QMCPACK Training Program

Argonne National Laboratory

July 15, 2014

Quantum Monte Carlo Methods

Wave-function Optimization

Miguel A. Morales
Lawrence Livermore National Laboratory

Lawrence Livermore National Laboratory

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Outline

- Basics
- Cost Functions
- Optimization Algorithms
  - Algorithms
  - Estimators
- QMCPACK specifics

VMC Review

VMC: Calculate matrix elements with a given wave-function using MC integration

Energy

$$E_V[\Psi_T,\{a_i\}] = \frac{\langle \Psi_T | \hat{H} | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} = \frac{\int d\vec{r}\ \Psi_T^* \hat{H} \Psi_T}{\int d\vec{r}\ \Psi_T^* \Psi_T} = \int d\vec{r}\ \pi(\vec{r}) E_L(\vec{r})$$

Variance

$$\sigma^2 = \frac{\langle \Psi_T | (\hat{H} - E_V)^2 | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} = \langle \hat{H}^2 \rangle - E_V^2$$

VMC Distribution

$$\pi(\vec{r}) = \frac{|\Psi_T(\vec{r})|^2}{\int d\vec{r}\,' |\Psi_T|^2}$$

Local Energy

$$E_L(\vec{r}) = \frac{\langle \vec{r} | \hat{H} | \Psi_T \rangle}{\langle \vec{r} | \Psi_T \rangle}$$

- Variational Principle: Rigorous upper bound to the energy (= only for exact wfn)

$$E_V[\Psi_T] >= E_{exact}$$

- Zero-Variance Principle: (Only!!!) Exact solution has zero variance

$$E_V \xrightarrow[\sigma^2 \to 0]{} E_{exact}$$

Wave-function Optimization

- Wave-function optimization used to be a complicated and time consuming task. Recent developments led to robust and efficient methods, e.g. linear method.
- QMC can use complicated wave-functions!
  - Include terms which depend on inter-electron distances, three-body terms, etc.
    - Very hard in deterministic methods, leads to high-dimensional integrals!
- Typical strategy:
  - Add variational freedom to the trial wave-function.
  - Optimize with VMC.
  - If result is not accurate enough, increase flexibility of your trial wave-function.

Type of variational parameters

Standard wave-function form: $\Psi_T(\vec{r}) = A(\vec{r})e^{J(\vec{r})}$

Anti-symmetric term

- Typically a linear combination of Slater Determinants of single particle orbitals
- Anti-symmetric under particle exchange
- Variational parameters can include:
  - Linear coefficients
  - Parameters in the single particle orbitals
  - Parameters in the basis set
    - gaussian exponents

Multi-Particle Correlation

- Best known as the Jastrow
- Symmetric under particle exchange
  - Does not affect the nodes!
  - In principle, only affects efficiency in DMC! (Very important nonetheless)

$$J(\vec{r}) = \sum_i u^{e-I}(\vec{r}_i) + \sum_{i>j} u^{e-e}(\vec{r}_i, \vec{r}_j) + \ldots$$

- Many forms of variational parameters!

More on trial wave-functions tomorrow!

Cost Functions – What to optimize?

- 2 main alternatives: **Energy** and **variance** minimization
  - For a given wave-function form, both conditions can lead to different wave-functions.
    - Parameters that lead to energy minimum do not lead to variance minimum.
  - Energy minimization is typically preferred, but it is more challenging.
  - Variance minimization is usually faster and more stable.
  - Other conditions can be used, e.g. maximize overlap with exact wave-function, minimize mean absolute difference of local energy, etc.
- Most codes (including QMCPACK) can optimize a mixture of energy and variance.

$$Cost = \alpha E + \beta\sigma^2$$

Optimization Methods

Gradient and hessian of the cost function

$$g_i^t = \left(\frac{\partial C(\vec{a})}{\partial a_i} \right) \Bigg|_{\vec{a}=\vec{a}_t} \qquad h_{ij}(\vec{a}_t) = \left(\frac{\partial^2 C(\vec{a})}{\partial a_i \partial a_j} \right) \Bigg|_{\vec{a}=\vec{a}_t}$$

Derivatives of the wave-function

$$\Psi_i(\vec{a}) = \left(\frac{\partial \Psi(\vec{a})}{\partial a_i} \right) \Bigg|_{\vec{a}=\vec{a}_t} \qquad \Psi_{ij}(\vec{a}) = \left(\frac{\partial^2 \Psi(\vec{a})}{\partial a_i \partial a_j} \right) \Bigg|_{\vec{a}=\vec{a}_t}$$

Steepest Descents

$$\vec{a}^{t+1} = \vec{a}^t - \gamma \vec{g}^t$$

- Will converge to the minimum, but slowly.
- Different type of parameters vary on very different length scales
  - linear parameters in the determinant expansion vs parameters in the exponent
- Small values of γ are typically required for smooth convergence.
- No need for hessians!

