Wave Function Optimization in Quantum Monte Carlo

Arne Lüchow

Theoretical Chemistry Group, RWTH Aachen University

Autumn School on Correlated Electrons, FZ Jülich Sept. 16, 2019

Outline

- Brief overview of quantum Monte Carlo methods (QMC)
- Optimization of wave functions
- Examples
- Recent developments

Quantum Monte Carlo – a brief overview

Quantum Monte Carlo methods

Numerical methods to solve the Schrödinger equation by means of Monte Carlo techniques (i.e. using random sampling) are called *quantum Monte Carlo* methods (QMC)

• electron structure QMC for *electronic Schrödinger equation*

Quantum Monte Carlo – a brief overview

Quantum Monte Carlo methods

Numerical methods to solve the Schrödinger equation by means of Monte Carlo techniques (i.e. using random sampling) are called *quantum Monte Carlo* methods (QMC)

• electron structure QMC for *electronic Schrödinger equation*

Two main variants:

 solve stationary Schrödinger equation with stochastic process: Diffusion Monte Carlo (DMC)
 Simulate stochastic process with random numbers

Quantum Monte Carlo – a brief overview

Quantum Monte Carlo methods

Numerical methods to solve the Schrödinger equation by means of Monte Carlo techniques (i.e. using random sampling) are called *quantum Monte Carlo* methods (QMC)

• electron structure QMC for electronic Schrödinger equation

Two main variants:

- solve stationary Schrödinger equation with stochastic process: Diffusion Monte Carlo (DMC)
 Simulate stochastic process with random numbers
- evaluate $E = \langle \Psi | H | \Psi \rangle$ with Monte Carlo integration: Variational Monte Carlo (VMC) Sample the integrand with random numbers

Diffusion Monte Carlo (DMC)

The starting point:

Schrödinger equation in imaginary time au = it

$$\frac{\partial}{\partial \tau}\Psi = -H\Psi = \frac{1}{2}\Delta\Psi - V\Psi$$

 \bullet Long-time solution is the ground state wave function Ψ_0

$$\lim_{\tau\to\infty}e^{-H\tau}\Psi\propto\Psi_0$$

¹A. Lüchow, WIREs Comput Mol Sci 2011, 1, 388-402

Diffusion Monte Carlo (DMC)

The starting point:

Schrödinger equation in imaginary time au = it

$$\frac{\partial}{\partial \tau}\Psi = -H\Psi = \frac{1}{2}\Delta\Psi - V\Psi$$

 $\bullet\,$ Long-time solution is the ground state wave function Ψ_0

$$\lim_{\tau\to\infty}e^{-H\tau}\Psi\propto\Psi_0$$

- Mathematically, this is a diffusion equation
- In diffusion Monte Carlo (DMC), this diffusion process is simulated with a stochastic process¹

¹A. Lüchow, WIREs Comput Mol Sci 2011, 1, 388-402

Importance Sampling: Drift-Diffusion process

• discretized stochastic process: time step $\Delta \tau$

$$X_{k+1} = X_k + b(X_k)\Delta au + \sqrt{\Delta au} \xi$$

(drift-diffusion process)

• drift term $b(x) = \nabla \Psi_G / \Psi_G$: non-random moves to large values of $|\Psi_G|$ (guide function Ψ_G).

Importance Sampling: Drift-Diffusion process

• discretized stochastic process: time step $\Delta \tau$

$$X_{k+1} = X_k + b(X_k)\Delta au + \sqrt{\Delta au} \xi$$

(drift-diffusion process)

- drift term $b(x) = \nabla \Psi_G / \Psi_G$: non-random moves to large values of $|\Psi_G|$ (guide function Ψ_G).
- stationary distribution of drift-diffusion process is $|\Psi_G|^2$ for $\Delta\tau \to 0.$

Importance Sampling: Drift-Diffusion process

• discretized stochastic process: time step $\Delta \tau$

$$X_{k+1} = X_k + b(X_k)\Delta \tau + \sqrt{\Delta \tau} \xi$$

(drift-diffusion process)

- drift term $b(x) = \nabla \Psi_G / \Psi_G$: non-random moves to large values of $|\Psi_G|$ (guide function Ψ_G).
- stationary distribution of drift-diffusion process is $|\Psi_G|^2$ for $\Delta\tau \to 0.$
- DMC: weights

$$W(\Delta\tau, x, x') = e^{-\left[\frac{1}{2}(E_L(x) + E_L(x')) - E_{ref}\right]\Delta\tau}$$

fluctuate around 1 if local energies $E_L = \hat{H} \Psi_G / \Psi_G$ are close to E_{ref} (nonsingular!)

