Metropolis Method, C. J. Umrigar, PRL **71** 408, (1993).

The point of the above exercise was not the particular problem treated, but rather to provide a concrete example of the ideas that enable making large moves with high acceptance, thereby achieving $T_{\text{corr}} \approx 1$.

Estimation of Errors
Autocorrelation time

N Monte Carlo steps = N_b blocks $\times$ N_s steps/block
If N_s is large enough the block averages are nearly independent.

$$\bar{E} \quad = \quad \text{average of } E_{\mathrm{L}} \text{ over the } N \text{ Monte Carlo steps}$$
$$\sigma \quad = \quad \text{rms fluctuations of individual } E_{\mathrm{L}}$$
$$\sigma_b \quad = \quad \text{rms fluctuations of block averages of } E_{\mathrm{L}}$$

Need to estimate T_{corr} to make sure $N_b \gg T_{\mathrm{corr}}$.
$N_{\mathrm{eff}} = N/T_{\mathrm{corr}}$ independent measurements of E_{L}, so get T_{corr} from:

$$\mathrm{err}(\bar{E}) = \frac{\sigma}{\sqrt{N_b \times N_s}} \sqrt{T_{\mathrm{corr}}} = \frac{\sigma_b}{\sqrt{N_b}}$$

$\Rightarrow$ $\boxed{T_{\mathrm{corr}} = N_s \left(\frac{\sigma_b}{\sigma} \right)^2}$ Choose $N_s \gg T_{\mathrm{corr}}$, say, 100 T_{corr}.
If $N_s \approx 10\,T_{\mathrm{corr}}$, T_{corr} underest. $\approx 10\%$.

Functional form of Trial Wave Function

Other methods: Restrictions on the form of the wavefn.:

1. Many-body wavefn. expanded in determinants of single-particle orbitals.
2. Single-particle orbitals are expanded in planewaves or gaussians.

   occasionally wavelets etc.

QMC: Great freedom in form of the wavefn. — use physics/chemistry intuition:

1. Multideterminant times Jastrow. Ceperley, many others

2. Antisymmetrized Geminal Power times Jastrow. Sorella, Casula
   $$\mathcal{A}\left[\Phi(\mathbf{r}_1^\uparrow, \mathbf{r}_1^\downarrow)\, \Phi(\mathbf{r}_2^\uparrow, \mathbf{r}_2^\downarrow)\, \cdots\, \Phi(\mathbf{r}_{N/2}^\uparrow, \mathbf{r}_{N/2}^\downarrow)\right]$$

3. Pfaffians times Jastrow. Schmidt, Mitas, Wagner and coworkers
   $$\mathcal{A}\left[\Phi(\mathbf{r}_1, s_1; \mathbf{r}_2, s_2)\, \Phi(\mathbf{r}_3, s_3; \mathbf{r}_4, s_4)\, \cdots\, \Phi(\mathbf{r}_{N-1}, s_{N-1}; \mathbf{r}_N, s_N)\right]$$

4. Backflow times Jastrow. Needs and coworkers, Moroni (extension of Feynman)

5. Laughlin and Composite Fermion. Jeon, Güclu, CJU and Jain

Multideterminant $\times$ Jastrow form of Trial Wavefunction

$$\Psi_T = \left(\sum_n d_n D_n^\uparrow D_n^\downarrow \right) \times \mathcal{J}(r_i, r_j, r_{ij})$$

- Determinants: $\sum_n d_n D_n^\uparrow D_n^\downarrow$

$D^\uparrow$ and $D^\downarrow$ are determinants of single-particle orbitals ϕ for up ($\uparrow$) and down ($\downarrow$) spin electrons respectively.

The single-particle orbitals ϕ are given by:

$$\phi(\mathbf{r}_i) = \sum_{\alpha k_\alpha} c_{k_\alpha} N_{k_\alpha} r_{i\alpha}^{n_{k_\alpha}-1} e^{-\zeta_{k_\alpha} r_{i\alpha}} Y_{l_{k_\alpha} m_{k_\alpha}}(\hat{r}_{i\alpha})$$

- Jastrow: $\mathcal{J}(r_i, r_j, r_{ij}) = \prod_{\alpha i} \exp(A_{\alpha i}) \prod_{ij} \exp(B_{ij}) \prod_{\alpha ij} \exp(C_{\alpha ij})$

  $A_{\alpha i} \Rightarrow$ electron-ion correlation $\qquad \sim N_{\text{atomtype}}$ of $\mathcal{J}$ parms.

  $B_{ij} \Rightarrow$ electron-electron correlation $\qquad \sim N_{\text{atomtype}}$ of ζ_{k_α} parms.

  $C_{\alpha ij} \Rightarrow$ electron-electron-ion correlation $\qquad \sim N_{\text{atom}}^2$ of c_{k_α} parms.

  $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \sim e^{N_{\text{atom}}}$ of d_n parms.

d_n, c_{k_α}, ζ_{k_α} and parms in $\mathcal{J}$ are optimized. $\qquad$ Power of QMC:

$\mathcal{J}$ parms. replace many d_n parms.

Cyrus J. Umrigar

Optimization of many-body wavefunctions

Cyrus J. Umrigar

Almost all errors reduced by optimizing trial wavefunctions

1. Statistical error (both the rms fluctuations of E_L and the autocorrelation time)
2. E_VMC
3. Fixed-node error in E_DMC (nodes move during optimization). Fixed node errors can be LARGE. For C_2, FN error for 1-det wavefn is 1.3 eV for total energy and 0.7 eV for well-depth. However, optimized multidet. wavefn has FN error that is better than chemical accuracy (1 kcal/mole = 0.043 eV/molecule).
4. Time-step error in DMC
5. Population control error in PMC
6. Pseudopotential locality error in DMC when using nonlocal pseudopotentials
7. Error of observables that do not commute with the Hamiltonian (mixed estimators, $\langle \Psi_0 | \hat{A} | \Psi_T \rangle$ not exact even for nodeless ψ_0, ψ_T) if one does not use forward/side walking.

Choices to be made when optimizing trial wavefunctions

1. What precisely do we want to optimize – the objective function or measure of goodness?

2. What method do we use to do the optimization? If more than one method is applied to the same objective function, they will of course give the same wavefunction, but the efficiency with which we arrive at the solution may be much different.

