1 Introduction

A mathematical description of a molecule (say LiH) is provided by a solution to the following SCHRÖDINGER EQUATION (an EIGENVALUE-eigenfunction PROBLEM):

$$-\frac{1}{2}\sum_{i}\boldsymbol{\nabla}_i^2\phi + V\phi = E\ \phi \tag{1}$$

where E (the eigenvalue) yields the molecule's energy, and ϕ^2 provides the distribution of the electrons' locations. The i summation is over all electrons, and

$$V \equiv \sum_{i,\alpha}\frac{-Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \sum_{i<j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{\alpha<\beta}\frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|} \tag{2}$$

is the molecule's ELECTROSTATIC POTENTIAL (summing over all electron-electron, electron-nucleus and nucleus-nucleus pairs; i, j label electrons, α, β label nuclei), where Z_α are the nuclei's charges (also called ATOMIC NUMBERS), and $\mathbf{r}_i$ and $\mathbf{R}_\alpha$ are the electrons' and nuclei's locations, respectively.

In this article, we are concerned only with the lowest (GROUND-STATE) value of E (denoted E_0) of electrically *neutral* molecules (of zero total charge). All quantities are expressed in atomic units, defined by $e = \hbar = m_e = 1$ (in these units, the electron's charge and mass, and Planck's constant, are equal to 1).

Also, the following **approximations** have been used:

- The non-relativistic limit (Eq. 1 has no spin-orbit and similar relativistic terms), is generally regarded as adequate when concerned with chemical properties.
- The nuclei are kept fixed (or CLAMPED) at their equilibrium locations, which results in the so called BORN-OPPENHEIMER approximation. When interested in the nuclei's VIBRATIONAL ENERGIES, we apply the clamped-nuclei technique to several different geometries (inter-nuclei distances) to find out how E_0 varies with these (at the equilibrium location, E_0 should have a minimum). We can then let the nuclei vibrate in the corresponding potential 'bowl'.

The solution to (1) is subject to very simple BOUNDARY CONDITIONS, namely that $\phi(\infty) = 0$; whenever one or more electrons are moved away from the molecule, the solution tends to zero (this is usually a fast, exponential decrease). There is yet another (but crucial) requirement that ϕ must change sign whenever two electrons of the *same spin* are interchanged (the so called PAULI EXCLUSION PRINCIPLE). Mathematically, this can be written (for a 4 electron molecule) as

$$\phi(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3,\mathbf{r}_4) = -\phi(\mathbf{r}_2,\mathbf{r}_1,\mathbf{r}_3,\mathbf{r}_4) = -\phi(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_4,\mathbf{r}_3) \tag{3}$$

assuming that the first two electrons have spin up, the last two spin down. Note that having equal number of electrons of each spin leads to the lowest value of E_0.

Eq. 3 implies that ϕ cannot be everywhere positive - clearly, half the CONFIGURATION SPACE (12 dimensional space of all possible $\mathbf{r}_1,\ \mathbf{r}_2,\ \mathbf{r}_3$ and $\mathbf{r}_4$ values) results in $\phi < 0$.

This is one of the main reasons for having to modify (1) before we can design a Monte Carlo procedure for solving it - statistically, we can generate a sample of CONFIGURATIONS (each is a set of $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4$ values) only from a distribution with *non-negative* PROBABILITY DENSITY FUNCTION (PDF for short). Such a modification requires having first an approximate (but 'analytic', i.e. *computable*) solution to (1). This solution will be denoted Ψ and called the TRIAL SOLUTION or the GUIDING FUNCTION.

Later on, we indicate how Monte Carlo can help us build a good trial function, but at this point we take an existence of Ψ for granted (there are many analytical techniques which can supply it).

Our main goal is to design a Monte Carlo procedure capable of improving (usually substantially) the results provided by such trial solution. But first we must take a small detour to explain the most common (for our purpose) form of Ψ.

2 Guiding function

First, we define the so called Slater-type ATOMIC ORBITALS, which have the form of

$$h^{[\ell,i,j,k]}(x,y,z;\xi) \equiv r^{\ell} x^i y^j z^k \ e^{-\xi r} \tag{4}$$

where x, y and z are components of an electron's location, *relative* to a specific nucleus, $r = \sqrt{x^2+y^2+z^2}$, ℓ, i, j and k are (usually small) integers, and ξ is a constant (different, in general, for each orbital).

