Optimisation of Many-Electron Wave Functions

QMC in the Apuan Alps 2017

Ryan J. Hunt
Lancaster University

2$^{\text{nd}}$ August 2017

Outline

Introduction

- Trial wave functions are at the heart of QMC. **Good** trial wave functions are at the heart of **good** QMC studies.

- QMC wave functions take some approximate starting point[1] (usually from a DFT code), and add some **exploitable variational freedom**.

- The QMC wave function fundamentally differs from the DFT wave function - it is **explicitly correlated**.

- In this talk, I aim to expose:
    1. How we add **variational freedom**.
    2. How we **exploit** it to make our trial wave functions better.

---

[1] Or, in rare cases, a known-good pen-and-paper starting point.

Outline

Antisymmetric Wave Functions

▶ Consider a general many-electron wave function

$$\Psi_S(\mathbf{X}) = \Psi(\{\mathbf{r}_1, s_1^z\}, \ldots, \{\mathbf{r}_{N_e}, s_{N_e}^z\}), \tag{1}$$

for a system of N_e electrons, having positions $\{\mathbf{r}_i\}$ and spin projections $\{s_i^z\}$. Let $\Psi_S(\mathbf{X})$ be an eigenfunction of $\hat{S} = \sum_i \hat{s}_i^z$.[2]

▶ Exact ground-states are eigenfunctions of $\hat{S}$ all cases where $\hat{\mathcal{H}}$ is **spin-independent**.

▶ A generic[3] expectation value then reads

$$\mathcal{O} = \frac{\langle \Psi_S | \hat{\mathcal{O}} | \Psi_S \rangle}{\langle \Psi_S | \Psi_S \rangle} = \frac{\sum_{\{s^z\}} \int d\mathbf{R} \, \Psi_S^\star(\mathbf{X}) \hat{\mathcal{O}}(\mathbf{R}) \Psi_S(\mathbf{X})}{\sum_{\{s^z\}} \int d\mathbf{R} \, |\Psi_S(\mathbf{X})|^2}. \tag{2}$$

---

[2] With eigenvalue $\sum_i s_i^z = (N_\uparrow - N_\downarrow)/2$.

[3] OK, $\hat{\mathcal{O}}$ has to be spin independent.

Antisymmetric Wave Functions (II)

▸ Remember, $\Psi_S(\mathbf{X})$ is antisymmetric under exchange of two *identical* electrons (same-spin electrons).

▸ We can order the arguments of Ψ_S any way we like: let's choose a new argument

$$\mathbf{X}' = (\{\mathbf{r}_{i_1}, \uparrow\}, \ldots, \{\mathbf{r}_{N_\uparrow}, \uparrow\}, \{\mathbf{r}_{N_\uparrow+1}, \downarrow\}, \ldots, \{\mathbf{r}_{N_e}, \downarrow\}), \tag{3}$$

leaves expectation values unchanged (permutation to return to $\mathbf{X}$ has parity ± 1 - both square to one).

▸ We can relabel the $i_n \to n$, as these are integration variables, and cancel the spin sums to obtain

$$\mathcal{O} = \frac{\int d\mathbf{R}\ \Psi^\star(\mathbf{R})\hat{\mathcal{O}}(\mathbf{R})\Psi(\mathbf{R})}{\int d\mathbf{R}\ |\Psi(\mathbf{R})|^2}, \qquad \underbrace{\Psi(\mathbf{R}) \equiv \Psi_S(\mathbf{X}')}_{\text{"The trial wave function"}} \tag{4}$$

Antisymmetric Wave Functions (III)

- $\Psi(\mathbf{R})$ is only antisymmetric w.r.t. exchange of **indistinguishable** electrons.
- $\Psi(\mathbf{R})$ is **easier to work with**, and we suffer no loss of generality by doing this (given $\Psi(\mathbf{R})$, we can recover $\Psi_S(\mathbf{X})$).
- Non-collinear spins? See Sec. 37.2 CASINO manual, and "3D_fluid_sdw" CASINO example.

Antisymmetric Wave Functions (IV)

- Is $\Psi(\mathbf{R})$ real? **Generally, no**. However, if the system you are studying has time-reversal symmetry ($\hat{\mathcal{H}}$ is real, with sensible b.c.'s), you can always **choose** $\Psi(\mathbf{R})$ real.

- CASINO can handle both ("complex_wf" keyword). Real wave functions are usually preferred, because real arithmetic is **faster**.

- Cases where one might want a complex wave function include:
  - Magnetic field $\mathbf{B} \neq \mathbf{0}$.
  - Calculations for solids where $\mathbf{k}$-point grid has no inversion symmetry.
  - Speed testing?

Slater-Jastrow Wave Functions

▶ So far we have considered generalities. Most QMC studies employ the
  Slater-Jastrow wave function form

$$\Psi_{SJ}(\mathbf{R}) = \exp\left[\mathcal{J}(\mathbf{R})\right] \cdot \underbrace{\sum_j \alpha_j \mathcal{D}_j^{\uparrow}(\mathbf{R})\mathcal{D}_j^{\downarrow}(\mathbf{R})}_{\text{"The Slater Part"}}, \tag{5}$$

where $\exp\left[\mathcal{J}(\mathbf{R})\right]$ is the **Jastrow factor**, $\{\alpha_j\}$ are (potential)
multideterminantal expansion coefficients, and $\mathcal{D}_j^{\uparrow/\downarrow}$ are up/down-spin
Slater determinants.

▶ CASINO can also use backflow functions and geminal (pairing) wave
  functions (later talk!).

