Projector Monte Carlo

Saverio Moroni
DEMOCRITOS IOM-CNR and SISSA, Trieste

a stochastic implementation of the power method for:

- exact ground state of many-boson systems in polynomial time
- exact ground state of many-fermion systems in exponential time
- “supervariational” ground state of many-fermion systems in polynomial time

for given Ψ, better than $\langle \Psi | H | \Psi \rangle$

- (excited states)

Outilne

- Representative projection Monte Carlo simulations
- The power method
- Random walks, Metropolis algorithm
- Variational Monte Carlo, correlated wave functions
- Projection Monte Carlo
  - diffusion Monte Carlo
  - the sign problem
  - fixed-node approximation
  - variational path integral
- Variational and diffusion Monte Carlo simulation of the 2D electron gas

PHYSICAL REVIEW A VOLUME 9, NUMBER 5 MAY 1974

Helium at zero temperature with hard-sphere and other forces

M. H. Kalos *
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012

D. Levesque and L. Verlet
Laboratoire de Physique Théorique et Hautes Energies, Orsay, France †
(Received 22 August 1973)

Various theoretical and numerical problems relating to heliumlike systems in their ground states are treated. New developments in the numerical solution of the Schrödinger equation permit the solution of 256-body systems with hard-sphere forces. Using periodic boundary

A typical simulation took 20 hours on a CDC6600 (1CPU @ 10MHz, 1Mbyte memory)

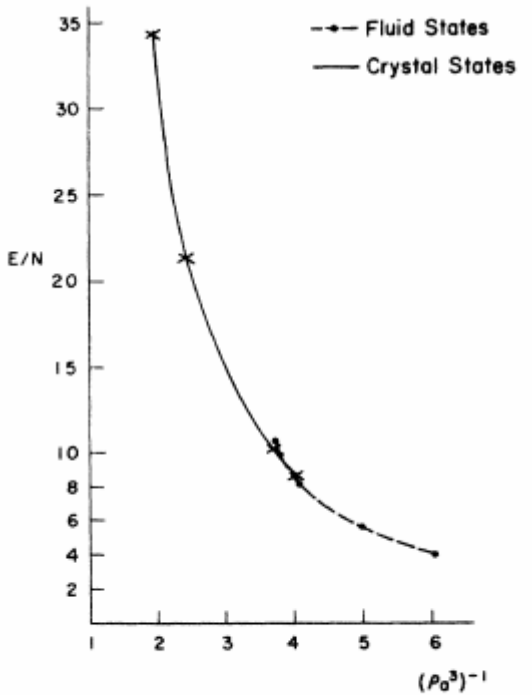


FIG. 2. "Exact" energy per particle as a function of $1/\rho a^3$.

Ground State of the Electron Gas by a Stochastic Method

D. M. Ceperley

National Resource for Computation in Chemistry, Lawrence Berkeley Laboratory, Berkeley, California 94720

and

B. J. Alder

Lawrence Livermore Laboratory, University of California, Livermore, California 94550

(Received 16 April 1980)

An exact stochastic simulation of the Schroedinger equation for charged bosons and fermions has been used to calculate the correlation energies, to locate the transitions to their respective crystal phases at zero temperature within 10%, and to establish the stability at intermediate densities of a ferromagnetic fluid of electrons.

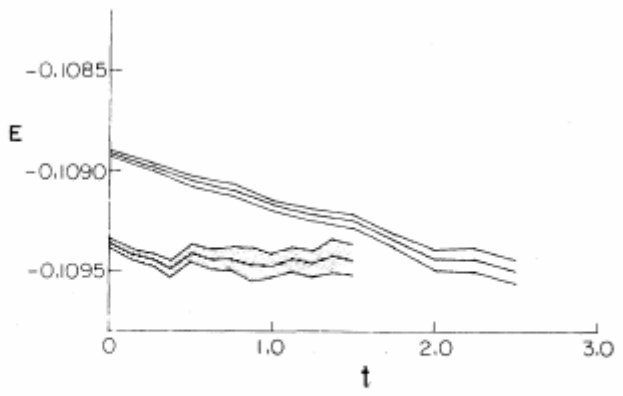


FIG. 1. The energy in rydbergs per particle of a 38-electron system at the density $r_s = 10$ vs diffusion time (in inverse Rydbergs) from removal of the fixed nodes.

