

Wanders in Diffusion Monte Carlo

Nick Underwood
Department of Physics
University of York
Heslington
United Kingdom
YO10 5DD

June 9, 2008

Abstract

We outline the theory behind Diffusion Monte Carlo, and provide a firm link to its between the concept of a projector algorithm and its stochastic implementation. We critically analyse a previous developed DMC code, and highlight and correct several important mistakes within it. We develop a run-time graphical analysis tool with which we use to assess the internal consistency of the code.

Contents

0.1	Introduction	3
0.1.1	Background	3
0.1.2	Diffusion Monte Carlo key ideas	3
	DMC as a projection algorithm	3
	The Green's function for a simple diffusion equation	5
	The Stochastic simulation of the itSe	6
	Normalisation cured	8
0.2	Fermions, the FNA, and Importance sampling	9
0.2.1	The need for a real positive wavefunction	9
	Properties of Fermion nodes	9
	The tiling theorem	10
	The fixed node approximation (FNA)	10
	The fixed node variational principle	10
0.2.2	Importance Sampling	10
0.2.3	Trial functions	12
	The problem in hand	12
	Slater-Jastrow trial functions	12
	More sophisticated trial functions	14
0.3	Testing of the code	15
	Evaluation of the coulomb potential	15
	Population instability	16
	The need for, and the implementation of real-time graphical analysis	16
0.3.1	Run-Time Analysis	17
0.3.2	Kinetic energy and internal consistency	19
	Evaluation of the kinetic energy	22
	Debugging the kinetic energy	22
	Confusion over the Gaussian random deviates	23
	Further investigation into the Internal consistency	23
	A comment on the Yukawa pseudopotential	24
0.4	Conclusions	26
	Derivation of the isitSe	27
	Derivation of the importance sampled Green's func- tion in 1D	27

0.1 Introduction

0.1.1 Background

The Group of methods collectively known as quantum Monte Carlo (QMC) occupy a very important niche in *ab initio* condensed matter theory. Typically achieving an order of magnitude greater precision than density functional theory [Foulkes et al., 2001], QMC methods are used when the need for precision is paramount. Monte Carlo methods are particularly applicable to many-body problems due to their efficiency when dealing with many-dimensional problems. The configurational space of a non-relativistic many-body problem is a Euclidean space with the dimensionality of the product of the number of particles and the dimensionality of the space, hereby written \mathcal{E}_{nd} . The wavefunction is defined within this in general high dimensional space, and most of this wavefunction will be *extremely* small. A simple metropolis procedure will seek out the regions in configurational space which have a significant contribution to the integration being performed. QMC comes in many different flavours, each with their pros and cons, however the two most prominent of these are called diffusion Monte Carlo (DMC) and Variational Monte Carlo (VMC).

Variational Monte Carlo is essentially a simple extension of the Rayleigh-Ritz method to situations in which the integral

$$\langle E \rangle = \frac{\langle \psi_T^\alpha | \hat{H} | \psi_T^\alpha \rangle}{\langle \psi_T^\alpha | \psi_T^\alpha \rangle} \quad (1)$$

cannot be performed analytically. Instead it is performed with a familiar metropolis sampling algorithm imposed upon a system of typically 100+ random walkers, where each walker occupies a position in \mathcal{E}_{nd} . The trial function being used represents an upper bound to the energy, and is as such optimised with respect to usually ~ 1000 variational parameters α . VMC, like Rayleigh-Ritz, depends entirely on the quality of the trial function being used, and despite the large number of parameters, it can rarely reach the accuracy of a projector algorithm like DMC. That said, it provides an invaluable role as importance sampled DMC makes use of this trial function. Also, recently more sophisticated trial functions when used in VMC, notably those with backflow effects have begun to rival DMC runs [Drummond et al., 2006].

0.1.2 Diffusion Monte Carlo key ideas

DMC as a projection algorithm

A projection map works in the following way:

$$pr_i : (x_1, x_2, \dots, x_n) \mapsto x_i \quad (2)$$

Simply, a DMC algorithm may be thought of as a projection operator acting on an initial trialfunction to project out the groundstate wavefunction[†] Firstly we introduce Eq. (3a), the imaginary time Schrödinger equation (itSe) in natural units, with an energy offset E_T the significance of which will be explained in due course.

$$(\hat{H} - E_T)\phi(\mathbf{R}, t) = -\frac{\partial}{\partial t}\phi(\mathbf{R}, t) \quad (3a)$$

$$\left(-\frac{1}{2}\nabla^2 + V(\mathbf{R}) - E_T\right)\phi(\mathbf{R}, t) = -\frac{\partial}{\partial t}\phi(\mathbf{R}, t) \quad (3b)$$

$$-\hat{L}\phi(\mathbf{R}, t) = -\frac{\partial}{\partial t}\phi(\mathbf{R}, t) \quad (3c)$$

Where $\nabla = (\nabla_1, \nabla_2 \dots, \nabla_n)$ and $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n)$; the analogues of the derivative operator, and the position coordinate are as expected. This is seen to be a diffusion equation in \mathcal{E}_{nd} , the equivalent integral equation of which is

$$\phi(\mathbf{R}, t + \tau) = \int G(\mathbf{R}, \mathbf{R}', \tau)\phi(\mathbf{R}', t)d\mathbf{R}'. \quad (4)$$

The Green's function of a diffusion equation can be formally written as Eq. (5) [Thijssen, 2007].

$$G(\mathbf{R}, \mathbf{R}'; t) = \langle \mathbf{R} | e^{t\hat{L}} | \mathbf{R}' \rangle \quad (5)$$

Now if look more closely at Eq. (4) we can find a very important property of the itSe. We may decompose the exponential operator down into an expansion in terms of eigenfunctions and their corresponding eigenvalues [Foulkes et al., 2001], viz $exp(-\tau\hat{L}) = \sum_i |\psi_i\rangle exp(-\tau E_i) \langle \psi_i|$, then we may write Eq. (4) as

$$\phi(\mathbf{R}, t + \tau) = \sum_i |\psi_i\rangle exp(-\tau(E_i - E_T))\langle \psi_i|\phi(t)\rangle. \quad (6a)$$

Where it is easily seen that the exponential factor in each term of the integral may be used as a damping factor by moving $E_T = E_0$ So that if we take the large τ limit, we may suppress all excited states, and project out the ground state wavefunction.

$$\lim_{\tau \rightarrow \infty} \phi(\mathbf{R}, t + \tau) = \psi_0(\mathbf{R})\langle \psi_0|\phi(t)\rangle \quad (7)$$

So it is seen that as in Eq. (2), if we are able to calculate this Green's function in the large τ limit then we will be able to project out the groundstate wavefunction. It will turn out that we are only able to get an approximate Green's function, valid only for small τ . This however is not a problem as repeated iterations of Eq. (4) lead to the same result. Indeed it will negate the

[†]extensions to excited states have been made, though extra complications are introduced [Ortiz et al., 1993]

problem of the inner product as this will be equal to unity. We are left with one problem however, as we have not defined a form of the wavefunction. Unless we can find a simple expression for the wavefunction then we have no way of expressing it for more than a few particles as numerical storage of the form would require values to be stored over the entirety of configurational space. This is completely unfeasible as the amount of data would scale exponentially with the number of particles. For instance if we were to consider a crude 3D space lattice model with ten lattice sites in each dimension, and 10 Bosonic particles then our configurational space would be \mathcal{E}_{30} . Hence we would have to store 10^{30} real numbers, equating to $\sim 3.7 \times 10^{21}$ GB of data storage in single precision [Godby, 2007]. Hence the full many body wavefunction for any more than a few particles is utterly unfeasible to even store on any computer available for the foreseeable future. Lets ignore these problems for now and concentrate on trying to find a form for the Green's function.