For a given molecule, we then select a group of such atomic orbitals to be centered on its first nucleus, another group centered on the second nucleus and so on, for the total of s atomic orbitals (subsequently, these are denoted h_1, h_2, ... h_s for short). In our LiH example, we will place two atomic orbitals (namely $h^{[0,0,0,0]}$ and $h^{[0,0,0,1]}$) on Li, and the same two atomic orbitals (with different ξ's) on H. Here, we assume that the locations of the individual nuclei (the molecule's 'geometry') are known (for LiH, we place Li at the origin, and H at 3 atomic units along the z axis).

If the molecule consists of even number (say $2m$) electrons, we designate the first half of them as spin-up, and the second half as spin-down electrons. This will result in achieving the lowest ground-state energy. For molecules with odd number of electrons (say $2m+1$) this would require having $m+1$ of them to be spin-up and the rest spin-down (or the other way round), with only a minor modification of the trial function. In the following, we assume the 'even-split' or the so called CLOSED-SHELL case.

The s atomic orbitals are then used to create m (matching the number of spin-up and spin-down electrons) sets of MOLECULAR ORBITALS, each as a *different* linear combination of the *same* atomic orbitals (it is very important to have chosen $s \geq m$), thus:

$$g_i = \sum_{j=1}^{s} C_{ij}\ h_j \qquad\qquad i = 1,\ 2,\ ..., m \tag{5}$$

The elements of the C_{ij} matrix (which is m by s) are usually found with the help of an optimization (described below). Note that, without a loss of generality, its first m by m block can be set to be the unit matrix (only the remaining elements need to be optimized). Each molecular orbital is thus a function of x, y and z (a location, in a convenient coordinate system, of a *single* electron), of the nuclei's locations, and of the ξ and C_{ij} parameters.

A SLATER DETERMINANT is then computed by evaluating each of the m molecular orbitals at the location of each spin-up electron, and taking the determinant of the resulting m by m matrix. A second such determinant is similarly computed, based on the locations of spin-down electrons. Finally, the two determinants are multiplied, resulting in the following (Hartree-Fock) trial solution (obviously antisymmetric with respect to the interchange of any two electrons of the same spin):

$$\Psi = \begin{vmatrix} g_1(\mathbf{r}_1) & g_2(\mathbf{r}_1) & \dots & g_m(\mathbf{r}_1) \\ g_1(\mathbf{r}_2) & g_3(\mathbf{r}_2) & \dots & g_m(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ g_1(\mathbf{r}_m) & g_2(\mathbf{r}_m) & \dots & g_m(\mathbf{r}_m) \end{vmatrix} \cdot \begin{vmatrix} g_1(\mathbf{r}_{m+1}) & g_2(\mathbf{r}_{m+1}) & \dots & g_m(\mathbf{r}_{m+1}) \\ g_1(\mathbf{r}_{m+2}) & g_3(\mathbf{r}_{m+2}) & \dots & g_m(\mathbf{r}_{m+2}) \\ \vdots & \vdots & \ddots & \vdots \\ g_1(\mathbf{r}_{2m}) & g_2(\mathbf{r}_{2m}) & \dots & g_m(\mathbf{r}_{2m}) \end{vmatrix} \tag{6}$$

This trial function can be further improved by multiplying it by

$$\prod_{i<j} \exp\left(\frac{a r_{ij}}{1 + b r_{ij}}\right) \tag{7}$$

called a JASTROW FACTOR, where a and b are two additional parameters, and r_{ij} is the distance between electrons i and j (the product is over all such pairs, regardless of spin). Note that this factor (being *symmetric* with respect to any two-electron interchange) preserves the original anti-symmetry of the Hartree-Fock solution.

The final form of the trial function can be then optimized (with respect to the C_{ij}, ξ's, and a and b parameters) using the so called VARIATIONAL PRINCIPLE, namely:

For *any* properly antisymmetric (in the sense of Eq. 3) Ψ, the following inequality holds:

$$\frac{\int \Psi \mathcal{H} \Psi d\mathbf{r}}{\int \Psi^2 d\mathbf{r}} \geq E_0 \tag{8}$$

where E_0 is the exact ground-state eigenvalue of (1),

$$\mathcal{H} \equiv -\frac{1}{2} \sum_i \boldsymbol{\nabla}_i^2 + V \tag{9}$$

is the corresponding (left-hand-side) operator, and $\int \ldots \, d\mathbf{r}$ stands for multivariate integration over *all* electrons. The equal sign in (8) can be achieved only by the exact solution to Eq. 1 (denoted ϕ); for approximate solutions, the smaller we make the left-hand side of (8), the closer we are to the exact solution. Our parameters are thus found by *minimizing* this left-hand side (getting a reasonable estimate of E_0 as a by-product).

This is often a non-trivial exercise since, for most trial functions (including our example), the multidimensional integration *cannot* be carried out analytically, and has to be done with the help of Monte Carlo itself - the details are discussed in Section 4.