Fermion sign problem

Perfect for vibrational ground state (vibrational DMC). BUT:

- lowest eigenstate of H is *not* the *electronic* ground state (but a bosonic state) for n > 2
- The stochastic process is *forced* to an antisymmetric state by *fixed node approximation*

Fermion sign problem

Perfect for vibrational ground state (vibrational DMC). BUT:

- lowest eigenstate of H is *not* the *electronic* ground state (but a bosonic state) for n > 2
- The stochastic process is *forced* to an antisymmetric state by *fixed node approximation*
- Schrödinger equation is solved with *additional boundary* condition given by nodes of a guide function Ψ_G : $\Psi_G(\mathbf{R}) = 0$

• **R**: all electron coords: $\mathbf{R} = (x_1, y_1, z_1, \dots, z_n)!$

• Nodes (nodal set): 3n-1 hypersurface, unrelated to orbital nodes

Variational Monte Carlo (VMC)

Variational energy of a *trial function* Ψ_T :

l

$$E_{VMC} = \frac{\langle \Psi_T | H | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} = \int E_L(\mathbf{R}) \cdot p(\mathbf{R}) d\mathbf{R}$$

with *local energy:*

$$\mathsf{E}_L(\mathsf{R}) = \frac{H\Psi_T(\mathsf{R})}{\Psi_T(\mathsf{R})}$$

and probability distribution:

$$p(\mathbf{R}) = \frac{|\Psi(\mathbf{R})|^2}{\int |\Psi(\mathbf{R})|^2 d\mathbf{R}}$$

Variational Monte Carlo (VMC)

Variational energy of a *trial function* Ψ_T :

$$E_{VMC} = \frac{\langle \Psi_T | H | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} = \int E_L(\mathbf{R}) \cdot p(\mathbf{R}) d\mathbf{R}$$

with local energy:

$$E_L(\mathbf{R}) = \frac{H\Psi_T(\mathbf{R})}{\Psi_T(\mathbf{R})}$$

and probability distribution:

$$p(\mathbf{R}) = rac{|\Psi(\mathbf{R})|^2}{\int |\Psi(\mathbf{R})|^2 d\mathbf{R}}$$

Monte Carlo integration with sample $\{\mathbf{R}_k\} \sim p(\mathbf{R})$:

$$E_{VMC} = \langle E_L \rangle_{P_{VMC}} = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^N E_L(\mathbf{R}_k)$$

• Metropolis-Hastings often: drift-diffusion step

VMC and DMC energies

Diffusion Monte Carlo (DMC):

- weighted/branched drift-diffusion: $p_{DMC}(\mathbf{R}) \propto \Psi_G \Psi_0^{(FN)}$
- DMC energy: average of local energies

$$E_{DMC} = \langle E_L \rangle_{P_{DMC}} = \frac{\langle \Psi_0^{(FN)} | H | \Psi_G \rangle}{\langle \Psi_0^{(FN)} | \Psi_G \rangle}$$

• (after time step extrapolation to $\Delta au
ightarrow 0)$

VMC and DMC energies

Diffusion Monte Carlo (DMC):

- weighted/branched drift-diffusion: $p_{DMC}(\mathbf{R}) \propto \Psi_G \Psi_0^{(FN)}$
- DMC energy: average of local energies

$$E_{DMC} = \langle E_L \rangle_{P_{DMC}} = \frac{\langle \Psi_0^{(FN)} | H | \Psi_G \rangle}{\langle \Psi_0^{(FN)} | \Psi_G \rangle}$$

- (after time step extrapolation to $\Delta \tau \rightarrow 0$) Variational quantum Monte Carlo (VMC):
 - VMC energy: average of local energies

$$E_{VMC} = \langle E_L \rangle_{P_{VMC}} = \frac{\langle \Psi_T | H | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle}$$

• DMC more accurate, VMC more efficient

Computational Considerations

- VMC and DMC require guide or trial function Ψ for local energy
- no integrals! Only 2nd derivatives!