Slater Determinants

▶ Slater determinants have the generic form

$$\mathcal{D}(\mathbf{R}) = \frac{1}{\sqrt{N_e}} \begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_2(\mathbf{r}_1) & \cdots & \phi_N(\mathbf{r}_1) \\ \phi_1(\mathbf{r}_2) & \phi_2(\mathbf{r}_2) & \cdots & \phi_N(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{r}_N) & \phi_2(\mathbf{r}_N) & \cdots & \phi_N(\mathbf{r}_N) \end{vmatrix}, \tag{6}$$

where the $\{\phi_i\}$ are a set of single-particle orbitals.[4]

▶ These orbitals **almost always** come from DFT, but can also be taken from HF. These orbitals
  - ▶ Can sometimes be **optimised** in QMC.
  - ▶ May be complex, but the determinant can **still be real**.
  - ▶ Can be represented in PW, Gaussian, . . . bases (see xwfn.data specifications).

---

[4] There will generally be two sets of $\{\phi_i\}$, one for each spin.

Slater Determinants (II)

Where can you acquire orbitals?

- ▶ Gaussians:
    - ▶ CRYSTAL
    - ▶ GAUSSIAN
    - ▶ GAMESS-US
    - ▶ Turbomole
    - ▶ ⟨any code which can write in the MOLDEN format⟩
- ▶ Plane Waves (to be converted to blips!):
    - ▶ QuantumEspresso (pwscf)
    - ▶ CASTEP
    - ▶ ABINIT
    - ▶ MCEXX
- ▶ Blips
    - ▶ Direct from pwscf.
- ▶ NAO / Slater orbitals: Others.

Slater Determinants (III)

The rest of the Slater part…

- If you like, you can use multideterminant expansions. The coefficients $\{\alpha_j\}$ can be
  - **optimised** in QMC (as variational parameters), or
  - taken as is from a **post-HF method**.[5]
- Why might you **not want to** do this?
  - Determinants are not a compact means of recovering **correlation energy**.
  - Solids (which are perhaps most interesting to QMC people) don't tend to have lots of **multi-reference character**.

---

[5] OR, I suppose, fixed by transformation properties of the desired trial state in an excited state calculation…

Figures of Merit: E_{corr}

- The **correlation energy**, E_{corr}, is a figure of merit we can use to judge how good a given wave function form is. Let the VMC energy of the wave function in question be E_{VMC}.

- If E_{HF} is the Hartree-Fock energy of the system,[6] and E_0 is the true ground state energy, we can define η as follows

$$\eta = \frac{E_{HF} - E_{VMC}}{E_{HF} - E_0}. \tag{7}$$

- η is the fraction of correlation energy retrieved by that given wave function.

- $\eta = 1$ is "in our dreams", $\eta > 0.8$ is **reasonable**.

- Because we don't know E_0, we often approximate it by the **fixed-node DMC energy**, E_{DMC}.

---

[6] Which is the lowest energy we can get with a single determinant!

Outline

Jastrow Factors

- The Jastrow factor, $\exp[\mathcal{J}(\mathbf{R})]$, is an explicit function of inter-particle distances.
- $\mathcal{J}$ is real, symmetric, and contains **optimisable parameters**.
- $\mathcal{J}$ offers a **compact** parametrisation of correlation effects. The number of parameters necessary to achieve a given η is **system-size independent**.
- If the wave function is to be antisymmetric, $\exp[\mathcal{J}] > 0$. Corollary: The Jastrow factor does not change the **nodal surface** of the wave function.[7]

---

[7] The nodal surface is the set of points on which the wave function is zero.

Jastrow Factors (II) - Why do we need them?

- Jastrow factors
  - Allow us to satisfy the **Kato cusp conditions**.[8]
  - **Lessen the extent** of systematic biases in the DMC method.
  - Allow us to obtain **smaller error bars** in VMC and DMC for the same amount of computational expense.
  - Improve the collection of quantities which are obtained from **extrapolated estimates**.[9]
  - **Reduce** the potency of population fluctuations (and related issues) in DMC calculations.
  - Allow us to obtain **even better** nodal surfaces when we optimise in parallel with parameters that affect the nodal surface.

---

[8] T. Kato, Communications on Pure and Applied Mathematics **10** (1957), R. T. Pack and W. B. Brown, The Journal of Chemical Physics **45** (1966).

[9] This requires a good VMC wave function.

Jastrow Factors (II) - Why do we need them?

- Jastrow factors
  - Allow us to satisfy the **Kato cusp conditions**.[8]
  - **Lessen the extent** of systematic biases in the DMC method.
  - Allow us to obtain **smaller error bars** in VMC and DMC for the same amount of computational expense.
  - Improve the collection of quantities which are obtained from **extrapolated estimates**.[9]
  - **Reduce** the potency of population fluctuations (and related issues) in DMC calculations.
  - Allow us to obtain **even better** nodal surfaces when we optimise in parallel with parameters that affect the nodal surface.
  - **Are worth the pain and headaches they might cause us when we struggle to optimise them.**

[8] T. Kato, Communications on Pure and Applied Mathematics **10** (1957), R. T. Pack and W. B. Brown, The Journal of Chemical Physics **45** (1966).

[9] This requires a good VMC wave function.

Kato Cusp Conditions (I) - Antiparallel Spins

- Consider two spins[10], having *opposite spins* ($\uparrow$, $\downarrow$), reduced mass $\mu = \frac{1}{2}$, and separated by a vector $\mathbf{r}$.

