Consider a trial wave function

$$\Psi = \lim_{c \to 0} \Psi_0 + c\Psi_1 \tag{1}$$

where Ψ_0 is the ground state and Ψ_1 is some excited state of a Hamiltonian H with energies E_0, E_1 respectively.

As $c \to 0$ the nodes of Ψ match up with the nodes of Ψ_0. Let's consider one such nodal point X where $\Psi(X) = \Psi_0(X) + c\Psi_1(X) \to 0$. We want to calculate the following quantity:

$$\lim_{r \to 0} \left[\langle E_L \frac{\partial_c \Psi}{\Psi} \rangle - \langle E_L \rangle \langle \frac{\partial_c \Psi}{\Psi} \rangle \right]_{D(X,r)}, \tag{2}$$

where the average $\langle \rangle$ is conducted by sampling Ψ^2 but with the positions sampled R restricted to the disk $R \in D(X,r)$ centered at the nodal point X with radius r.

1   Local energy

First we will expand the local energy in powers of c, and then in powers of r.

$$E_L = \frac{H\Psi}{\Psi} = \lim_{c \to 0} \frac{E_0\Psi_0 + cE_1\Psi_1}{\Psi_0 + c\Psi_1} \sim \lim_{c \to 0}(E_0\Psi_0 + cE_1\Psi_1) * (1 - c\Psi_1/\Psi_0)/\Psi_0 =$$

$$E_L(X + r) = \lim_{c \to 0} E_0 + c(E_1 - E_0)\frac{\Psi_1}{\Psi_0} + O(c^2)$$

Now expanding in terms of r we get:

$$\frac{\Psi_1(X + r)}{\Psi_0(X + r)} = \frac{\Psi_1(X) + \Psi_1'(X)r + O(r^2)}{\Psi_0(X) + \Psi_0'(X)r + O(r^2)}.$$

Note that $\Psi_0(X) = 0, \Psi_1'(X) = \Psi_1''(X) = ... = 0$, namely that near the nodes Ψ_0 is linear in r whereas Ψ_1 is a *constant*. Using these facts we can reduce the above expression to lowest order in r as: $\frac{\Psi_1}{\Psi_0'}(X)\frac{1}{r}$. The final expression we get is then:

$$\boxed{\lim_{r \to 0} E_L = \lim_{c,r \to 0} E_0 + c(E_1 - E_0)\frac{\Psi_1(X)}{\Psi_0'(X)}\frac{1}{r} + O(c^2) + O(r^{-2}).} \tag{3}$$

This is an expression for the local energy in a disk of radius r about the nodal point X for the wave function in equation (1).

2 Wave function derivative

Now we will expand the wave function derivative in powers of c, then in powers of r.

$$\frac{\partial_c \Psi}{\Psi} = \lim_{c \to 0} \frac{\Psi_1}{\Psi_0 + c\Psi_1} = \lim_{c \to 0} \frac{\Psi_1}{\Psi_0}(1 - c\frac{\Psi_1}{\Psi_0}) + O(c^2).$$

Using the expression above for $\Psi_1(X+r)/\Psi_0(X+r)$ and substituting we get that

$$\boxed{\lim_{r \to 0} \frac{\partial_c \Psi}{\Psi}(X+r) = \lim_{c,r \to 0} \frac{\Psi_1(X)}{\Psi_0'(X)}\frac{1}{r} - c(\frac{\Psi_1(X)}{\Psi_0'(X)}\frac{1}{r})^2 + O(c^2) + O(r^{-3})}. \quad (4)$$

3 Equation (2)

We can now combine equations (3) and (4) to get the value of (2) that we were interested in. The first term in equation (2) takes the form:

$$\lim_{c,r \to 0} \langle E_0 \frac{\Psi_1(X)}{\Psi_0'(X)}\frac{1}{r}\rangle + \langle c(E_1 - 2E_0)(\frac{\Psi_1(X)}{\Psi_0'(X)}\frac{1}{r})^2\rangle + O(c^2) + O(r^{-3})$$

and the second term takes form:

$$\lim_{c,r \to 0} \langle E_0\rangle\langle \frac{\Psi_1(X)}{\Psi_0'(X)}\frac{1}{r}\rangle - \langle c(E_1-E_0)\frac{\Psi_1(X)}{\Psi_0'(X)}\frac{1}{r}\rangle\langle\frac{\Psi_1(X)}{\Psi_0'(X)}\frac{1}{r}\rangle - \langle E_0\rangle\langle c(\frac{\Psi_1(X)}{\Psi_0'(X)}\frac{1}{r})^2\rangle + O(c^2) + O(r^{-3})$$

subtracting these two terms we come to the conclusion:

$$(2) = c(E_1 - E_0)\Big(\langle(\frac{\Psi_1(X)}{\Psi_0'(X)}\frac{1}{r})^2\rangle - \langle\frac{\Psi_1(X)}{\Psi_0'(X)}\frac{1}{r}\rangle^2\Big) + O(c^2) + O(r^{-3}) \to$$

$$\boxed{(2) \lim_{c,r \to 0} c(E_1 - E_0)\text{Var}[\frac{\partial_c \Psi}{\Psi}]_{D(X,r)} + O(c^2) + O(r^{-3})} \quad (5)$$

where equation 4 was substituted in to lowest order in c,r, and the Monte Carlo estimator for the variance is evaluated on samples of position R drawn from Ψ^2 but restricted to the disk $R \in D(X,r)$ centered at the nodal point X with radius $r \to 0$.

4 $\partial E/\partial c$

In order to conduct the total parameter derivative we need to sample positions from the entire wave function, which we can write in the following way:

$$\frac{\partial E}{\partial c} = \left[\langle E_L \frac{\partial_c \Psi}{\Psi} \rangle - \langle E_L \rangle \langle \frac{\partial_c \Psi}{\Psi} \rangle \right]$$

where the expectation is over the entire domain Ω of Ψ and points are sampled from Ψ^2. We can decompose this into parts which deal with the nodes, and parts which don't, as:

$$\frac{\partial E}{\partial c} = \left[\langle E_L \frac{\partial_c \Psi}{\Psi} \rangle - \langle E_L \rangle \langle \frac{\partial_c \Psi}{\Psi} \rangle \right]_{\Omega/nodes} + \int dX(2)(X). \tag{6}$$

The first term here is the Monte Carlo average calculated on all points excluding the nodal surface, and the second term is the average calculated on the nodal surface: here the integral over X is over the nodal hypersurface. The first term is well behaved as seen empirically, and from the theory calculation above is also well behaved as away from the nodes Ψ_0 has a lowest order constant term, removing all divergences w.r.t r in expression (5).

On an infinite sample the variance term in (2) is well behaved since $\langle 1/r^2 \rangle \sim \int \Psi_0^2 (1/r^2) \sim \int r^2/r^2$ does not diverge near the nodes. However, if we take a finite sample, it's easy to see that the *variance of the estimator for the variance (sample variance)* is not well behaved.

$$\text{Var}[s^2] \sim \int (1/r^4)\Psi_0^2 dr \sim \int (1/r^2) dr.$$

This expression does diverge near the nodes, where s^2 is a finite sample estimator of $\text{Var}[\frac{\partial_c \Psi}{\Psi}]$ and the integration is conducted about D(X,r). Therefore we come to our final conclusion:

> On a finite MC sample, term 1 in equation (6) has a finite variance

> while term 2 in equation (6) has a large (possibly infinite) variance.

> On an infinite MC sample both terms are finite with zero variance.

5   More general derivation

Consider a trial wave function of the form:

$$\Psi = \Psi_0 + c^p \Psi_p$$

where Ψ_0, Ψ_p are not eigenstates of the Hamiltonian H. We want to calculate the functional dependence of the following quantities near nodes:

$$\frac{H\Psi}{\Psi}, \frac{\partial_p \Psi}{\Psi}$$

5.1 Local energy

Let's consider the denominator of the local energy first, namely let's Taylor expand Ψ near the node $\mathbf{N}$ as

$$\Psi(\mathbf{N} + \mathbf{r}) = 0 + \nabla\Psi(\mathbf{N}) \cdot (\mathbf{r} - \mathbf{N}) + \mathcal{O}((\mathbf{r} - \mathbf{N})^2)$$

The numerator can be considered in the same way. The numerator can be evaluated at the node by noting that $H\Psi$ is a wave function in our Hilbert space which generally not have the same nodes as Ψ unless Ψ happens to be an eigenstate. Foregoing the chance that Ψ is an eigenstate, we can Taylor expand the numerator as

$$H\Psi(\mathbf{N} + \mathbf{r}) = C_H + \nabla(H\Psi)(\mathbf{N}) \cdot (\mathbf{r} - \mathbf{N}) + \mathcal{O}((\mathbf{r} - \mathbf{N})^2)$$

Taking the ratio and limit as we get arbitrarily close to the nodes

$$\lim_{||\mathbf{r}-\mathbf{N}||\to 0} E_L(\mathbf{r}) = \frac{C_H + \nabla(H\Psi)(\mathbf{N}) \cdot (\mathbf{r} - \mathbf{N}) + \mathcal{O}((\mathbf{r} - \mathbf{N})^2)}{0 + \nabla\Psi(\mathbf{N}) \cdot (\mathbf{r} - \mathbf{N}) + \mathcal{O}((\mathbf{r} - \mathbf{N})^2)} = \frac{C_H}{\nabla\Psi(\mathbf{N}) \cdot (\mathbf{r} - \mathbf{N})} + \mathcal{O}((\mathbf{r}-\mathbf{N})^0)$$