To give an example of how a result of such an optimization may look like, here is what one gets for LiH, using the following four atomic orbitals (note their particular order):

$$\begin{aligned} &\exp(-2.89 r_{\mathrm{Li}}) \\ &\exp(-0.87 r_{\mathrm{H}}) \\ &z \exp(-2.85 r_{\mathrm{Li}}) \\ &(z-3) \exp(-0.95 r_{\mathrm{H}}) \end{aligned}$$

where r_{Li} and r_{H} are distances from the Li and the H nucleus, respectively.

The C_{ij} coefficients then turn out to be

$$\begin{bmatrix} 1 & 0 & 0.05 & 0 \\ 0 & 1 & 0.38 & -0.22 \end{bmatrix}$$

and $a = 0.5$, $b = 0.6$.

3 Monte Carlo simulation

The Monte Carlo technique deals with problems, which require finding an unknown function, by interpreting it as a PDF of a statistical distribution and generating a sample (the larger the better) from it. This can work only when the function is non-negative, and integrates (over all possible values of its arguments) to 1. Unfortunately, the eigenfunctions of Eq. 1 are half positive, half negative. We thus first need to re-formulate (1), to make its solution non-negative.

This can be achieved by pre-multiplying it by a trial (i.e. approximate but analytic) solution Ψ, getting:

$$-\frac{1}{2}\sum_i \boldsymbol{\nabla}_i^2 \, f + \sum_i \boldsymbol{\nabla}_i \bullet \mathbf{F}_i \, f + (E_L - E_0) \, f = 0 \tag{10}$$

where $f \equiv \phi \cdot \Psi$,

$$\mathbf{F}_i \equiv \frac{\boldsymbol{\nabla}_i \Psi}{\Psi} \tag{11}$$

and

$$E_L \equiv -\frac{1}{2} \cdot \frac{\sum_i \nabla_i^2 \Psi}{\Psi} + V = \frac{\mathcal{H}\Psi}{\Psi} \tag{12}$$

The $\mathbf{F}_i$'s are called DRIFT functions and E_L is know as LOCAL ENERGY. Note that ∇_i is understood to operate on *all* functions to its right.

All it takes to verify the correctness of (10) is two applications of the product rule, namely:

$$\nabla_i \Psi\phi = \phi\ \nabla_i \Psi + \Psi\ \nabla_i \phi \tag{13}$$

and

$$\nabla_i^2 \Psi\phi = \phi\ \nabla_i^2 \Psi + \Psi\ \nabla_i^2 \phi + 2(\nabla_i \Psi) \bullet \nabla_i \phi \tag{14}$$

(the range of ∇_i is restricted by parentheses).

It is clear that f is now non-negative, *provided* that ϕ has the same NODES (locations of $\phi = 0$) as Ψ. Function ϕ thus represents the exact solution to (1) with slightly incorrect boundary conditions, inherited from the approximate solution Ψ. This is the nature of the so-called FIXED NODE approximation, resulting in E_0 which is still slightly higher than the exact eigenvalue, but which is, at the same time, a major improvement over the variational estimate of Eq. 8. Since (10) is linear (in f) and homogeneous, we can always re-scale f to yield a multivariate PDF (having the total probability of 1). In subsequent formulas, we assume that the arbitrary multiplicative constant of ϕ has been chosen to yield

$$\int \Psi\phi\ d\mathbf{r} = 1 \tag{15}$$

To solve (10) by a Monte Carlo technique, we have to introduce an extra independent variable t (we call it 'time', even though it is unrelated to the actual physics of the problem), and change the right-hand side of the equation thus

$$-\frac{1}{2}\sum_i \nabla_i^2\ f + \sum_i \nabla_i \bullet \mathbf{F}_i\ f + (E_L - E_0)\ f = -\frac{\partial f}{\partial t} \tag{16}$$

where f is now a function of not only the individual electrons' locations, but also of t.

If we can solve (16), we can clearly also solve (10): we start with *any* INITIAL DISTRIBUTION (f at time $t = 0$), then keep on increasing t until we reach the STATIONARY solution of Eq. 16 (f no longer changes in time, i.e. $\frac{\partial f}{\partial t} = 0$), which is thus also a solution to (10).

Equations of type (16) are usually solved by finding the corresponding Green's function $G(\mathbf{r}_0 \rightarrow \mathbf{r},\ t)$ ($\mathbf{r}$ is the variable, $\mathbf{r}_0$ is a parameter) which solves (16) while meeting a rather special initial condition, namely $G(\mathbf{r}_0 \rightarrow \mathbf{r}, 0) = \delta(\mathbf{r}-\mathbf{r}_0)$, i.e. at time $t = 0$ the whole distribution is concentrated at $\mathbf{r}_0$. In our context, both $\mathbf{r}$ and $\mathbf{r}_0$ represent the individual locations of *all* electrons (for 4 electrons, this becomes a 12 dimensional vector which we have already named a CONFIGURATION).

