

Optimisation of Many-Electron Wave Functions



Neil D. Drummond

Department of Physics, Lancaster University

QMC and the CASINO program, TTI, Vallico Sotto, Italy

Tuesday 6th August, 2013

Introduction

- First stage of any VMC or DMC calculation: *generate a trial wave function*.
 - Basic approximation comes from a **DFT** or **quantum chemistry** code (e.g., CASTEP or Gaussian) or is known on theoretical grounds (e.g. plane-wave orbitals for the fluid phase of an electron gas).
 - We need better approximations in QMC, so we add **variational freedom** to the basic approximation.
- Parameters in the explicitly correlated wave function are optimised within VMC.
 - *Most difficult technical aspect of QMC.*
 - But. . . **Good wave functions are vital for successful QMC work.**
- In first half of this talk I will discuss the required properties and usual form of QMC trial wave functions.
- In the second half I will give an overview of optimisation methods.
- **Throughout:** Hartree atomic units. $\hbar = |e| = m_e = 4\pi\epsilon_0 = 1$.

Wave-Function Antisymmetry (I)

- Consider a many-electron wave function $\Psi_S(\mathbf{X})$, where $\mathbf{X} = (\{\mathbf{r}_1, \sigma_1\}, \dots, \{\mathbf{r}_N, \sigma_N\})$, \mathbf{r}_i is electron position and σ_i ($=\uparrow$ or \downarrow) is spin.
- Assume that Ψ_S is an eigenfunction of total spin operator $\hat{S}_z = \sum_i \hat{s}_{zi}$.
 - Let the eigenvalue of \hat{S}_z be $(N_\uparrow - N_\downarrow)/2$, where $N_\uparrow + N_\downarrow = N$.
 - Electron spins are *collinear*: N_\uparrow up-spin electrons and N_\downarrow down-spin electrons.
 - Restricts form of Ψ_S , but being an eigenfunction of \hat{S}_z is a property of the exact ground-state wave function if the Hamiltonian is spin-independent.
- Electrons are **fermions**, so Ψ_S is **antisymmetric** under exchange of electrons.
- Expectation value of a **spin-independent** operator \hat{A} with respect to Ψ_S :

$$\langle \hat{A} \rangle_{\Psi_S} = \frac{\langle \Psi_S | \hat{A} | \Psi_S \rangle}{\langle \Psi_S | \Psi_S \rangle} = \frac{\sum_{\{\sigma\}} \int \Psi_S^*(\mathbf{X}) \hat{A}(\mathbf{R}) \Psi_S(\mathbf{X}) d\mathbf{R}}{\sum_{\{\sigma\}} \int |\Psi_S(\mathbf{X})|^2 d\mathbf{R}},$$

where $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$ and the sums run over all spin configurations such that the numbers of spin-up and spin-down electrons are N_\uparrow and N_\downarrow , respectively.

Wave-Function Antisymmetry (II)

- \mathbf{X} may be replaced by $\mathbf{X}' = \left(\{\mathbf{r}_{i_1}, \uparrow\}, \dots, \{\mathbf{r}_{i_{N_\uparrow}}, \uparrow\}, \{\mathbf{r}_{i_{N_\uparrow}+1}, \downarrow\}, \dots, \{\mathbf{r}_{i_N}, \downarrow\} \right)$ without altering $\langle \hat{A} \rangle_{\Psi_S}$ (due to the antisymmetry of Ψ_S).
- Relabel dummy integration variables: $\mathbf{r}_{i_1} \rightarrow \mathbf{r}_1$, etc.
 - Integrals are clearly the same for each spin configuration $\{\sigma\}$ such that the number of spin-up electrons is N_\uparrow ; hence the spin sums cancel.
- So the expectation of \hat{A} may be evaluated as

$$\langle \hat{A} \rangle_{\Psi_S} = \frac{\int \Psi^*(\mathbf{R}) \hat{A}(\mathbf{R}) \Psi(\mathbf{R}) d\mathbf{R}}{\int |\Psi(\mathbf{R})|^2 d\mathbf{R}},$$

where the **spatial wave function**

$$\Psi(\mathbf{R}) \equiv \Psi_S(\{\mathbf{r}_1, \uparrow\}, \dots, \{\mathbf{r}_{N_\uparrow}, \uparrow\}, \{\mathbf{r}_{N_\uparrow+1}, \downarrow\}, \dots, \{\mathbf{r}_N, \downarrow\})$$

is **only antisymmetric with respect to exchanges of positions of same-spin electrons.**

Wave-Function Antisymmetry (III)

- In effect, electrons of different spin are treated as **distinguishable** fermions.
- If we know $\Psi(\mathbf{R})$ then we can recover $\Psi_S(\mathbf{X})$ and *vice versa*.
 - Much easier just to work with $\Psi(\mathbf{R})$.
 - Henceforth, Ψ will just be called “the (trial) wave function”.
- For example, variational principle is

$$\frac{\langle \Psi_S | \hat{H} | \Psi_S \rangle}{\langle \Psi_S | \Psi_S \rangle} = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \geq E_0,$$

where \hat{H} is the Hamiltonian and E_0 is the ground-state energy.

- Apply variational principle to optimise $\Psi(\mathbf{R})$.
- **Noncollinear** spins can be used in VMC in CASINO (at present just for spiral spin-density waves in the homogeneous electron gas).
 - If you want to study spin-orbit coupling, someone will need to implement this. . .

Real or Complex Wave Functions

- The trial wave function can always be chosen to be real for systems with **time-reversal symmetry** (i.e., a real Hamiltonian with appropriate boundary conditions).¹
- Either real or complex wave functions can be used in QMC calculations.
 - Former are **faster**, but latter are necessary if “**twisted**” boundary conditions are used or a **magnetic field** is present.
 - In CASINO, use the **complex_wf** keyword to choose between real (F) or complex (T) wave functions.

¹ Proof: $\hat{H}\Psi = E_0\Psi$ if and only if $\hat{H}\Psi^* = E_0\Psi^*$, if Hamiltonian \hat{H} is real. So either Ψ and Ψ^* differ only by a phase factor or Ψ and Ψ^* are degenerate. In either case $\Psi + \Psi^*$ is a real eigenfunction of \hat{H} with eigenvalue E_0 , unless Ψ is purely imaginary, in which case $i\Psi$ is a real eigenfunction.

Slater–Jastrow Wave Functions

- Most QMC calculations use *Slater–Jastrow* trial wave functions:

$$\Psi(\mathbf{R}) = \exp[J(\mathbf{R})] \sum_n c_n D_n^\uparrow(\mathbf{R}) D_n^\downarrow(\mathbf{R}),$$

where $\{D_n^\uparrow\}$ and $\{D_n^\downarrow\}$ are **Slater determinants** for spin-up and down electrons, $\exp(J)$ is a **Jastrow factor** and the $\{c_n\}$ are determinant expansion coefficients.

- The $\sum_n c_n D_n^\uparrow(\mathbf{R}) D_n^\downarrow(\mathbf{R})$ factor is called the *Slater wave function*.
- CASINO can work with more exotic wave-function forms such as backflow wave functions or pairing (geminal) wave functions.
 - See Pablo’s lecture later in the week.

Slater Wave Function (I)

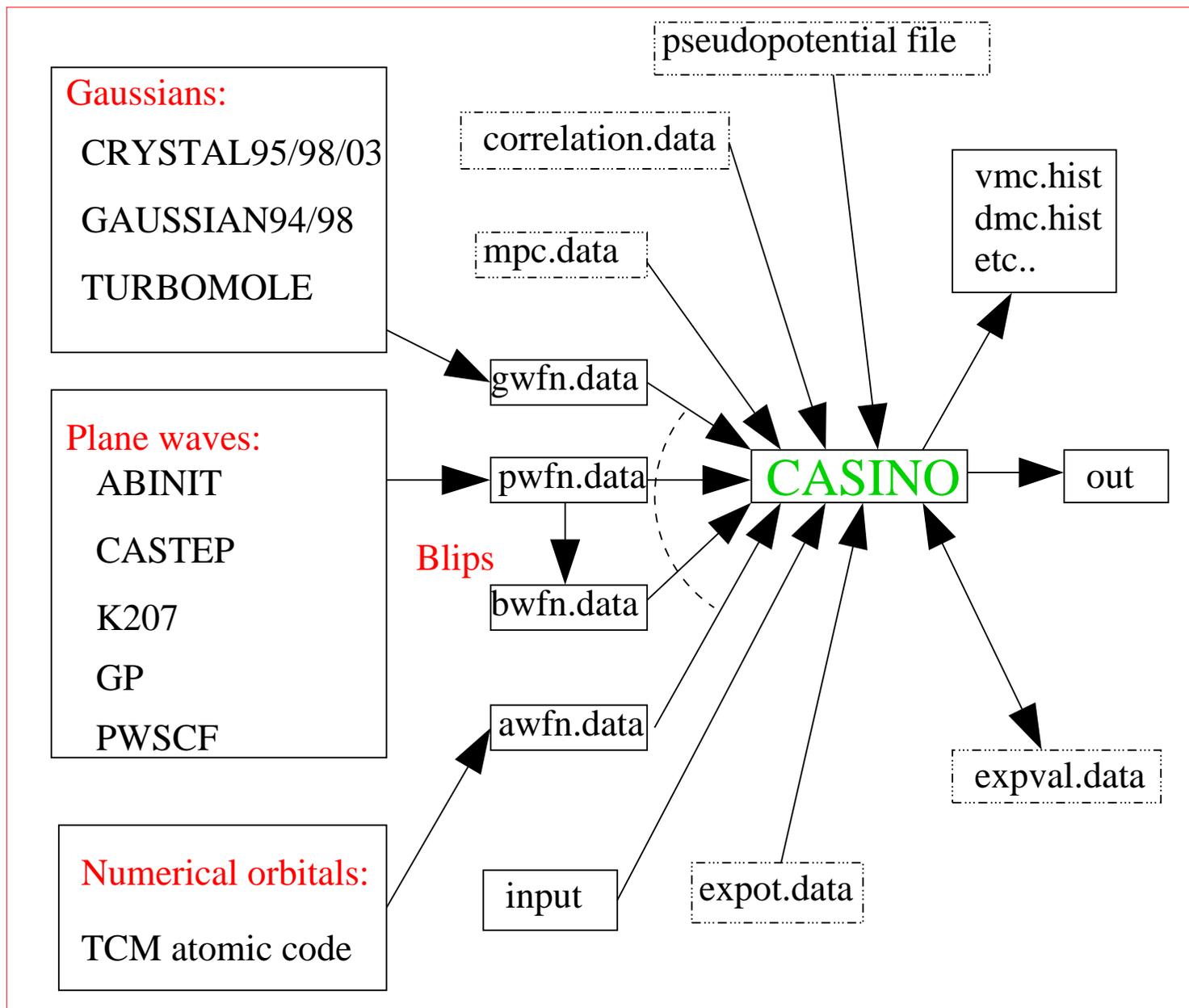
- Each Slater determinant is of the form