- We can write the two-particle wave function[11] for these two spins as

$$\psi(\mathbf{r}) = \sum_{\ell=0}^{\infty} \sum_{m_\ell=-\ell}^{\ell} \psi_{\ell,m_\ell}(r) \cdot \mathcal{Y}_{\ell,m_\ell}(\theta, \phi) \cdot r^\ell. \tag{8}$$

- Now we ask the question: what is the contribution to the local energy from this configuration?

$$E_L = \frac{-\nabla^2 \psi}{\psi} + \frac{1}{r} = \frac{-\nabla^2 \psi_{00}(r)}{\psi_{00}(r)} + \frac{1}{r} + \mathcal{O}(r). \tag{9}$$

- We care about small r values. Here, $\psi_{00}(r) \approx \psi_{00}(0) + r \cdot \left[\frac{\partial \psi_{00}(r)}{\partial r} \right]_{r=0}$

---

[10] In an isolated system, or as a subset of another.

[11] Which **doesn't** have to be spatially antisymmetric! I.e. we can use **both** odd and even ℓ.

Kato Cusp Conditions (II) - Antiparallel Spins cont.

- Substituting the Taylor expansion of ψ_{00} into the local energy, we find that

$$E_L = \frac{-2}{r\psi_{00}(0)}\left[\frac{\partial\psi_{00}(r)}{\partial r}\right]_{r=0} + \frac{1}{r} + \mathcal{O}(r). \tag{10}$$

- **Important bit:** E_L should not diverge as $r \to 0$. For an eigenfunction, it should be a constant (equal to the energy eigenvalue of that eigenfunction). Therefore, we require

$$\boxed{\left[\frac{\partial\psi_{00}}{\partial r}\right]_{r=0} = \frac{1}{2}\psi_{00}(0)} \tag{11}$$

$\underbrace{}$

"Antiparallel Kato Cusp Condition"

Kato Cusp Conditions (III) - Parallel Spins

▶ What changes when the spins are **parallel**? We have to make the spatial wave function manifestly **antisymmetric** - we only include odd ℓ!

$$\psi(\mathbf{r}) = \sum_{\text{odd } \ell} \sum_{m_\ell=-\ell}^{\ell} \psi_{\ell,m_\ell}(r) \cdot \mathcal{Y}_{\ell,m_\ell}(\theta,\phi) \cdot r^\ell. \tag{12}$$

▶ A similar analysis yields a set (one for each $-1 \leq m_\ell \leq 1$) of parallel-spin Kato conditions

$$\left[\frac{\partial \psi_{1,m_\ell}}{\partial r} \right]_{r=0} = \frac{1}{4}\psi_{1,m_\ell}(0) \tag{13}$$

"Parallel Kato Cusp Condition(s)"

Kato Cusp Conditions (IV)

▶ The cusp conditions mean that the local energy has discontinuities.

Figure 1: The effect of applying cusp conditions to a pair of coalescing electrons. These electrons were part of a family of 16, in an all-electron calculation for the O_2 dimer. The remnant electrons have been frozen in identical positions for both plots.
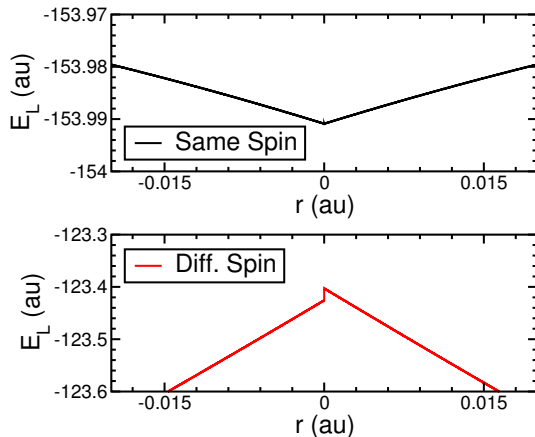
Kato Cusp Conditions (V)

▶ To really make the point - if you don't apply cusp conditions, you see this:
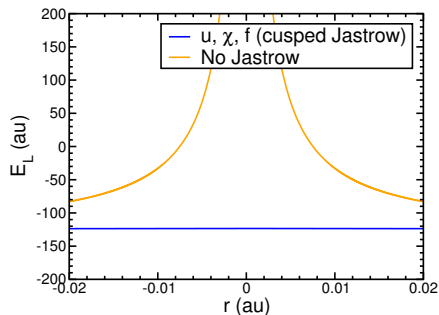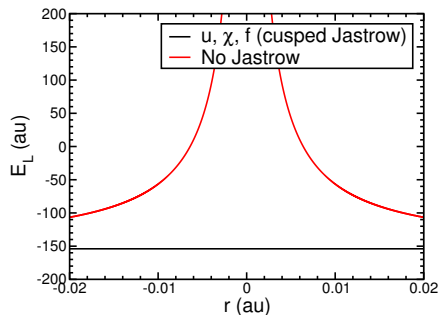

Figure 2: Different spins.

Figure 3: Same spins.

Electron-nucleus Cusp Conditions

▶ There are other cases where the local energy might diverge. I just showed you plots from **an all-electron calculation**.

▶ This means **bare nuclei** are present, and that the local energy may also diverge as an electron moves into a nucleus (of nuclear charge Z, let's say).

▶ In this case, the *spherical average* of $\psi_1(\mathbf{r})$[12] obeys

$$\left[\frac{\partial \langle \psi_1 \rangle_{\theta,\phi}}{\partial r} \right]_{r=0} = -Z \langle \psi_1 \rangle_{\theta,\phi}(0), \tag{14}$$

with

$$\langle a \rangle_{\theta,\phi}(r) \propto \frac{1}{4\pi r^2} \frac{\int d\mathbf{r}\, a(\mathbf{r}) \cdot P(\mathbf{r})}{\int d\mathbf{r}\, P(\mathbf{r})}. \tag{15}$$

---

[12] Which is a 1-electron wave function, with **r** centred on the nucleus (i.e. not a relative coordinate, like before)!