Having the Green's function and *any* initial distribution $f(\mathbf{r},\ 0)$, one can then find $f(\mathbf{r},\ t)$ by

$$\int f(\mathbf{r}_0, 0) \cdot G(\mathbf{r}_0 \to \mathbf{r},\ t)\ d\mathbf{r}_0 \tag{17}$$

Similarly, having a solution $f(\mathbf{r},\ t)$ at time a t, one can advance it to $t+\tau$ thus:

$$f(\mathbf{r},\ t+\tau) = \int f(\mathbf{r}_0, t) \cdot G(\mathbf{r}_0 \to \mathbf{r},\ \tau)\ d\mathbf{r}_0 \tag{18}$$

Unfortunately, (16) is much too complex to have an analytic Green's function. But, if we break it into three separate equations, namely

$$-\frac{1}{2}\sum_i \boldsymbol{\nabla}_i^2\ f \quad = \quad -\frac{\partial f}{\partial t} \tag{19}$$

$$\sum_i \boldsymbol{\nabla}_i \bullet \mathbf{F}_i\ f \quad = \quad -\frac{\partial f}{\partial t} \tag{20}$$

$$(E_L - E_0)\ f \quad = \quad -\frac{\partial f}{\partial t} \tag{21}$$

the individual Green's functions are, respectively:

$$G_{\text{diff}}(\mathbf{r}_0 \quad \to \quad \mathbf{r},\ \tau) = (2\pi\tau)^{-3m} \exp\left[-\frac{(\mathbf{r}-\mathbf{r}_0)^2}{2\tau}\right] \tag{22}$$

$$G_{\text{dr}}(\mathbf{r}_0 \quad \to \quad \mathbf{r},\ \tau) = \delta[\mathbf{r} - \tilde{\mathbf{r}}(\tau)] \text{ where } \frac{d\tilde{\mathbf{r}}(\tau)}{d\tau} = \mathbf{F}[\tilde{\mathbf{r}}(\tau)] \text{ and } \tilde{\mathbf{r}}(0) = \mathbf{r}_0 \tag{23}$$

$$G_{\text{br}}(\mathbf{r}_0 \quad \to \quad \mathbf{r},\ \tau) = \delta(\mathbf{r}-\mathbf{r}_0) \exp\{-\tau[E_L(\mathbf{r}_0) - E_0]\} \tag{24}$$

which can be verified by direct substitution (reminder: $2m$ is the number of electrons, $\mathbf{r}$, $\mathbf{r}_0$ and $\mathbf{F}$ are thus $6m$-dimensional vectors; the last one being a collection of the individual, 3-dimensional $\mathbf{F}_i$'s).

One can show that the Green's function to (16) can be built as a CONVOLUTION of the above three (effectively, applying them to f, in the manner of Eq. 18, one after the other). Except that, such Green's function will no longer be exact, but it will have an error proportional to τ (and higher powers of τ), being thus correct only in the $\tau \to 0$ limit. This implies that we can advance our solution only a *small* time step at a time, obtaining results which will also have a bias, proportional - in the $O(\tau)$ sense - to the STEP SIZE τ. Luckily, such bias can be easily removed, as discussed shortly.

But first we have to have a closer look at (22) to (24), and figure out what do these represent when applied to a *sample* of configurations from f (called, in this context, an ENSEMBLE). To a statistician, this is an easy task: convolution of two statistical distributions represents *adding* the corresponding random variables.

We can thus interpret G_{diff}, G_{dr} and G_{br}, individually, as follows:

Since (22) is a PDF of an $6m$-dimensional, uncorrelated Normal distribution with zero mean and standard deviation of $\sqrt{\tau}$, all we have to do to simulate its effect on an ensemble of configurations is to add, to each component of $\mathbf{r}$, a normally distributed random number (this process is known as DIFFUSION).

Similarly, G_{dr} (the so called DRIFT) represents moving each electron in a velocity field $\mathbf{F}_i$ for time τ. Note that, to our $O(\tau)$ accuracy, it is sufficient to evaluate $\mathbf{F}$ (collection of all the $\mathbf{F}_i$'s) at the current configuration $\mathbf{r}$, and then add $\tau\mathbf{F}(\mathbf{r})$ to $\mathbf{r}$.