$$D^\uparrow(\mathbf{R}) = \begin{vmatrix} \psi_1^\uparrow(\mathbf{r}_1) & \cdots & \psi_{N_\uparrow}^\uparrow(\mathbf{r}_1) \\ \vdots & & \vdots \\ \psi_1^\uparrow(\mathbf{r}_{N_\uparrow}) & \cdots & \psi_{N_\uparrow}^\uparrow(\mathbf{r}_{N_\uparrow}) \end{vmatrix},$$

and similarly for spin-down determinants.

- The $\{\psi_i^\sigma\}$ are single-particle spatial orbitals, which are usually generated in either DFT or Hartree–Fock calculations.
 - It is sometimes possible to optimise orbitals within QMC.
 - For periodic systems with time-reversal symmetry, even if the individual orbitals are complex Bloch orbitals, the Slater determinant is real. See tomorrow’s talk.
 - Orbitals may be represented in different ways (plane waves, Gaussians, “blips” (B-splines), Slater functions, . . .).

Slater Wave Function (II)



Slater Wave Function (III)

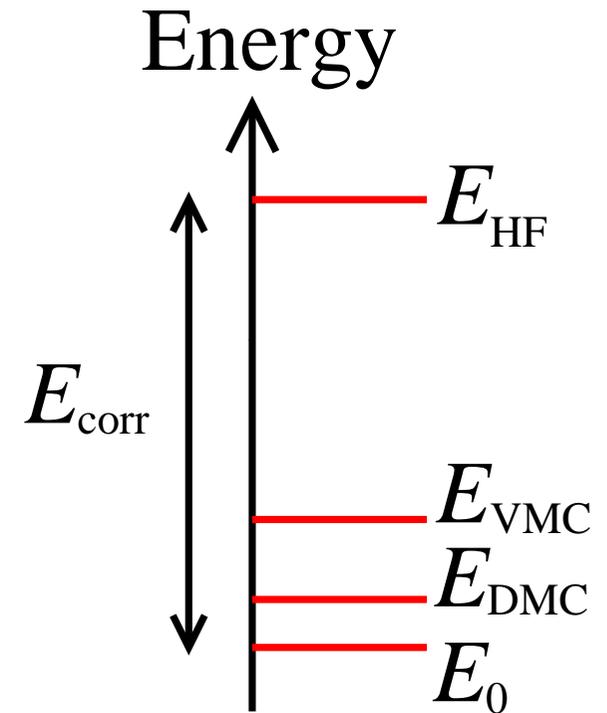
- Multideterminant expansion coefficients $\{c_n\}$ can be taken from post-Hartree–Fock calculations, or can be optimised within QMC.
- Actually, most QMC calculations use only one term in the multideterminant expansion.
 - Having large numbers of determinants is most important for atoms or small molecules where there is a high degree of “static correlation” (multireference character).
 - The number of determinants required to retrieve a given fraction of the correlation energy increases **exponentially** with system size: *we do not normally use multideterminant wave functions for periodic systems.*

Correlation Energy

- Let E_{HF} , E_{VMC} and E_0 be the Hartree–Fock energy, the energy expectation with a Slater–Jastrow wave function and the ground-state energy, respectively.
- **Correlation energy:** $E_{\text{corr}} \equiv E_0 - E_{\text{HF}}$.
- Fraction of correlation energy retrieved by a Slater–Jastrow wave function:

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

- If assessing the quality of a Jastrow factor, one often uses the fixed-node DMC energy E_{DMC} instead of E_0 .
- Typically, Jastrow factor retrieves 90–99% of correlation energy with respect to DMC for atoms and 80–95% for molecules and solids.



Jastrow Factor

- The Jastrow factor is an explicit function of the interparticle distances.
 - Hence the Jastrow factor can be **compactly** parametrised, *unlike, e.g., multideterminant calculations.*
 - *The number of parameters in the Jastrow factor required to retrieve a given fraction of correlation energy is essentially independent of the system size.*
- The Slater wave function has the required **antisymmetry**, so J must be **symmetric** under the exchange of same-spin electrons.
- J contains free parameters, to be determined by an optimisation method.
- The Jastrow factor does not alter the nodal surface of the wave function, because $\exp(J) > 0$.

Need for a Good Jastrow Factor

- The Jastrow factor can satisfy the electron–electron cusp conditions.²
- *In principle, the fixed-node DMC energy is independent of the Jastrow factor.*
However:
 - The Jastrow factor reduces DMC time-step and population-control biases.
 - Better wave functions give smaller statistical error bars in both VMC and DMC.
 - Good wave functions reduce errors resulting from the *locality approximation* in DMC calculations with nonlocal pseudopotentials.
 - *Extrapolated estimation* of quantities that don't commute with the Hamiltonian requires a good wave function.
 - If parameters that affect the nodal surface are optimised at the same time as the Jastrow factor then the resulting wave function and nodal surface will be better.
 - Fluctuations in the DMC configuration population and the associated difficulties are greatly reduced.
- **So. . . make sure your wave function is highly optimised!**

² T. Kato, Commun. Pure Appl. Math. **10**, 151 (1957); R. T. Pack & W. B. Brown, J. Chem. Phys. **45**, 556 (1966).

Antiparallel-Spin Electron–Electron Kato Cusp Condition (I)

- Reminder: *local energy* is $E_L = \Psi^{-1} \hat{H} \Psi$.
- Suppose two antiparallel-spin electrons i and j approach one another, with separation \mathbf{r}_{ij} . The reduced mass of the pair is $1/2$ a.u.
- Fix other coordinates and write

$$\Psi(\mathbf{r}_{ij}) = \sum_{l=0}^{\infty} \sum_{m=-l}^l \Psi_{lm}(r_{ij}) r_{ij}^l Y_{lm}(\theta, \phi),$$

where $Y_{lm}(\theta, \phi)$ is the (l, m) th spherical harmonic.

- Also,

$$\Psi_{lm}(r_{ij}) = \Psi_{lm}(0) + r_{ij} \left(\frac{\partial \Psi_{lm}}{\partial r_{ij}} \right)_{r_{ij}=0} + \mathcal{O}(r_{ij}^2).$$

Antiparallel-Spin Electron–Electron Kato Cusp Condition (II)

- Contribution to the local energy from coalescing pair is

$$\begin{aligned} E_L^{ij} &= -\frac{2\nabla_{ij}^2 \Psi}{2\Psi} + \frac{1}{r_{ij}} = -\frac{\nabla_{ij}^2 \Psi_{00}(r_{ij})}{\Psi_{00}(r_{ij})} + \mathcal{O}(r_{ij}^0) + \frac{1}{r_{ij}} \\ &= \frac{-2}{r_{ij}\Psi_{00}(0)} \left(\frac{\partial \Psi_{00}}{\partial r_{ij}} \right)_{r_{ij}=0} + \mathcal{O}(r_{ij}^0) + \frac{1}{r_{ij}}. \end{aligned}$$