Cusp Conditions - Notes

- It is easy to see how conditions on the wave functions become conditions on the Jastrow factor (which we can fix ourselves!). Examples:[13]

$$\text{Antiparallel e-e cusp} \implies \left[\frac{\partial \mathcal{J}}{\partial r_2}\right]_{r_2=0} = \frac{1}{2}. \qquad (16)$$

$$\text{Parallel e-e cusp} \implies \left[\frac{\partial \mathcal{J}}{\partial r_2}\right]_{r_2=0} = \frac{1}{4}. \qquad (17)$$

$$\text{Electron-nucleus cusp} \implies \left[\frac{\partial \mathcal{J}}{\partial r_1}\right]_{r_1=0} = -Z. \qquad (18)$$

- These become **trivial constraints** on certain parameters in our Jastrow factor.

---

[13] All of these assume the Slater part of the corresponding wave functions is smooth at coalescence points!

Cusp Conditions - Notes (II)

- If you use a non-divergent pseudopotential, you **don't have a cusp**, and hence, you don't need the electron-nucleus cusp conditions.
  - Trying to apply cusp conditions in this case will lead to **bad behaviour** in VMC and (if you manage to optimise anything...) also in DMC. Trust me ☺.
- If your orbitals (the Slater part of the wave function) satisfy the cusp conditions, then the Jastrow factor **doesn't need to**.
  - Each orbital must separately satisfy the electron-nucleus cusp conditions for this to be the case!
- It is **impossible** to describe cusps with a finite number of Gaussian[14] / Plane Wave / Blip basis functions.
  - See **use_gpcc** and **cusp_correction** CASINO keywords.

---

[14] A. Ma et al., The Journal of Chemical Physics **122** (2005).

CASINO's Jastrow Factor(s)

- CASINO has two means of describing a Jastrow factor:
  - The old, fast way: the correlation.data file & pjastrow.f90.
  - The new, slower (but more general) way: the parameters.casl file & gjastrow.f90.
- I'll talk about the former, because it's *what I know*, and it's not massively restrictive.[15] If you want to **play around**, you can switch schemes and have CASINO turn a correlation.data file into a parameters.casl file.[16]

---

[15] A notable example is the ν-Jastrow factor, which is only available within the general scheme.
[16] I keep saying I will do this...

CASINO's Jastrow Factor (II)

- ▶ Regardless of how you access it, the "standard form" of CASINO's Jastrow factor[17] is

$$\mathcal{J}(\mathbf{R}) = \sum_{\substack{\text{e-e pairs} \\ i,j}} u(r_{ij}) + \sum_{\substack{\text{e-N pairs} \\ i,I}} \chi_I(r_{iI})$$

$$+ \sum_{\substack{\text{e-e-N triplets} \\ i,j,I}} f_I(r_{iI}, r_{jI}, r_{ij}) + \sum_{\substack{\text{e-e pairs} \\ i,j}} p(\mathbf{r}_{i,j}) + \text{ maybe a few others} \quad (19)$$

- ▶ $u(\chi)$ is an isotropic function of e-e(e-N) separations; $r_{ij}(r_{iI})$.
- ▶ f_I is an isotropic function of e-N and e-e separations; $r_{iI},\ r_{jI},\ r_{ij}$.
- ▶ p is a plane-wave expansion in e-e vectors; $\mathbf{r}_{ij}$
- ▶ Distances are evaluated in the **minimum image convention**.

---

[17] N. D. Drummond et al., Physical Review B **70** (2004).

CASINO's Jastrow Factor (III) - Gory Details

▶ Each of these contains **optimisable parameters**. Here, C is the truncation order (2 or 3 common), and $\mathbf{G}$ are simulation-cell reciprocal lattice vectors.

$$u(r_{ij}) = \Theta(L_u - r_{ij})(r_{ij} - L_u)^C \sum_{l=0}^{N_u} \alpha_l r_{ij}^l$$

$$\chi_I(r_{iI}) = \Theta(L_{\chi_I} - r_{iI})(r_{iI} - l_{\chi_I})^C \sum_{m=0}^{N_{\chi_I}} \beta_m r_{iI}^m$$

$$f_I(r_{iI}, r_{jI}, r_{ij}) = \Theta(L_{f_I} - r_{iI})\Theta(L_{f_I} - r_{jI})(r_{iI} - L_{f_I})^C (r_{jI} - L_{f_I})^C \times$$

$$\sum_{l=0,m=0}^{N_{f_I}^{eN}} \sum_{n=0}^{N_{f_I}^{ee}} \gamma_{lmn} r_{iI}^l r_{jI}^m r_{ij}^n$$

$$p(\mathbf{r}_{ij}) = \sum_{A,\mathbf{G}} a_A \cos\left(\mathbf{G} \cdot \mathbf{r}_{ij}\right) \tag{20}$$

CASINO's Jastrow Factor (IV) - Further Gory Details

- In the previous, the α_l, β_m, γ_{lmn} and a_A are **optimisable** parameters.
- The $L_{\langle\text{something}\rangle}$ are cut-off lengths, at these distances, the polynomials determined by the optimisable parameters go **smoothly** to zero.
- Cut-off lengths *can* be optimised,[18] but advice is: make L_u as big as possible, and make L_χ and L_f equal to around 1-2 bond lengths, say.
- A value of $C = n$ means that the $(n-1)^{\text{th}}$ derivative of the wave function is continuous everywhere. $C = 3$ is typical.
- We know different spin configurations have different cusp conditions. We can have **different parameters for each spin** if we wish (u-term).

---

[18] Only by variance minimisation!