Finally, G_{br} (BRANCHING) increases or decreases the WEIGHT of each configuration according to $\exp\{-\tau[E_L(\mathbf{r}_0) - E_0]\}$, without changing the electrons' location. The resulting weights are subsequently used to average E_L (and other quantities).

Monte Carlo technique thus works as follows: we create (randomly) an ENSEMBLE of many (typically, thousands of) configurations, representing an initial (arbitrary) distribution. To advance each configuration of this ensemble by a TIME STEP τ, we

1. add, to each component of every configuration, a random number having the Normal distribution with zero mean and standard deviation equal to $\sqrt{\tau}$ (these must be independent of each other),

2. for each configuration, we evaluate $\mathbf{F}_i = \frac{\boldsymbol{\nabla}_i \Psi}{\Psi}$, and add $\tau\mathbf{F}_i$ to the location of electron i (done with all electrons),

3. we also compute E_L (this can be done, together with $\mathbf{F}_i$, in the previous step), and convert it to $\exp\{-\tau[E_L(\mathbf{r}) - \overline{E_L}]\}$ to be used for computing the configuration's weight (see below).

Note that, in the last step, E_0 of Eq. 24 (the exact but unknown ground-state eigenvalue) had to be replaced by $\overline{E_L}$ (the ensemble's *weighted* average of individual local energies E_L).

The previous three steps constitute one ITERATION of the process, advancing our solution by one time step. We then repeat the whole cycle (of an iteration) as many times as possible.

The weight of each configuration should be computed as a *cumulative product* of the $\exp\{-\tau[E_L(\mathbf{r}) - \overline{E_L}]\}$ factors from all past iterations. If unchecked, this would ultimately lead to instability (one configuration accumulating a bigger weight than the rest of the ensemble put together) - to fix this, we compute the configurations' weights as a product of the $\exp\{-\tau[E_L(\mathbf{r}) - \overline{E_L}]\}$'s from the *last* ℓ iterations only, where ℓ is a judiciously chosen function of τ (introducing yet another $O(\tau)$ error).

As soon as we establish that there is no systematic increase or decrease of $\overline{E_L}$ values (the process has stabilized and reached its stationary solution - only random statistical fluctuations remain), these can be used as estimates of the ground-state energy E_0. It is to our advantage (to reduce the statistical error of the final estimate) to collect as many individual iteration averages as we can, eventually averaging these as well (creating a single, GRAND AVERAGE of local energies, over both configurations *and* iterations). At the same time, the corresponding ensembles (*including weights*) provide a sample from $\phi\Psi$ (the so called MIXED distribution).

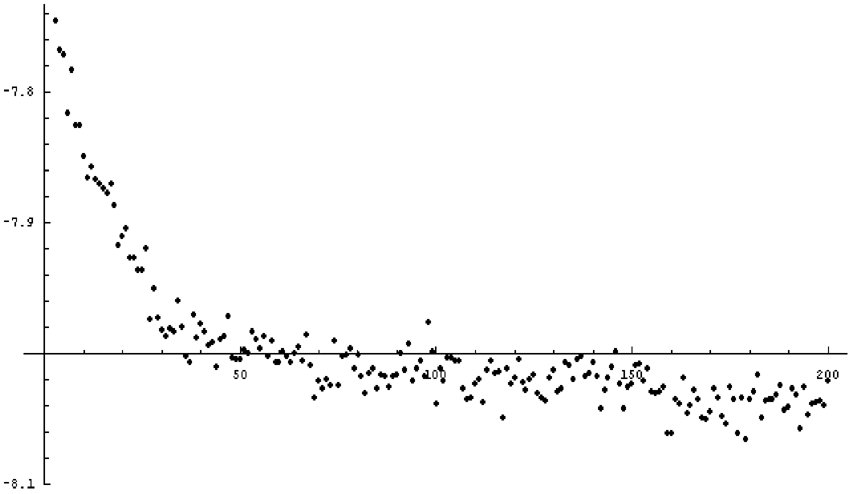
To understand why $\overline{E_L}$ estimates E_0, consider this (read $\simeq$ as 'is an estimate of'):

$$\overline{E_L} \simeq \int \phi\Psi E_L d\mathbf{r} = \int \phi\mathcal{H}\Psi d\mathbf{r} = \int \Psi\mathcal{H}\phi d\mathbf{r} = \int \Psi E_0 \phi d\mathbf{r} = E_0 \tag{25}$$

based on (12), $\mathcal{H}$ being Hermitian, (1), and the fact that E_0 is a constant.