Measures of goodness of variational wave functions

$$\min \ E_{\text{VMC}} \ = \ \frac{\langle \psi_{\text{T}} | H | \psi_{\text{T}} \rangle}{\langle \psi_{\text{T}} | \psi_{\text{T}} \rangle} \ = \ \langle E_{\text{L}} \rangle_{|\psi_{\text{T}}|^2}$$

$$\min \ \sigma^2_{\text{VMC}} \ = \ \frac{\langle \psi_{\text{T}} | (H - E_{\text{T}})^2 | \psi_{\text{T}} \rangle}{\langle \psi_{\text{T}} | \psi_{\text{T}} \rangle} \ = \ \langle E_{\text{L}}^2(\mathbf{R}_i) \rangle_{|\psi_{\text{T}}|^2} - \langle E_{\text{L}}(\mathbf{R}_i) \rangle^2_{|\psi_{\text{T}}|^2}$$

$$\max \ \ \Omega^2 \ = \ \frac{|\langle \psi_{\text{FN}} | \psi_{\text{T}} \rangle|^2}{\langle \psi_{\text{FN}} | \psi_{\text{FN}} \rangle \langle \psi_{\text{T}} | \psi_{\text{T}} \rangle} \ = \ \frac{\left\langle \frac{\psi_{\text{FN}}}{\psi_{\text{T}}} \right\rangle^2_{|\psi_{\text{T}}|^2}}{\left\langle \left| \frac{\psi_{\text{FN}}}{\psi_{\text{T}}} \right|^2 \right\rangle_{|\psi_{\text{T}}|^2}}$$

$$\min \ E_{\text{DMC}} \ = \ \frac{\langle \psi_{\text{FN}} | H | \psi_{\text{T}} \rangle}{\langle \psi_{\text{FN}} | \psi_{\text{T}} \rangle} \ = \ \langle E_{\text{L}} \rangle_{|\psi_{\text{FN}} \psi_{\text{T}}|}$$

For an infinitely flexible wave function all optimizations will yield the exact wavefunction (except that minimizing σ could yield an excited state) but for the imperfect functional forms used in practice they differ.
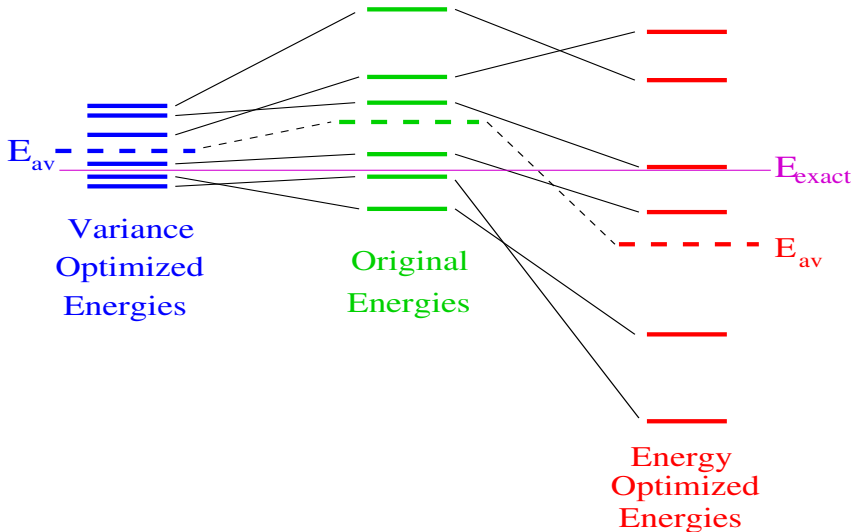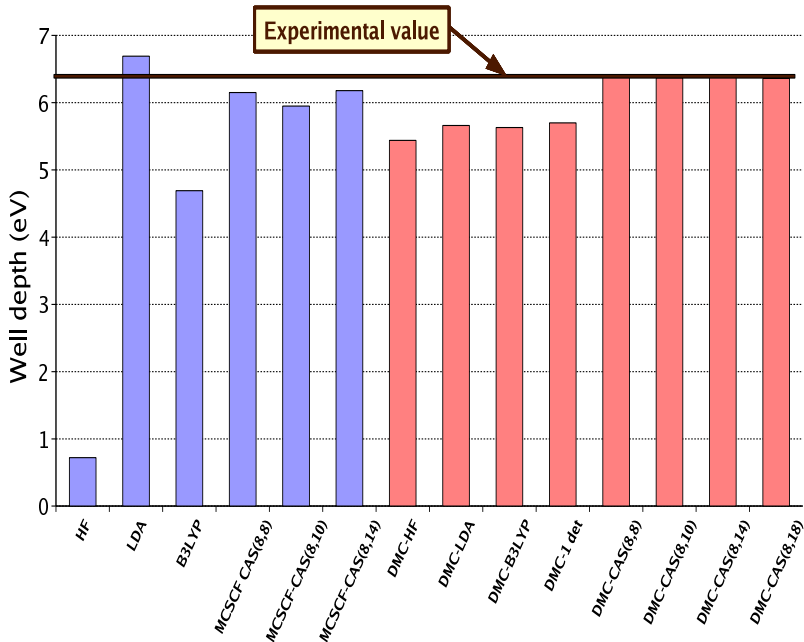
Cyrus J. Umrigar