Optimization Methods

Newton's Method

$$C(\vec{a}) = C(\vec{a}_0) + \vec{g}^{\,t} \bullet \Delta\vec{a} + \frac{1}{2}\Delta\vec{a} \bullet \overleftrightarrow{h} \bullet \Delta\vec{a}$$

$$\vec{a}^{\,t+1} = \vec{a}^{\,t} - \overleftrightarrow{h} \bullet \vec{g}^{\,t}$$

- Efficient method, typically requires less iterations than SD.
- Needs second derivatives!
  - These can be hard to implement, many potential cross terms!

Linear Method

1. Expand the wave-function to first order around the current set of parameters

$$\overline{\Psi}_{lin}(\vec{a}) = \Psi_0 + \sum_{i=1}^{N_{opt}} \Delta a_i \overline{\Psi}_i$$

Orthogonalized Derivatives

$$\overline{\Psi}(\vec{a}) = \frac{\Psi(\vec{a})}{\sqrt{\langle \Psi(\vec{a}) | \Psi(\vec{a}) \rangle}}$$

$$\overline{\Psi}_i = \left(\frac{\partial \overline{\Psi}(\vec{a})}{\partial a_i} \right)\Bigg|_{\vec{a}=\vec{a}_0} = \Psi_i - \langle \Psi_0 | \Psi_i \rangle \Psi_0 = \Psi_i - S_{0,i}\Psi_0$$

2. Minimization of the linear energy leads to a generalized eigenvalue problem:

$$\min E_{lin}(\vec{a}) = \min \frac{\langle \overline{\Psi}_{\min}(\vec{a}) | \hat{H} | \overline{\Psi}_{\min}(\vec{a}) \rangle}{\langle \overline{\Psi}_{\min}(\vec{a}) | \overline{\Psi}_{\min}(\vec{a}) \rangle}$$

$$\overline{H} \bullet \Delta\vec{a} = E_{lin} \overline{S} \bullet \Delta\vec{a}$$

$$\overline{H}_{i,j} = \langle \overline{\Psi}_i | \hat{H} | \overline{\Psi}_j \rangle$$

$$\overline{S}_{i,j} = \langle \overline{\Psi}_i | \overline{\Psi}_j \rangle$$

Hamiltonian and overlap matrix over N_{opt}+1 basis functions $\{\Psi_0, \overline{\Psi}_1, \ldots, \overline{\Psi}_{N_{opt}}\}$

J. Toulouse and C. Umrigar, "*Optimization of quantum Monte Carlo wave functions by energy minimization* ",J. Chem. Phys. **126**, 084102 (2007).

Gradients and Hessian of the VMC Energy

$$\frac{dE}{da_i} = \frac{d}{da_i}\left[\frac{\int d\vec{R}\,\pi(\vec{R})E_L(\vec{R})}{\int d\vec{R}\,\pi(\vec{R})}\right] = 2\left\langle \frac{\Psi_i}{\Psi}\left(E_L - \langle E_L\rangle\right)\right\rangle$$

$$\frac{d^2E}{da_i da_j} = 2\left\langle\left(\frac{\Psi_{ij}}{\Psi} - \frac{\Psi_i}{\Psi}\frac{\Psi_j}{\Psi}\right)\left(E_L - \langle E_L\rangle\right)\right\rangle + 4\left\langle\left(\frac{\Psi_i}{\Psi} - \left\langle\frac{\Psi_i}{\Psi}\right\rangle\right)\left(\frac{\Psi_j}{\Psi} - \left\langle\frac{\Psi_j}{\Psi}\right\rangle\right)\left(E_L - \langle E_L\rangle\right)\right\rangle$$
$$+\left\langle\frac{\Psi_i}{\Psi}\left(E_{L,j} - \langle E_{L,j}\rangle\right) + \frac{\Psi_j}{\Psi}\left(E_{L,i} - \langle E_{L,i}\rangle\right)\right\rangle$$

Gradient of the local energy

$$E_{L,i}(\vec{R}) = \frac{\partial E_L(\vec{R})}{\partial a_i}$$

Notice that $\left\langle E_{L,i}(\vec{R})\right\rangle = 0$

All quantities necessary to construct the hessian and gradients can be easily calculated from a standard VMC calculation.

J. Toulouse and C. Umrigar, "*Optimization of quantum Monte Carlo wave functions by energy minimization* ",J. Chem. Phys. **126**, 084102 (2007).