And the last important technicality: a configuration may occasionally land so close to a singularity that both $\mathbf{F}$ and E_L become (in absolute value) extremely large. Even though such occurrences are quite rare, they are capable of seriously disturbing our simulation. A simple fix (introducing yet another τ-proportional error) is not to allow any component of $\mathbf{F}$ to have a magnitude bigger than $\frac{0.1}{\tau}$ (truncating it accordingly, when it does), and not to allow E_L to deviate from its grand average (as computed so far) by more than $\frac{1}{\tau}$ (with a similar truncation).

Using the trial function of the previous section with $\tau = 0.025$ and $\ell = 145$ (it should be matched to the equilibration time - see below), we have advanced 1000 configurations (the initial distribution was generated randomly from a unit cube centered on Li for electrons 1 and 3 and on H for electrons 2 and 4) for a total of 12000 iterations. The following graph displays $\overline{E_L}$ for the first 200 iterations



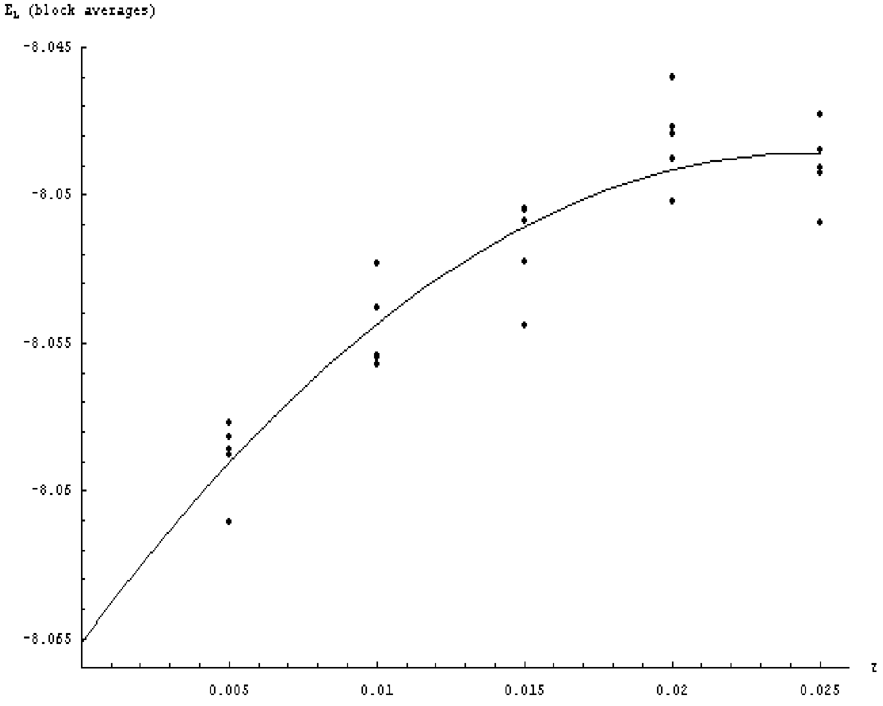
$\overline{E_L}$ against the iteration number

One can observe that the process has stabilized (EQUILIBRATED) after approximately 150 iterations, which indicates that consecutive values of $\overline{E_L}$ are *highly correlated*, since it takes roughly 150 iterations for this serial correlation to become negligible. This creates a bit of a problem when computing the standard error of the corresponding grand average (we cannot treat the individual $\overline{E_L}$'s as independent). To expediently resolve this, we divide the whole run into 'blocks' of 2000 consecutive iterations each, compute the resulting block averages (which *are*, to a good approximation, statistically independent), and only then find the corresponding sample mean and standard deviation of these, in our case 5 values (we discarded the first block of 2000 to allow, rather generously, for equilibration). This resulted in the value of -8.0589, with the standard error of 0.0006.

Unfortunately, we cannot take this to be an adequate estimate of E_0, since the estimate contains a τ-dependent bias of our several approximations. To remove this bias, we have to repeat the whole procedure using several different values of τ (0.020, 0.015, 0.010 and 0.005 in our case), then extrapolate the answers to $\tau = 0$ by performing a polynomial regression on the resulting block averages, with τ as the independent variable.

In general, the serial correlation increases with $\frac{1}{\tau}$ (this is the reason why we cannot choose our τ's too small). It is thus a good idea to increase the block size accordingly (to 2500, 3333, 5000 and 10000 respectively). Similarly (but slightly *faster* than $\frac{1}{\tau}$, not to introduce a systematic bias), we have to keep on increasing the value of ℓ. The nearest integer higher than $\frac{2-\log_{10}\tau}{\tau}$ proved to be a good choice.