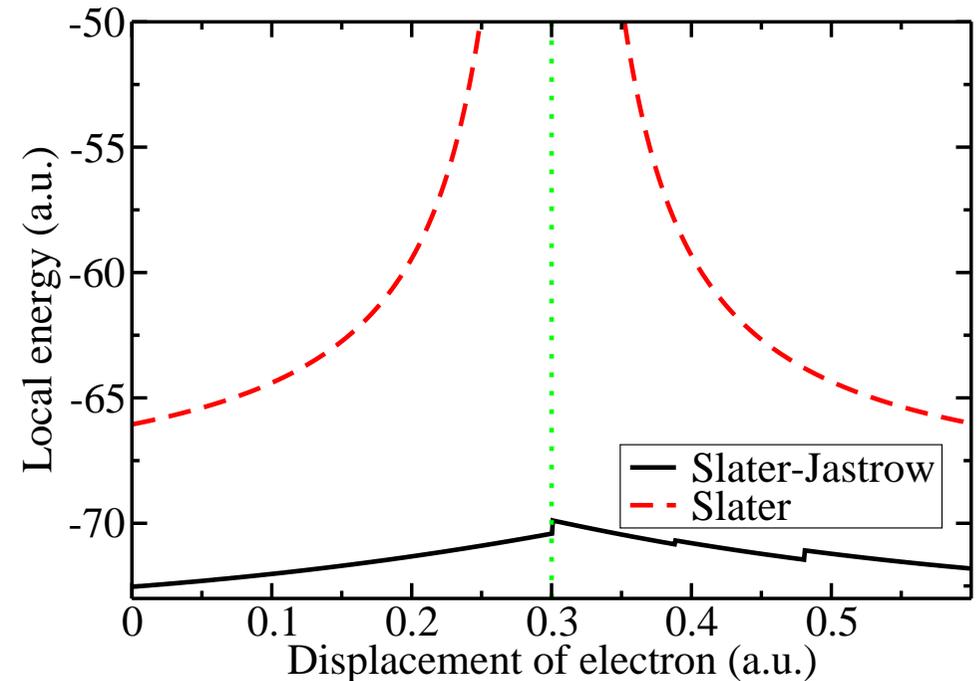
- *The local energy should not diverge.* (In fact the local energy is constant if Ψ is an eigenfunction of \hat{H} .)
 - Demand that the local energy does not diverge at the coalescence point.
 - So $\mathcal{O}(r_{ij}^{-1})$ terms must vanish.
 - This leads to the antiparallel-spin Kato cusp condition:

$$\left(\frac{\partial \Psi_{00}}{\partial r_{ij}} \right)_{r_{ij}=0} = \frac{1}{2} \Psi_{00}(0).$$

Antiparallel-Spin Electron–Electron Kato Cusp Condition (III)

- Hence $(\partial \log(\Psi_{00})/\partial r_{ij})_{r_{ij}=0} = 1/2$.
- Suppose the Jastrow factor is isotropic at short range.
- So, if the Slater wave function is smooth at coalescence points, then

$$\left(\frac{\partial J}{\partial r_{ij}}\right)_{r_{ij}=0} = \frac{1}{2}.$$



Local energy as one electron is moved straight through another of opposite spin in H_2O . Jastrow factor imposes Kato cusp condition.

Parallel-Spin Electron–Electron Kato Cusp Condition (I)

- Consider two parallel-spin electrons i and j , separation r_{ij} , approaching each other.
- Wave function is antisymmetric. So, omitting other electron coordinates,

$$\Psi(\mathbf{r}_{ij}) = \sum_{l \text{ odd}} \sum_{m=-l}^l \Psi_{lm}(r_{ij}) r_{ij}^l Y_{lm}(\theta, \phi).$$

- Contribution to local energy from coalescing pair is

$$\begin{aligned} E_L^{ij} &= -\frac{2\nabla_{ij}^2 \Psi}{2\Psi} + \frac{1}{r_{ij}} = -\frac{\nabla_{ij}^2 \left[r_{ij} \sum_{m=-1}^1 \Psi_{1m}(r_{ij}) Y_{1m}(\theta, \phi) \right]}{r_{ij} \sum_{m=-1}^1 \Psi_{1m}(r_{ij}) Y_{1m}(\theta, \phi)} + \mathcal{O}(r_{ij}^0) + \frac{1}{r_{ij}} \\ &= -\frac{4 \sum_{m=-1}^1 [\partial \Psi_{1m} / \partial r_{ij}]_{r_{ij}=0} Y_{1m}(\theta, \phi)}{r_{ij} \sum_{m=-1}^1 \Psi_{1m}(0) Y_{1m}(\theta, \phi)} + \mathcal{O}(r_{ij}^0) + \frac{1}{r_{ij}} \end{aligned}$$

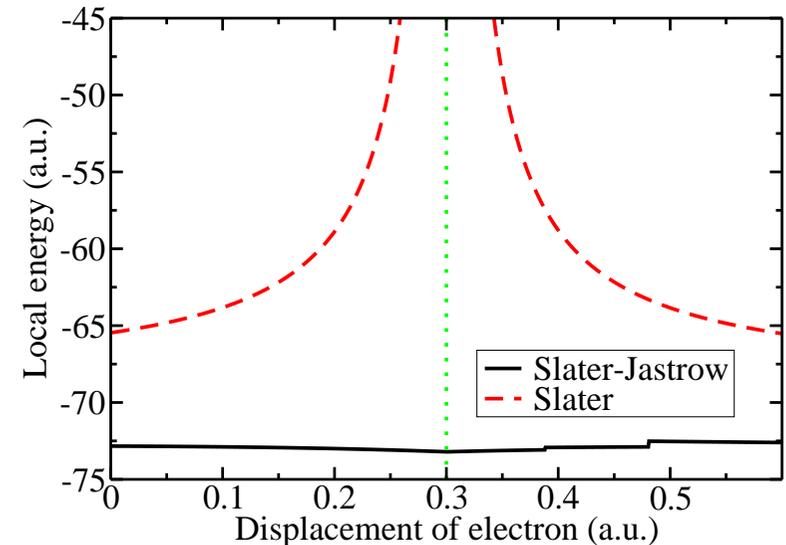
Parallel-Spin Electron–Electron Kato Cusp Condition (II)

- Demand that the local energy does not diverge at the coalescence point, i.e. that $\mathcal{O}(r_{ij}^{-1})$ term is zero $\forall \theta, \phi$. Gives the parallel-spin Kato cusp condition:

$$\left(\frac{\partial \Psi_{1m}}{\partial r_{ij}} \right)_{r_{ij}=0} = \frac{1}{4} \Psi_{1m}(0).$$

- So $(\partial \log(\Psi_{1m}) / \partial r_{ij})_{r_{ij}=0} = 1/4$.
- But $\Psi_{1m}(r_{ij}) = \exp[J(r_{ij})] S_{1m}(r_{ij})$, where S is the Slater wave function. If the Slater wave fn. is smooth at coalescence points then

$$\left(\frac{\partial J}{\partial r_{ij}} \right)_{r_{ij}=0} = \frac{1}{4},$$



Local energy as one electron is moved straight through another of the same spin in H_2O . Jastrow factor imposes Kato cusp condition.

Electron–Nucleus Cusp Condition (I)

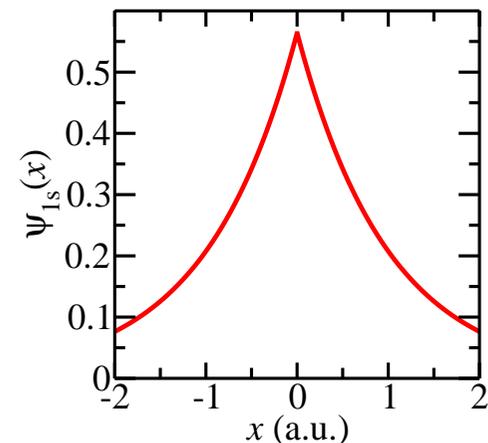
- **Kato**: spherical average of $\Psi(\mathbf{r}_i)$ about a nucleus of atomic number Z must obey

$$\left(\frac{\partial \bar{\Psi}}{\partial r_i}\right)_{r_i=0} = -Z \bar{\Psi}_{r_i=0}.$$

- Proof: similar to antiparallel-spin cusp condition.
- If the Slater wave function is continuously differentiable at the nucleus then the Jastrow factor must satisfy