5.2 Wave function derivative

The denominator of this term is the same as the local energy, the numerator is just Ψ_p. We note that near a node Ψ_p is generally (assuming the nodes of the different determinants don't line up!) non-zero and has the Taylor expansion:

$$\Psi_p = C_P + \nabla\Psi_p(\mathbf{N}) \cdot (\mathbf{r} - \mathbf{N}) + \mathcal{O}((\mathbf{r} - \mathbf{N})^2)$$

Again, taking the ratio and limit as we get arbitrarily close to the nodes

$$\lim_{||\mathbf{r}-\mathbf{N}||\to 0} \frac{\partial_p \Psi}{\Psi} = \frac{C_P + \nabla\Psi_p(\mathbf{N}) \cdot (\mathbf{r} - \mathbf{N}) + \mathcal{O}((\mathbf{r} - \mathbf{N})^2)}{0 + \nabla\Psi(\mathbf{N}) \cdot (\mathbf{r} - \mathbf{N}) + \mathcal{O}((\mathbf{r} - \mathbf{N})^2)} = \frac{C_P}{\nabla\Psi(\mathbf{N}) \cdot (\mathbf{r} - \mathbf{N})} + \mathcal{O}((\mathbf{r}-\mathbf{N})^0)$$

5.3 "Local" $\partial E/\partial p$

We can define a quantity which I'll call the "local dE/dp" which is defined in the following way:

$$\frac{\partial E}{\partial p}(r, \mathbf{r}) = \left[\langle E_L \frac{\partial_p \Psi}{\Psi} \rangle - \langle E_L \rangle \langle \frac{\partial_p \Psi}{\Psi} \rangle \right]_{D(r,\mathbf{r})}$$

where the brackets $\langle\rangle$ are Monte Carlo estimates using samples drawn from Ψ^2 drawn from the disk of radius r around the point $\mathbf{r}$.

The "local dE/dp" expression near the nodes can then be written as:

$$\lim_{r\to 0}\frac{\partial E}{\partial p}(r,\mathbf{N}) = C_H C_P \mathrm{Var}\Big[\frac{1}{\nabla\Psi(\mathbf{N})\cdot(\mathbf{r}-\mathbf{N})}\Big]_{D(r,\mathbf{N})} + \mathcal{O}(r^0) \qquad (7)$$

Near any non-nodal point of Ψ we can note that the zeroth order term in the Taylor expansion of Ψ does not vanish so that.

$$\lim_{||\mathbf{r}-\mathbf{x}||\to 0} E_L(\mathbf{r}) = \frac{C_H}{\Psi(\mathbf{x})} + \mathcal{O}(\mathbf{r}-\mathbf{x}), \quad \lim_{||\mathbf{r}-\mathbf{x}||\to 0}\frac{\partial_p\Psi}{\Psi}(\mathbf{r}) = \frac{C_P}{\Psi(\mathbf{x})} + \mathcal{O}(\mathbf{r}-\mathbf{x})$$

Therefore for a point nearby a non-nodal point $\mathbf{x}$ the local energy and parameter derivative are constants. The "local dE/dp" expression near non-nodal points $\mathbf{x}$ can then be written as:

$$\lim_{r\to 0}\frac{\partial E}{\partial p}(r,\mathbf{x}) = 0 + \mathcal{O}(r) \qquad (8)$$

which makes sense because the evaluating dE/dp requires integrating over all space.

5.4   Finite and infinite MC samples

Let's consider equation (7) with both "infinite" (meaning we sample with infinite points and across the entire wave function domain) and finite number of MC samples.

In the first case, the expectation values can be converted to analytic integrals as:

$$\lim_{r\to 0}\frac{\partial E}{\partial p}(r,\mathbf{N}) = \lim_{r\to 0} C_H C_P\left[\int_{D(\mathbf{N},r)} dr\Psi^2(r)(\frac{1}{\nabla\Psi(\mathbf{N})\cdot r})^2 - \Big(\int_{D(\mathbf{N},r)} dr\Psi^2(r)\frac{1}{\nabla\Psi(\mathbf{N})\cdot r}\Big)^2\right]$$

substituting in $\Psi \sim \nabla\Psi(\mathbf{N})\cdot r$ near the nodes, we get:

$$\lim_{r\to 0} C_H C_P\left[\int_{D(\mathbf{N},r)} dr\ 1 - \Big(\nabla\Psi(\mathbf{N})\cdot\int_{D(\mathbf{N},r)} dr\ r\Big)^2\right] = 0$$

Both of these terms vanish in the limit since the domain of integration $D(\mathbf{N},r)$ vanishes and the integrands have no poles. Hence we see that near the nodes we get the same behavior away from the nodes if we use an analytic expression.