Guide and Trial Functions

Compact and physically motivated approximations to exact wave function. Explicitly correlated!

- Scaling of local energy $E_L = H\Psi/\Psi$ with number of electrons
- Variance $\langle (E_L \langle E_L \rangle)^2 \rangle$ determines sample size
- Scales well to 10,000 of cpu cores

Optimization of wave functions: energy minimization

- large sample of electron configurations $\{\mathbf{R}_k\}, \ k = 1, ..., K$ with $p(\mathbf{R}) = |\Psi(\mathbf{R})|^2 / \int |\Psi(\mathbf{R})|^2 d\tau$
- parameter vector p
- wave function $\Psi(\mathbf{R},\mathbf{p})\equiv\Psi(\mathbf{p})$
- local energy $E_L(\mathbf{R},\mathbf{p}) \equiv E_L(\mathbf{p})$

$$E_{\text{VMC}} = E_m(\mathbf{p}) = \frac{1}{K} \sum_{k=1}^{K} E_L(\mathbf{R}_k, \mathbf{p}).$$

Optimization of wave functions: energy minimization

- large sample of electron configurations $\{\mathbf{R}_k\}, \ k = 1, ..., K$ with $p(\mathbf{R}) = |\Psi(\mathbf{R})|^2 / \int |\Psi(\mathbf{R})|^2 d\tau$
- parameter vector p
- wave function $\Psi(\mathbf{R},\mathbf{p})\equiv\Psi(\mathbf{p})$
- local energy $E_L(\mathbf{R},\mathbf{p}) \equiv E_L(\mathbf{p})$

$$E_{\text{VMC}} = E_m(\mathbf{p}) = \frac{1}{K} \sum_{k=1}^{K} E_L(\mathbf{R}_k, \mathbf{p}).$$

Small parameter changes \mathbf{p}_0 to \mathbf{p} : weighted mean

$$E_{\text{VMC}} = E_{wm}(\mathbf{p}) = \frac{\sum_{k=1}^{K} w_k E_L(\mathbf{R}_k, \mathbf{p})}{\sum_{k=1}^{K} w_k}, \qquad w_k = \frac{|\Psi(\mathbf{R}_k, \mathbf{p})|^2}{|\Psi(\mathbf{R}_k, \mathbf{p}_0)|^2}.$$

• Stochastic optimization of wave function: minimization of energy Optimization of wave functions: variance minimization

variance of the local energy

$$V_m(\mathbf{p}) = rac{1}{\kappa} \sum_{k=1}^{\kappa} (E_L(\mathbf{R}_k, \mathbf{p}) - E_m)^2$$

• exact eigenfunctions of H have $V_m = 0$

• Advantage: stable, each term positive Variance minimization:

- very stable and efficient
- less accurate energies (VMC and DMC)

VMC vs DMC wave function optimization

• DMC more accurate than VMC: DMC optimization desired!

VMC vs DMC wave function optimization

• DMC more accurate than VMC: DMC optimization desired! BUT:

- projected $\Psi_0^{(FN)}$ depends on (guide) function parameters
- $p_{\text{DMC}} = \Psi_0^{(FN)} \Psi_G$ only statistically known
- derivatives?

VMC vs DMC wave function optimization

• DMC more accurate than VMC: DMC optimization desired! BUT:

- projected $\Psi_0^{(FN)}$ depends on (guide) function parameters
- $p_{\text{DMC}} = \Psi_0^{(FN)} \Psi_G$ only statistically known
- derivatives?

AND:

- use optimized VMC trial function as guide functions in DMC!
- VMC optimization can optimize DMC energy
- parameters modifying nodes of wave function required!
- FN-DMC energy: optimal wf with given nodes
- flexible wf: optimal VMC energy close to DMC energy

Notation

Notation for parameter derivatives

$$\Psi_i = \frac{\partial \Psi(\mathbf{p})}{\partial p_i}, \qquad \Psi_{ij} = \frac{\partial^2 \Psi(\mathbf{p})}{\partial p_i \partial p_j}, \qquad E_{L,i} = \frac{\partial E_L(\mathbf{p})}{\partial p_i}.$$