$$(\partial J / \partial r_i)_{r_i=0} = -Z.$$

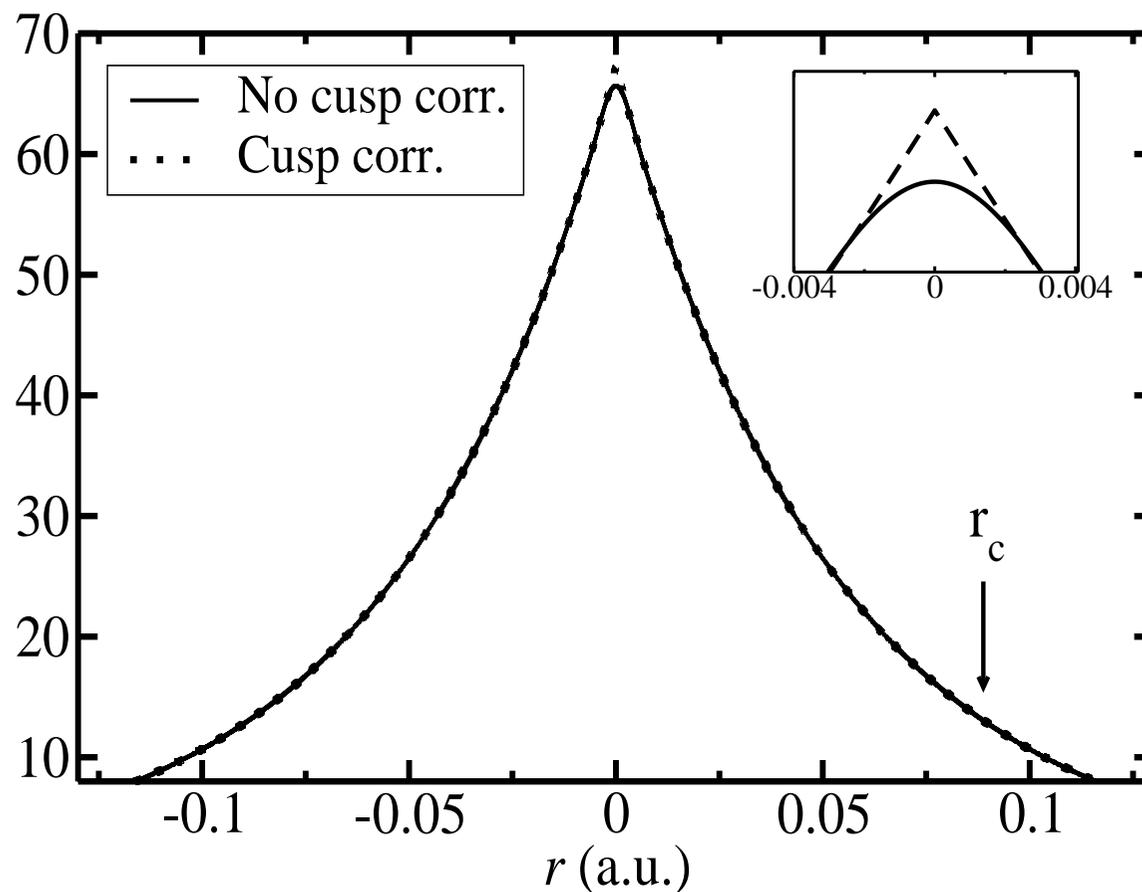
- If the Slater wave function satisfies the e–n cusp condition, or a nondivergent pseudopotential is used, J should be cusplless at nuclei, i.e. $Z = 0$ in above equation.
- The Slater wave function satisfies the electron–nucleus Kato cusp condition if and only if each orbital satisfies the cusp condition.



Hydrogenic 1s orbital
 $\psi = (1/\sqrt{\pi}) \exp(-r)$
clearly satisfies Kato cusp condition.

Electron–Nucleus Cusp Condition (II)

- It is impossible to describe cusps with finite numbers of Gaussians or plane waves. In CASINO appropriate cusps can be added to Gaussian³ or plane-waves/blip⁴ orbitals.



³ A. Ma *et al.*, J. Chem. Phys. **122**, 224322 (2005). Activate with the **cusp_correction** keyword. On by default.

⁴ Activate in CASINO with the **use_gpcc** keyword.

CASINO's Jastrow Factor (I)

- CASINO's Jastrow factor⁵ consists of 6 types of term:
 1. Isotropic function of electron–electron separations $u(r_{ij})$.
 2. Isotropic functions of electron–ion separations $\chi(r_{iI})$.
 3. Isotropic functions of electron–electron and electron–ion separations $f(r_{iI}, r_{jI}, r_{ij})$.
 4. Plane-wave expansions in electron–electron separations $p(\mathbf{r}_{ij})$.
 5. Plane-wave expansions in electron positions $q(\mathbf{r}_i)$.
 6. Isotropic three-electron term H_{3B} .
- The Jastrow factor can be written as:

$$\begin{aligned} J(\mathbf{R}) = & \sum_{i=1}^{N-1} \sum_{j=i+1}^N u(r_{ij}) + \sum_{I=1}^{N_{\text{ions}}} \sum_{i=1}^N \chi_I(r_{iI}) + \sum_{I=1}^{N_{\text{ions}}} \sum_{i=1}^{N-1} \sum_{j=i+1}^N f_I(r_{iI}, r_{jI}, r_{ij}) \\ & + \sum_{i=1}^{N-1} \sum_{j=i+1}^N p(\mathbf{r}_{ij}) + \sum_{i=1}^N q(\mathbf{r}_i) + H_{3B}(\mathbf{R}). \end{aligned}$$

⁵ N. D. Drummond *et al.*, Phys. Rev. B **70**, 235119 (2004).

CASINO's Jastrow Factor (II)

where

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

$$\chi_I(r_{iI}) = (r_{iI} - L_{\chi I})^C \times \Theta(L_{\chi I} - r_{iI}) \times \sum_{m=0}^{N_{\chi}} \beta_{mI} r_{iI}^m;$$

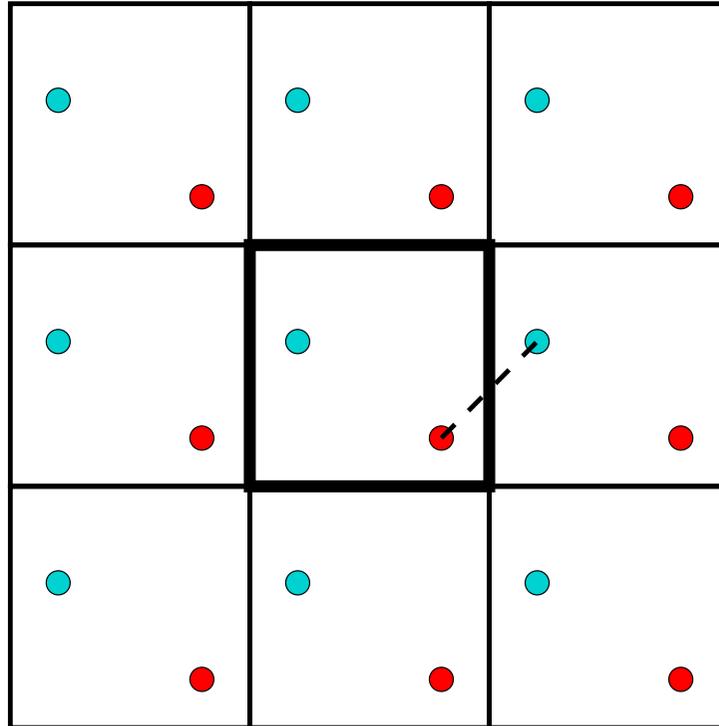
$$\begin{aligned} f_I(r_{iI}, r_{jI}, r_{ij}) &= (r_{iI} - L_{fI})^C (r_{jI} - L_{fI})^C \Theta(L_{fI} - r_{iI}) \Theta(L_{fI} - r_{jI}) \\ &\quad \times \sum_{l=0}^{N_{fI}^{eN}} \sum_{m=0}^{N_{fI}^{eN}} \sum_{n=0}^{N_{fI}^{ee}} \gamma_{lmnI} r_{iI}^l r_{jI}^m r_{ij}^n; \end{aligned}$$

$$p(\mathbf{r}_{ij}) = \sum_A a_A \sum_{\mathbf{G}_A^+} \cos(\mathbf{G}_A \cdot \mathbf{r}_{ij});$$

$$q(\mathbf{r}_i) = \sum_B b_B \sum_{\mathbf{G}_B^+} \cos(\mathbf{G}_B \cdot \mathbf{r}_i).$$

CASINO's Jastrow Factor (III)

- In periodic systems, r_{ij} and r_{iI} are evaluated under the **minimum-image** convention.



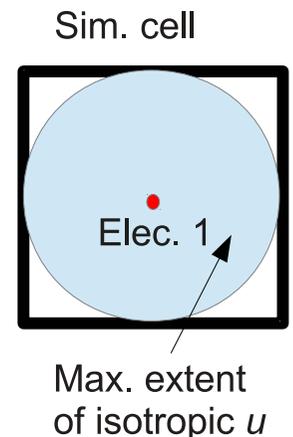
Minimum-image distance between red and blue particles in simulation cell. Used in evaluation of Jastrow factor.

CASINO's Jastrow Factor (IV)