Practical use of CASINO's JF

- When creating a new Jastrow factor, the easiest place to start is **with an existing one**. Can either find a similar one from previous work, or copy a close example from the CASINO examples directory.

- Sometimes, one only needs a certain subset of the available Jastrow functions. Terms describing higher correlations (f, H for example) **don't usually change much**.

- All parameters that are not cut-off lengths are "linear parameters", and we usually start knowing **none** of these.

- CASINO knows how to "fill in blanks" with reasonable defaults (i.e. defaults that **satisfy the cusp conditions**).

- Blank p terms can be generated with the **make_p_stars** utility.

Outline

Optimisation

- ▶ The **hardest** part of any QMC calculation is usually wave function optimisation.
- ▶ This problem defies all attempts at automation, and (especially if you consider systems that are capable of being unbound) can be quite **annoying**.
- ▶ However, the **rewards** are numerous, and we should all go to great lengths to make sure we have the **best wave function we can possibly have**.

Energy Minimisation

▶ Consider the exact eigenstates of a Schrödinger equation

$$\hat{\mathcal{H}}\Phi_i = E_i\Phi_i, \qquad E_0 < E_1 < E_2 < \ldots, \qquad \langle\Phi_i|\Phi_j\rangle = \delta_{i,j} \qquad (21)$$

▶ Any arbitrary wave function Ψ can be written as $\Psi = \sum_i c_i\Phi_i$, and we can define an energy functional

$$E[\Psi] = \frac{\langle\Psi|\hat{\mathcal{H}}|\Psi\rangle}{\langle\Psi|\Psi\rangle} = \underbrace{\frac{\sum_m|c_m|^2 E_m}{\sum_n|c_n|^2} \geq \frac{E_0\sum_m|c_m|^2}{\sum_n|c_n|^2}}_{\text{"the variational principle"}} = E_0. \qquad (22)$$

▶ Essentially, our $\{\alpha,\ \beta,\ \gamma,\ \ldots\}$ variational parameters control (with limited capacity) the particular coefficients c_i appearing in the expansion of our trial wave function.

▶ The idea of optimisation (generally) is to vary these such that some goal is met. In energy minimisation, the goal is to **minimise the energy**.

Energy Minimisation (II)

- We can't just do this by repeated VMC calculation. Why? **Noise**.
- We need optimisation methods that **don't require derivatives**, and that can **cope with noise**. We also want to minimise the number of individual "function evaluations" (here, each one of those would correspond to its own VMC calculation!).
- Before we do this, let's look at a general tool for use in statistical physics, **correlated sampling**.

Correlated Sampling

- Suppose that Ψ^A and Ψ^B are two different many-electron wave functions. The expectation value of an operator $\hat{\mathcal{O}}$ with respect to Ψ^A is

$$\langle \hat{\mathcal{O}} \rangle = \frac{\int d\mathbf{R} \, |\Psi^A(\mathbf{R})|^2 \times \left[\hat{\mathcal{O}} \Psi^A(\mathbf{R}) / \Psi^A(\mathbf{R}) \right]}{\int d\mathbf{R} \, |\Psi^A(\mathbf{R})|^2}$$

$$= \frac{\int d\mathbf{R} \, |\Psi^B(\mathbf{R})|^2 \times \left[|\Psi^A(\mathbf{R})|^2 / |\Psi^B(\mathbf{R})|^2 \right] \times \left[\hat{\mathcal{O}} \Psi^A(\mathbf{R}) / \Psi^A(\mathbf{R}) \right]}{\int d\mathbf{R} \, |\Psi^B(\mathbf{R})|^2 \times \left[|\Psi^A(\mathbf{R})|^2 / |\Psi^B(\mathbf{R})|^2 \right]} \quad (23)$$

Correlated Sampling (II)

- So, the expectation value $\langle \hat{\mathcal{O}} \rangle$ is the average of $\hat{\mathcal{O}}_L = \Psi^{A-1}\hat{\mathcal{H}}\Psi^A$...
- Over configurations which are distributed as $|\Psi^B|^2$...
- Weighted by $W_B^A = [|\Psi^A(\mathbf{R})|^2 / |\Psi^B(\mathbf{R})|^2]$...
- So what? We can calculate an expectation value with respect to one state (Ψ^A, say) by using configurations generated by Ψ^B.
- Estimates of $\langle \hat{\mathcal{O}} \rangle$ are a smooth function of any parameters in Ψ^A, for a given configuration set.

Correlated Sampling (II)

- So, the expectation value $\langle \hat{\mathcal{O}} \rangle$ is the average of $\hat{\mathcal{O}}_L = \Psi^{A-1} \hat{\mathcal{H}} \Psi^A$...

- Over configurations which are distributed as $|\Psi^B|^2$...

- Weighted by $W_B^A = [|\Psi^A(\mathbf{R})|^2 / |\Psi^B(\mathbf{R})|^2]$...

- So what? We can calculate an expectation value with respect to one state (Ψ^A, say) by using configurations generated by Ψ^B.

- Estimates of $\langle \hat{\mathcal{O}} \rangle$ are a smooth function of any parameters in Ψ^A, for a given configuration set.

- **This is a useful, general tool, that happens (in this case) to be handy for optimising wave functions!**

Reweighted Energy Minimisation

- Suppose we have an initial parameter set, $\{\mathbf{s}_0\}$, and initial wave function $\Psi_{\{\mathbf{s}_0\}}$.
- Generate a set of configs. $\{\mathbf{R}\}$ distributed according to $|\Psi_{\{\mathbf{s}_0\}}|^2$ in VMC.
- Using correlated sampling, we can estimate the energy expectation value of a wavefunction $\Psi_{\{\mathbf{s}\}}$ by the weighted mean of the local energy $\Psi_{\{\mathbf{s}\}}^{-1}\hat{\mathcal{H}}\Psi_{\{\mathbf{s}\}}$ - taken with the configurations $\{\mathbf{R}\}$ and weighted by $|\Psi_{\{\mathbf{s}\}}|^2/|\Psi_{\{\mathbf{s}_0\}}|^2$.
- **Why can't we use this method?**
    - We sample a **finite** number of configurations. The error in the reweighted energy gets large when $\Psi_{\{\mathbf{s}_0\}}$ and $\Psi_{\{\mathbf{s}\}}$ differ appreciably.
    - The reweighted energy has **false minima**.