In our case, this resulted in (each point represents a block average, the line is the best least-square fit):



Quadratic extrapolation to $\tau = 0$

Performing a quadratic polynomial regression (usually adequate), we found the intercept (our ultimate estimate of E_0), to have the value of -8.0651 ± 0.0015. Considering our rather crude trial function (ordinary trial functions usually employ dozens of atomic orbitals), the agreement with the exact value of -8.0700 is fairly good.

4 Trial-function optimization

First we observe that, by removing the last term on the left-hand side of (16), we get a new, simpler equation whose *stationary* solution (in terms of f) is Ψ^2 (up to a multiplicative constant). This is obvious from

$$\boldsymbol{\nabla}_i \Psi^2 = 2\Psi \ \boldsymbol{\nabla}_i \Psi \tag{26}$$

and

$$\boldsymbol{\nabla}_i^2 \Psi^2 = 2\Psi \ \boldsymbol{\nabla}_i^2 \Psi + 2(\boldsymbol{\nabla}_i \Psi) \bullet \boldsymbol{\nabla}_i \Psi \tag{27}$$

Ψ^2 is clearly a non-negative function which, when normalized to

$$\frac{\Psi^2}{\int \Psi^2 d\mathbf{r}} \tag{28}$$

defines a multivariate PDF of the so called VARIATIONAL DISTRIBUTION. Based on the arguments of the previous section, we can generate a sample from this distribution if we *drop the branching step* (performing only diffusion and drift) of the Monte Carlo simulation described above. The result we get (in terms of the $\overline{E_L}$ estimate) is not as good as the one we obtain by sampling the mixed distribution (*with* branching), but this variational sampling enables us to optimize the parameters of Ψ by *minimizing* the left-hand side of (8) or, more accurately, its estimate

$$\overline{E_L} \simeq \frac{\int E_L \Psi^2 d\mathbf{r}}{\int \Psi^2 d\mathbf{r}} = \frac{\int \Psi \mathcal{H} \Psi d\mathbf{r}}{\int \Psi^2 d\mathbf{r}} \tag{29}$$

Furthermore, there is a modification of the Monte Carlo procedure which enables us, in this case, to *entirely remove* the τ-dependent error of the simulation, and thus avoid the rather expensive $\tau \to 0$ extrapolation. It is called METROPOLIS SAMPLING, and it deserves a detailed description:

We can always generate a sample from a known distribution (normalized Ψ^2 in our case) by advancing each configuration $\mathbf{r}_0$ according to an *arbitrarily* Green's function, say

$$G(\mathbf{r}_0 \to \mathbf{r}, \ \tau) \tag{30}$$

and then accepting this 'MOVE' with the probability of

$$T \equiv \frac{\Psi^2(\mathbf{r})}{\Psi^2(\mathbf{r}_0)} \cdot \frac{G(\mathbf{r} \to \mathbf{r}_0, \ \tau)}{G(\mathbf{r}_0 \to \mathbf{r}, \ \tau)} \tag{31}$$

(accepting *automatically* when this quantity is bigger than 1). When a move is rejected, we take $\mathbf{r} = \mathbf{r}_0$ (the configuration stays put).

Mathematically, we can express the overall conditional PDF of the new $\mathbf{r}$, given that we start at $\mathbf{r}_0$, as follows

$$P(\mathbf{r}_0 \to \mathbf{r}) = \begin{cases} G(\mathbf{r}_0 \to \mathbf{r},\ \tau) \quad \text{when} \quad T > 1 \\ \frac{\Psi^2(\mathbf{r})}{\Psi^2(\mathbf{r}_0)} \cdot G(\mathbf{r} \to \mathbf{r}_0,\ \tau) + c \cdot \delta(\mathbf{r} - \mathbf{r}_0) \quad \text{when} \quad T \le 1 \end{cases} \tag{32}$$

where $\delta(\mathbf{r} - \mathbf{r}_0)$ stands for the stay-where-you-are distribution (the δ-function equals 0 when $\mathbf{r} \neq \mathbf{r}_0$, thus contributing only in the $T = 1$ case), and c is a constant such that $\int P(\mathbf{r}_0 \to \mathbf{r})\ d\mathbf{r}_0 = 1$ (note that $\int \delta(\mathbf{r} - \mathbf{r}_0)\ d\mathbf{r} = 1$).

One can easily verify that

$$\Psi^2(\mathbf{r}_0)P(\mathbf{r}_0 \to \mathbf{r}) = \Psi^2(\mathbf{r})P(\mathbf{r} \to \mathbf{r}_0) \tag{33}$$

which implies that

- the process, after reaching equilibrium, is REVERSIBLE (its Green's function remains the same when the direction of time is reversed),
- Ψ^2 is its stationary distribution.