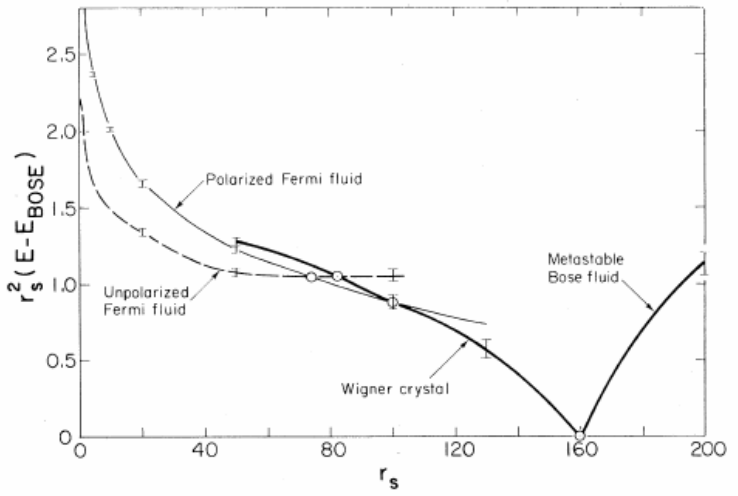


FIG. 2. The energy of the four phases studied relative to that of the lowest boson state times r_s^2 in rydbergs vs

energy vs. variance of the local energy
for wave functions of different quality

PHYSICAL REVIEW LETTERS **124,** 206404 (2020)

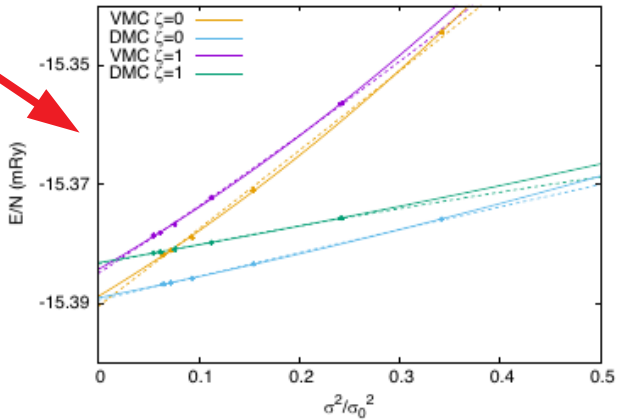
Itinerant-Electron Magnetism: The Importance of Many-Body Correlations

Markus Holzmann
Univ. Grenoble Alpes, CNRS, LPMMC, 38000 Grenoble, France
and Institut Laue Langevin, BP 156, F-38042 Grenoble Cedex 9, France

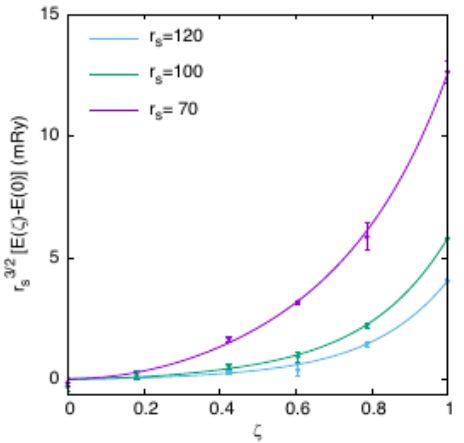
Saverio Moroni
CNR-IOM DEMOCRITOS, Istituto Officina dei Materiali, and SISSA Scuola Internazionale Superiore di Studi Avanzati, Via Bonomea 265, I-34136 Trieste, Italy

(Received 8 October 2019; accepted 5 May 2020; published 21 May 2020)

Do electrons become ferromagnetic just because of their repulsive Coulomb interaction? Our calculations on the three-dimensional electron gas imply that itinerant ferromagnetism of delocalized electrons without lattice and band structure, the most basic model considered by Stoner, is suppressed due to many-body correlations as speculated already by Wigner, and a possible ferromagnetic transition lowering the density is precluded by the formation of the Wigner crystal.



polarization energy



improved structure and optimization of wave functions,
better control of size effects, faster computers

The power method

$M\mathbf{v}_i = \lambda_i \mathbf{v}_i$ M is a $N \times N$ matrix with eigenvalues λ_i and eigenvectors $\mathbf{v}_i$