The Green's function for a simple diffusion equation

If we, for clarity, momentarily put $V(R) = E_T = 0$ and $\mathbf{R} = x$ then Eq. (3a) becomes a simple diffusion equation in one dimension.

$$-\frac{1}{2}\nabla^2\phi(x, t) + \frac{\partial}{\partial t}\phi(x, t) = 0 \quad (8)$$

then, using a unitary operator of the form of an integral over projection operators: $\int dp |p\rangle \langle p| = 1$, and $\langle x|p\rangle = \frac{1}{\sqrt{2\pi}}e^{ipx}$ we may calculate the Green's function explicitly as

$$G(x, x'; \tau) = \langle x| e^{-\tau\hat{p}^2/2} |x'\rangle \quad (9a)$$

$$G(x, x'; \tau) = \int dp \langle x|p\rangle \langle p| e^{-\tau\hat{p}^2/2} |p\rangle \langle x'|p\rangle^* \quad (9b)$$

$$= \frac{1}{2\pi} \int dp e^{-\frac{1}{2}\tau p^2} e^{ip(x-x')} \quad (9c)$$

Then completing the square in the exponent of the integrand of Eq. (9c) yields

$$G(x, x'; \tau) = \frac{1}{2\pi} e^{-\frac{1}{2\tau}(x-x')^2} \int_{-\infty}^{\infty} e^{-\frac{\tau}{2}(p-\frac{1}{\tau}(x-x'))^2} dp \quad (10a)$$

$$= \frac{1}{\pi\sqrt{2\tau}} e^{-\frac{1}{2\tau}(x-x')^2} \int_{-\infty}^{\infty} e^{-u^2} du \quad (10b)$$

$$= \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{1}{2\tau}(x-x')^2} \quad (10c)$$

Where the standard improper Gaussian integral $\int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{\pi}$ has been used. Eq. (10c) is a nicely normalised Gaussian for $\tau > 0$ with the property

that $\lim_{\tau \rightarrow 0} G(x, x'; \tau) = \delta(x - x')$. We must now remember that $V(R) \neq 0$, $E_T \neq 0$. In order to build this into the Green's function, we may take one of two different routes, both resulting in $\mathcal{O}[\tau^2]$ errors. We will follow Foulkes et al. [2001] in the use of the Trotter-Suzuki formula

$$e^{-\tau(\hat{A}+\hat{B})} = e^{-\tau\hat{B}/2}e^{-\tau\hat{A}}e^{-\tau\hat{B}/2} + \mathcal{O}[\tau^3] \quad (11)$$

So with $\hat{H} = \hat{p}^2/2$ and $\hat{B} = V(x) - E_T$

$$G(x, x'; \tau) = \langle x | e^{-\tau(\hat{p}^2/2 + V(x) - E_T)} | x' \rangle \quad (12a)$$

$$\approx e^{-\tau[V(x) - E_T]/2} \langle x | e^{-\tau(\hat{p}^2/2)} | x' \rangle e^{-\tau[V(x') - E_T]/2} \quad (12b)$$

$$= \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{1}{2\tau}(x-x')^2} e^{-\tau[V(x)+V(x')-E_T]/2} \quad (12c)$$

A simple extension to \mathcal{E}_{nd} results in

$$G(\mathbf{R}, \mathbf{R}', \tau) \approx \underbrace{\frac{1}{(2\pi\tau)^{3N/2}} \exp\left[-\frac{(\mathbf{R} - \mathbf{R}')^2}{2\Delta\tau}\right]}_Q \times \underbrace{\exp[-\Delta\tau(V(\mathbf{R}) + V(\mathbf{R}') - 2E_T)/2]}_P + \mathcal{O}[\tau^3] \quad (13)$$

This is our Green's function which we can use to project out the ground state many body wavefunction. As noted before, this must be done in repeated iterations of the integral itSe (Eq. (4)) as Eq. (13) is a short time approximation. It should be noted now that the factor Q as a Gaussian in \mathcal{E}_{nd} is well normalised, as was Eq. (10c), however the factor P ruins our normalisation. This will become significant in our subsequent analysis. We must now turn to the problem highlighted earlier with regard to the expressing of a many body-interacting wavefunction, and for this we require a little revision of Markovian chains. Unfortunately there isn't enough time to go satisfactorily into stochastic analysis here, however with due disregard for rigour a qualitative picture may be drawn. For more information see my previous essay [Underwood, 2007], or for a full exposition Feller [1962] come highly recommended.

The Stochastic simulation of the itSe

A Markovian chain is a sequence of trials defined by

$$P\{(X_1, X_2, \dots, X_n)\} = a_{X_1} p_{X_1 \rightarrow X_2} p_{X_2 \rightarrow X_3} p_{X_3 \rightarrow X_4} \dots p_{X_{n-2} \rightarrow X_{n-1}} p_{X_{n-1} \rightarrow X_n}. \quad (14)$$

Where $P\{(X_0, X_1, \dots, X_n)\}$ is the probability of a system starting from an initial state X_0 and being in the state X_1 after the first trial, and the state X_2 after the second etc. $p_{X_i \rightarrow X_{i+1}}$ is the probability that the system will

occupy the state X_{i+1} if it was in the state X_i before the previous trial, and a_{X_i} the absolute probability of the system being in the state X_i at any one time. This Markov chain is customarily called a walk, and the state of the system performing the walk a walker. Also each trial will here be considered a step in time, though this interpretation is only valid in some situations. The main characteristic of a Markov chain is that each trial is only dependent upon the current state of the system. We must ensure that this rule is not broken[†], at least not so much that the effects of which become statistically significant. A single step in a Markov chain is described by Eq. (15)[‡], the so called master equation of a Markov chain.

$$\rho(X, t + 1) - \rho(X, t) = - \int [p_{X \rightarrow X'} \rho(X, t) + p_{X' \rightarrow X} \rho(X', t)] dX' \quad (15)$$

Where $\rho(X, t)$ is a probability distribution of walkers in space and time. The master equation is satisfied over a large number of independent trials, whether these be from one walker over a large period of time, or over many independent walkers. We are specifically interested in ergodic chains, and stochastic theory prescribes two more conditions for a Markovian chain to be ergodic, viz.

Irreducibility We define the state of the system to be in a set C such that no state outside C may be reached from within C , or

$$C \quad \text{is closed iff} \quad P_{jk} = 0 \quad \forall \quad j \in C, k \notin C. \quad (16)$$

Such a set is called closed. We furthermore specify that all states within C within have a finite probability of occupation

Aperiodicity Periodicity is the condition that a walker may not return to a given state until after a finite time. We require the opposite.

Finite mean recurrence times For completeness, the requirement of finite mean recurrence times is a slight extension on Aperiodicity

It may be proven [Feller, 1962] that if a Markovian chain is ergodic then there exists a unique stationary distribution of states $\{u_k\}$. Furthermore the absolute probabilities $\{a_k\}$ tend to $\{u_k\}$ with increasing time. For this stationary situation $\rho(X, t) = \rho(X)$ and so the master equation becomes

$$\int p_{X \rightarrow X'} \rho(X) dX' = \int p_{X' \rightarrow X} \rho(X') dX'. \quad (17)$$

The parallels between the projection operator and the an ergodic Markov chain should now be extremely evident, and indeed we are now in a position

[†]It will prove necessary to break the rule in dealing with the nodal surface

[‡]This is the continuum space version

to tackle the problem stated at the end of section 0.1.2. Let us state the integral form of the itSe again

$$\phi(\mathbf{R}, t + \tau) = \int G(\mathbf{R}, \mathbf{R}', \tau) \phi(\mathbf{R}', t) d\mathbf{R}'. \quad (18)$$

But we already know that repeated iterations of this equation will project out the ground state, ie. $\phi(\mathbf{R}, t + \tau) = \phi(\mathbf{R}, t)$ for large t . Also worth noting now is the fact that all the Green's functions calculated so far have symmetric in \mathbf{R} and \mathbf{R}' . This is a property of all (exact) Green's functions. Substituting these two facts back into the RHS of Eq. (18) we find

$$\int G(\mathbf{R}, \mathbf{R}', \tau) \phi(\mathbf{R}') d\mathbf{R}' = \int G(\mathbf{R}', \mathbf{R}, \tau) \phi(\mathbf{R}) d\mathbf{R}'. \quad (19)$$

ie. an equation of exactly the same form as Eq. (17). So we simulate the itSe stochastically by interpreting $\phi(\mathbf{R}, t)$ to be a probability distribution which our walkers will move according to. Then, the Greens function may be considered to be the probability of a walker at \mathbf{R} in configurational space moving to \mathbf{R}' in a step through our imaginary time variable t of the size τ . In this way we never need ever express the wavefunction explicitly, and thus avoid the associated problems. All we need then do is average the quantity we are interested in over the walkers walks. I would like to note that I have not seen the stochastic interpretation presented in this way in any of the DMC literature I have read. This way of presentation greatly helped me understand the underlying theory behind the DMC method.