For infinite MC samples

$$\lim_{r\to 0}\frac{\partial E}{\partial p}(r,\mathbf{N}) = 0 \qquad (9)$$

Further, immediately before equation (8) we showed that the local energy and derivatives are constants near a non-nodal point, meaning that:

For infinite MC samples

$$\lim_{r \to 0} \frac{\partial E}{\partial p}(r, \mathbf{x}) = 0$$

(10)

Hence for "infinite" MC samples the "local dE/dp" behaves as expected for all points in the domain. Once integrated together we will get a non-divergent and well behaved value of dE/dp. (Duh)

In the latter case the situation is slightly more difficult. Equation (8) still holds in the finite case, so

For finite MC samples

$$\lim_{r \to 0} \frac{\partial E}{\partial p}(r, \mathbf{x}) = 0, \text{Var} = 0.$$

(11)

Equation (7) does not behave particularly nicely when you have a finite number of samples. In particular, consider an estimator for the variance which is just the sample variance s^2 of the quantity $\frac{1}{\nabla\Psi(\mathbf{N})(\mathbf{r}-\mathbf{N})}$. We can calculate properties of the sample variance, for sample $E[s^2]$ is the population variance, which is what we want in equation (10). We can also calculate the variance of this quantity analytically:

$$Var[s^2] \sim \lim_{r \to 0} \int_{D(r,\mathbf{N})} dr\, \Psi^2 \left(\frac{1}{\nabla\Psi(\mathbf{N}) \cdot r}\right)^4 \sim \lim_{r \to 0} \int_{D(r,\mathbf{N})} dr \left(\frac{1}{\nabla\Psi(\mathbf{N}) \cdot r}\right)^2 \to \infty.$$

(12)

This variance of our estimator diverges! Hence we see that

For finite MC samples, (7) has mean 0 but diverging variance.

This seems to be the core problem, that the $1/r^2$ behavior of dE/dp near the nodes makes estimating the value of dE/dp using a finite MC sample difficult.

6   Estimator for derivative

Noting that the derivative $\partial_p \Psi$ diverges near the node because $\Psi_p(\mathbf{N}) \neq 0$, we can simply subtract off that term to ensure that the divergence is gone near the nodes. Therefore, consider the following estimator (conditions described later).

$$\hat{\theta} = \begin{cases} \frac{\Psi_p(\mathbf{r})}{\Psi(\mathbf{r})} & \text{away from node} \\ \frac{\Psi_p(\mathbf{r}) - \Psi_p(\mathbf{N})}{\Psi(\mathbf{r})} & \text{near node} \end{cases}$$

(13)

There are a couple additional facts that are useful here. Since Ψ is linear near the node across which the sign of the wave function changes, it follows that

$$\Psi(\mathbf{r} + \mathbf{N}) = -\Psi(\mathbf{N} - \mathbf{r}) \text{ as } ||\mathbf{r} - \mathbf{N}|| \to 0 \tag{14}$$

meaning that if we integrate in a small ball around the nodal point then the integral of the wave function itself should vanish!

Let's then consider breaking up our integration domain into nodal "balls" of radius r and the rest of the domain. Therefore in the conditional of equation (13) "away from node" means not contained in *any* nodal balls, and "near node" means contained within the nodal ball centered at $\mathbf{N}$. We can then consider the expectation value of the estimator in (13):

$$E[\hat{\theta}] = \int_\Omega d\mathbf{r} \frac{\Psi_p}{\Psi} |\Psi|^2 - \int d\mathbf{N} \int_{B(\mathbf{N},r)} d\mathbf{r} \frac{\Psi_p(\mathbf{N})}{\Psi} |\Psi|^2 =$$

$$E[\frac{\Psi_p}{\Psi}] - \int d\mathbf{N} \Psi_p(\mathbf{N}) \int_{B(\mathbf{N},r)} d\mathbf{r} \Psi(\mathbf{r}) = E[\frac{\Psi_p}{\Psi}].$$

$$\boxed{E[\hat{\theta}] = E[\frac{\Psi_p}{\Psi}], \text{PDF} = |\Psi|^2} \tag{15}$$

Hence we see that the estimator we have in (13) exactly reproduces the expectation value of the parameter derivative but also removes the singularity near the nodes. The only question left is then how to implement this estimator, which involves thinking about how to define a ball of radius r, figure out whether a sample is near a node, and how to figure out which node the sample is near! **For fun:** the collection of "balls" attached to the nodal surface can be thought of like a Fiber bundle!