Variance vs. Energy

$$\sigma^2 = \sum_{i=1}^{N_{\text{conf}}} \left(\frac{\mathcal{H}\Psi_{\text{T}}(\mathbf{R}_i)}{\Psi_{\text{T}}(\mathbf{R}_i)} - \bar{E} \right)^2$$
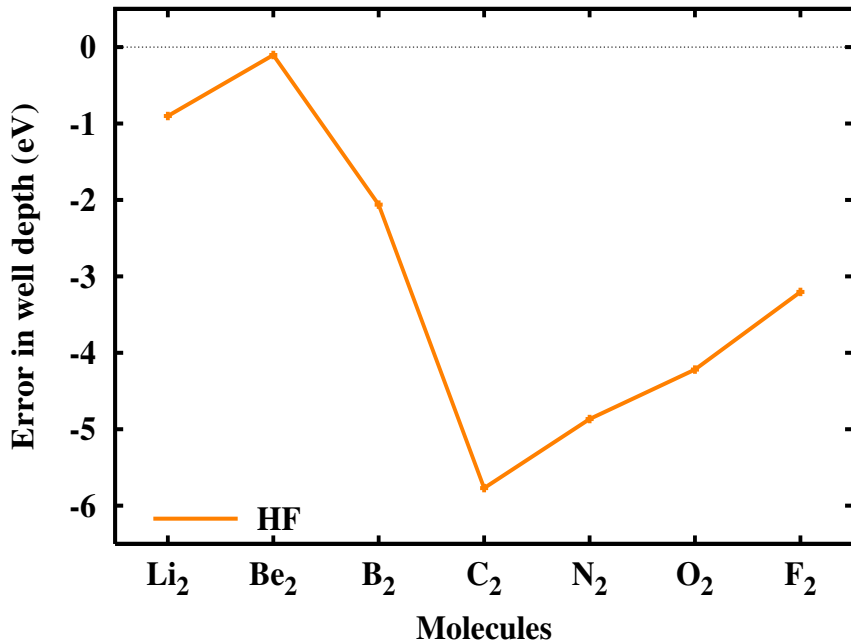
$$\bar{E} = \sum_{i=1}^{N_{\text{conf}}} \frac{\mathcal{H}\Psi_{\text{T}}(\mathbf{R}_i)}{\Psi_{\text{T}}(\mathbf{R}_i)}$$
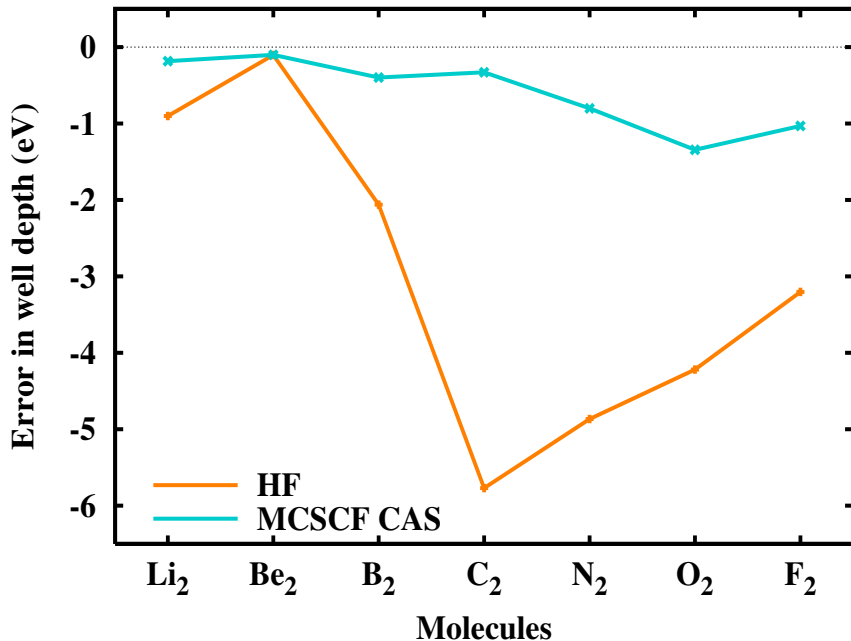
E_{av}

E_{exact}

E_{av}

Variance Optimized Energies

Original Energies

Energy Optimized Energies

Cyrus J. Umrigar

Well-depth of C$_2$

Experimental value

Well depth (eV)

HF, LDA, B3LYP, MCSCF CAS(8,8), MCSCF-CAS(8,10), MCSCF-CAS(8,14), DMC-HF, DMC-LDA, DMC-B3LYP, DMC-1 det, DMC-CAS(8,8), DMC-CAS(8,10), DMC-CAS(8,14), DMC-CAS(8,18)

Cyrus J. Umrigar

Error in Well-Depth of 1st-Row Diatomic Molecules

Cyrus J. Umrigar

Error in Well-Depth of 1st-Row Diatomic Molecules

Error in well depth (eV) vs Molecules (Li_2, Be_2, B_2, C_2, N_2, O_2, F_2)

HF

MCSCF CAS

Cyrus J. Umrigar

Error in Well-Depth of 1st-Row Diatomic Molecules

Cyrus J. Umrigar

Error in Well-Depth of 1st-Row Diatomic Molecules

Cyrus J. Umrigar

Error in Well-Depth of 1st-Row Diatomic Molecules

Error in well depth (eV) vs **Molecules** (Li_2, Be_2, B_2, C_2, N_2, O_2, F_2)

Legend:
- HF
- MCSCF CAS
- VMC J×det
- DMC J×det
- VMC J×CAS

Cyrus J. Umrigar

Error in Well-Depth of 1st-Row Diatomic Molecules

Error in well depth (eV) vs Molecules (Li_2, Be_2, B_2, C_2, N_2, O_2, F_2)