Normalisation cured

To round off our discussion of a simple DMC algorithm we must go back to the problem of the normalisation of the Green's function of Eq. (13) and the factor $P = \exp[-\Delta\tau(V(\mathbf{R}) + V(\mathbf{R}') - 2E_T)/2]$. If we are to treat the Eq. (13) as a probability distribution of a walker in \mathcal{E}_{nd} then we will have to account for an unnormalised probability. The assignment of walkers with weights would be the simplest extension to the algorithm. As noted previously however the wavefunction is very small in the majority of configurational space, and as such most walkers will become statistically insignificant, while still taking up valuable computer resources. An alternative solution is branching, or to extend our metaphor, a birth/death step for our walkers. This entails treating the excess of P over the largest integer a it is greater than as the probability that a walker will be there in the next step in addition to a that will be there for certain. This is conveniently a one line procedure in code Foulkes et al. [2001]

$$\text{number of walkers} = INT(P + \eta). \quad (20)$$

Where η is a uniformly distributed random deviate in the interval $0 \rightarrow 1$. This method wouldn't be practicable unless we could control the population in order to be computationally viable. By consideration of Eq. (3a) it is plain to see that by controlling E_T we may control the time derivative of our probability distribution. We must however take care that the population of walkers does not vary too rapidly, as after a walker has branched into two, these two take some time to become statistically independent of each other. This introduces a (small) systematic bias, the population control bias [Umrigar et al., 1993]. We may correct for any random deviations from a desired population with a negative feedback algorithm. This will be explained in more detail in section 0.3

0.2 Fermions, the FNA, and Importance sampling

0.2.1 The need for a real positive wavefunction

Our primitive DMC algorithm suffers from a few, as yet unhighlighted problems. The diffusion process described thus far will map out a probability distribution equal to $\phi_0(\mathbf{R})$. Unfortunately for us a wavefunction may be both negative and complex, and a probability distribution may not[†]. In the absence of a magnetic field the groundstate wavefunction is real [Foulkes et al., 2001], and so this is not difficult to deal with as will be highlighted in section 0.2.3. It is a simple proof to show that the Bosonic ground state wavefunction is positive everywhere. Fermions however obey the Pauli principle, and as such are antisymmetric under the exchange of both position and spin of any two particles. We will assume from now on, for reasons that will become apparent in section 0.2.3, that we have fixed the spin and as such the wavefunction is now only antisymmetric under that exchange of particles of the same spin.

Properties of Fermion nodes

Eq. (3a) is a smoothly varying function in \mathcal{E}_{nd} , and so $\phi_0(\mathbf{R}) = 0$ will define in general a $nd - 1$ surface in \mathcal{E}_{nd} . This surface defines the boundary between positive and negative regions of $\phi_0(\mathbf{R}) = 0$, and is what is called the nodal surface. If $d = 1$, so that $nd - 1 = n - 1$ then the nodal surface will be of the same dimensionality as the space each particle moves within. It then becomes plausible that the nodal surface will be exhausted by the $n - 1$ dimensional surface defined by any two particles occupying the same position. This unfortunately does not scale to higher dimensions, and so in general we do not know the form of these abstract hypersurfaces. Two theorems serve to mitigate this problem to a significant extent.

[†]For a discussion of complex wavefunctions in QMC see Ortiz et al. [1993]

The tiling theorem

A nodal pocket is defined as the region of configurational space $\Omega(\mathbf{R})$ which may be reached from any point within $\Omega(\mathbf{R})$ without crossing a node. The Tiling theorem of Ceperley [1991] states that all nodal pockets may be reached by permutation of the particle coordinates \mathbf{r}_i , $i = 1, \dots, n$. Alternatively stated; $\sum_P \Omega(P\mathbf{R}) = \text{all configurational space}$, where P is a permutation. Since we are dealing with identical particles this also implies that all the nodal pockets are equivalent, and we needn't worry which one a walker is in.

The fixed node approximation (FNA)

If we knew the positions of the nodes to a acceptable degree of accuracy the tiling theorem implies that conducting our DMC process inside a single nodal pocket would be equivalent to conducting it in the whole of configurational space. As long as the wavefunction did not change sign then it would not matter whether the wavefunction was positive or negative, as they could still both be simulated by our diffusion process. Walkers which strayed across the nodal surface would be judged to have a P reweighting factor of zero and would be subsequently deleted. This approximation was first suggested by Anderson [1975], and has become enormously famous, both because of its success in terms of results, and its place as the one great hurdle in the way of a truly exact theory.

The fixed node variational principle

The fixed node approximation owes a great deal of its success to the fact that it represents an upper bound to the true ground state energy. In DMC literature reference is often made to an improvement of the nodal surface. Unfortunately it is very difficult to say much about the actual form of these nodal hypersurfaces due to the sheer complexity involved. We interpret a lowering of the total energy calculated as an improvement of the nodal surface. The proof of the fixed node variational principle is given in Ceperley [1991].

0.2.2 Importance Sampling

Importance sampling is a regularly used technique in Monte Carlo integration used to improve efficiency, and usually to great effect. Importance sampling will improve the efficiency of the algorithm as it stands by several orders of magnitude. Crucially, it will also prove to be a practical way to impose the fixed node approximation. As explained already importance sampling involves the sampling of the product of the distribution

in question and a trial function of a similar shape. Here we shall sample $f(\mathbf{R}, t) = \phi_0(\mathbf{R}, t)\psi_T(\mathbf{R})$. We proceed by multiplying the itSe (3a) by $\psi_T(\mathbf{R})$, and by several implementations of the chain rule we may derive what we shall call from now on the importance sampled imaginary time Schrödinger equation (isitSe)[†].

$$\psi_T(\mathbf{R})(\hat{H} - E_T)\psi(\mathbf{R})^{-1}f(\mathbf{R}, t) = \underbrace{-\frac{1}{2}\nabla^2 f(\mathbf{R}, t)}_A + \underbrace{\nabla \cdot (\mathbf{v}_D(\mathbf{R})f(\mathbf{R}, t))}_B + \underbrace{(E_L(\mathbf{R}) - E_T)f(\mathbf{R}, t)}_C = -\frac{\partial f(\mathbf{R}, t)}{\partial t} \quad (21)$$

Where

$$\begin{aligned} v_D(\mathbf{R}) &= \frac{\nabla\psi_T(\mathbf{R})}{\psi_T(\mathbf{R})}, \\ E_L &= \frac{\hat{H}\psi_T(\mathbf{R})}{\psi_T(\mathbf{R})} = \frac{-\nabla^2\psi_T}{2\psi_T} + V(\mathbf{R}) \end{aligned} \quad (22)$$

are the drift velocity and local energy respectively. We shall deal with these two quantities a great deal. The three terms A , B and C in Eq. (21) describe diffusion, drift and growth/decay respectively [Umrigar et al., 1993]. This is reflected in the Green's function modified for importance sampling.

$$G(\mathbf{R}, \mathbf{R}', \tau) \approx \underbrace{\frac{1}{(2\pi t)^{3N/2}} e^{-\frac{\tau}{2}(E_L(\mathbf{R}) + E_L(\mathbf{R}') - 2E_T)}}_{G_b} \overbrace{e^{-(\mathbf{R} - \mathbf{R}' - \tau\mathbf{v}_D(\mathbf{R}))^2/2\tau}}^{G_d} \quad (23)$$

So now the algorithm works in the same way as before except as G_d suggests, the Gaussian term responsible for diffusion has been displaced by an amount $\tau\mathbf{v}_D(\mathbf{R})$. Now we must impose an nd dimensional drift term on the movement of our walkers. This drift term serves to push our walkers towards areas of large $\phi_0(\mathbf{R})$, ie. away from the nodal surface, significantly reducing the associated errors. Indeed, in the small τ limit, for a trial function with the correct nodal surface this would automatically impose the fixed node constraint. In order to deal with the few that will try to cross the surface then all we need do is evaluate the trial function at the attempted move, and check whether it has changed sign.

The final result of the DMC run will then depend entirely upon the quality of the nodal surface of the trial function. Trial functions may come from As is apparent there has deal a great deal of effort put toward trying to improve the trial functions nodal surface, exacerbated how little is actually

[†]A full derivation of the isitSe is given in the appendices

known about the surface itself. Recently however there has been a point of contention over one method of mitigation, though in order to appreciate this we need to understand something about form of the trial functions commonly used.

0.2.3 Trial functions

The problem in hand

I have made the discussion so far very general, but it is plain that the trial function must be specific to the problem being tackled. This project has been the continuation of a masters project of last year by O'Rourke. The code he developed was designed to tackle the two dimensional homogeneous electron gas in periodic boundary conditions. In addition to this it was designed to make the use of a relatively more sophisticated trial function involving a linear combination of determinants in the same way as configuration interaction functions are formed.