Notation for expectation and sample mean:

$$\langle A \rangle = \int A(\mathbf{R}) \rho(\mathbf{R}) d\mathbf{R}, \qquad \rho(\mathbf{R}) = \frac{|\Psi(\mathbf{R})|^2}{\int |\Psi(\mathbf{R})|^2 d\mathbf{R}},$$

corresponding estimator is the sample mean

$$\langle A \rangle = \lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} A(\mathbf{R}_k), \qquad \mathbf{R}_k \sim p(\mathbf{R}).$$

Newton methods

Expand $E_{VMC}(\mathbf{p})$ around \mathbf{p}_0 with $\mathbf{p} = \mathbf{p}_0 + \Delta \mathbf{p}$:

$$\Delta \mathbf{p} = -\mathbf{h}^{-1}\mathbf{g}$$

with the gradient and Hessian matrix of the VMC energy

$$g_i = rac{\partial E_{VMC}}{\partial p_i}, \qquad h_{ij} = rac{\partial^2 E_{VMC}}{\partial p_i \partial p_j}$$

How to obtain gradient and Hessian?

Gradient of VMC energy

Be careful!

$$\frac{\partial E_{\text{VMC}}}{\partial p_{i}} = \frac{\partial \langle E_{L} \rangle}{\partial p_{i}} \neq \langle E_{L,i} \rangle = \left\langle \frac{\partial E_{L}}{\partial p_{i}} \right\rangle$$

• Distribution depends on parameters: $p(\mathbf{R}) = p(\mathbf{R}, \mathbf{p})!$

Gradient of VMC energy

Be careful!

$$\frac{\partial E_{\rm VMC}}{\partial p_i} = \frac{\partial \langle E_L \rangle}{\partial p_i} \neq \langle E_{L,i} \rangle = \left\langle \frac{\partial E_L}{\partial p_i} \right\rangle$$

• Distribution depends on parameters: $p(\mathbf{R}) = p(\mathbf{R}, \mathbf{p})!$

$$\frac{\partial}{\partial p_{i}} \langle E_{L} \rangle = \frac{\partial}{\partial p_{i}} \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle}
= \frac{\partial}{\partial p_{i}} \langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} - \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \frac{\partial}{\partial p_{i}} \langle \Psi | \Psi \rangle}{\langle \Psi | \Psi \rangle}
= \frac{\langle \Psi_{i} | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} + \frac{\langle \Psi | H | \Psi_{i} \rangle}{\langle \Psi | \Psi \rangle} - 2 \langle E_{L} \rangle \frac{\langle \Psi_{i} | \Psi \rangle}{\langle \Psi | \Psi \rangle}
= \left\langle \frac{\Psi_{i}}{\Psi} E_{L} \right\rangle + \left\langle \frac{H\Psi_{i}}{\Psi} \right\rangle - 2 \langle E_{L} \rangle \left\langle \frac{\Psi_{i}}{\Psi} \right\rangle$$
(1)

cont'd

Comparing

$$E_{L,i} = \frac{H\Psi_i}{\Psi} - E_L \frac{\Psi_i}{\Psi} \implies \langle E_{L,i} \rangle = \left\langle \frac{H\Psi_i}{\Psi} \right\rangle - \left\langle \frac{\Psi_i}{\Psi} E_L \right\rangle$$
$$\frac{\partial}{\partial p_i} \langle E_L \rangle = 2 \left[\left\langle \frac{\Psi_i}{\Psi} E_L \right\rangle - \left\langle \frac{\Psi_i}{\Psi} \right\rangle \langle E_L \rangle \right] + \langle E_{L,i} \rangle$$

use Hermicity of H

$$\langle E_{L,i} \rangle = \left\langle \frac{H\Psi_i}{\Psi} \right\rangle - \left\langle \frac{\Psi_i}{\Psi} \frac{H\Psi}{\Psi} \right\rangle = \frac{\langle \Psi | H | \Psi_i \rangle}{\langle \Psi | \Psi \rangle} - \frac{\langle \Psi_i | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} = 0$$

finally

$$\frac{\partial E_{\text{VMC}}}{\partial p_i} = 2 \left[\left\langle \frac{\Psi_i(\mathbf{R})}{\Psi(\mathbf{R})} E_L(\mathbf{R}) \right\rangle - \left\langle \frac{\Psi_i(\mathbf{R})}{\Psi(\mathbf{R})} \right\rangle \langle E_L(\mathbf{R}) \rangle \right]$$

- covariance form: $\langle xy \rangle \langle x \rangle \langle y \rangle$
- no derivative of local energy $E_L!$