Legend:
- HF
- MCSCF CAS
- VMC J×det
- DMC J×det
- VMC J×CAS
- DMC J×CAS

Cyrus J. Umrigar

Projector MC

Cyrus J. Umrigar
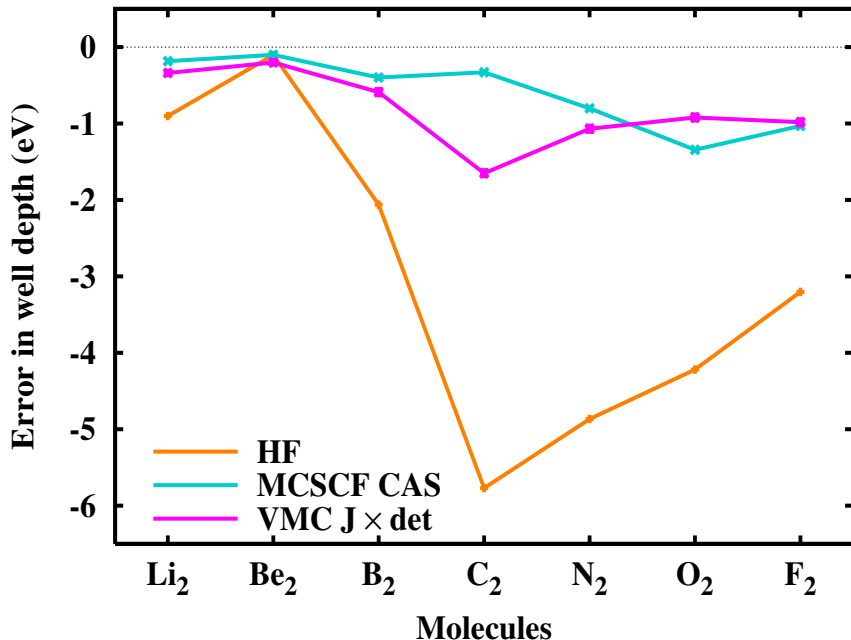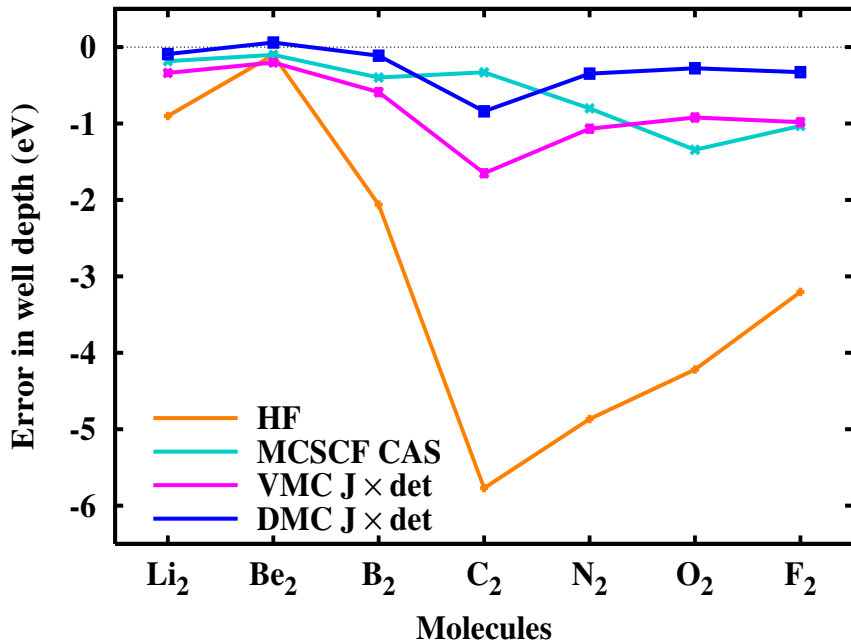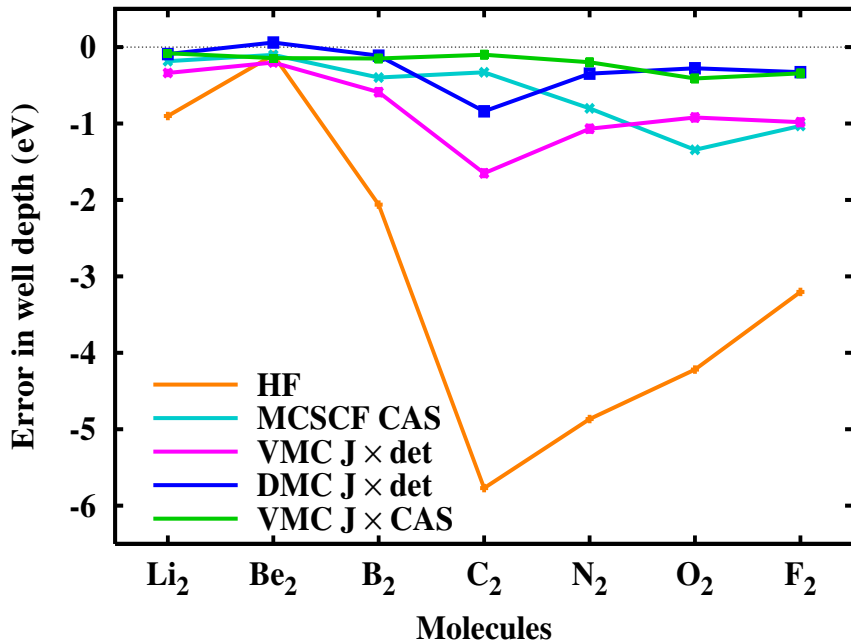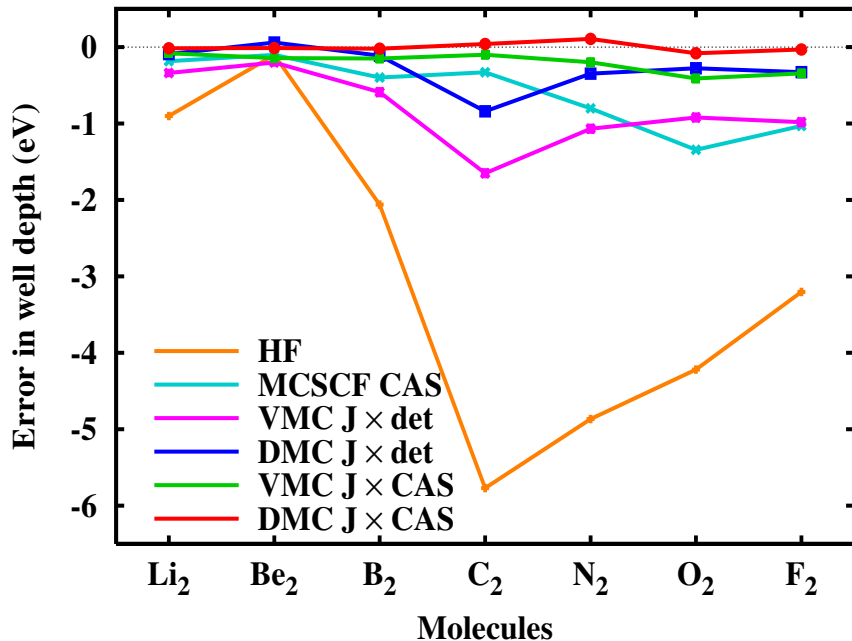
Projector: $\quad |\Psi_0\rangle = \hat{P}(\infty) |\Psi_T\rangle = \lim_{n\to\infty} \hat{P}^n(\tau) |\Psi_T\rangle$

Projector is any function of the Hamiltonian that maps the ground state eigenvalue of $\hat{H}$ to 1, and the higher eigenvalues of $\hat{H}$ to absolute values that are < 1 (preferably close to 0).