Slater-Jastrow trial functions

The most common type of trial function is of the Slater-Jastrow form [Jastrow, 1955]. This, put simply consists of the completely anti-symmetrised sum of the product one-particle orbitals in the form of a Slater determinant d . This then is multiplied by the Jastrow factor Ψ_J . The Jastrow factor is repulsive, and is an approximate way of accounting for electron correlation.

$$\psi_T = \Psi_J d = \Psi_J \begin{vmatrix} \psi_1(\mathbf{r}_1, \sigma_1) & \psi_2(\mathbf{r}_1, \sigma_1) & \dots & \psi_n(\mathbf{r}_1, \sigma_1) \\ \psi_1(\mathbf{r}_2, \sigma_2) & \psi_2(\mathbf{r}_2, \sigma_2) & \dots & \psi_n(\mathbf{r}_2, \sigma_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{r}_n, \sigma_n) & \psi_2(\mathbf{r}_n, \sigma_n) & \dots & \psi_n(\mathbf{r}_n, \sigma_n) \end{vmatrix} \quad (24)$$

The evaluation of a determinant is in general $\propto n^3$ in computer time, and as already noted we have fixed the spin of individual electrons, so it is general practice in QMC calculations to separate d into a product of determinants for up and down electrons.

$$\psi_T(\mathbf{R}) = \Psi_J d_\uparrow(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{n_\uparrow}) d_\downarrow(\mathbf{r}_{n_\downarrow+1}, \mathbf{r}_{n_\downarrow+2}, \dots, \mathbf{r}_n) \quad (25)$$

Although this trial function is not antisymmetric under exchange of spin antiparallel electrons, it can be shown to give the same expectation value as (24) for spin independent operators [Towler, 2006]. The Jastrow factor is of an exponential form which would normally be a warning sign in terms of

computational expense, however this will be taken care of in section 0.3.2

$$\Psi_J = \exp \left[- \sum_{i>j} u_{\sigma_i \sigma_j}(r_{ij}) \right] \quad (26a)$$

$$u = \frac{A_{\sigma_i \sigma_j} \left(1 - e^{r_{ij}/F_{\sigma_i \sigma_j}} \right)}{r_{ij}} \quad (26b)$$

The random phase approximation (RPA) requires that $u(r_{ij}) \rightarrow 1/r_{ij}$ for large r_{ij} . Also, though $V(r_{ij})$ diverges as $r_{ij} \rightarrow 0$, cancelling with the kinetic part of the Hamiltonian means that $u_{r_{ij}}$ is finite [Dunn and Broyles, 1966]. Becker et al. suggested the simple functional form (26b) in order to satisfy these constraints, and this was used by Ceperley [1978], Ceperley and Alder [1980] to produce some very accurate results for the HEG.

This is the $u(r_{ij})$ that was chosen by Schofield [2001], O'Rourke [2007], and is called the Yukawa pseudopotential, or difference of Yukawas. It was specifically chosen first for its simplicity, though also Ceperley [1978] has some good reliable results for the 2D HEG, useful for testing purposes.

The parameters $F_{\sigma_i \sigma_j}$, are spin independent and are chosen to meet the cusp conditions. These were derived in Schofield [2001] to be

$$F_{\uparrow\uparrow} = \sqrt{\frac{3}{2}A_{\uparrow\uparrow}} \quad F_{\uparrow\downarrow} = \sqrt{\frac{3}{2}A_{\uparrow\downarrow}} \quad (27)$$

with the further specification that $A_{\uparrow\uparrow} = A_{\downarrow\downarrow} = A$ the one spare free variable A was then taken from Ceperley [1978]. There is one important observation that we must now make. As the Jastrow factor is always positive, it does not affect the nodal surface at all. This means that though the convergence of a DMC algorithm will be very much affected by different Jastrow factors, the final answer should be the same. This proved to be extremely useful when testing for errors concerning the the evaluation of derivatives of Ψ_J . The VMC energy will of course be very much dependent upon Ψ_J . The one particle orbitals are usually obtained from other *ab initio* calculations, but as we are dealing with a simple homogeneous system we may use simple planes waves.

$$\psi(x, y) = \begin{cases} \cos(k_x x) \cos(k_y y) & \forall k \\ \sin(k_x x) \cos(k_y y) & \text{if } k_x \neq 0 \\ \cos(k_x x) \sin(k_y y) & \text{if } k_y \neq 0 \\ \sin(k_x x) \sin(k_y y) & \text{if } k_x \neq 0, k_y \neq 0 \end{cases} \quad (28)$$

where,

$$\mathbf{k} = \left(\frac{2\pi n_x}{L_x}, \frac{2\pi n_y}{L_y} \right) \quad n_x, n_y \in \{0, \mathbb{N}\} \quad (29)$$

In this way we have avoided the complication of a complex probability distribution. As we have used real functions instead of complex exponentials we must further specify that we have closed shells in order to satisfy the requirement of translational invariance of $\sum_n |\psi_i|^2$. For $n_\uparrow = n_\downarrow$ and in a square simulation cell ($L_x = L_y$) the allowed system sizes are outlined in table 1.

n_x	n_y	$\cos(k_x)\cos(k_y)$	$\cos(k_x)\sin(k_y)$	$\sin(k_x)\cos(k_y)$	$\sin(k_x)\sin(k_y)$	N
0	0	$\downarrow\uparrow$				2
0	1	$\downarrow\uparrow$	$\downarrow\uparrow$			10
1	0	$\downarrow\uparrow$		$\downarrow\uparrow$		10
1	1	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	18
0	2	$\downarrow\uparrow$	$\downarrow\uparrow$			26
2	0	$\downarrow\uparrow$		$\downarrow\uparrow$		26
2	1	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	42
1	2	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	42
2	2	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	50
0	3	$\downarrow\uparrow$	$\downarrow\uparrow$			58
3	0	$\downarrow\uparrow$		$\downarrow\uparrow$		58
3	1	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	74
1	3	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	$\downarrow\uparrow$	74
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

Table 1: Allowed configurations for $n_\uparrow = n_\downarrow$, $L_x = L_y$. We are of course, able to simulate other configurations by breaking the symmetry of the system.

More sophisticated trial functions

Configuration interaction (CI) is a member of a group of so called post-Hartree Fock methods regularly employed in quantum chemistry to attempt to take into account the effects of electron correlation. It is based around expanding a variational trial wavefunction into a linear combination of configuration state functions (CSFs). A configuration state function is itself a linear combination of Slater determinants of single particle orbitals. This trial function is then variationally optimised, which as is evident, is a very time consuming procedure and is only applicable to small systems. The key idea being that an inclusion of more of the systems Hilbert space will account for some of the effects of correlation. These optimised functions can of course be inserted into a DMC code. While however CI is a systematic approach to improving Hartree Fock results, there is no guarantee that the nodal surface of the function will have improved.

Recently this has developed into a slight point of contention. Umrigar et al. [2007] found in a study of Si_2 and C_2 molecules a monotonic improvement

in the DMC energy, despite the parameter optimisation being performed in VMC. Drummond et al. [2006] found however in a study of Ne and the Ne^+ ion that though significantly lower VMC energies could be achieved than with a single determinant function, that the DMC energy increased. Therein lies the interest in our investigation. The functionality of the DMC code developed by O'Rourke [2007] could shed new light on the matter, through the application on a very well investigated system, the homogeneous electron gas.

0.3 Testing of the code

The code was acquired without any documentation on how it had been tested, or how long a standard run would take, so it was necessary to perform my own testing. The first run was set off with the 'out of the box settings' supplied on a standard Pentium 4 workstation. This ran for four hours and then segmentation faulted.

This turned out to be due to an array bounds error in the calculation of the coulomb potential.

Evaluation of the coulomb potential

This is calculated using a standard Ewald imaging technique, splitting the potential into real, and reciprocal space, designed to speed up the rate of convergence of the sum. The 2D version derived by Schofield [2001](30) follows the the general formula [Ceperley, 1999], and has been modified to account for the 2D system.

$$V(\mathbf{R}) = \sum_{i<j}^n \frac{\text{erfc}(\kappa r_{ij})}{r_{ij}} + \frac{\pi}{L^2} \left[\sum_{\mathbf{k}} \frac{\text{erfc}(k/2\kappa)}{k} \left| \sum_{i=1}^n \exp(i\mathbf{k}\cdot\mathbf{r}_i) \right|^2 \right] - \frac{n\kappa}{\sqrt{\pi}} - \frac{n^2\sqrt{\pi}}{\kappa L^2} \quad (30)$$

The first term in (30) is the real space term, the second the reciprocal space term, and the last two account for the electrons interaction with there own images. The parameter κ is determined determined by the level of convergence desired. A simple compiler check on run-time conditions reveal the problem to be an array bounds error in part of the calculation of the imaginary term. The Ewald routine was the thoroughly checked, and proved to be no more problem throughout the investigation.