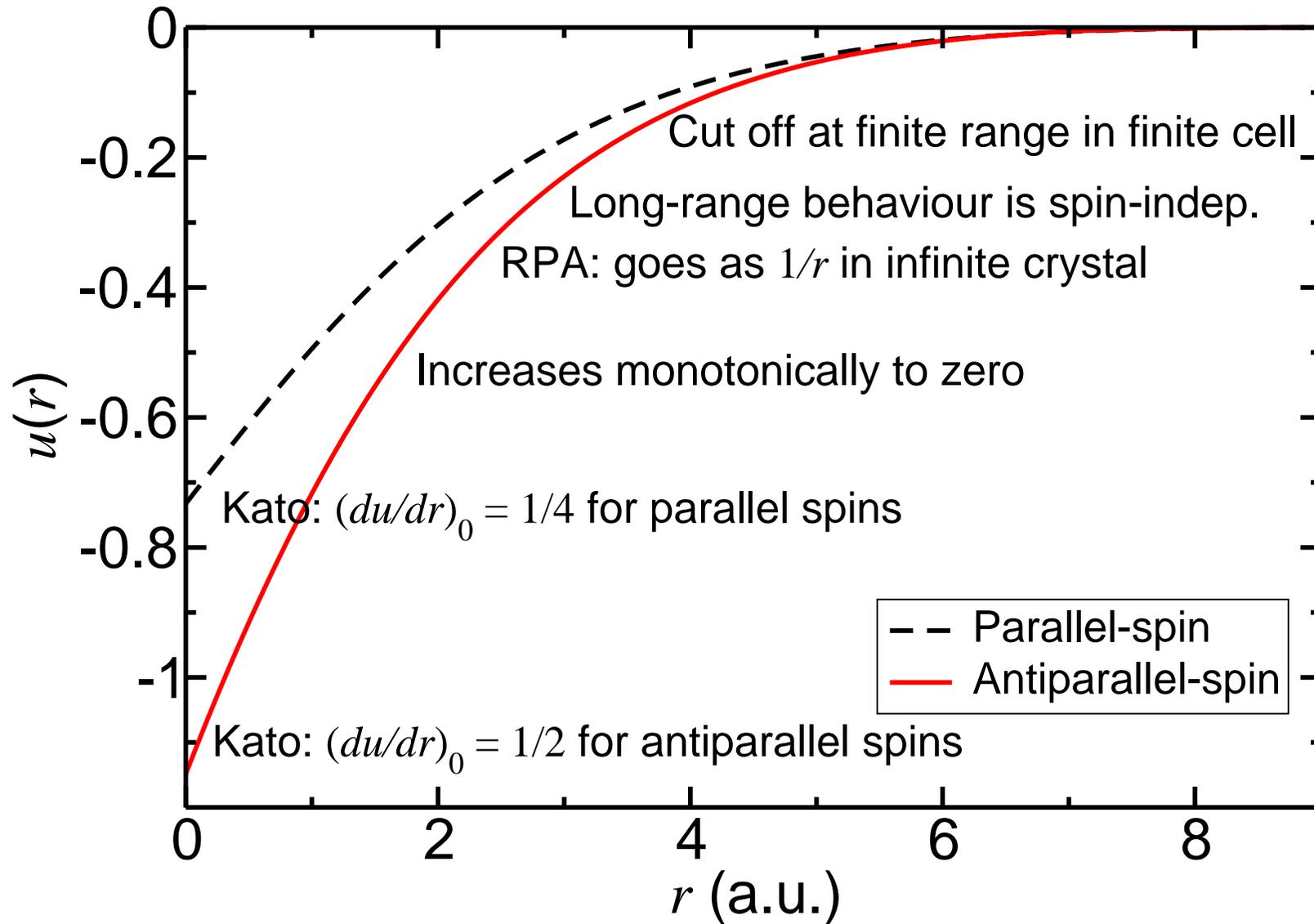
- u , χ and f are **polynomials**; the polynomial coefficients are optimisable parameters (or are determined by the cusp conditions).
 - Polynomials are **quick** to evaluate; polynomials are **flexible**.
 - The polynomials go smoothly to zero at **cutoff lengths** (L_u , L_χ and L_f), which are themselves optimisable.
 - In a periodic system, the cutoff lengths cannot exceed the radius of the largest sphere that can be inscribed in the Wigner–Seitz cell of the simulation cell, otherwise the derivatives of the Jastrow factor would be discontinuous.
 - Usually choose to have a continuous second derivative of the wave function at cutoff lengths (**truncation order** $C = 3$), to aid optimisation.
- p and q are plane-wave expansions; expansion coefficients are optimisable parameters.
- **Spin-dependence**: have different parameters for different combinations of spins.
 - Parallel and antiparallel spins have different cusp conditions; usually have different u terms for parallel- and antiparallel-spin electrons.
 - For spin-polarised systems we have different χ terms for up- & down-spin electrons.

CASINO's Jastrow Factor (V)

- The electron–electron term u describes both short-range (Kato cusp conditions) and long-range **two-body correlations**.
 - u is primarily responsible for setting up the correlation contribution to the “**exchange–correlation hole**”.
 - However, in an inhomogeneous system, u can distort the charge density.
- The electron–ion term χ “**undoes the effect of u on the Hartree–Fock charge density**”.
 - It is generally preferable to have the electron–nucleus cusp in the orbitals.
- f is like a **density-dependent** u term. Don't usually need spin-dependence.
- p describes similar correlations to u , but adds variational freedom in “corners” of sim. cell and describes anisotropic correlations.
 - p terms are necessary to obtain correct two-body correlations at large separations. Spin-dependence is unimportant.
- q and H_{3B} are rarely used for real systems.



Typical Plot of Two-Body Jastrow Term for Extended System



Paramagnetic 3D electron gas at density parameter $r_s = 2$.

Cutoff Lengths and Scaling

- Usually find that the optimal cutoff length L_u is proportional to the system size and the optimal cutoff lengths of L_χ and L_f are independent of system size.
 - The u term describes both long- and short-range correlations.
 - χ_I and f_I describe inhomogeneous correlations that are spherically symmetric about atom I . Cannot describe inhomogeneous correlations far from I .
- Hence the cost of updating the u term after each electron move scales as $\mathcal{O}(N)$, while cost of updating the χ and f terms is independent of system size.
- **Optimising cutoffs is difficult** (see later). Choosing good initial values helps.
 - Should aim to make L_u equal to an appropriate linear size for a system, e.g., the “diameter” of a molecule or the radius of the largest sphere that can be inscribed in the Wigner–Seitz cell of the simulation cell in a periodic system.
 - Should choose L_χ and L_f to be between one and two bond lengths.
 - Cutoff lengths should only be optimised with variance minimisation, and should be fixed before the final optimisation (by energy minimisation or variance minimisation).

Using CASINO's Jastrow Factor

- *Copy a correlation.data file from a relevant example.* Most changes needed should (hopefully) be fairly self-explanatory. May need to change the number of sets of χ or f functions (one set for each chemically distinct set of ions): just copy and paste, and alter the ion lists in each set as appropriate. Not all terms need to be present: e.g., one can use just u and χ .
- Last block of parameters in each set are the linear parameters.
- Usual starting point is a *blank* Jastrow factor: delete all the lines holding linear parameters in all sets in the file. E.g.,

```
Parameter values ; Optimizable (0=NO; 1=YES)  
END SET 1
```

If CASINO reads a blank Jastrow factor then it will use appropriate default behaviour.⁶

- To construct a p term, use the [MAKE_P_STARS](#) utility.⁷

⁶ For molecules and 1D or 2D inhomogeneous systems, no Jastrow factor is used if the Jastrow is blank; otherwise a simple Jastrow factor that satisfies the Kato cusp conditions is used by default.

⁷ Type *make_p_stars* and answer the questions.

Wave-Function Optimisation

I have nothing to offer but blood, toil, tears and sweat.

Sir Winston Churchill



Energy Minimisation: Variational Principle

- Let Ψ be a many-electron wave function, \hat{H} the Hamiltonian and E_i the i th eigenenergy.
- Write $\Psi = \sum_i c_i \phi_i$, where ϕ_i is the i th eigenfunction of \hat{H} . NB, $\langle \phi_i | \phi_j \rangle = \delta_{ij}$.

- Then

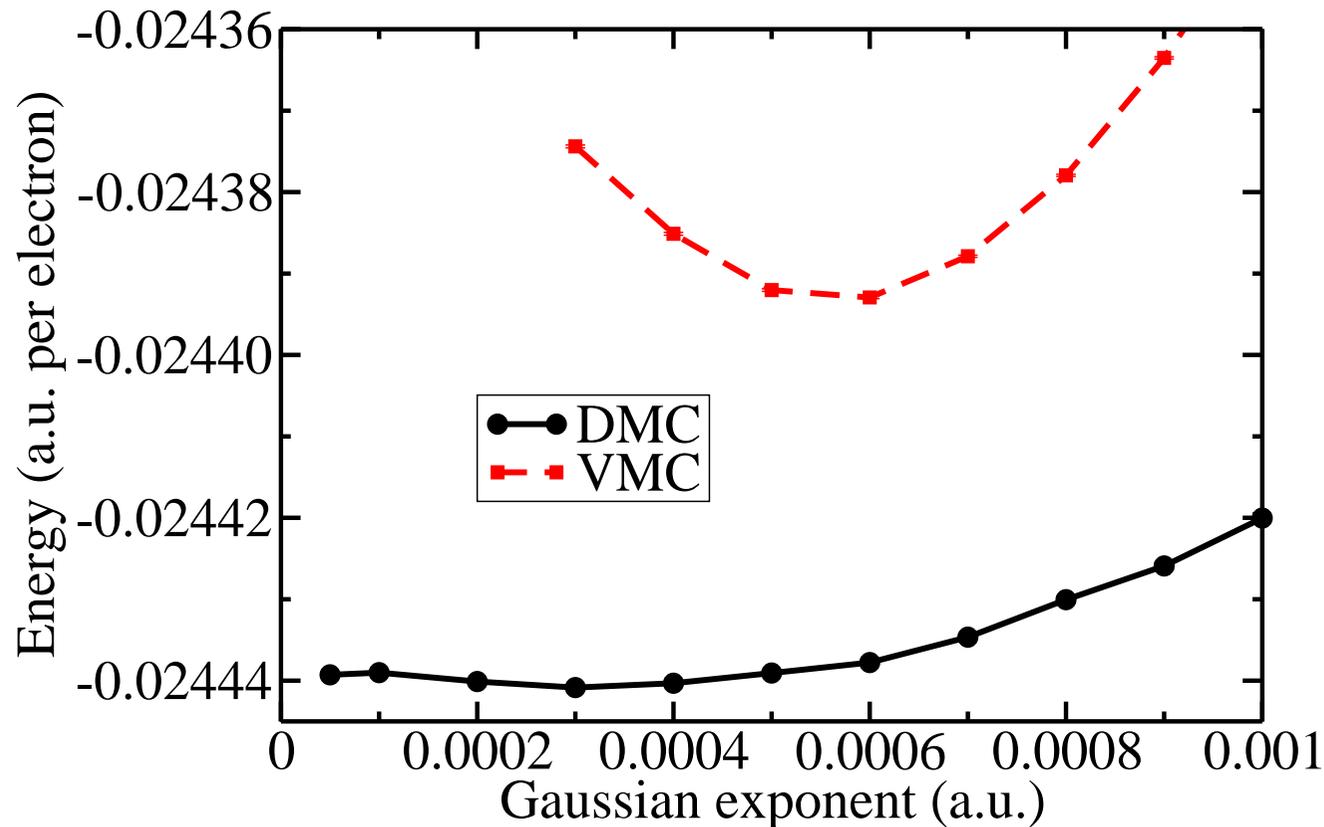
$$\frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\sum_i |c_i|^2 E_i}{\sum_i |c_i|^2} \geq \frac{\sum_i |c_i|^2 E_0}{\sum_i |c_i|^2} = E_0.$$

This is the (hopefully very familiar) **variational principle** for the energy. So . . .