$\max\{\lambda_i\} = \lambda_N$ λ_N is the largest eigenvalue

$\mathbf{v} = \sum_i c_i \mathbf{v}_i$ expansion of a generic state in eigenstates of M

The application of M^P to $\mathbf{v}$ projects out the eigenvector with the largest eigenvalue:

$$M^P \mathbf{v} = \sum_i c_i \lambda_i^P \mathbf{v}_i = c_N \lambda_N^P \left[\mathbf{v}_N + \sum_{i<N} \frac{c_i}{c_N} \left(\frac{\lambda_i}{\lambda_N} \right)^P \mathbf{v}_i \right] \qquad \lim_{P \to \infty} \left(\frac{\lambda_i}{\lambda_N} \right)^P = 0$$

This can be done iteratively, so we only need to (repeatedly) apply M to a state:

$$M^P \mathbf{v} = M \left(M^{P-1} \mathbf{v} \right) = M \mathbf{v}'$$

The power method

iterative eigensolvers:

deterministic	stochastic
Examples: Lanczos, Davidson	Examples: PMC, random walk sampling
given $\mathbf{v}$ calculate $M\mathbf{v}$	given a sample from $\mathbf{v}$ get a sample from $M\mathbf{v}$
problem: the full many-body state Has $\sim \exp(N_P)$ components	problem: can't sample from a signed solution

Monte Carlo integration

suppose we want to calculate $\langle f \rangle = \int \widetilde{\pi}(x) f(x) dx$

$\begin{cases} x & \text{a configuration in } D \text{ dimensions} \\ \widetilde{\pi}(x) \geq 0, \quad \int \widetilde{\pi}(x) dx = 1 & \\ \widetilde{\pi}(x) & \text{a probability distribution} \end{cases}$

example: $\widetilde{\pi} = e^{-\beta V}/Z, \qquad \langle f \rangle = \frac{1}{Z} \int e^{-\beta V} f dx$ canonical thermal average

- sample a set $\{x_1, x_2, \ldots, x_K\}$ of configurations from the distribution $\widetilde{\pi}(x)$
- the sample average $\overline{f} = \frac{1}{K} \sum_{k=1}^{K} f_k$ is an unbiased estimate of $\langle f \rangle$
- if the variance $\langle (f - \langle f \rangle)^2 \rangle$ of f is finite $\overline{f}$ is normally distributed (for large K);

  the standard deviation of $\overline{f}$ is estimated from the data as $\epsilon = \sqrt{cV/K}$

  where $V = \frac{1}{K-1} \sum_{k=1}^{K} (f_k - \overline{f})^2$

  $c = 1 + 2 \sum_{i=1}^{i_{\text{cut}}} c(i)$

  $c(i) = \frac{1}{V} \frac{1}{K-1} \sum_{k=1}^{k_2 - i} (f_k - \overline{f})(f_{k+i} - \overline{f})$

$\propto 1/\sqrt{K}$ scaling
of statistical error independent of *D*

Random walk sampling gives correlated data ($c > 1$) but can sample arbitrary distributions

Random walks (i.e. how to sample from an arbitrary probability using the power method)

In a random walk we change the configuration s of our system (e.g. the coordinates of all the particles) to a configuration s' randomly sampled with probability $M(s, s')$, thus generating a sequence $\{s_0, s_1, s_2, \ldots\}$ ← Markov chain

M is a probability, hence $M(s, s') \geq 0$ and $\sum_{s'} M(s, s') = 1$: it is a "stochastic matrix".

If a stochastic matrix is ergodic, it has a non-negative (left) eigenstate $\pi(s)$ with eigenvalue 1,

$$\sum_s \pi(s) M(s, s') = \pi(s')$$ ← unique stationary state

and all the other eigenvectors $\mathbf{v}_i(s)$ have eigenvalues $\lambda_i < 1$, hence

$$\mathbf{v}^{(P)} = \mathbf{v} M^P = \pi + \sum_i c_i \lambda_i^P \mathbf{v}_i \xrightarrow[P \to \infty]{} \pi$$

If s_0 is sampled from the (positive) state $\mathbf{v}$, the configuration s_1 will be sampled from $\mathbf{v}^{(1)}$ and so on, and after the stationary state is reached for P large enough all configurations will be sampled from π.