The code was then submitted, again with the out of the box settings, on a Beowolf cluster running a portable batch system. The program reached the limit of its request for 12 hours CPU time, and was subsequently terminated. The same thing happened when 24 hours CPU time was requested

Population instability

Analysis of output files showed that the problem was due to a rather dramatic instability in the population in the transition from VMC to DMC. Email correspondence with O'Rourke confirmed that he and Schofield had had great difficulty with population control. Indeed, Schofield had resorted to culling walkers at random in order to meet the population target. The population had exploded, and few steps had been completed before the program was terminated. A drop in the calculated local energy E_L had resulted in an abnormally large renormalisation factor, G_b , which had not been controlled properly(23).

$$G_b = \frac{1}{(2\pi t)^{3N/2}} e^{-\frac{\tau}{2}(E_L(\mathbf{R})+E_L(\mathbf{R}')-2E_T)} \quad (31)$$

Random drifts in the population in DMC are accounted for with a relatively arbitrary negative feedback procedure to control the trial energy E_T . Detailed analysis of errors lead Umrigar et al. [1993] to suggest

$$E_T(t + \tau) = E_{est}(t) - \frac{1}{g\tau} \ln \left(\frac{W(t)}{W_{target}} \right) \quad (32)$$

Where $E_{est}(t)$ is the estimate of the total energy averaged over all walkers, for the time t . $W(t)$ is the number of walkers at time t , and W_{target} is the desired average number of walkers. This procedure is such that the random deviations will be corrected in a mean time g later.

This was a very simple problem to solve. A mix up in notation had meant that W_{target} in Eq. (32) had been replaced by $W(t - \tau)$.

The practical implications of this were that whereas a run of 6 electrons previously took hours to complete ~ 1000 time steps, it now took seconds. This meant that reasonable results could be obtained for systems of > 50 electrons, and finite size effects could be estimated.

The need for, and the implementation of real-time graphical analysis

The code developed by O'Rourke was designed to work in four stages

1. Initial equilibration[†]
Meant primarily to stabilise system, and reduce the chance of persistent configurations [O'Rourke, 2007].
2. VMC
Performed the Fokker-Planck VMC algorithm, see Thijssen [2007] for an outline of this.

[†]This equilibration period proved unnecessary once further corrections had been made (see section 0.3.2)

3. DMC equilibration

A DMC equilibration period - no results were output, and E_T updated at every step as in (32) in order to cope with the drop of E_L in the transition from VMC to DMC.

4. DMC full run

Once the DMC run had reached equilibrium (i.e. once the Markovian chain was mapping out the stationary distribution $\phi_0(\mathbf{R})$) then the population control algorithm update period was relaxed, and sampling commenced.

It was realised that full analysis of the system would require the analysis of a large amount of data. The following quantities would have to be monitored:

1. Population
2. Trial Energy E_T
3. Local Energy E_L
4. Kinetic Energy
5. Potential Energy
6. Effective time step
7. Average drift velocity (see section 0.3.2)
8. Average $T_i = -\frac{1}{4}\nabla_i^2\ln|\phi|$ (also see section 0.3.2)

This of course would be a very time consuming procedure if each were analysed separately, and graphs were formatted each and every time, so it was judged a good use of time to develop a automated procedure.

0.3.1 Run-Time Analysis

The quick and efficient analysis of run output, would be best handled by a run-time graphical output facility. The PGPLOT graphics subroutine library is a FORTRAN callable library with which I have had a small amount of experience. It is also very flexible, offering many different formats of output. Importantly it offers both run-time output, and the plotting of multiple axes in a single window, or multiple windows. It also offers some basic WIMP functionality that could have been useful if there had developed an aspiration for run-time tweaking of parameters, or display variables. This addition proved absolutely vital in the continued analysis of the code. Figure 1, as an early example of the routine, illustrates this.

This showed that the observed drop in energy was primarily in the kinetic energy. This generally had a discontinuity in the move from VMC to DMC

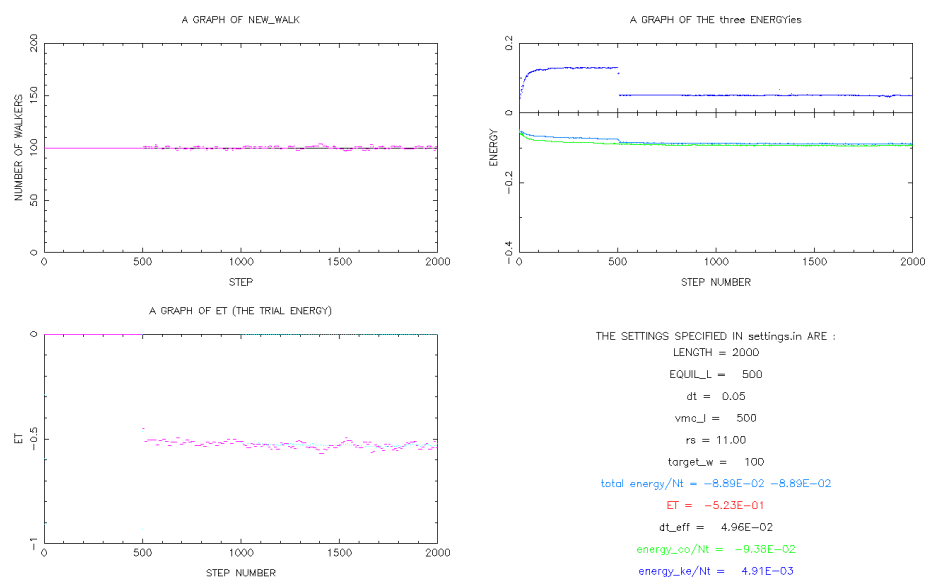


Figure 1: The output generated from a single short run. VMC length:500 steps, DMC equilibration:500 steps, DMC full:1000 steps. Note the periodic varying of the trial energy keeping the population at a steady level within DMC, and the stated discontinuity observed in the Kinetic energy. Note that the kinetic energy is displayed a factor of ten larger than its true value. This output is updated after every step, and is drawn throughout the run.

(see figure 1, though also unexpected discontinuities midway through a VMC run also took place. Rather more disconcerting however, was the fact that on a large proportion of runs the kinetic energy turned negative. It was very clear that the estimators of the kinetic energy must be looked into.

0.3.2 Kinetic energy and internal consistency

The total energy of the system was calculated using a simple though reliable method employed in the majority of DMC codes called the mixed estimator. The mixed estimator is a simple average of the local energy E_L taken over all walkers and time.

$$E_L(\mathbf{R}) = \frac{\hat{H}\psi_T(\mathbf{R})}{\psi_T(\mathbf{R})} \quad (33)$$

Better estimators exist for energy calculation, see Umrigar et al. [1993] for a detailed discussion of these. The mixed estimator has the significant advantage that it is common to both VMC and DMC, and such greatly reduces programming time and complexity. We will now split the Hamiltonian into kinetic and potential terms $\hat{H} = \hat{K} + \hat{V}$, where $\hat{K} = -\frac{1}{2}\nabla^2$ and we'll concern ourselves with this from now on. Reminding ourselves that $\nabla = (\nabla_1, \nabla_2 \dots, \nabla_n)$, we now split the kinetic energy estimator into terms involving the gradient and Laplacian of the logarithm of the wavefunction for each electron (34d).

$$-\frac{1}{2} \frac{\nabla_i^2 \psi_T}{\psi_T} = -\frac{1}{2} \left[\frac{\psi_T \nabla_i^2 \psi_T - (\nabla_i \psi_T)^2}{\psi_T^2} + \left(\frac{\nabla_i \psi_T}{\psi_T} \right)^2 \right] \quad (34a)$$

$$= -\frac{1}{2} \left[\nabla_i \cdot \left(\frac{\nabla_i \psi_T}{\psi_T} \right) + \left(\frac{\nabla_i \psi_T}{\psi_T} \right)^2 \right] \quad (34b)$$

$$= -\frac{1}{2} \nabla_i^2 \ln \psi_T - \left(\frac{1}{\sqrt{2}} \nabla_i \ln \psi_T \right)^2 \quad (34c)$$

$$K_i = -\frac{1}{2} T_i - \left| \frac{1}{\sqrt{2}} \mathbf{v}_{iD} \right|^2 \quad (34d)$$

Where Eq. (34b) is found by use the quotient rule backwards. We have defined two new quantities viz.

$$T_i = \nabla_i^2 \ln \psi_T = \frac{\nabla^2 \psi_T}{\psi_T} - \left(\frac{\nabla \psi_T}{\psi_T} \right)^2 \quad (35)$$

$$\mathbf{v}_{iD} = \nabla_i \ln \psi_T = \frac{\nabla \psi_T}{\psi_T} \quad (36)$$

The reason we do this is two-fold. Firstly \mathbf{v}_{iD} is the drift velocity for the electron i , and as such must be calculated anyway. The logarithm in both quantities allows for the factorisation of the Slater-Jastrow trial function into

contributions from from each part, easing computational expense. Secondly, and crucially we can derive an expression that will allow us to evaluate the internal consistency of the system.