- *Minimise the energy expectation value to optimise the wave function.*

Naïve Energy Minimisation (I)

- For any given Ψ_α , we can estimate $\langle \Psi_\alpha | \hat{H} | \Psi_\alpha \rangle / \langle \Psi_\alpha | \Psi_\alpha \rangle$ using VMC.
- *Why not minimise the VMC energy with respect to wave-function parameters α ?*



VMC and DMC energy of a 2D hexagonal Wigner crystal at $r_s = 40$ a.u. against the Gaussian exponent in the orbitals.

Naïve Energy Minimisation (II)

- Can successfully minimise one or two parameters at a time by performing series of VMC simulations with different parameter values.
- However, this brute-force approach is infeasible for multiple parameters.
 - VMC data come with statistical **noise**.
 - Taking derivatives of the VMC energy is problematic: the noise will swamp the energy difference in the limit of a small change in the parameters.
 - The noise invalidates most standard minimisation techniques.
 - Robust optimisation methods that don't require derivatives and can cope with noise typically need hundreds of thousands of function evaluations. But each VMC calculation can be very expensive.
- *Clearly need a different approach. . .*

Correlated Sampling

- Let Ψ_A and Ψ_B be two different wave functions and \hat{O} be a quantum-mechanical operator. Then

$$\begin{aligned}\langle \hat{O} \rangle_{\Psi_A} &= \frac{\int \Psi_A^*(\mathbf{R}) \hat{O} \Psi_A(\mathbf{R}) d\mathbf{R}}{\int |\Psi_A(\mathbf{R})|^2 d\mathbf{R}} = \frac{\int |\Psi_A(\mathbf{R})|^2 \frac{\hat{O} \Psi_A(\mathbf{R})}{\Psi_A(\mathbf{R})} d\mathbf{R}}{\int |\Psi_A(\mathbf{R})|^2 d\mathbf{R}} \\ &= \frac{\int |\Psi_B(\mathbf{R})|^2 \frac{|\Psi_A(\mathbf{R})|^2}{|\Psi_B(\mathbf{R})|^2} \frac{\hat{O} \Psi_A(\mathbf{R})}{\Psi_A(\mathbf{R})} d\mathbf{R}}{\int |\Psi_B(\mathbf{R})|^2 \frac{|\Psi_A(\mathbf{R})|^2}{|\Psi_B(\mathbf{R})|^2} d\mathbf{R}}.\end{aligned}$$

- So $\langle \hat{O} \rangle_{\Psi_A}$ can be evaluated as the average of $O_L^A = \Psi_A^{-1} \hat{O} \Psi_A$ over configurations distributed according to $|\Psi_B(\mathbf{R})|^2$, weighted by $w_B^A = |\Psi_A(\mathbf{R})|^2 / |\Psi_B(\mathbf{R})|^2$.
 - Hence we can calculate expectation values with respect to wave function Ψ_A using the configuration set generated by VMC using Ψ_B .
 - NB, estimate of $\langle \hat{O} \rangle_{\Psi_A}$ is a **smooth** function of the parameters in Ψ_A for a given configuration set.
 - Generally, can use this approach to obtain accurate energy differences, etc.

Reweighted Energy Minimisation

- Reweighted energy minimisation:
 - Suppose we have an initial parameter set α_0 . Generate a set of configurations $\{\mathbf{R}\}$ distributed according to $|\Psi_{\alpha_0}|^2$ using VMC.
 - Correlated sampling: for any parameter set α , the energy expectation value $\langle \hat{H} \rangle_{\Psi_\alpha}$ may be estimated by the weighted mean of the *local energy* $\Psi_\alpha^{-1} \hat{H} \Psi_\alpha$ over the configurations, where the configuration weights are $|\Psi_\alpha|^2 / |\Psi_{\alpha_0}|^2$.
 - Vary α to minimise the estimate of $\langle \hat{H} \rangle_{\Psi_\alpha}$.
- In principle this might work if one started very close to the optimal wave function.
- However, the method is unstable and unusable in practice.
 - Finite-sampling error in the reweighted energy rapidly gets very large when Ψ_α and Ψ_{α_0} differ substantially, so that the objective function is only a good estimate of the energy expectation value very close to the initial parameter set.
 - *Objective function is badly behaved, with spurious minima.*

Reweighted Variance Minimisation

- Energy variance: $\sigma_{\Psi}^2 \equiv \langle \hat{H}^2 \rangle_{\Psi} - \langle \hat{H} \rangle_{\Psi}^2$.
 - $\sigma_{\Psi}^2 \geq 0$. In general $\sigma_{\Psi}^2 > 0$.
 - If Ψ is an eigenfn. of the Hamiltonian with eigenvalue E then $\sigma_{\Psi}^2 = E^2 - E^2 = 0$.
 - So, can minimise the variance to optimise the wave function.⁸
- Reweighted variance minimisation:
 - Generate a set of configurations distributed according to $|\Psi_{\alpha_0}|^2$ using VMC, where α_0 is an initial parameter set.
 - Correlated sampling: associate a weight of $|\Psi_{\alpha}|^2/|\Psi_{\alpha_0}|^2$ with each configuration.
 - **Reweighted variance** $\sigma_{w\alpha}^2$ of the local energies $\Psi_{\alpha}^{-1}\hat{H}\Psi_{\alpha}$ estimates $\sigma_{\Psi_{\alpha}}^2$.
 - Vary α to minimise $\sigma_{w\alpha}^2$ over the fixed configuration set.
- This optimisation method works (although not always very well).
- To minimise finite-sampling errors, we usually do multiple cycles of (i) **VMC configuration-generation** and (ii) **optimisation**.

⁸ C. J. Umrigar *et al.*, Phys. Rev. Lett. **60**, 1719 (1988); P. R. C. Kent *et al.*, Phys. Rev. B **59**, 12344 (1999); N. D. Drummond & R. J. Needs, Phys. Rev. B **72**, 085124 (2005).

Unweighted Variance Minimisation

- Ψ is an eigenstate of \hat{H} if and only if $E_L = \Psi^{-1}\hat{H}\Psi$ is **constant in config. space**.
- Can therefore optimise Ψ_α by minimising the **unweighted variance**

$$\sigma_{u\alpha}^2 = \frac{1}{N_C - 1} \sum_{\mathbf{R}} [E_{L\alpha}(\mathbf{R}) - \bar{E}_\alpha]^2,$$

of the local energies of *any* fixed set of N_C configurations $\{\mathbf{R}\}$. [$\bar{E}_\alpha = (1/N_C) \sum_{\mathbf{R}} E_{L\alpha}(\mathbf{R})$ is the average local energy.]

- Unweighted variance minimisation has a strong similarity to least-squares fitting.
- Relatively small numbers of configurations are needed.
- Suppose the configurations are distributed as $|\Psi_{\alpha_0}|^2$ (using VMC):
 - In limit of perfect sampling, $\sigma_{w\alpha}^2$ is independent of Ψ_{α_0} (it is equal to $\sigma_{\Psi_\alpha}^2 = \langle \hat{H}^2 \rangle_{\Psi_\alpha} - \langle \hat{H} \rangle_{\Psi_\alpha}^2$). **This is not true of $\sigma_{u\alpha}^2$** , which still depends on α_0 .
 - Can perform multiple cycles of unweighted variance minimisation until **self-consistency** (where the optimised α is equal to α_0) is achieved.

Linear Jastrow Parameters (I)

- Suppose the Jastrow exponent is **linear** in P parameters α :

$$J(\mathbf{R}) = \sum_{i=1}^P f_i(\mathbf{R})\alpha_i + J_0(\mathbf{R}).$$

- Then the local energy $E_L = \Psi^{-1}\hat{H}\Psi$ is a quadratic function of the parameters.
- Suppose N_C configurations are distributed according to $|\Psi_{\alpha_0}|^2$.
 - The unweighted average local energy is also quadratic in α . It has a global maximum at $\alpha = \alpha_0$.
 - The unweighted variance is a **quartic** function of the parameters.⁹
 - Sums over the configurations are not required, so $\sigma_{u\alpha}^2$ can be evaluated *extremely* rapidly.