Random walks (i.e. how to sample from an arbitrary probability using the power method)

Now, we don't want to sample from the stationary state π of a given M, but from a given π

- A useful property: detailed balance

$$\pi(s)M(s,s') = \pi(s')M(s',s) \implies \sum_s \pi(s)M(s,s') = \pi(s')$$

- Metropolis algorithm: $M(s,s') = T(s,s')A(s,s')$
  - choose a probability $T(s,s')$ of proposing the move
    - ergodic
    - can be directly sampled
  - calculate the probability $A(s,s')$ of accepting the move using the given π:

$$A(s,s') = \min\left\{1, \frac{\pi(s')}{\pi(s)}\frac{T(s',s)}{T(s,s')}\right\}$$

    - only ratio of π's needed
  - the resulting $M(s,s')$ does verify the detailed balance condition, hence it does have the given π as its unique stationary state

Variational Monte Carlo (in coordinate representation, $R = \{\mathbf{r}_1, \ldots, \mathbf{r}_{N_P}\}$)

$E_V = \langle\Psi|H|\Psi\rangle / \langle\Psi|\Psi\rangle \geq E_0$ definition of the variational energy

$\langle\Psi|H|\Psi\rangle = \int \frac{\langle R|\Psi\rangle}{\langle R|\Psi\rangle} \langle\Psi|R\rangle\langle R|H|\Psi\rangle dR = \int \pi(R) E_L(R) dR$

$E_L(R) = \langle R|H|\Psi\rangle / \langle R|\Psi\rangle = [(-\frac{\hbar^2}{2m}\nabla^2 + V)\Psi(R)]/\Psi(R)$ local energy

$\pi(R) = |\Psi(R)|^2$ (unnormalized) probability density

$$E_V = \frac{\int \pi(R) E_L(R) dR}{\int \pi(R) dR} = \overline{E}_L \pm \sqrt{\frac{c(\overline{E_L^2} - \overline{E}_L^2)}{K-1}}$$

Monte Carlo estimate with statistical error

$\overline{f} \equiv 1/K \sum_{k=1}^{K} f(R_k)$ sample average

$\{R_1, \ldots, R_K\}$ configurations sampled from $\pi(R)$ after equilibration

c is the autocorrelation time of the data $\{f(R_1), \ldots, f(R_K)\}$

(we have K/c independent samples)

Correlated wave functions

For the sake of practical reference, we consider the homogeneous electron gas (energy units of Ry and length units of $r_s a_0$, where $r_s = (3/4\pi\rho)^{1/3}/a_0$)

$$H = -\frac{1}{r_s^2}\nabla^2 + \frac{2}{r_s}\sum_{i<j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \text{const}$$

We start from a mean-field solution and construct a hierarchy of improved wave function making the local energy smoother and smoother. The rationale is that for an exact eigenstate Φ_i of H the local energy $E_L(R) = \langle R|H|\Phi_i\rangle / \langle R|\Phi_i\rangle = E_i$ is a constant

- level 0 of the hierarchy: Hartree wave function $\Psi_0 = \prod_{i=1}^{N_P} \exp(i\mathbf{k}_i \cdot \mathbf{r}_i)$
  - we assume spin-polarized electrons
  - the antisymmetrization can be done afterwards, on the improved wave function
  - the local energy is $1/r_s^2 \sum_i k_i^2 + 2/r_s \sum_{i<j} 1/r_{ij}$
  - No choice of one-particle orbitals can modify the red term: we need a pair-product (Jastrow) factor in the wave function.

Correlated wave functions

- level 1: Jastrow wave function $\Psi_1 = \prod_{i=1}^{N_P} \exp(i\mathbf{k}_i \cdot \mathbf{r}_i) \exp[-\sum_{i<j} u(r_{ij})]$
  - the divergence of the potential is removed from the local energy by the cusp condition,
    $$\lim_{r\to 0} du(r)/dr = -1/4r_s$$
  - the long-range behavior is given by the zero-point motion of the plasmon,
    $$\lim_{r\to \infty} u(r) = \sqrt{r_s/3}/r$$
  - the RPA pseudopotential interpolates between these two limits
    $$2u_{RPA}(k) = -1/S_0(k) + \left[1/S_0(k)^2 + 2V(k)r_s^2/k^2\right]^{1/2}$$
    (otherwise one can parametrize *u(r)* and optimize numerically)