Exponential projector: $\qquad \hat{P} = e^{\tau(E_T\hat{\mathbf{1}} - \hat{H})}$ $\qquad$ (usually has time-step error)

Linear projector: $\qquad \hat{P} = \hat{\mathbf{1}} + \tau(E_T\hat{\mathbf{1}} - \hat{H})$ $\qquad (\tau < \frac{2}{E_{\max} - E_0})$

Green's function projector: $\hat{P} = \dfrac{1}{\hat{\mathbf{1}} - \tau(E_T\hat{\mathbf{1}} - \hat{H})}$

Taxonomy of Projector Monte Carlo Methods

The amplitudes of Ψ_0 in the chosen basis are obtained by using a "Projector", $\hat{P}$, that is a function of the Hamiltonian, $\hat{H}$, and has Ψ_0 as its dominant state.

Various Projector Monte Carlo Methods differ in:
a) form of the projector, and,
b) space in which the walk is done (single-particle basis and quantization).
(1st-quantized $\equiv$ unsymmetrized basis, 2nd-quantized $\equiv$ antisymmetrized basis.)

Method	Projector	SP Basis	Quantiz
Diffusion Monte Carlo	$e^{\tau(E_T\hat{1}-\hat{H})}$	$\mathbf{r}$	1st
GFMC (Kalos, Ceperley, Schmidt)	$e^{\tau(E_T\hat{1}-\hat{H})}$ (samp. τ)	$\mathbf{r}$	1st
LRDMC (Sorella, Casula)	$e^{\tau(E_T\hat{1}-\hat{H})}$ (samp. τ)	$\mathbf{r}_i$	1st
FCIQMC (Alavi, Booth)	$\hat{1} + \tau(E_T\hat{1} - \hat{H})$	ϕ_i^{orthog}	2nd
phaseless AFQMC (Zhang, Krakauer)	$e^{\tau(E_T\hat{1}-\hat{H})}$	$\phi_i^{\mathrm{nonorthog}}$	2nd

$1 + \tau(E_T\hat{1} - \hat{H})$ can be used only if the spectrum of $\hat{H}$ is bounded, $\tau < \frac{2}{E_{\max}-E_0}$.

Cyrus J. Umrigar

Diffusion Monte Carlo

i.e., $\hat{P}(\tau) = \exp(\tau(E_T - \hat{H})), \quad |\phi_i\rangle = |\mathbf{R}\rangle$ **, walkers are** 1st**-quantized**

$$G(\mathbf{R}', \mathbf{R}, \tau) \equiv P(\mathbf{R}', \mathbf{R}, \tau) = \langle \mathbf{R}' | e^{\tau(E_T - \hat{H})} | \mathbf{R} \rangle.$$

Diffusion Monte Carlo – Importance Sampled Green's Function

Importance sampling: Multiply imaginary-time Schrödinger equation

$$-\frac{1}{2}\nabla^2\Psi(\mathbf{R},t) + (\mathcal{V}(\mathbf{R}) - E_{\mathrm{T}})\Psi(\mathbf{R},t) = -\frac{\partial\Psi(\mathbf{R},t)}{\partial t}$$

by $\Psi_{\mathrm{T}}(\mathbf{R})$ and rearranging terms we obtain

$$-\frac{\nabla^2}{2}(\Psi\Psi_{\mathrm{T}}) + \nabla\cdot\left(\frac{\nabla\Psi_{\mathrm{T}}}{\Psi_{\mathrm{T}}}\Psi\Psi_{\mathrm{T}}\right) + \left(\underbrace{\frac{-\nabla^2\Psi_{\mathrm{T}}}{2\Psi_{\mathrm{T}}} + \mathcal{V}}_{E_{\mathrm{L}}(\mathbf{R})} - E_{\mathrm{T}}\right)(\Psi\Psi_{\mathrm{T}}) = -\frac{\partial(\Psi\Psi_{\mathrm{T}})}{\partial t}$$

defining $f(\mathbf{R},t) = \Psi(\mathbf{R},t)\Psi_{\mathrm{T}}(\mathbf{R})$, this is

$$\underbrace{-\frac{1}{2}\nabla^2 f}_{\text{diffusion}} + \underbrace{\nabla\cdot\left(\frac{\nabla\Psi_{\mathrm{T}}}{\Psi_{\mathrm{T}}}f\right)}_{\text{drift}} + \underbrace{(E_{\mathrm{L}}(\mathbf{R}) - E_{\mathrm{T}})f}_{\text{growth/decay}} = -\frac{\partial f}{\partial t}$$

Since we know the exact Green function for any one term on LHS, an approximation is:

$$\tilde{G}(\mathbf{R}',\mathbf{R},\tau) \approx \frac{1}{(2\pi\tau)^{3N/2}} e^{\left[-\frac{(\mathbf{R}'-\mathbf{R}-\mathbf{V}(\mathbf{R})\tau)^2}{2\tau} + \left\{E_{\mathrm{T}} - \frac{(E_{\mathrm{L}}(\mathbf{R}')+E_{\mathrm{L}}(\mathbf{R}))}{2}\right\}\tau\right]}$$

Singularities of Green's function

CJU, Nightingale, Runge, JCP 1993

Region	Local energy E_{L}	Velocity $\mathbf{V}$
Nodes	$E_{\mathrm{L}} \sim \pm\frac{1}{R_{\perp}}$ for Ψ_{T}	$V \sim \frac{1}{R_{\perp}}$
	$E_{\mathrm{L}} = E_0$ for Ψ_0	for both Ψ_{T} and Ψ_0
e-n and e-e	$E_{\mathrm{L}} \sim \frac{1}{x}$ if cusps not imposed	V has a discontinuity
coincidences	E_{L} finite if cusps are imposed	for both Ψ_{T} and Ψ_0
	$E_{\mathrm{L}} = E_0$ for Ψ_0	

All the above infinities and discontinuities cause problems, e.g.,

$$\int_0^a dx\, E_{\mathrm{L}} = \int_0^a dx\, \left(\frac{1}{x}\right) = \pm\infty$$

$$\int_0^a dx\, E_{\mathrm{L}}^2 = \int_0^a dx\, \left(\frac{1}{x}\right)^2 = \infty$$

Modify Green's function, by approximately integrating E_{L} and $\mathbf{V}$ over path, taking account of the singularities, at no additional computational cost.