Terms required for Newtons method

• estimators for Hessian similar (see refs) required:

$$E_{L,i}(\mathbf{R}), \quad \frac{\Psi_i(\mathbf{R})}{\Psi(\mathbf{R})}, \quad \frac{\Psi_{ij}(\mathbf{R})}{\Psi(\mathbf{R})}$$

- normalization factors cancel
- local energy derivative required in Hessian
- Hessian can also be cast in covariance-like terms

Linear method

• expand wave function to first order in n_p parameters $\Psi^{(1)}(\mathbf{p}) = \Psi^{(1)}(\mathbf{p}_0 + \Delta \mathbf{p}) = \Psi(\mathbf{p}_0) + \sum_{i=1}^{n_p} \Delta p_i \Psi_i(\mathbf{p}_0)$

• basis
$$\Psi_i = \Psi_i(\mathbf{p}_0)$$
, $i = 0, \dots, n_p$, with $\Psi_0 \equiv \Psi(\mathbf{p}_0)$

diagonalize Hamiltonian in this basis

$$\mathbf{H} \Delta \mathbf{p} = E \mathbf{S} \Delta \mathbf{p}$$

with $H_{ij} = \langle \Psi_i | H | \Psi_j \rangle$, $S_{ij} = \langle \Psi_i | \Psi_j \rangle$.

- parameter changes from lowest eigenvector with $\Delta p_0 = 1$
- H_{ij}, S_{ij} analogously to Newton gradient and Hessian
- do not symmetrize H_{ij}, use unsymmetric eigenvalue code
- iterate until convergence

Stochastic reconfiguration method and variants

- Use basis $\Psi \equiv \Psi_0$ and derivatives Ψ_i
- apply DMC propagator $\exp(-tH) pprox 1 Ht$ to Ψ
- project result into vector space of basis

$$\langle \Psi_i | 1 - tH | \Psi
angle = \sum_{j=0}^{n_p} lpha_j \langle \Psi_i | \Psi_j
angle$$

- parameter change $\Delta p_i = lpha_i / lpha_0$
- requires S_{ij} and $H_{i0} = \frac{1}{2} \frac{\partial E_{VMC}}{\partial p_i}$
- no local energy derivatives necessary
- iterate with suitable t

energy fluctuation potential and perturbative method

Variant of stochastic reconfiguration method:

$$\Delta p_i = -\frac{1}{\Delta \varepsilon_i} \sum_{j=1}^{n_p} (\mathbf{S}^{-1})_{ij} H_{j0}$$

with

$$\Delta \varepsilon_i = \frac{H_{ii}}{S_{ii}} - H_{00}$$

•
$$H_{j0} = \frac{1}{2} \frac{\partial E_{VMC}}{\partial p_j}$$

Newton-like!

• local energy derivative $E_{L,i}$ required for $\Delta \varepsilon_i$

Wave functions to optimize

Slater-Jastrow form:

$$\Psi(\mathbf{R}) = e^{U(\mathbf{R})} \sum_{d} c_{d} \Phi^{(d)}(\mathbf{R})$$

with Slater determinants

$$\Phi^{(d)}(\mathbf{R}) = \det(\phi_1(1)\phi_2(2)\cdots\phi_n(n))$$

based on molecular (spin) orbitals ϕ_i and Jastrow correlation factor

$$U(\mathbf{R}) = U_2(\{r_{ij}\}) + U_3(\{r_{ij}, R_{li}, R_{lj}\}) + \cdots$$

electron distances $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ electron-nucleus distances $R_{li} = |\mathbf{R}_l - \mathbf{r}_i|$

Wave functions to optimize

Slater-Jastrow form:

$$\Psi(\mathbf{R}) = e^{U(\mathbf{R})} \sum_{d} c_{d} \Phi^{(d)}(\mathbf{R})$$

with Slater determinants

$$\Phi^{(d)}(\mathbf{R}) = \det(\phi_1(1)\phi_2(2)\cdots\phi_n(n))$$

based on molecular (spin) orbitals ϕ_i and Jastrow correlation factor

$$U(\mathbf{R}) = U_2(\{r_{ij}\}) + U_3(\{r_{ij}, R_{li}, R_{lj}\}) + \cdots$$

electron distances $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ electron-nucleus distances $R_{li} = |\mathbf{R}_l - \mathbf{r}_i|$