  - however no choice of *u(r)* can eliminate new terms in the local energy such as
    $u'(r_{ij})u'(r_{ik})\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{r}}_{ik}$ or $u'(r_{ij})\mathbf{k}_i \cdot \hat{\mathbf{r}}_{ij}$

    we need three-body correlations and combination of orbitals with pair correlations

Correlated wave functions

- level 2: Jastrow-three-body-backflow wave function

$$\Psi_2 = \prod_{i=1}^{N_P} \exp(i\mathbf{k}_i \cdot \mathbf{x}_i) \exp[-\sum_{i<j} u(r_{ij}) - \sum_i \mathbf{G}_i \cdot \mathbf{G}_i]$$

$$\mathbf{x}_i = \mathbf{r}_i + \sum_{j \neq i} \eta(r_{ij})(\mathbf{r}_i - \mathbf{r}_j)$$

$$\mathbf{G}_i = \sum_{j \neq i} \xi(r_{ij})(\mathbf{r}_i - \mathbf{r}_j)$$

- ...

many other options (different mean-field starting points such as geminals or pfaffians, multideterminants, generalized and/or iterated backflow, neural networks, ...)

Projection Monte Carlo

$\mathbf{v}(s) \longrightarrow \Psi(R)$ a trial function

$M \longrightarrow e^{-\tau(H-E_T)}$ a function of the Hamiltonian which projects out the ground state

evolution in imaginary time (other functions of *H* can be used)

a step of the power method: $\Psi^{(p+1)}(R) = \int G(R,R',\tau)\Psi^{(p)}(R')dR'$

$\langle R|e^{-\tau(H-E_T)}|R'\rangle$

$\lim_{P\to\infty} \Psi^{(P)}(R) = c_0 e^{-P\tau(E_0-E_T)}\Phi_0(R)$

exact ground state energy and wave function

Stochastic implementation: like in a Markov chain (assuming a positive wave function)
given R_P sampled from $\Psi^{(P)}(R)$ we get a sample R_{P+1} of $\Psi^{(p+1)}(R)$
using the matrix G

Projection Monte Carlo: diffusion Monte Carlo

- $\lambda = \hbar^2/2m$
- ξ uniform deviate in (0,1)
- N_w current n. of walkers
- E_{e} current est. of energy

in coordinate representation a short-time approximation of G is

$$G(R, R', \tau) = e^{-(R-R')^2/4\lambda\tau} e^{-\tau[V(R)-E_T]} + \mathcal{O}(\tau^2)$$

- the first term (diffusion) implies a gaussian displacement which can be directly sampled
- the second term (growth/decay) implies multiplication of the distribution by a positive weight

The combined process is simulated with a branching random walk:

- initialize N_w^* walkers $\{R_{i;0},\ \ i = 1, \ldots, N_w^*\}$ sampled from $\Psi(R)$ (the 0th generation)
- iterate: advance each walker $R_{i;k}$ to the $(k+1)^{\mathrm{th}}$ generation:

  displace $R_{i;k}$ to a position R' sampled from $\exp[-(R - R_{i;k})^2/4\lambda\tau]$ (diffusion)
  make a number $\mathrm{int}\{\xi + \exp[-\tau(V(R') - E_T)]\}$ of copies of R' (branching)

- every so often adjust $E_T \longrightarrow E_{\mathrm{e}} + \alpha \ln(N_w^*/N_w)$ to control the population
- after P generations (stationary distribution reached) accumulate averages

The extrapolation of the results to $\tau \to 0$ and $N_w^* \to \infty$ is exact within the statistical error

Projection Monte Carlo: importance sampled diffusion Monte Carlo

The scheme described suffers from large fluctuations of the potential in the branching factor.
Diffusion Monte Carlo is made vastly more efficient using importance sampling

$$\Psi^{(p+1)}(R) = \int G(R,R',\tau)\Psi^{(p)}(R')dR'$$
$$\Psi(R)\Psi^{(p+1)}(R) = \int \Psi(R)G(R,R',\tau)/\Psi(R')\Psi(R')\Psi^{(p)}(R')dR'$$
$$f^{(p+1)}(R) = \int \widetilde{G}(R,R',\tau)f^{(p)}(R')dR'$$