Let us consider the expectation value for the operator \hat{T}_i .

$$\langle T_i \rangle = \langle \psi_T | \hat{T}_i | \psi_T \rangle \quad (37a)$$

$$= \int \psi_T \left\{ \frac{\nabla^2 \psi_T}{\psi_T} - \left(\frac{\nabla \psi_T}{\psi_T} \right)^2 \right\} \psi_T d\tau \quad (37b)$$

$$= \int \left(\nabla^2 \psi_T - (\nabla \psi_T)^2 \right) d\tau \quad (37c)$$

$$= \int \nabla \cdot (\psi_T \nabla \psi_T) - 2 (\nabla \psi_T)^2 d\tau \quad (37d)$$

$$= \int \psi_T (\nabla \psi_T) \cdot \hat{n} dS - 2 \int \psi_T \left(\frac{\nabla \psi_T}{\psi_T} \right)^2 \psi_T d\tau \quad (37e)$$

Since we are in a periodic system any surface integral will of course be identically zero, meaning (37e) will reduce to $\langle T \rangle = -2 \langle v_D^2 \rangle$. It may be shown by similar means that $\langle K \rangle = \frac{1}{2} \langle v_D^2 \rangle$ resulting in

$$\langle K \rangle = \frac{1}{2} \langle v_D^2 \rangle = -\frac{1}{4} \langle T \rangle = \left\langle -\frac{1}{2} (T + v_D^2) \right\rangle \quad (38)$$

The importance of this identity was summed up by Fahy et al. [1990], who stated "When working with the complicated numerical wavefunctions of real systems they are the only exact analytic check of internal consistency" This identity implies that we may use any one of these three estimators, however, as shall be shown the variance of K is significantly lower than that of the other two. For this reason we shall call K the primary estimator, and $\frac{1}{2}v_D^2$ and $-\frac{1}{4}T$ secondary estimators.

Internal consistency was mentioned briefly by Schofield [2001], however an error in his derivation meant that the wrong identity was arrived at.

The program, though seeming well behaved for some runs, and not for others, never fulfilled the criteria of Eq. (38). By this time it was become increasingly apparent that there was something seriously wrong with the code. The results outputted were heavily, and unpredictably dependent on the number of walkers, and since the only way walkers affect each other is indirectly through their averages changing E_T , were serious problems. For instance it was noticed that a particular run with seven hundred walkers outputted reasonable looking results though when the same parameters were used except with an additional one walker the kinetic energy stabilised at a vastly negative number.

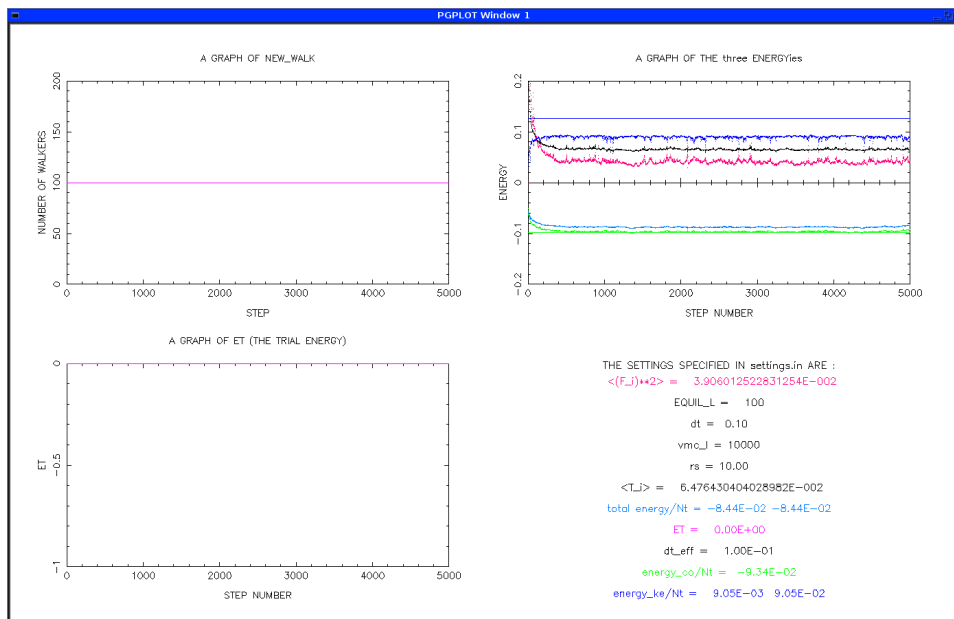


Figure 2: An entirely VMC run showing the lack of internal consistency in the code. The values named as $\langle (F_i)^2 \rangle$, $\langle T_i \rangle$ and $energy_{ke}/Nt$ are the three estimators for the kinetic energy following the notation conventions of Foulkes et al. [2001], and as such should be equal to each other. The horizontal blue line is the Ceperley [1978] value for the kinetic energy at $r_s = 10$ (again note a factor of ten to assist viewing)

Evaluation of the kinetic energy

As stated we calculate the kinetic energy by $K = -\frac{1}{2}(T + \mathbf{v}_D^2)$, and so we must evaluate both the gradient and Laplacian of the logarithms of the trial function. The logarithms provide an invaluable service in this in the following way.

$$\nabla_i \ln(\psi_T) = \nabla_i \ln(\Psi_J d) = \nabla_i \ln \Psi_J + \nabla_i \ln d = \nabla_i \left[-\frac{1}{2} \sum_{j \neq i} u_{\sigma_i \sigma_j}(r_{ij}) \right] + \frac{1}{d} \nabla_i d \quad (39a)$$

The form for the Jastrow factors contribution may be derived analytically from (26b), avoiding the need for numerical differentiation. Fahy et al. [1990] developed a very efficient way of calculating $\frac{1}{d} \nabla_i d$. Standard matrix theory dictates that the inverse of the transpose of a matrix is equal to the matrix of cofactors over its determinant, or $(A_{jk}^T)^{-1} = \frac{1}{|A|} C_{jk}$. Since the each element in the matrix of cofactors does not depend upon either the row or the column in which it lies and the movement of an electron changes only one row of the Slater determinant, we find

$$\frac{d_{new}}{d_{old}} = \sum_{j=1}^n (D_{ji,old}^T)^{-1} D_{ji,new} = q_i. \quad (40)$$

Where D is the Slater matrix, and the i th electron has been moved. In this way we may calculate the final term in Eq. (39a) from the ratio of old to new determinants, thus avoiding the $\propto n^3$ explicit evaluation of the determinant. The inverse transpose Slater matrix, which we shall call \tilde{D} , must stored and updated each step according to Fahy et al. [1990]

$$\tilde{D}_{jk} = \begin{cases} \tilde{D}_{jk}/q, & k = i \\ \tilde{D}_{jk} - \frac{\tilde{D}_{ji}}{q} \left[\sum_{l=1}^n \tilde{D}_{lk} \psi_l(\mathbf{r}_i) \right], & k \neq i \end{cases} \quad (41)$$

So in there entirety the two secondary estimators took the form

$$\nabla_i \ln(\psi_T) = -\frac{1}{2} \sum_{j \neq i} \nabla_i u_{\sigma_i \sigma_j}(r_{ij}) + \frac{1}{q} \sum_{j=1}^n \tilde{D}_{ji,old} \nabla_i \psi_j(\mathbf{r}_i) \quad (42a)$$

$$\nabla_i^2 \ln(\psi_T) = -\frac{1}{2} \sum_{j \neq i} \nabla_i^2 u_{\sigma_i \sigma_j}(r_{ij}) - \left[\frac{1}{q} \sum_{j=1}^n \tilde{D}_{ji,old} \nabla_i \psi_j(\mathbf{r}_i) \right]^2 + \frac{1}{q} \sum_{j=1}^n \tilde{D}_{ji,old} \nabla_i^2 \psi_j(\mathbf{r}_i) \quad (42b)$$

Debugging the kinetic energy

My own derivations of the analytic derivative of the Jastrow contributions (0.3.2) agreed with Schofield [2001], and I could find only a couple of minor

mix ups between the parallel and anti-parallel A parameters used in the code.

$$\nabla(-u) = \frac{A\mathbf{r}}{r^3} \left[e^{-\frac{r}{F}} \left(1 + \frac{r}{F} \right) - 1 \right] \quad (43)$$

$$\nabla^2(-u) = \frac{A}{r^3} \left[e^{-\frac{r}{F}} \left(1 + \frac{r}{F} + \frac{r^2}{F^2} \right) - 1 \right] \quad (44)$$

It is plain to see that Eqs. (41,42a and 42b) are minefield of programming errors. So simple verification subroutine was written in which the inverse transpose of the Slater matrix would be calculated explicitly using the LAPACK routines *dgetrf* and *dgetri*, and the associated quantities could subsequently be (inefficiently) calculated for the use of comparison. This showed that there was an error in the updating of the inverse transpose Slater matrix, with the ultimate consequence that the movement of the walkers was effectively *completely random*.