The last assertion can be easily proven by integrating the previous equation with respect to $\mathbf{r}_0$, resulting in

$$\int \Psi^2(\mathbf{r}_0)P(\mathbf{r}_0 \to \mathbf{r})\ d\mathbf{r}_0 = \Psi^2(\mathbf{r}) \tag{34}$$

Note that, in our case

$$G(\mathbf{r}_0 \to \mathbf{r},\ \tau) = \exp\left(-\frac{[(\mathbf{r} - \mathbf{r}_0 - \tau\mathbf{F}(\mathbf{r}_0)]^2}{2\tau}\right) \tag{35}$$

(combining drift and diffusion), where every component of $\tau\mathbf{F}$ needs to be truncated to ± 0.1, to prevent configurations from getting 'stuck' (to have all attempted moves rejected). Luckily, this does not interfere with the technique being exact, as our $G(\mathbf{r}_0 \to \mathbf{r},\ \tau)$ is arbitrary (up to a point - it is to our advantage to use G whose stationary distribution is *close* to Ψ^2, thus minimizing the number of move rejections - (35) achieves that nicely).

Using this technique, one can easily show that the trial function of our example yields -8.027 as the *variational* estimate of E_0 (we have already seen that the corresponding *mixed-distribution* estimate was 9 times closer to the correct value - an order of magnitude improvement is quite typical here).

Returning to the issue of Ψ-optimization: it is relatively easy to compute (preferably analytically, numerically if necessary) the first derivative of Ψ (and of $\frac{\boldsymbol{\nabla}_i^2\Psi}{\Psi}$, and consequently of E_L) with respect to any of its parameters (i.e. C_{ij}, ξ, a and b - see the first section) - we will denote this derivative, rather symbolically, Ψ'. Considering the parameter to be 'variable', we can easily differentiate the left-hand side of (8), namely

$$\frac{\int E_L \Psi^2 d\mathbf{r}}{\int \Psi^2 d\mathbf{r}} \tag{36}$$

with respect to this parameter, getting

$$\frac{\int E_L' \Psi^2 d\mathbf{r}}{\int \Psi^2 d\mathbf{r}} + \frac{2\int E_L \Psi \Psi' d\mathbf{r}}{\int \Psi^2 d\mathbf{r}} - \frac{2\int \Psi \Psi' d\mathbf{r} \cdot \int E_L \Psi^2 d\mathbf{r}}{\left(\int \Psi^2 d\mathbf{r}\right)^2} \tag{37}$$

The last expression can be estimated by

$$\overline{E_L'} + 2\overline{E_L \cdot \tfrac{\Psi'}{\Psi}} - 2\overline{E_L} \cdot \overline{\tfrac{\Psi'}{\Psi}} \equiv \overline{E_L'} + 2\mathrm{Cov}(E_L, \tfrac{\Psi'}{\Psi}) \tag{38}$$

where 'Cov' denotes the usual SAMPLE COVARIANCE between the values of E_L and $\frac{\Psi'}{\Psi}$.

We can then use Monte Carlo to generate a sample of configurations from (28) and, for each of the Ψ's parameters, compute (38) and the corresponding standard error. When there is a statistically significant indication (the magnitude of (38) is bigger than twice its standard error) that the derivative is non-zero, we increase or decrease the corresponding parameter accordingly (to *minimize* (36), we increase when (38) is *negative*, decrease for (38) positive), by some small amount (say 0.001). We then take a few extra iterations to re-equilibrate, re-compute (38) and modify the parameters of Ψ again. This, after many iterations and parameter adjustments, will ultimately lead to optimizing the trial function (this is how we have obtained Ψ of our example).

This approach can also be employed to optimize the internuclear distance(s) - for our trial function, this has lead to the value of 3.020 ± 0.002 for the distance between the Li and H nuclei (in excellent, and probably slightly fortuitous, agreement with the exact value of 3.015). Unlike the original parameters of Ψ, internuclear distances also affect the value of V - this needs to be considered in the computation of E_L'.

Internuclear distances (and other similar molecular properties) can be estimated even more accurately based on the 'exact' results of a Mote Carlo simulation with branching. These and related issues will be discussed in a future article.

References

[1] Kolos,W and Wolniewicz L, *J. Chem. Phys.* **46**, 1426 (1967)

[2] Kalos M H, Levesque D and Verlet L, *Phys. Rev. A,* **1**, 2178 (1978)

[3] Anderson J B, *J. Chem. Phys.,* **73**, 3897 (1980)

[4] Reynolds P J, Ceperley D M, Adler B J and Lester W A Jr, *J. Chem. Phys.,* **77**, 5593 (1982)

[5] Ceperley D and Alder B, *Science,* **231**, 555 (1986)