Hamiltonian and Overlap Matrices

$$\bar{S}_{00} = 1 \qquad \bar{S}_{i0} = \bar{S}_{0j} = 0$$

$$\bar{S}_{i,j} = \left\langle \frac{\Psi_i}{\Psi}\frac{\Psi_j}{\Psi} \right\rangle - \left\langle \frac{\Psi_i}{\Psi} \right\rangle \left\langle \frac{\Psi_j}{\Psi} \right\rangle$$

$$\bar{H}_{0,0} = \langle E_L \rangle$$

$$\bar{H}_{0,j} = \left\langle \frac{\Psi_j}{\Psi}\left(E_L - \langle E_L \rangle\right) \right\rangle + \langle E_{L,j} \rangle$$

$$\bar{H}_{i,0} = \left\langle \frac{\Psi_i}{\Psi}\left(E_L - \langle E_L \rangle\right) \right\rangle$$

$$\bar{H}_{i,j} = \left\langle \frac{\Psi_i}{\Psi}\frac{\Psi_j}{\Psi}E_L \right\rangle - \left\langle \frac{\Psi_i}{\Psi} \right\rangle \left\langle \frac{\Psi_j}{\Psi}E_L \right\rangle - \left\langle \frac{\Psi_j}{\Psi} \right\rangle \left\langle \frac{\Psi_i}{\Psi}E_L \right\rangle + \left\langle \frac{\Psi_i}{\Psi} \right\rangle \left\langle \frac{\Psi_j}{\Psi} \right\rangle \langle E_L \rangle - \left\langle \frac{\Psi_i}{\Psi}\left(E_{L,j} - \langle E_{L,j} \rangle\right) \right\rangle$$

Non-symmetric H matrix leads to lower variance!

J. Toulouse and C. Umrigar, "*Optimization of quantum Monte Carlo wave functions by energy minimization* ",J. Chem. Phys. **126**, 084102 (2007).

Solving 1-D Minimization Problem

- Due to the strong non-linearity of many parameter types, it is usually necessary to perform a 1-D minimization along the direction produced by the previous methods. Some options:
  - Use a fixed (and small) time-step.
    - Leads to very slow convergence.
  - Evaluate the cost function at various points along the chosen direction and fit a polynomial
    - Cost function can be evaluated with a short VMC calculation, with correlated sampling or using reweighting.
  - Perform a direct line-minimization using reweighting.
    - Careful with wave-function overlap (reweighting efficiency) when parameters get far.
- C. Umrigar, et al. suggest various rescaling procedures to obtain reasonable step lengths, thus avoiding the line minimization step.
  - C.Umrigar, et al., "*Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions*" PRL 98, 110201 (2007).

Reweighting

Monte Carlo Integration

$$Q = \int f(x)\, g(x)\, dx \;=\; \frac{1}{N}\sum_i f(x_i) + O\left[\sqrt{\frac{\sigma_f^2}{N}}\right]$$

The points in the set $\{x_i\}$ are distributed according to g(x).

What happens when we want to perform the integral with respect to a different distribution g'(x), but we already have a set of points sampled from g(x).
Can we estimate the new integral using the old set? **YES, use reweighting!!!**

$$Q' = \int f(x)\, g'(x)\, dx \;= \int \left[f(x)\frac{g'(x)}{g(x)} \right] g(x)\, dx \;= \frac{1}{N}\sum_i \left[f(x_i)\frac{g'(x_i)}{g(x_i)} \right] + O\left[\sqrt{\frac{\sigma^2_{f\frac{g'}{g}}}{N}}\right]$$