⁹ N. D. Drummond & R. J. Needs, Phys. Rev. B **72**, 085124 (2005).

Linear Jastrow Parameters (II)

- The coefficients of the quartic variance can be accumulated in VMC.
 - Along any line in parameter space, $\sigma_{u\alpha}^2$ is a quartic polynomial of a single parameter.
 - The global minimum of a quartic polynomial can be determined analytically.
 - Hence $\sigma_{u\alpha}^2$ can be minimised rapidly, exactly and globally along lines in parameter space.
- After using a standard minimisation technique such as Broyden–Fletcher–Goldfarb–Shanno to find a local minimum of $\sigma_{u\alpha}^2$, one can perform billions of analytic line minimisations along random directions to look for lower minima.

Scaling

- Standard variance minimisation scales as

$$\mathcal{O}(N^2 P^2 N_C).$$

- The variance minimisation method for linear Jastrow parameters scales as

$$\mathcal{O}(P^4).$$

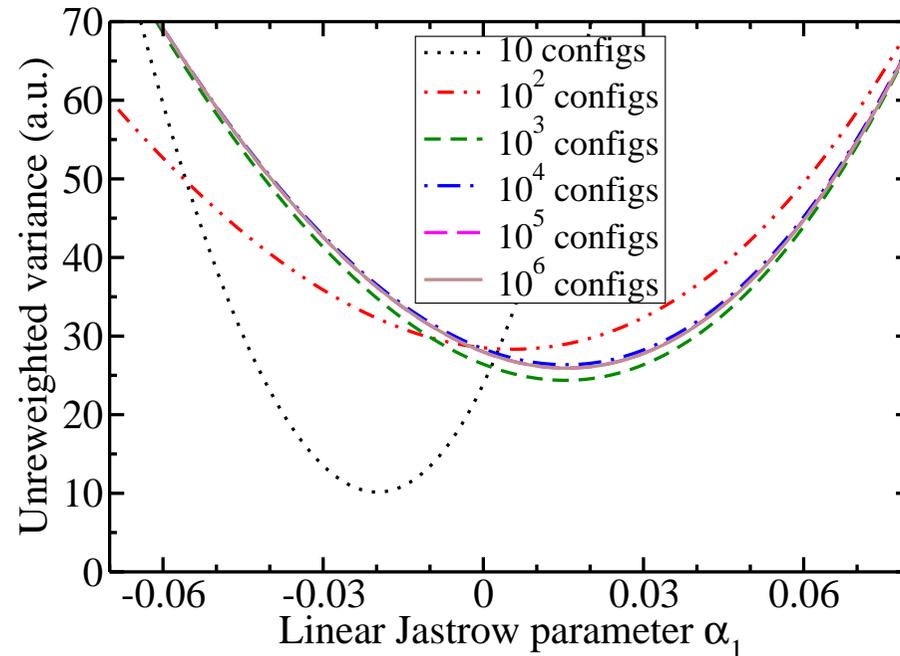
- Evaluating the “basis functions” $f_i(\mathbf{R})$ in VMC scales as

$$\mathcal{O}(PN^2 N_C).$$

- Evaluating the quartic coefficients in VMC scales as

$$\mathcal{O}(P^4 N_C).$$

Behaviour of the Variance



All-electron neon atom: $\sigma_{u\alpha}^2$ plotted against a linear Jastrow parameter. Configurations are distributed according to square of Hartree–Fock wave function.

- Unweighted variance converges rapidly with the number of configurations. Typically use more than **10,000** configurations.
- Nonglobal minima of $\sigma_{u\alpha}^2$ are *only* found for very poor samplings of config. space.
- Position of minimum is relatively insensitive to configuration distribution.

Reweighted v. Weighted Variance Minimisation

Self-consistent unweighted variance minimisation is better than reweighted variance minimisation for optimising Jastrow factors because:

1. In the limit of infinite sampling, the resulting energy is generally lower;
2. For small samplings of configuration space, it is more stable;
3. The unweighted variance is a quartic function of linear Jastrow parameters, enabling rapid evaluation;
4. For any reasonable sampling of configuration space, the unweighted variance does not have nonglobal minima in the space of linear Jastrow parameters.

Using Standard Variance Minimisation (I)

- Choose **opt_method** = “varmin” and **runtype** = “vmc_opt”.
- Number of configurations to be used in optimisation is determined by **vmc_nconfig_write**. (NB, must have **vmc_nstep** \geq **vmc_nconfig_write**.)
- Typically use 10,000 or more configurations.
- You may wish to set **vmc_nstep** to a higher value than **vmc_nconfig_write**, so that reported VMC energies are highly accurate.
- Select whether you want to optimise Jastrow factor, backflow function, orbitals or determinant coefficients by setting **opt_jastrow**, **opt_backflow**, **opt_orbitals** and **opt_detcoeff** to T or F as appropriate.
- **vmc_decorr_period** (number of moves between write-outs of configurations) should be large enough that configurations are uncorrelated. Values between 10 and 20 are typical.

Using Standard Variance Minimisation (II)

- Number of configuration-generation / optimisation cycles is specified by **opt_cycles** keyword. Between 4 and 8 is typical.
- Generally find that most of the energy reduction occurs in first optimisation step; further cycles simply “bounce around”.
- However, for large systems cutoff lengths may increase slowly over several cycles. [NB, self-consistent variance minimisation may converge to the wrong place if one starts with a large cutoff and the Hartree–Fock distribution of configurations.]
- The optimised wave functions are written to `correlation.out.x`.
- Type `ve` to obtain a summary of the optimisation.
- Choose `correlation.out.x` file that gives lowest energy to be the optimised `correlation.data`.

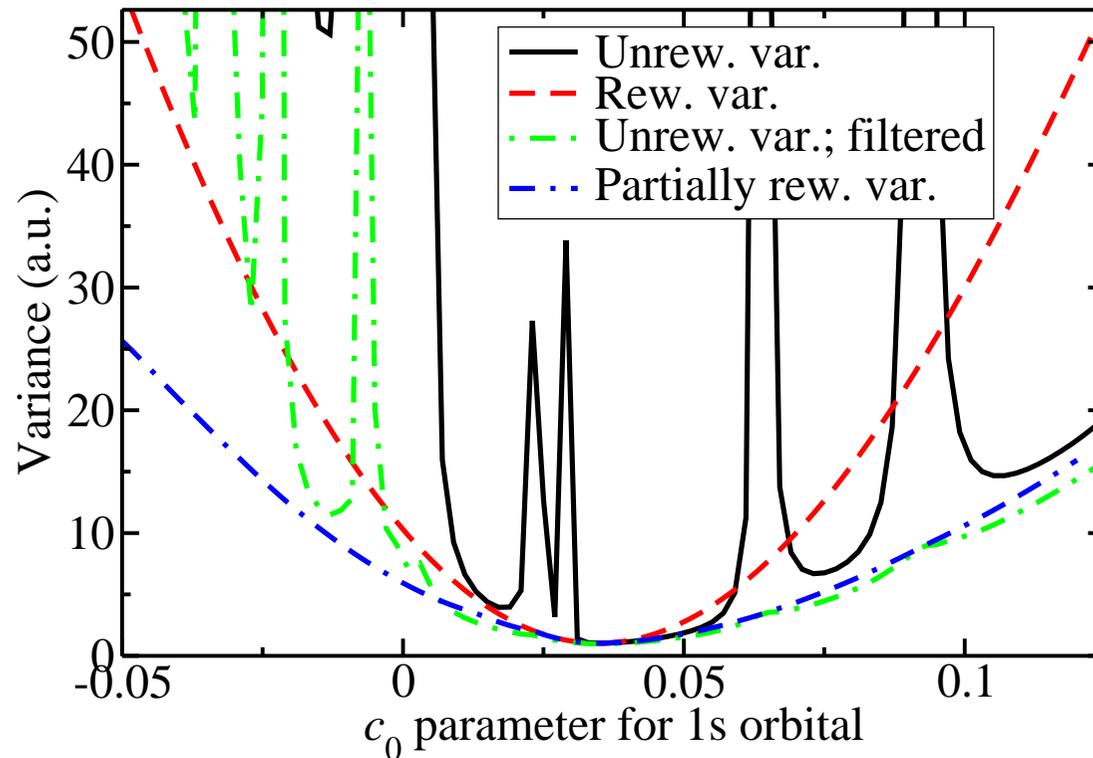
Using the Fast Method for Linear Jastrow Parameters

Exactly the same as standard variance minimisation, except:

- Choose **opt_method** = "varmin_linjas".
- Can only optimise linear parameters in Jastrow; **opt_backflow** etc. should be F.
- Cutoff lengths will not be optimised (even if their "optimizable" flags are set to "1" in `correlation.data`).

Parameters that Affect the Nodal Surface (I)

- Parameters that affect the nodal surface: multideterminant expansion coefficients, parameters in orbitals and backflow functions.
- Optimisation of such parameters is difficult because the local energy $E_L = \Psi^{-1}\hat{H}\Psi$ diverges at nodal surface.



All-electron neon atom: variance against a parameter in the orbitals.

Parameters that Affect the Nodal Surface (II)

- The unweighted variance is not well-defined in the limit of perfect sampling.
- The variance is defined for a finite sample, but optimisation is more “hit-or-miss”.
- The unweighted variance diverges whenever the nodal surface moves through a configuration.
- Often find relatively severe mismatches between energy and variance minima.