- strong correlation: multi Slater (e.g. CAS)
- dynamic correlation (electron cusp, vdW): Jastrow

Parameters to optimize

Slater-Jastrow wave function:

$$\Psi(\mathsf{R};\mathsf{c},\kappa,\mathsf{q}) = \Phi(\mathsf{R};\mathsf{c},\kappa)e^{U(\mathsf{R};\mathsf{q})}$$

with

$$\Phi(\mathbf{R};\mathbf{c},\kappa) \equiv \Phi(\mathbf{c},\kappa) = \sum_{d=1}^{n_{\mathrm{det}}} c_d \Phi^{(d)}(\kappa)$$

- Jastrow parameters q (FN invariant!)
- $\bullet\,$ configuration interaction (CI) parameters c
- orbital rotation parameters (MO) κ

Parameters to optimize

Slater-Jastrow wave function:

$$\Psi(\mathsf{R};\mathsf{c},\kappa,\mathsf{q}) = \Phi(\mathsf{R};\mathsf{c},\kappa)e^{U(\mathsf{R};\mathsf{q})}$$

with

$$\Phi(\mathbf{R};\mathbf{c},\kappa) \equiv \Phi(\mathbf{c},\kappa) = \sum_{d=1}^{n_{\text{det}}} c_d \Phi^{(d)}(\kappa)$$

- Jastrow parameters q (FN invariant!)
- \bullet configuration interaction (CI) parameters ${\bf c}$
- orbital rotation parameters (MO) κ

Note (known from MCSCF):

- Slater dets from set of *orthogonal molecular orbitals* $\{\phi_i(\mathbf{r})\}$
- orbital rotation: pairwise mixing, retain orthogonality
- orthogonal transformation: $\mathbf{O} = \exp(\mathbf{X})$, rotation angles x_{ij}

Derivatives w.r.t. orbital rotations

orbital rotation: mixing of ϕ_i and ϕ_j by angle x_{ij}

$$\begin{pmatrix} \phi_i' \\ \phi_j' \end{pmatrix} = \begin{pmatrix} \cos x_{ij} & -\sin x_{ij} \\ \sin x_{ij} & \cos x_{ij} \end{pmatrix} \begin{pmatrix} \phi_i \\ \phi_j \end{pmatrix}$$

with $\phi_i, \phi_j, \phi'_i, \phi'_j$ vectors in atomic basis set.

Derivatives w.r.t. orbital rotations

orbital rotation: mixing of ϕ_i and ϕ_j by angle x_{ij}

$$\begin{pmatrix} \phi_i' \\ \phi_j' \end{pmatrix} = \begin{pmatrix} \cos x_{ij} & -\sin x_{ij} \\ \sin x_{ij} & \cos x_{ij} \end{pmatrix} \begin{pmatrix} \phi_i \\ \phi_j \end{pmatrix}$$

with $\phi_i, \phi_j, \phi'_i, \phi'_j$ vectors in atomic basis set. For every Slater determinant Φ of orthogonal orbital set :

$$\hat{x}_{ij}\Phi = \cos x_{ij}\Phi + \sin x_{ij}\left(\hat{a}_i^{\dagger}\hat{a}_j - \hat{a}_j^{\dagger}\hat{a}_i\right)\Phi.$$

Derivatives w.r.t. orbital rotations

orbital rotation: mixing of ϕ_i and ϕ_j by angle x_{ij}

$$\begin{pmatrix} \phi_i' \\ \phi_j' \end{pmatrix} = \begin{pmatrix} \cos x_{ij} & -\sin x_{ij} \\ \sin x_{ij} & \cos x_{ij} \end{pmatrix} \begin{pmatrix} \phi_i \\ \phi_j \end{pmatrix}$$

with $\phi_i, \phi_j, \phi'_i, \phi'_j$ vectors in atomic basis set. For every Slater determinant Φ of orthogonal orbital set :

$$\hat{x}_{ij}\Phi = \cos x_{ij}\Phi + \sin x_{ij}\left(\hat{a}_i^{\dagger}\hat{a}_j - \hat{a}_j^{\dagger}\hat{a}_i\right)\Phi$$