Reweighted Variance Minimisation

- We can play the same game with a different objective function

$$\sigma^2_{RW}[\Psi] = \langle \hat{\mathcal{H}}^2 \rangle[\Psi] - \left(\langle \hat{\mathcal{H}} \rangle[\Psi] \right)^2. \tag{24}$$

- If Ψ were an eigenstate, σ^2_{RW} would be **zero**. We can therefore use σ^2_{RW} as an objective function, and try to minimise it![19]

- **Key difference:** this method often works in practice.

- The finite sampling errors persist, however, and we have to find **ways to mitigate them**. A common strategy is to perform **multiple cycles** of (i) configuration generation and (ii) optimisation.

---

[19] C. J. Umrigar et al., Physical Review Letters **60** (1988), P. R. C. Kent et al., Physical Review B **59** (1999), N. D. Drummond and R. J. Needs, Physical Review B **72** (2005).

Unreweighted Variance Minimisation

- If $\Psi(\mathbf{R})$ is an eigenstate, then $E_L(\mathbf{R}) = \Psi^{-1}(\mathbf{R})\hat{\mathcal{H}}\Psi(\mathbf{R})$ is a constant.

- Can optimise a trial wave function $\Psi_{\{\mathbf{s}\}}(\mathbf{R})$ by minimising the *unreweighted* variance

$$\sigma^2_{U,\{\mathbf{s}\}} = \frac{1}{N-1} \sum_{\mathbf{R}} \left[E_{L,\{\mathbf{s}\}}(\mathbf{R}) - \langle E_{L,\{\mathbf{s}\}} \rangle \right]^2, \tag{25}$$

  with N the **number of configurations** sampled.

- This minimisation procedure is like least-squares fitting, and N doesn't have to be too large. But...

- Suppose we take configurations from VMC (sample $\Psi_{\{\mathbf{s}_0\}}(\mathbf{R})$):
  - If we have infinitely many samples, σ^2_{RW} is independent of the starting trial function.
  - However, σ^2_U still **depends on** $\{\mathbf{s}_0\}$. In practice, we have to cycle URW variance minimisation until a **self-consistent** set of $\{\mathbf{s}\}$ are obtained.

Linear Parameters in $\mathcal{J}$ - "linjas" methods[20]

- ▶ Many Jastrow parameters are "linear": they appear as multipliers of non-linear functions of inter-particle distances.

- ▶ Suppose we have a Jastrow factor which is part linear (L linear parameters, α_l)

$$\mathcal{J}(\mathbf{R}) = \mathcal{J}_{NL}(\mathbf{R}) + \sum_{l=0}^{L} \alpha_l \cdot j_l(\mathbf{R}) \tag{26}$$

- ▶ **Important fact:** E_L is a **second order polynomial** in the linear parameters. Proof outline:
    - ▶ Important E_L contribution from $\Psi^{-1}\nabla^2\Psi$
    - ▶ Product rules on Laplacian
    - ▶ Has to be a term going like $\mathcal{J}'^2$
    - ▶ Terms have to be **trivial offsets** to $\mathcal{J}$ not to contribute quadratically to E_L

20 N. D. Drummond and R. J. Needs, Physical Review B **72** (2005).

Linear Parameters (II)

- ► Why is this neat? σ_U^2 is now a fourth order polynomial, and we **don't have to sum over configurations to optimise it**. We can find global minima of quartics fast, and repeatedly, along lines in parameter space.
- ► This is a **very** fast method, but only works for linear parameters.
- ► Happily, **most parameters in the Jastrow factor are linear**.
- ► If you have already given up on optimising cut-off lengths, and can sensibly fix them, you will benefit from using this method **greatly**.
- ► Other non-linear parameters: see other talks! (I'll mention these briefly later.)

Variance Minimisation URW vs. RW

- Generally speaking, one always wants to use URW variance minimisation. Why?
  - In the limit of large numbers of configurations, it usually leads to **lower energies** (i.e. objectively *better* wave functions).
  - In the limit of low numbers of configurations, it is usually **more stable**.
  - I just showed you that we can optimise the URW variance for linear Jastrow factors **very efficiently**.
  - The URW variance tends to have **fewer spurious minima** in parameter space (you can be fairly sure that obtaining the global minimum is possible and likely for a well-behaved system).

Practical Use in CASINO

Standard (URW) Variance Minimisation

- Set opt_method to "varmin", and run_type to "vmc_opt".
- Typically want number of configs (set by "vmc_nconfig_write") of order 100,000 (system-dependent).
- In order to ensure VMC data are not serially correlated, can either use "vmc_nstep" $\gg$ "vmc_nconfig_write".[21]
- Set what you want to optimise with the "opt_<x>" keywords.

Fast (Linear Param.) Variance Minimisation

- Set opt_method to "varmin_linjas" - only "opt_jastrow"!
- Cut-off lengths will be ignored (regardless of flag setting 0 or 1).

---

[21]OR set "vmc_decorr_period" (sets decorrelation loop length).