Cyrus J. Umrigar

Combining with Metropolis to reduce time-step error

Reynolds, Ceperley, Alder, Lester, JCP 1982

$$\underbrace{-\frac{1}{2}\nabla^2 f}_{\text{diffusion}} + \underbrace{\nabla \cdot \left(\frac{\nabla \psi_{\text{T}}}{\psi_{\text{T}}} f\right)}_{\text{drift}} + \underbrace{(E_{\text{L}}(\mathbf{R}) - E_{\text{T}}) f}_{\text{growth/decay}} = -\frac{\partial f}{\partial t}$$

If we omit the growth/decay term then $|\Psi_{\text{T}}|^2$ is the solution.

$$-\frac{1}{2}\nabla^2 \psi_{\text{T}}^2(\mathbf{R}) + \nabla \cdot \left(\frac{\nabla \psi_{\text{T}}}{\psi_{\text{T}}} \psi_{\text{T}}^2(\mathbf{R})\right) = 0$$

We can sample $|\Psi_{\text{T}}|^2$ exactly using Metropolis-Hastings! So, view $G(\mathbf{R}', \mathbf{R}, t)$ as being the proposal matrix $T(\mathbf{R}', \mathbf{R})$ and introduce accept-reject step after drift and diffusion steps.

Since some of the moves are rejected, account for that approximately by reducing the time step in the reweighting factor from τ to τ_{eff}.

If accept/reject is done after each 1-electron move, then

$$\tau_{\text{eff}} = \tau \frac{R_{\text{accep}}^2}{R_{\text{prop}}^2}$$

Branching random walks
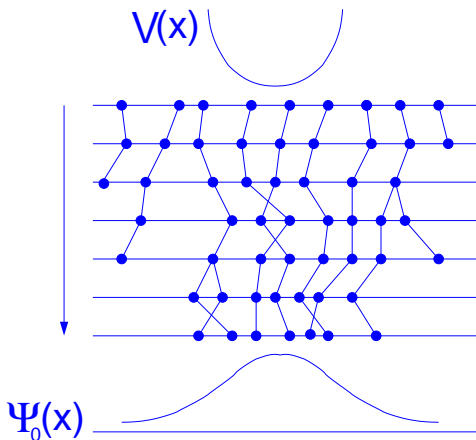
Walkers multiply/die in regions of lower/higher V than E_{T} (no imp. sampling)
Walkers multiply/die in regions of lower/higher E_{L} than E_{T} (with imp. sampling)

Sign Problem in PMC

The sign problem differs for the various PMC methods.
However, in all PMC methods the underlying cause is that a state other than the desired state grows exponentially compared to the desired state, combined with the fact that since we are sampling states, cancellations of opposite sign contributions are relatively ineffective.

In DMC we saw that we sample not $\Psi_0(\mathbf{R})$ but $\Psi_T(\mathbf{R})\Psi_0(\mathbf{R})$ using the importance-sampled projector.
This sneaks in the fixed-node approximation, since we are projecting onto the lowest state that has the same nodes $\Psi_T(\mathbf{R})$ rather than the global ground state.

Fermion Nodes - a simple case

Consider a He atom in its 1^1S ground state. What are its nodes?

Fermion Nodes - a simple case

Consider a He atom in its 1^1S ground state. What are its nodes?
It has none!

Consider a He atom in its 1^3S state. What are its nodes?

Fermion Nodes - a simple case

Consider a He atom in its 1^1S ground state. What are its nodes?
It has none!

Consider a He atom in its 1^3S state. What are its nodes?
$r_1 = r_2$ ($\mathbf{r}_1 = \mathbf{r}_2$ is co-dimension 2 from $r_1 = r_2$.)
Proof: Suppose $r_1 = r_2$.
If we rotate by $180°$ about line joining nucleus to the midpoint of the 2 electrons, $\Psi \to \Psi$.
If we exchange the electrons, $\Psi \to -\Psi$.
So, $\Psi = 0$ when $r_1 = r_2$.

Sign Problem

Except for some special cases, there is a sign problem, and there is a FN error.

Diffusion Monte Carlo

Physical dimension of space	d
Number of parallel-spin electrons	N
Dimension of wavefunction	dN
Dimension of nodal surface	$dN - 1$
Dimension of particle coincidences	$dN - d$

So, in 1-d the nodal surface is known (when particles cross) and DMC does not have a sign problem.

Another special case: AFQMC does not have a sign problem for the 1/2-filled Hubbard model.

What if we use the projector without importance sampling?

Cyrus J. Umrigar

Sign Problem in DMC

$\hat{P}(\tau) = e^{\tau(E_T\hat{1}-\hat{H})}$. $|\phi_i\rangle = |\mathbf{R}\rangle$

$\langle\mathbf{R}|\hat{P}(\tau)|\mathbf{R}'\rangle \approx \dfrac{e^{\dfrac{-\left(\mathbf{R}-\mathbf{R}'\right)^2}{2\tau}+\left(E_T-\dfrac{\mathcal{V}(\mathbf{R})+\mathcal{V}(\mathbf{R}')}{2}\right)\tau}}{(2\pi\tau)^{3N/2}}$ is nonnegative.

So, where does the sign problem come from?