Derivative w.r.t. x_{ij} at zero angle for all determinants:

$$\left. \frac{\partial}{\partial x_{ij}} \hat{x}_{ij} \Phi^{(d)} \right|_{x_{ij}=0} = \left(\hat{a}_i^{\dagger} \hat{a}_j - \hat{a}_j^{\dagger} \hat{a}_i \right) \Phi^{(d)} = \hat{E}_{ij}^{-} \Phi^{(d)}$$

with excitation operator $\hat{E}_{ij}^- = \hat{a}_i^\dagger \hat{a}_j - \hat{a}_j^\dagger \hat{a}_i$

Parameter derivative of the local energy local energy

$$-\frac{1}{2}\frac{H\Psi}{\Psi}, \qquad H = T + V = -\frac{1}{2}\nabla^2 + V$$

parameter derivative:

$$E_{L,i} = -\frac{1}{2} \frac{\nabla^2 \Psi_i}{\Psi} + \frac{1}{2} \frac{\nabla^2 \Psi}{\Psi} \frac{\Psi_i}{\Psi} + \frac{\partial V_{\mathsf{nl}}}{\partial p_i}$$

Parameter derivative of the local energy local energy

$$-\frac{1}{2}\frac{H\Psi}{\Psi}, \qquad H = T + V = -\frac{1}{2}\nabla^2 + V$$

parameter derivative:

$$E_{L,i} = -\frac{1}{2} \frac{\nabla^2 \Psi_i}{\Psi} + \frac{1}{2} \frac{\nabla^2 \Psi}{\Psi} \frac{\Psi_i}{\Psi} + \frac{\partial V_{\mathsf{nl}}}{\partial p_i}$$

Careful with localized nonlocal pseudo potential \hat{W} :

$$\frac{\partial}{\partial p_i}\frac{\hat{W}\Psi}{\Psi} = \frac{\hat{W}\Psi_i}{\Psi} - \frac{\hat{W}\Psi}{\Psi}\frac{\Psi_i}{\Psi}$$

- Lebedev integration for nonlocal $|ml
 angle \langle ml| \Psi
 angle$
- $\hat{W}\Psi_i$ all excitations for MO parameters!
- local energy derivs expensive for MO optimization with PP

Example: MO optimization in C_2

• carbon dimer C_2 , full CAS + Jastrow, no PP



Simultaneous vs alternating optimization I

- \bullet carbon dimer C2, Jastrow+CI+MO opt, full CAS + Jastrow,
- no PP



Simultaneous vs alternating optimization II

- carbon dimer C₂, Jastrow+CI+MO opt, full CAS + Jastrow,
- with PP (QMC-PP)



Experience with parameter optimization methods

- linear method usually best!
- too expensive for MO opt with PP
- perturbative method/EFP with approximate $\Delta \varepsilon_i$!
- alternating optimization with pert.method for MO only

transition metal compound FeH

- CAS(9,7): 21, Jas: 69, MO: 159 parameters
- with PP
- \bullet Alternating optimization: Jas \rightarrow Cl \rightarrow MO



Results: dissociation energies of transition metal dimers

Compound	Ansatz	Orbitals	Optimization level	$D_0/{ m eV}$
ZnO	Single det	B3LYP opt	Jas Jas+MO	1.449(19) 1.566(21)
	CAS	CAS CAS opt	Jas Jas+Cl Jas+MO+Cl	1.253(19) 1.451(21) 1.691(19)
	exp.			1.61(4)
FeO	Single det	B3LYP opt	Jas Jas+MO	3.688(20) 3.826(20)
	CAS	CAS CAS opt	Jas Jas+Cl Jas+MO+Cl	3.266(20) 3.761(20) 4.112(20)
	exp.			4.18(1)

Results: dissociation energies of transition metal dimers

Compound	Ansatz	Orbitals	Optimization level	D_0
FeH	Single det	B3LYP opt	Jas Jas+MO	1.020(17) 1.020(17)
	CAS	CAS CAS opt	Jas Jas+CI Jas+MO+CI	1.099(17) 1.369(17) 1.791(17)
	exp.			1.63(8)
FeS	Single det	opt	Jas+MO	2.914(15)
	CAS	opt	Jas+MO+CI	3.159(15)
	exp.			3.31(15)

memory bottleneck:

- quadratic memory for all 3 methods (e.g. H,S)
- bottleneck for thousands of parameters
- way out: blocked linear method (Neuscamman, 2017)
- or: descent methods (linear memory), ADAM, AMSGrad