$\widetilde{G}(R,R',\tau) = \Psi(R)G(R,R',\tau)/\Psi(R')$ importance sampled Green's function

$f^{(P)}(R) = \Psi(R)\Psi^{(P)}(R) \underset{P\to\infty}{\longrightarrow} \Psi(R)\Phi_0(R)$ mixed distribution

short time approximation of $\widetilde{G}$:

$$\widetilde{G}(R,R',\tau) = e^{-(R-[R'+\lambda\tau\nabla \ln \Psi^2(R')])^2/4\lambda\tau} e^{-\tau[E_L(R)-E_T]} + \mathcal{O}(\tau^2)$$

new drift term

E_L replaces V in the branching factor

Projection Monte Carlo: importance sampled diffusion Monte Carlo

- The importance sampled process is simulated with a different branching random walk:
  - initialize N_w^* walkers $\{R_{i;0},\ \ i = 1, \ldots, N_w^*\}$ sampled from $\Psi^2(R)$ (the 0th generation)
  - iterate: advance each walker $R_{i;k}$ to the $(k+1)^{\text{th}}$ generation:

    displace $R_{i;k}$ to a position $R'' = R_{i;k} + \lambda\tau\nabla \ln \Psi^2(R_{i;k})$ (drift)
    further displace to a position R' sampled from $\exp[-(R - R'')^2/4\lambda\tau]$ (diffusion)
    make a number $\text{int}\{\xi + \exp[-\tau(E_L(R') - E_T)]\}$ of copies of (branching)

  - every so often adjust $E_T \longrightarrow E_e + \alpha \ln(N_w^*/N_w)$ to control the population
  - after P generations (stationary distribution reached) accumulate averages

- The stationary distribution of the random walk is $\Psi(R)\Phi_0(R)$

The extrapolation of the results to $\tau \to 0$ and $N_w^* \to \infty$ is exact within the statistical error

Projection Monte Carlo:calculating properties

Sampling the mixed distribution we can calcolate mixed estimators:

$$\langle\Psi|H|\Phi_0\rangle = \int \langle\Psi|H|R\rangle\langle R|\Phi_0\rangle dR = \int E_L(R)\Psi(R)\Phi_0(R)dR = E_0$$

this is good for the energy because of the low variance of E_L for good wave functions.

If $[A, H] \neq 0$ the mixed estimator is biased, .

$$A_{\text{mix}} = \langle\Psi|A|\Phi_0\rangle = \langle\Phi_0|A|\Phi_0\rangle + \mathcal{O}(\Psi - \Phi_0)$$

Extrapolated estimators, $2A_{\text{mix}} - A_{\text{var}}$ or $A^2_{\text{mix}}/A_{\text{var}}$ with $A_{\text{var}} = \langle\Psi|A|\Psi\rangle$ have a bias $\mathcal{O}((\Psi - \Phi_0)^2)$.

Unbiased estimates of $\langle\Phi_0|A|\Phi_0\rangle$ can be obtained with the “forward walking” technique, which however is numerically unstable, or with “variational path integral” methods, which however are not (yet) very popular.

Projection Monte Carlo: technicalities

Two examples of technical details that increase the efficiency significantly:

- Rejection step: walkers are displaced with the drift-diffusion part of $\widetilde{G}$,

$$\widetilde{G}_d = e^{-(R-[R'+\lambda\tau\nabla \ln \Psi^2(R')])^2/4\lambda\tau}$$

It is common practice to accept the move with probability $\min\left\{1, \frac{\Psi^2(R')}{\Psi^2(R)}\frac{\widetilde{G}_d(R,R',\tau)}{\widetilde{G}_d(R',R,\tau)}\right\}$
If the trial function is the exact ground state, this restores the detailed balance condition of the exact Green's function and eliminates the time step error. For approximate trial functions, it reduces the time step error allowing to use larger time steps with smaller autocorrelation times