Sign Problem in DMC
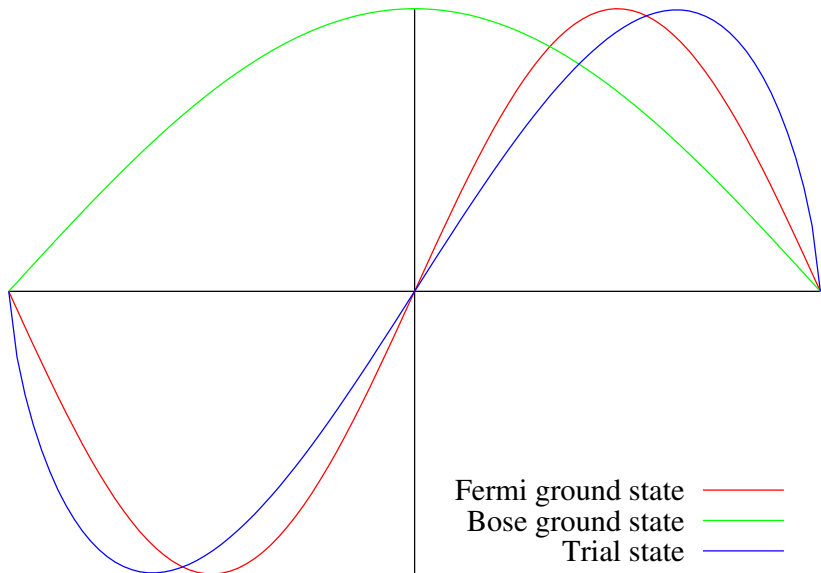
$\hat{P}(\tau) = e^{\tau(E_T \hat{1} - \hat{H})}. \; |\phi_i\rangle = |\mathbf{R}\rangle$

$\langle \mathbf{R} | \hat{P}(\tau) | \mathbf{R}' \rangle \approx \dfrac{e^{\dfrac{-\left(\mathbf{R} - \mathbf{R}'\right)^2}{2\tau} + \left(E_T - \dfrac{\mathcal{V}(\mathbf{R}) + \mathcal{V}(\mathbf{R}')}{2} \right)\tau}}{(2\pi\tau)^{3N/2}}$ is nonnegative.

So, where does the sign problem come from?

Problem: Since the Bosonic energy is always lower than the Fermionic energy, the projected state is the Bosonic ground state.
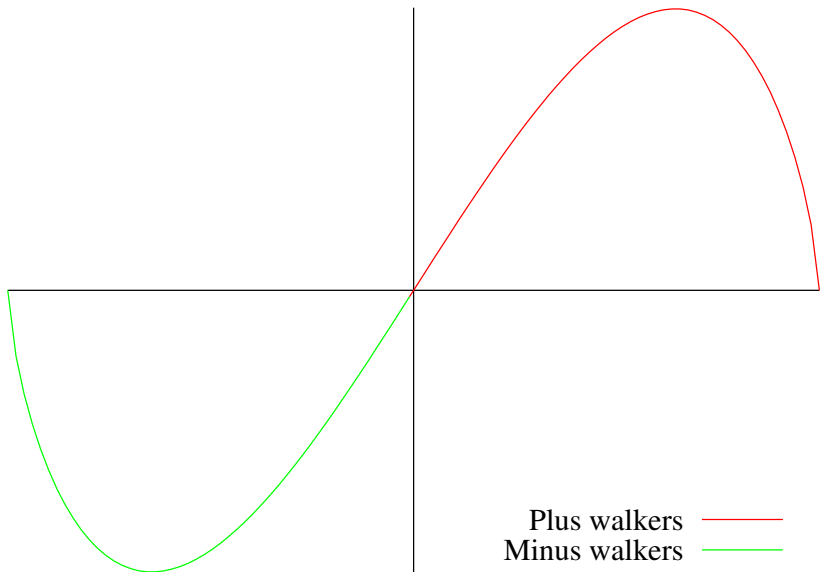
Fixed-node approximation

All except a few calculations (release-node, Ceperley) are done using FN approximation. Instead of doing a free projection, impose the boundary condition that the projected state has the same nodes as the trial state $\Psi_T(\mathbf{R})$.

This gives an upper bound to the energy and becomes exact in the limit that Ψ_T has the same nodes as Ψ_0.
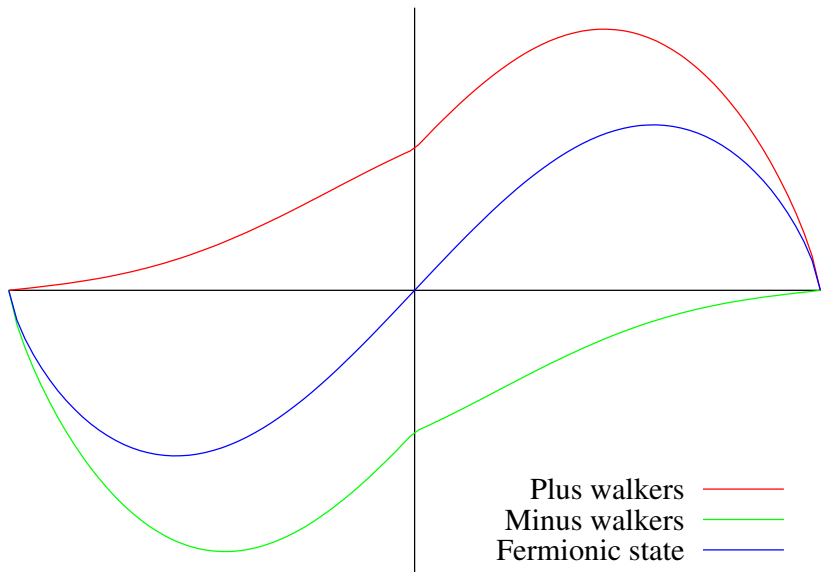
Cyrus J. Umrigar