Because of the intricate nature of the problem, and the fact that the code was complicated with the use of multiple determinants meant that a good deal of time had to be spent fixing this error. The main problem turned out to be in the calculation of the ratio of old and new determinants q . Thankfully the subroutine I had written served to aid this process, and the problem was finally fixed. The initial equilibration period was found to serve only one purpose. This was to initialise the drift velocity, as this is calculated the previous time step. Once the updating procedure for the inverse transpose slater matrix was fixed, there was no need for this period.

Confusion over the Gaussian random deviates

The Gaussian random deviate χ used for the diffusion step of the walkers move, should have a variance of the time step τ , in \mathcal{E}_{nd} . It was noticed, however that the code was apparently calculating deviate with a variance of $\sqrt{\tau}$. This had the effect of making the plots of all the energy estimators significantly smoother, as shown in figure 3. It also however made the drift velocity tend toward zero. Under closer scrutiny, the random deviate subroutine turned out be incorrectly labelled. It produced deviates with a standard deviation of its argument, and not as the comments in the code suggested the variance.

Further investigation into the Internal consistency

Despite all of the corrections made so far, the test for internal consistency was still failing. Further thought, and suggestions from others [Bokes, 2008], lead to the following logic. The Internal consistency is a mathematical identity, and should therefore hold whatever the system. In VMC correlation

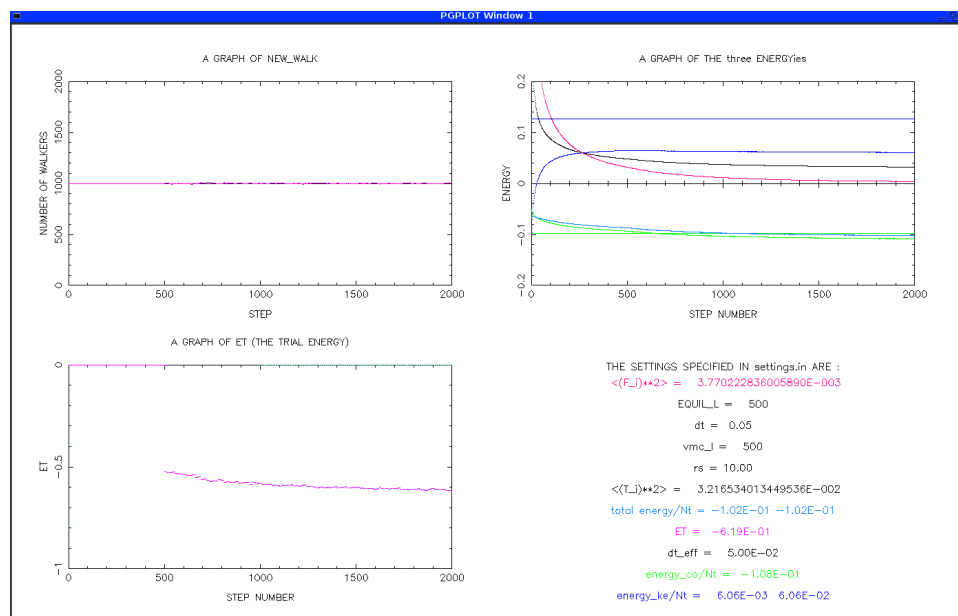


Figure 3: The system when the diffusion of the walkers was mistakenly reduced, note the smoothness of all estimators and that the drift velocity secondary estimator is significantly diminished

is entirely dependent upon the Jastrow factor, so if we remove the Jastrow factor we should have a completely non-interacting system. If we then neglect the potential terms in the Hamiltonian can calculate the solution of the one particle time independent Schrödinger equation, and average over all particles with a pen and paper. This result should agree with the codes calculated potential energy if set the Jastrow factor to one.

This was done, for a number of different system sizes, and the results were rather astonishing. Despite the secondary estimators $\frac{1}{2}v_D^2$ and $-\frac{1}{4}T$ at times having extremely large variances, and being orders of magnitude out, the primary estimator K was always near to exactly on the analytical value (even with only one walker). Figure 4 illustrates this very well.

A comment on the Yukawa pseudopotential

Before final conclusions are made, I would like to make an observation about the Jastrow factor used. Schofield [2001] and O'Rourke [2007] used a form of the Yukawa pseudopotential taken from Ceperley [1978]. In looking back at this paper however, I found that this form of the pseudopotential (26b) had only been used for the 3D homogeneous electron gas, and for the 2D

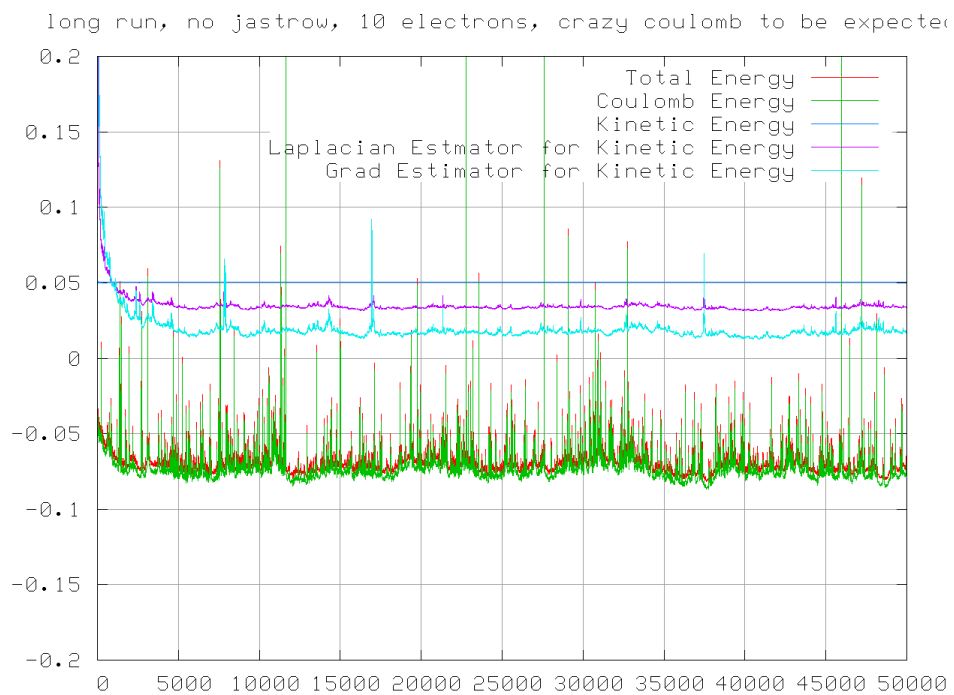


Figure 4: A VMC run when the Jastrow factor is turned off. The analytic kinetic energy is 0.05026548246 Hartrees, and the system achieves this at every time step to 3 significant figures. The Coulomb energy is to be expected to behave like this, as there are no correlation effects taken into account.

case a more general expression had been used (45)

$$u = \frac{A_{\sigma_i \sigma_j} e^{-B_{\sigma_i \sigma_j} r_{ij}} \left(1 - e^{r_{ij}/F_{\sigma_i \sigma_j}}\right)}{r_{ij}} \quad (45)$$

The analysis that was used to hypothesise these two forms is based upon the random phase approximation, and beyond my knowledge. For more information see Monnier [1972], Jr. and Pokrant [1973]. To avoid the risk that the analysis was not valid for the 2D gas, as Ceperley [1978] would suggest I instead used the form of the pseudopotential, of Eq. (45). I derived the derivatives to be

$$\nabla(-u) = \frac{-A\mathbf{r}}{r^3} e^{-Br} \left[e^{-\frac{r}{F}} \left(1 + Br + \frac{r}{F}\right) - 1 - Br \right] \quad (46)$$

$$\nabla^2(-u) = -e^{-\frac{r}{F}} e^{-Br} \left[\frac{A}{r^2 F} + \frac{2AB}{rF} + \frac{A}{rF^2} \right] \quad (47)$$

These reduce down to Eqs. (and) in the case where $B = 0$

0.4 Conclusions

Unfortunately time ran out before I could find out why the secondary estimators combined to produce such a perfect primary estimator. More unfortunately perhaps was the way that the project could not progress in the way that had been hoped at the start of the year. Every time one bug was fixed, a new one was found, and despite the fact that the code now performs reliably, and consistently, there is still an error somewhere. The importance of using a well tested code has been highlighted. I cannot help feeling that this would have, and future QMC projects would do better to start from the platform of one of the many open-source codes available, or from scratch.