Minimising Other Measures of the Spread

- The logic of unweighted variance minimisation applies to any measure of the spread of the distribution of local energies for a given sampling of configuration space.
- **Robust** measures of spread, such as the difference of the upper and lower quartiles of the distribution of local energies are defined, even if the variance diverges.
- **Outliers** (e.g. divergent local energies) do not have as much effect on robust measures of the spread as they do on the variance in a finite sample.
- Minimising the **mean absolute deviation from the median energy**,

$$M = \frac{1}{N_C} \sum_{\mathbf{R}} |E_L(\mathbf{R}) - E_M|,$$

where E_M is the median of $\{E_L\}$, works very well. *Minimising the MAD generally gives a lower energy than minimising the variance.*

- Set **opt_method** = “madmin” to minimise the MAD instead of the variance.

A Basis for Diagonalising the Hamiltonian

- Consider a small change to the parameter set: $\alpha_0 \rightarrow \alpha = \alpha_0 + \delta\alpha$. Then

$$\Psi_{\alpha} = \Psi_{\alpha_0} + \sum_{j=1}^P \delta\alpha_j \left(\frac{\partial \Psi}{\partial \alpha_j} \right)_{\alpha_0} + \mathcal{O}(\delta\alpha^2) \simeq \sum_{j=0}^P c_j \beta_j,$$

where

$$\beta_j = \begin{cases} \Psi_{\alpha_0} & \text{if } j = 0 \\ \left(\frac{\partial \Psi}{\partial \alpha_j} \right)_{\alpha_0} & \text{if } j > 0 \end{cases}$$
$$c_j = \begin{cases} 1 & \text{if } j = 0 \\ \delta\alpha_j & \text{if } j > 0 \end{cases}$$

- For multideterminant coefficients, this expansion is exact. Minimise the energy by **diagonalising** in the basis $\{\beta_i\}$.
- Method is valid in the limit of self-consistency, where $\delta\alpha$ is small.

“Standard” Diagonalisation

- Neglect the remainder term and demand that $\langle \Psi_\alpha | \hat{H} | \Psi_\alpha \rangle$ is minimised with respect to \mathbf{c} such that $\langle \Psi_\alpha | \Psi_\alpha \rangle$ is constant.¹⁰
- Hence $(H - ES)\mathbf{c} = \mathbf{0}$ where $H_{ij} = \langle \beta_i | \hat{H} | \beta_j \rangle$ and $S_{ij} = \langle \beta_i | \beta_j \rangle$ and E is an eigenvalue.
- Lowest eigenvalue gives the lowest-energy state that can be constructed from the basis functions, and the corresponding eigenvector gives the coefficients for this state.
- $c_0 = 1$ determines normalisation of eigenvector; other coefficients then give $\delta\alpha_i = c_i$.
- Matrices H and S can be estimated in VMC. One can enforce symmetry (but it turns out to be much better not to do this.)
- The method is exact in the limit of small steps $\delta\alpha$.

¹⁰ M. P. Nightingale and V. Melik-Alaverdian, Phys. Rev. Lett. **87**, 043401 (2001).

Linear Least-Squares Method (I)

- Seek eigenstates of \hat{H} by considering the action of \hat{H} on basis states $\{\beta_j\}$.¹¹
- Suppose the $\{\beta_i\}$ span an invariant subspace of \hat{H} , i.e., the action of \hat{H} on any basis state results in a linear combination of all the basis states. Then we can write

$$\hat{H}\beta_i(\mathbf{R}_m) = \sum_{j=0}^P \eta_{ji}\beta_j(\mathbf{R}_m).$$

for all N_C VMC-generated configurations $\{\mathbf{R}_m\}$ for some coefficients η_{ji} .

- Once η has been determined, one can solve $\eta\mathbf{c} = E\mathbf{c}$ to find $\delta\alpha$.
- If $\{\beta_j\}$ really spanned an invariant subspace of \hat{H} , one could simply take *any* $P + 1$ distinct points in configuration space $\{\mathbf{R}_m\}$ and solve for η .
- In practice, take $N_C \gg P$, so there are $N_C(P + 1)$ equations but only $(P + 1)^2$ unknowns $\{\eta_{ji}\}$.

¹¹ J. Toulouse & C. J. Umrigar, J. Chem. Phys. **126**, 084102 (2007).

Linear Least-Squares Method (II)

- Overdetermined linear problem. Seek least-squares fit by minimising

$$\chi^2 = \sum_{m=1}^{N_C} \sum_{i=0}^P \left| \frac{\hat{H} \beta_i(\mathbf{R}_m) - \sum_{j=0}^P \eta_{ji} \beta_j(\mathbf{R}_m)}{\Psi_{\alpha_0}(\mathbf{R}_m)} \right|^2$$

with respect to $\{\eta_{ji}\}$.

- Let $B_{mi} = \beta_i(\mathbf{R}_m)/\Psi_{\alpha_0}(\mathbf{R}_m)$ and $B_{mi}^H = [\hat{H} \beta_i(\mathbf{R}_m)]/\Psi_{\alpha_0}(\mathbf{R}_m)$.

- $\partial \chi^2 / \partial \eta_{pq} = 0$, so

$$\sum_{m=1}^{N_C} \left(B_{mi}^H - \sum_{j=0}^P \eta_{ji} B_{mj} \right) B_{mp}^* = 0,$$

and hence $\eta = (B^\dagger B)^{-1} B^\dagger B^H$.

- $B^\dagger B$ and $B^\dagger B^H$ are proportional to (non-symmetrised) VMC estimates of S and H .

Linear Least-Squares Method (III)

- So $\lim_{N_C \rightarrow \infty} \eta = S^{-1}H$.
- In the limit of perfect sampling, “standard diagonalisation” and linear-least-squares optimisation solve the same eigenproblem.
- But the least-squares algorithm is also exact in limit that the basis functions span an invariant subspace of Hamiltonian (even for finite sampling).
- In practice the least-squares method has much smaller finite-sampling errors.

Preventing Large Parameter-Value Changes

- **Semi-orthogonalisation**: we can improve the stability of the algorithm by choosing a parameter-dependent normalisation for Ψ_α .
 - The normalisation is such that the $\{\beta_i\}$ for $i \geq 1$ are orthogonal to a chosen wave function Λ .
 - Λ is chosen to be a linear combination of the old wave function and the linear approximation to the new one.
- Instabilities are usually associated with large steps in parameter space, which invalidate the linear approximation.
- If $H = B^\dagger B^H$ were diagonal with the top-left element having the lowest value then eigenvector with lowest eigenvalue would correspond to not changing wave function. (Recall that $\beta_0 = \Psi_{\alpha_0}$.)
 - Can improve stability by artificially increasing the on-diagonal elements. This is done by trial and error. A bit messy, but it usually works.

How and When to Use Linear-Least-Squares Energy Minimisation

- Procedure for energy min. is the same as for variance min. except that:
 - The **opt_method** keyword should be set to “emin”.
 - Usually need more configurations than variance minimisation. Ensure the statistical error bars on your VMC energy are smaller than your final target accuracy.
 - Numerically, cutoff lengths are difficult to optimise. Fix them for energy minimisation.
 - Energy minimisation struggles if the initial wave function is poor (e.g., a blank wave function).
- Energy minimisation is more likely to succeed if the initial wave function is good; it is nearly always a good idea to perform a preliminary “varmin”, “varmin-linjas” or “madmin” optimisation.
- Use varmin-linjas if rapid optimisation is required, e.g. for large systems or when deciding how many parameters are needed in the Jastrow factor.
- It is important to use energy minimisation if parameters that move the nodal surface are to be optimised: the DMC energy obtained is generally lower.

Example Numerical Results with Different Optimisation Methods

Method	Energy (a.u.)	Variance (a.u.)
Varmin (rewt.)	-5.4374(2)	0.0538
Varmin (unrewt.)	-5.4387(2)	0.0548
Madmin	-5.4401(2)	0.0622
Emin	-5.4414(2)	0.0757

VMC energies for a pseudocarbon atom with a Slater–Jastrow wave function, optimised using 100,000 configurations.

Method	Energy (a.u.)	Variance (a.u.)
Varmin (unrewt.)	-5.4406(2)	0.0520
Madmin	-5.4448(2)	0.0840
Emin	-5.4463(2)	0.0988

VMC energies for a pseudocarbon atom with a Slater–Jastrow–backflow wave function, optimised using 20,000 configurations.

Summary

- Despite enormous progress in the last few years, choosing and optimising a trial wave function remains the most difficult aspect of using QMC.
 - **The bad news:** one has to be prepared to try out different parametrisations and optimisation methods.
 - **The good news:** usually one only needs to carry out one such investigation in a QMC study of a set of similar molecules or solids.
- **An optimisation strategy for difficult cases (be prepared to try others!):**
 - Use quick “varmin-linjas” runs with just u and χ terms to obtain initial wave functions and to investigate the choice of cutoff length for u .
 - Use “varmin-linjas” to optimise a complete Jastrow factor (u , χ , f and p).
 - If you are not satisfied with the cutoff lengths, perform a longer “varmin” run with a large number of cycles in which the cutoff lengths are free.
 - Fix all the cutoff lengths, add a blank backflow function to `correlation.data` if desired, and optimise everything together in a large “emin” calculation.