- At the nodes the local energy diverges for approximate trial functions and the drift term diverges for both approximate and exact trial functions. Fluctuations, particularly in the branching factor, are reduced by replacing in the drift and in the branching

$$F = \nabla^2 \ln \Psi \longrightarrow \overline{F} = \frac{(1+\lambda\tau F^2)^{1/2}-1}{\lambda\tau F^2/2} F$$

and $E_L \longrightarrow \overline{E}_L = E_L \overline{F}/F$

$\lambda\tau\overline{F} \sim (4\lambda\tau)^{1/2}$ when $\lambda\tau F^2 \gg 1$

divergences of E_L and F cancel out

Sign problem

Fermionic wave functions have a sign. Exact Monte Carlo calculations are still possible but their computational cost scales exponentially with the system size.

$$E_F = \frac{\int \Psi' H \exp(-\beta H)\Psi dX}{\int \Psi' \exp(-\beta H)\Psi dX}$$

fermionic ground state energy.

Ψ and Ψ' are evaluated at configurations connected by a random walk and their signs S and S' may change. We can rewrite E_F in terms of positive quantities that can be sampled:

$$E_F = \frac{\int S'|\Psi'| H \exp(-\beta H)|\Psi| S dX / \int |\Psi'| \exp(-\beta H)|\Psi| dX}{\int S'|\Psi'| \exp(-\beta H)|\Psi| S dX / \int |\Psi'| \exp(-\beta H)|\Psi| dX}$$

both the numerator and the denominator can be calculated sampling the positive distribution $|\Psi'| \exp(-\beta H)|\Psi|$

Sign problem

analysis of the denominator:

$$\int S'|\Psi'|\exp(-\beta H)|\Psi|SdX / \int |\Psi'|\exp(-\beta H)|\Psi|dX$$

imaginary time evolution of a fermionic state: $\sim e^{-\beta E_F}$

imaginary time evolution of a bosonic state: $\sim e^{-\beta E_B}$

Our denominator is $\sim e^{-\beta(E_F - E_B)} \propto e^{-\beta N_p}$ exponentially small, and we need a correspondingly small statistical error ϵ to calculate E_F without dividing by ~zero .
The computational cost $\propto 1/\epsilon^2$ grows exponentially with the system size

Fixed Node Approximation

The nodal surface S of a fermion wave function Φ is the $(dN_P - 1)$-dimensional set defined by the equation $\Phi = 0$ where d is the dimension of the physical space.
For the ground state, S divides the dN_P-dimensional space of the configurations into equivalent nodal cells where Φ has the same sign, so that if the exact S were known one could calculate exact ground state properties without a sign problem restricting the random walk into one cell.

The FN approximation assumes that the nodes of the solution are the same as the nodes of the trial function Ψ and solves for Φ_{FN}, the lowest-energy eigenstate of H with that nodal surface.
It gives a lower upper bound to E_0 than the VMC energy, and it is exact if the nodes are exact.

It is easily implemented in Projection MC by rejecting the move if the sign of Ψ changes

Projection Monte Carlo: Variational Path Integral

$\Psi_\beta = e^{-\beta H}\Psi = \left(e^{-\tau H}\right)^P \Psi, \qquad \tau = \beta/P$ imaginary time evolution of the trial function

$\Psi_\beta(R_0) = \int dR_1 \ldots dR_P \left[\prod_{i=1}^{P} G(R_i, R_{i-1}, \tau)\right] \Psi(R_P)$ path integral representation

$\langle \Psi_\beta | A | \Psi_\beta \rangle = \int dR_0 \Psi_\beta(R_0) A(R_0) \Psi_\beta(R_0)$ ground-state matrix element

unbiased estimators obtained by Metropolis sampling in an enlarged configuration space:

$$A_\beta = \frac{\int dX \Psi(R_{-P}) \left[\prod_{i=-P+1}^{P} G(R_i,R_{i-1})\right] \Psi(R_P) A(R_0)}{\int dX \Psi(R_{-P}) \left[\prod_{i=-P+1}^{P} G(R_i,R_{i-1})\right] \Psi(R_P)} = \frac{\int dX \pi(X) A(R_0)}{\int dX \pi(X)} \qquad X=\{R_{-P},\ldots,R_P\}$$

π is positive for bosons and fixed-node fermions

results are exact (bar the FNA) in the limit of large β and small τ

- No population control error
- No importance sampling (an advantage for poor trial functions or large systems)
- Unbiased estimators, including nonlocal properties (OBDM)
- Imaginary time correlation functions: static and (some) dynamic linear response