Parameters That Affect Nodes

- Backflow parameters, multideterminant expansion coefficients, orbital parameters, and other such parameters **affect the nodal surface of a trial wave function**.
- These are traditionally **hard** to optimise, because the local energy diverges at the nodal surface (and because points close to the nodal surface are rarely sampled!).
- URW variance is poorly defined in the limit of infinite sampling, and diverges when a configuration moves through a node (because E_L does…)
- When optimising parameters that affect the nodal surface, we often find that energy and variance minima are **significantly separated**.

Other Measures of Spread

▶ The variance is not special. There are other moments and measures on the local energy.

▶ Even if the variance diverges, some of these other **robust** measures of spread are finite. Singular points don't affect these measures as strongly as they do the variance.

▶ A good example is the Mean Absolute Deviation from the median (a.k.a. the MAD)

$$\mathrm{MAD} = \frac{1}{N} \sum_{\mathbf{R}} |E_L(\mathbf{R}) - E_{1/2}| \tag{27}$$

with $E_{1/2}$ the median.[22] This usually results in an energy lower than that achieved by variance minimisation.

---

[22] Because it separates the lower half of the energies from the upper half!

Diagonalising the Hamiltonian

- How do we actually find optimal parameter values?[23]
- Suppose we propose a (small) change in parameter values $\mathbf{s}_0$

$$\mathbf{s}_0 \to \mathbf{s} = \mathbf{s}_0 + \delta\mathbf{s}, \quad \Psi_{\mathbf{s}} = \Psi_{\mathbf{s}_0} + \sum_{i=1}^{N_s} \delta s_i \left[\frac{\partial\Psi}{\partial s_i}\right]_{\mathbf{s}_0} + \mathcal{O}\left[(\delta\mathbf{s})^2\right] \simeq \sum_{i=0}^{N_s} c_i\beta_i, \quad (28)$$

where c_i and β_i are defined by

$$c_i = \begin{cases} 1 & \text{if } i = 0 \\ \delta s_i & \text{if } i > 0 \end{cases}, \quad (29)$$

and

$$\beta_i = \begin{cases} \Psi_{\mathbf{s}_0} & \text{if } i = 0 \\ \left[\frac{\partial\Psi}{\partial s_i}\right]_{\mathbf{s}_0} & \text{if } i > 0 \end{cases}. \quad (30)$$

[23] M. P. Nightingale and V. Melik-Alaverdian, Physical Review Letters **87** (2001).

Diagonalising the Hamiltonian (II)

- The previous slide is approximate for most optimisable parameters, but is exact if one is optimising multideterminantal expansion coefficients (expression of **total** derivative).

- Generally this method is valid when $\delta\mathbf{s}$ is small. We usually iterate to **self-consistency**.

- Ignoring the terms $\mathcal{O}\left[(\delta\mathbf{s})^2\right]$, we can minimise $\langle\hat{\mathcal{H}}\rangle$ with respect to the c_i, keeping the norm of Ψ fixed.[Nightingale] I.e. we can solve the **secular equation**

$$(\mathcal{H} - ES)_{ij} \cdot c_j = 0_i \tag{31}$$

with

$$\mathcal{H}_{ij} = \langle\beta_i|\hat{\mathcal{H}}|\beta_j\rangle, \quad S_{ij} = \langle\beta_i|\beta_j\rangle. \tag{32}$$

Diagonalising the Hamiltonian (III)

- E is an eigenvalue. The lowest E corresponds to the set of parameters c_i which lead to the lowest energy.
- How do we estimate the values of the high-dimensional integrals appearing in $\mathcal{H}_{ij}$ and S_{ij}? **You tell me.**

Diagonalising the Hamiltonian (III)

- E is an eigenvalue. The lowest E corresponds to the set of parameters c_i which lead to the lowest energy.
- How do we estimate the values of the high-dimensional integrals appearing in $\mathcal{H}_{ij}$ and S_{ij}? **You tell me.**
- **VMC!**

Least Squares Method

- We could also consider diagonalising this Hamiltonian by another means.[24]

- If (and we will assume) the basis states $\{\beta_i\}$ span an **invariant subspace** of $\hat{\mathcal{H}}$, then

$$\hat{\mathcal{H}}\beta_i(\mathbf{R}) = \sum_{j=0}^{N_s} \eta_{ji}\beta_j(\mathbf{R}) \tag{33}$$

  for all of the (N total) VMC-generated configurations $\mathbf{R}$ and with some coefficients η_{ji}.

- If we know η, then we can solve a related problem on η ($\sum_j \eta_{ij} \cdot c_j = Ec_i$) for the $\delta\mathbf{s}$ (i.e. the eigenvector c_i).[25]

---

[24] J. Toulouse and C. J. Umrigar, The Journal of Chemical Physics **126** (2007).

[25] If c solves this equation, then you can show that c solves a Schrödinger-like equation. The method becomes equivalent to the previous one, with a sampling caveat.

Least Squares Method (II)

- If the $\{\beta_i\}$ did span an invariant subspace, we could take **any** $N_s + 1$ points in configuration space ($N_s + 1$ different configs **R**) and solve for η (a linear problem, with equal number of equations and unknowns).

- Usually, $N \gg N_s$ - so we have **far more equations** than we have **unknowns**. The problem is *overdetermined*.

- We can solve this problem by least-squares, we minimise

$$\chi^2 = \sum_{k=1}^{N} \sum_{l=0}^{N_s} \left| \frac{\hat{\mathcal{H}}\beta_l(\mathbf{R}_k) - \sum_{j=0}^{N_s} \eta_{jl}\beta_j(\mathbf{R}_k)}{\Psi_{\{\mathbf{s}_0\}}(\mathbf{R}_k)} \right|^2 \tag{34}$$

with respect to η_{ji}.