I would like however to highlight the run-time graphical analysis. This was absolutely instrumental in the conduct of the project. I would also like to make one final critique of the results of the previous two projects [O'Rourke, 2007], and Schofield [2001]. Both authors found their results to be lower than anticipated. This was said by O'Rourke [2007] to be due to a combination of finite size and the use of a less sophisticated trial function. As discussed in section 0.2.1, this cannot be the case of a less sophisticated trial function as the fixed node variational principle would imply the opposite.

Derivation of the isitSe

$$-\psi_T(\mathbf{R}) \frac{\partial}{\partial t} \phi(\mathbf{R}, t) = \psi_T(\mathbf{R}) (\hat{H} - E_T) \psi(\mathbf{R})^{-1} f(\mathbf{R}, t) \quad (48a)$$

$$= -\frac{1}{2} \nabla^2 f(\mathbf{R}, t) + \frac{1}{2} \nabla \cdot (\mathbf{v}_D(\mathbf{R}) f(\mathbf{R}, t)) + \frac{1}{2} \nabla \psi_T(\mathbf{R}) \cdot \nabla \phi(\mathbf{R}, t) + (V(\mathbf{R}) - E_T) f(\mathbf{R}, t) \quad (48b)$$

$$= -\frac{1}{2} \nabla^2 f(\mathbf{R}, t) + \nabla \cdot (\mathbf{v}_D(\mathbf{R}) f(\mathbf{R}, t)) + \frac{1}{2} \nabla \psi_T(\mathbf{R}) \cdot \nabla \phi(\mathbf{R}, t) - \nabla \left(\frac{\nabla \psi_T(\mathbf{R})}{\psi_T(\mathbf{R})} \psi_T(\mathbf{R}) \phi(\mathbf{R}, t) \right) + (V(\mathbf{R}) - E_T) f(\mathbf{R}, t) \quad (48c)$$

$$= -\frac{1}{2} \nabla^2 f(\mathbf{R}, t) + \nabla \cdot (\mathbf{v}_D f(\mathbf{R}, t)) + (E_L - E_T) f(\mathbf{R}, t) \quad (48d)$$

where $E_L(\mathbf{R}) = -\frac{\nabla^2 \psi_T}{2\psi_T} + V(\mathbf{R})$, and

$$\psi_T(\mathbf{R}) \nabla^2 \phi(\mathbf{R}, t) = \nabla (\psi_T(\mathbf{R}) \nabla \phi) - \nabla \psi_T(\mathbf{R}) \cdot \nabla \phi(\mathbf{R}, t) \quad (49a)$$

$$= \nabla (\nabla (\psi_T(\mathbf{R}) \phi(\mathbf{R}, t)) - \phi(\mathbf{R}, t) \nabla \psi_T(\mathbf{R})) - \nabla \psi_T(\mathbf{R}) \cdot \nabla \phi(\mathbf{R}, t) \quad (49b)$$

$$= \nabla^2 f(\mathbf{R}, t) - \nabla \cdot (\mathbf{v}_D f(\mathbf{R}, t)) - \nabla \psi_T(\mathbf{R}) \cdot \nabla \phi(\mathbf{R}, t) \quad (49c)$$

has been used to get to step 48.

Derivation of the importance sampled Green's function in 1D

The Green's function may be formally written

$$G(x, x', \tau) = \langle x | e^{t\hat{L}} | x' \rangle \quad (50)$$

where, from the isitSe, $\hat{L} = -\frac{1}{2} \hat{p} [\hat{p} + 2i\mathbf{v}_D] + [E_L(x) - E_T]$. We will put aside the $[E_L(x) - E_T]$ term for now as we did with the potential term in the non importance sampled Green's function. So continuing,

$$\langle x | e^{-\frac{t}{2} \hat{p} [\hat{p} + 2i\mathbf{v}_D]} | x' \rangle \quad (51a)$$

$$= \frac{1}{2\pi} \int dp \langle p | e^{-\frac{t}{2} \hat{p} [\hat{p} + 2i\mathbf{v}_D]} | p \rangle e^{ip(x-x')} \quad (51b)$$

$$= \frac{1}{2\pi} \int dp e^{-\frac{t}{2} p [p + 2i\mathbf{v}_D] + ip(x-x')} \quad (51c)$$

Then, by completing the square p in the exponent of the integrand we find

$$= \frac{1}{2\pi} e^{\frac{1}{2t} (\mathbf{v}_D - (x-x'))^2} \int dp e^{-\frac{t}{2} [p + i(\mathbf{v}_D - (x-x'))/t]^2} \quad (52a)$$

$$\frac{1}{\pi \sqrt{2t}} e^{((x-x') - t\mathbf{v}_D)^2 / 2t} \int_{-\infty}^{\infty} e^{-u^2} du \quad (52b)$$

$$\frac{1}{\sqrt{2\pi t}} e^{((x-x') - t\mathbf{v}_D)^2 / 2t} \quad (52c)$$

By using the spectral expansion, we then build back in the $[E_L(x) - E_T]$

$$\begin{aligned}
G(x, x', \tau) &= \langle x | e^{t\hat{L}} | x' \rangle \\
&= e^{-\tau(E_L(x) - E_T)} \langle x | e^{-\frac{t}{2}\hat{p}[\hat{p} + 2i\mathbf{v}_D]} | x' \rangle e^{-\tau(E_L(x') - E_T)} \\
&= \frac{1}{\sqrt{2\pi t}} e^{-\tau(E_L(x) + E_L(x') - 2E_T)} e^{(x - x' - t\mathbf{v}_D)^2 / 2t}
\end{aligned} \tag{53}$$

which when extended to \mathcal{E}_{nd} becomes Eq. (23).

Bibliography

- J. B. Anderson. A random walk simulation of the schrodinger equation. *Journal of Chemical Physics*, 1975.
- M. S. Becker, A. A. Broyles, and T. Dunn. A parametric approach to the ground-state energy of an electron gas. *Physical Review*, 1968.
- P. Bokes, 2008. in conversation.
- D. Ceperley. Goundstate of the fermion one-component plasma: A monte carlo study in two and three dimensions. *Physical Review B*, 1978.
- D. Ceperley. Long-range potentials with the ewald image technique. Summer school lecture notes, <http://archive.ncsa.uiuc.edu/Apps/CMP/lectures/cms99/ewald.pdf>, 1999.
- D. M. Ceperley. Fermion nodes. *Journal of Statistical Physics*, 1991.
- D. M. Ceperley and B. J. Alder. Ground state of the electron gas by a stochastic method. *Physical Review Letters*, 1980.
- N. D. Drummond, P. L. Ríos, A. Ma, J. R. Trail, G. G. Spink, M. D. Towler, and R. J. Needs. Quantum monte carlo study of the ne atom and the ne⁺ ion. *The Journal of Chemical Physics*, 2006.
- T. Dunn and A. A. Broyles. Method for determining the thermodynamic properties of the quantum electron gas. *Physical Review*, 1966.
- S. Fahy, X. W. Wang, and S. G. Louie. Variational quantum monte carlo nonlocal pseudopotential approach to solids: Formulation and application to diamond, graphite, and silicon. *Physical Review B*, 1990.
- W. Feller. *An Introduction to probability Theory and Its Applications*. Wiley, 2 edition, 1962.
- W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum monte carlo simulations of solids. *Reviews of Modern Physics*, 2001.

- R. W. Godby, 2007. Colloquium given at University of York.
- R. J. Jastrow. Many body problem with strong forces. *Phys. Rev.* 98, 1479-1484, 1955.
- F. A. S. Jr. and M. A. Pokrant. Variational calculation of the electron-gas correlation energy. *Physical Review A*, 1973.
- R. Monnier. Monte carlo approach to the correlation energy of the electron gas. *Physical Review A*, 1972.
- C. O'Rourke. Fixed nodes in diffusion monte carlo. Master's thesis, University of York, 2007.
- G. Ortiz, D. M. Ceperley, and R. M. Martin. New stochastic method for systems with broken time-reversal symmetry: 2-d fermions in a magnetic field. *Physical Review Letters*, 1993.
- B. Schofield. Node fixing in diffusion monte carlo. Master's thesis, University of York, 2001.
- J. M. Thijssen. *Computational Physics*. Cambridge University Press, 2 edition, 2007.
- M. D. Towler. The quantum monte carlo method. *Physica status solidi*, 2006.
- C. J. Umrigar, M. P. Nightingale, and K. J. Runge. A diffusion monte carlo algorithm with very small time-step errors. *J. Chem. Phys.* 99, 2865, 1993.
- C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig. Alleviation of the fermion-sign problem by optimisation of many-body wave functions. *Physical Review Letters*, 2007.
- N. G. Underwood. Fermion problems in quantum monte carlo methods. Subject area review, previously submitted, 2007.