Sign Problem in 1st Quantization and R space

Fermi ground state
Bose ground state
Trial state

Cyrus J. Umrigar

Sign Problem in 1st Quantization and R space

Plus walkers ———
Minus walkers ———

Sign Problem in 1st Quantization and R space

Plus walkers
Minus walkers
Fermionic state

Cyrus J. Umrigar

Sign Problem in 1st Quantization and R space

Plus walkers
Minus walkers
Fermionic state

Sign Problem in 1st Quantization and R space

Plus walkers ——
Minus walkers ——
Fermionic state ——

Cyrus J. Umrigar
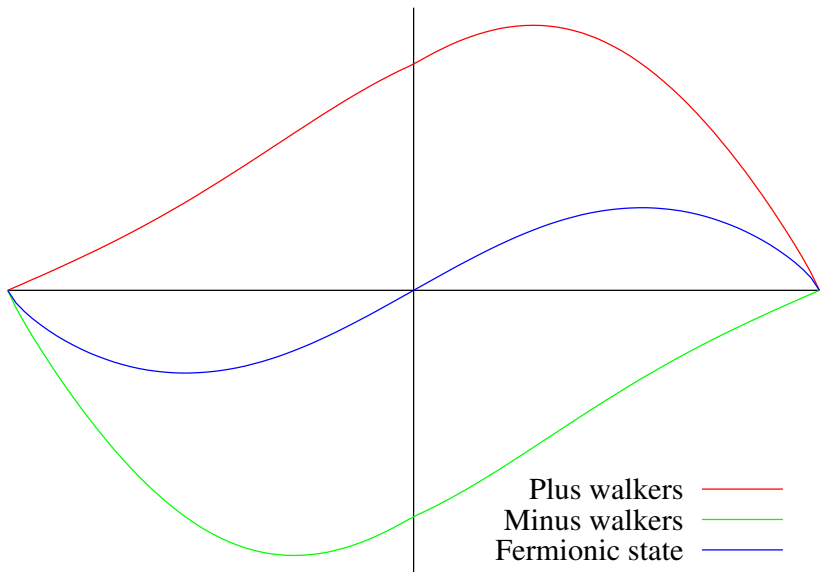
Sign Problem in 1st Quantization and R space
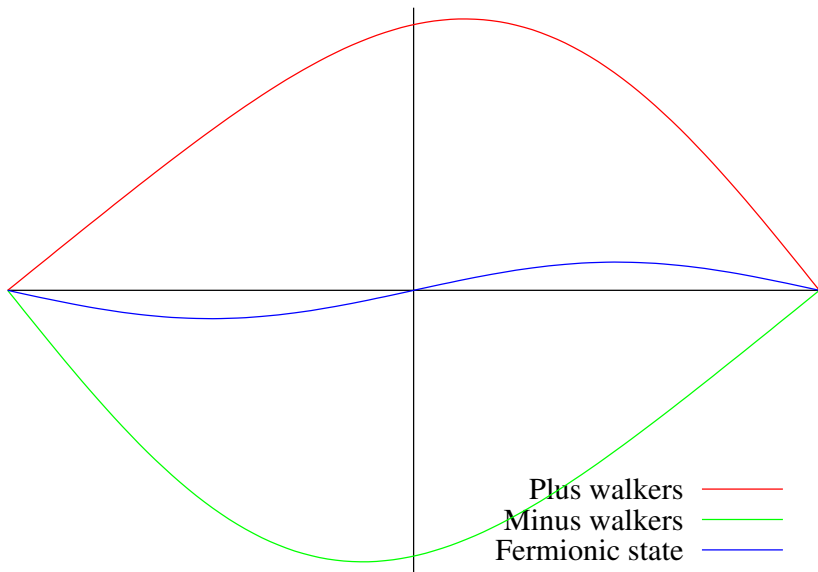
Plus walkers
Minus walkers
Fermionic state
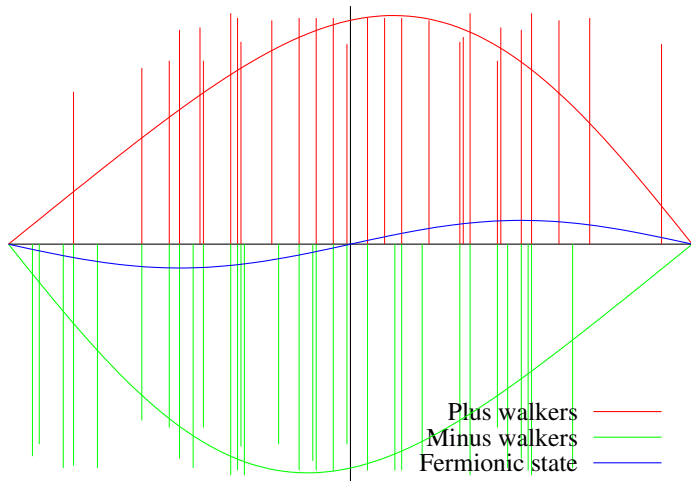
Problem: In large space walkers rarely meet and cancel, so tiny signal/noise! Further, if there are many cancellations, eventually there will be exclusively walkers of one sign only and a purely Bosonic distribution.

Cyrus J. Umrigar

Sign Problem in 2nd quantization

It would appear from the above discussion that one could eliminate the sign problem simply by using an antisymmetrized basis. In that case there are no Bosonic states or states of any other symmetry than Fermionic, so there is no possibility of getting noise from non-Fermionic states. Is that the case?

Sign Problem in 2nd quantization

It would appear from the above discussion that one could eliminate the sign problem simply by using an antisymmetrized basis. In that case there are no Bosonic states or states of any other symmetry than Fermionic, so there is no possibility of getting noise from non-Fermionic states. Is that the case?

No!

Sign Problem in 2nd quantization

Walk is done in the space of determinants.

Since Bosonic and other symmetry states are eliminated, there is some hope of having a stable signal to noise, but there is still a sign problem.

Problem: Paths leading from state i to state j can contribute with opposite sign. Further, Ψ and $-\Psi$ are equally good.

The projector in the chosen 2nd-quantized basis does not have a sign problem if:
The columns of the projector have the same sign structure aside from an overall sign, e.g.

$$P\Psi = \begin{bmatrix} + & - & + & + \\ - & + & - & - \\ + & - & + & + \\ + & - & + & + \end{bmatrix} \begin{bmatrix} + \\ - \\ + \\ + \end{bmatrix} = \begin{bmatrix} + \\ - \\ + \\ + \end{bmatrix}$$

or equivalently:
It is possible to find a set of sign changes of the basis functions such that all elements of the projector are nonnegative.

The sign problem is an issue only because of the stochastic nature of the algorithm.
Walkers of different signs can be spawned onto a given state in different MC generations.

Cyrus J. Umrigar