Least Squares Method (III)

How do we do this in practice?

► Define $B_{kl} = \beta_l(\mathbf{R}_k)/\Psi_{\{\mathbf{s}_0\}}(\mathbf{R}_k)$ and $B_{kl}^{\mathcal{H}} = \left[\hat{\mathcal{H}}\beta_l(\mathbf{R}_k)\right]/\Psi_{\{\mathbf{s}_0\}}(\mathbf{R}_k)$.

► The requirement that $\partial_{\eta_{ji}}\chi^2 = 0$ implies

$$\sum_{k=1}^{N}\left[B_{ki}^{\mathcal{H}} - \sum_{l=0}^{N_s}\eta_{li}B_{kl}\right]B_{kp}^{\star} = 0. \tag{35}$$

i.e.

$$\eta = \left(B^{\dagger}B\right)^{-1} - B^{\dagger}B^{\mathcal{H}}. \tag{36}$$

► Each of these matrix products are proportional to VMC estimates that **are known**.[26]

---

[26] Estimates of S and $\mathcal{H}$ to be exact.

Least Squares Method (IV)

- The previous diagonalisation method, and the least squares method, are equivalent in the limit of **infinite sampling**

$$\lim_{N \to \infty} \eta = S^{-1} \mathcal{H}. \tag{37}$$

- Least squares is also exact **if basis functions span an invariant subspace** of $\hat{\mathcal{H}}$ (useful example: if the $\{\mathbf{s}\}$ are multideterminantal expansion coefficients!). This is true even for finitely sampled configuration sets.

- **Point:** Least squares method is **usually better** - having much smaller finite sampling errors.

Our Approximations and Their Validity

- We've always stated the caveat that these methods are OK when parameters changes are *small*. We can try to prevent large parameter changes by **semi-orthogonalisation**.
- The idea is to pick a norm. for $\Psi_{\{s\}}$ which depends on parameters.
    - The norm. is such that $\{\beta_i\}$ ($i \geq 1$) are *orthogonal* to some wave function Φ.
    - Φ is chosen to be a mixed combination of the old wave function and (the linear approximation to) the new one.
- Instabilities are usually associated with large steps in parameter space. This scheme **lessens the potential potency** of such instabilities.[27]

---

[27] Should they occur…

Using Linear-Least-Squares Energy Minimisation

- The Linear-Least-Squares method is the default in CASINO, and is what will be used should you set **opt_method** to **emin**.
- Generally:
  - Energy minimisation requires **more configurations** than variance minimisation would.
  - Energy minimisation often **struggles** if the starting wave function is non-at-all optimised (i.e. if you have a blank correlation.data file).
  - Energy minimisation is almost always used as the **last step before DMC**.
  - If you have parameters that can affect the nodal surface, energy minimisation will usually **lower the energy of the trial wave function significantly**.

Outline

Summary

- Optimisation is an important part of a QMC study. It should be performed to the **best of our ability**, and it is almost always **worth the extra effort**.
- Optimisation can often be **tricky**, but on the other hand, we do it only a **few times** per project, and it takes up a **small fraction** of total CPU time.

A Strategy:

- **Fix cut-off lengths** to physically reasonable values.
- **Start with varmin / madmin** - emin struggles from nothing!
- If you have difficulty, **lower the number of parameters** / exclude less important terms (e.g. in a solid, if you have trouble, leave out the f term initially, add it in later and re-optimise).

Summary

- Optimisation is an important part of a QMC study. It should be performed to the **best of our ability**, and it is almost always **worth the extra effort**.
- Optimisation can often be **tricky**, but on the other hand, we do it only a **few times** per project, and it takes up a **small fraction** of total CPU time.

A Strategy:

- **Fix cut-off lengths** to physically reasonable values.
- **Start with varmin / madmin** - emin struggles from nothing!
- If you have difficulty, **lower the number of parameters** / exclude less important terms (e.g. in a solid, if you have trouble, leave out the f term initially, add it in later and re-optimise).
- **Don't be afraid to ignore this advice - we all find exceptions, sometimes regularly!**

Thank you all for listening!

References I

T. Kato, "On the eigenfunctions of many-particle systems in quantum mechanics", Communications on Pure and Applied Mathematics **10**, 151–177 (1957).

R. T. Pack and W. B. Brown, "Cusp conditions for molecular wavefunctions", The Journal of Chemical Physics **45**, 556–559 (1966).

A. Ma, M. D. Towler, N. D. Drummond, and R. J. Needs, "Scheme for adding electron–nucleus cusps to gaussian orbitals", The Journal of Chemical Physics **122**, 224322 (2005).

N. D. Drummond, M. D. Towler, and R. J. Needs, "Jastrow correlation factor for atoms, molecules, and solids", Physical Review B **70**, 235119 (2004).

C. J. Umrigar, K. G. Wilson, and J. W. Wilkins, "Optimized trial wave functions for quantum Monte Carlo calculations", Physical Review Letters **60**, 1719 (1988).

P. R. C. Kent, R. J. Needs, and G. Rajagopal, "Monte Carlo energy and variance-minimization techniques for optimizing many-body wave functions", Physical Review B **59**, 12344 (1999).

N. D. Drummond and R. J. Needs, "Variance-minimization scheme for optimizing Jastrow factors", Physical Review B **72**, 085124 (2005).

References II

📄 M. P. Nightingale and V. Melik-Alaverdian, "Optimization of ground-and excited-state wave functions and van der Waals clusters", Physical Review Letters **87**, 043401 (2001).

📄 J. Toulouse and C. J. Umrigar, "Optimization of quantum Monte Carlo wave functions by energy minimization", The Journal of Chemical Physics **126**, 084102 (2007).