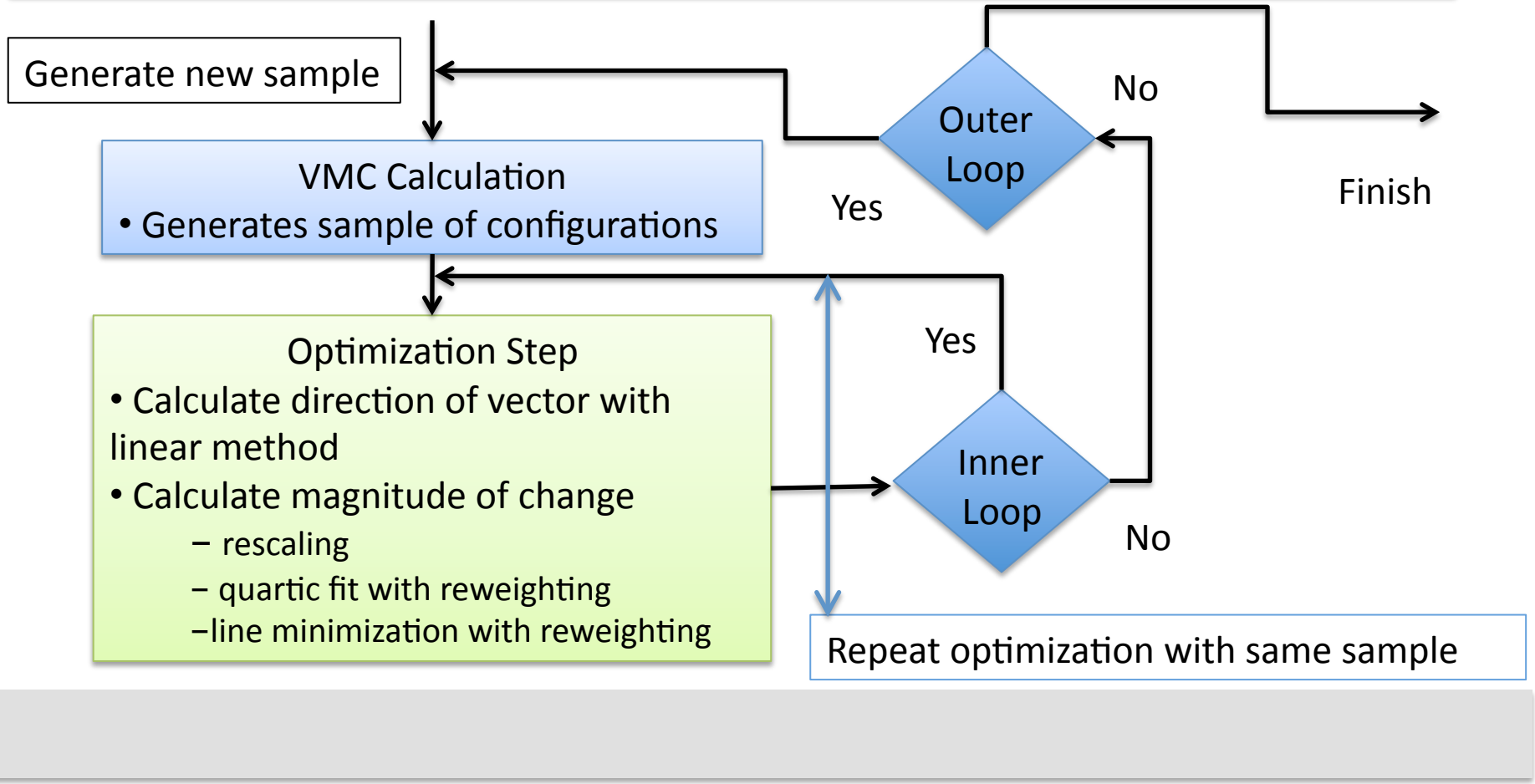
The efficiency of the procedure will depend on how close g and g' are.

Effective number of samples: $\dfrac{N_{eff}}{N} = \dfrac{< w >^2}{< w^2 >}$, where $\mathrm{w(x)} = \dfrac{\mathrm{g'(x)}}{\mathrm{g(x)}}$

QMCPACK Specifics - I

The linear method is the recommended optimization algorithm in QMCPACK.

- Robust: Can handle all type of parameters efficiently.
- Fast and simple: Only first derivatives of the wave-function and local energy are needed.



QMCPACK Specifics – Sample XML Block

Outer loop: Redo optimization with new sample

```
<loop max="10">
 <qmc method="linear" move="pbyp" checkpoint="-1" gpu="no">
   <parameter name="blocks">     10  </parameter>
   <parameter name="warmupsteps"> 25 </parameter>
   <parameter name="steps"> 1 </parameter>
   <parameter name="substeps"> 20 </parameter>
   <parameter name="timestep"> 0.5 </parameter>
   <parameter name="samples"> 10240  </parameter>
   <cost name="energy">                    0.95 </cost>
   <cost name="unreweightedvariance">      0.0 </cost>
   <cost name="reweightedvariance">        0.05 </cost>
   <parameter name="useDrift">  yes </parameter>
   <parameter name="bigchange">10.0</parameter>
   <estimator name="LocalEnergy" hdf5="no"/>
   <parameter name="usebuffer"> yes </parameter>
   <parameter name="nonlocalpp"> yes </parameter>
   <parameter name="MinMethod">quartic</parameter>
   <parameter name="exp0">-6</parameter>
   <parameter name="alloweddifference"> 1.0e-5 </parameter>
   <parameter name="stepsize">  0.15 </parameter>
   <parameter name="nstabilizers"> 1 </parameter>
 </qmc>
</loop>
```

VMC parameters

Cost function

Optimization parameters

QMCPACK Specifics

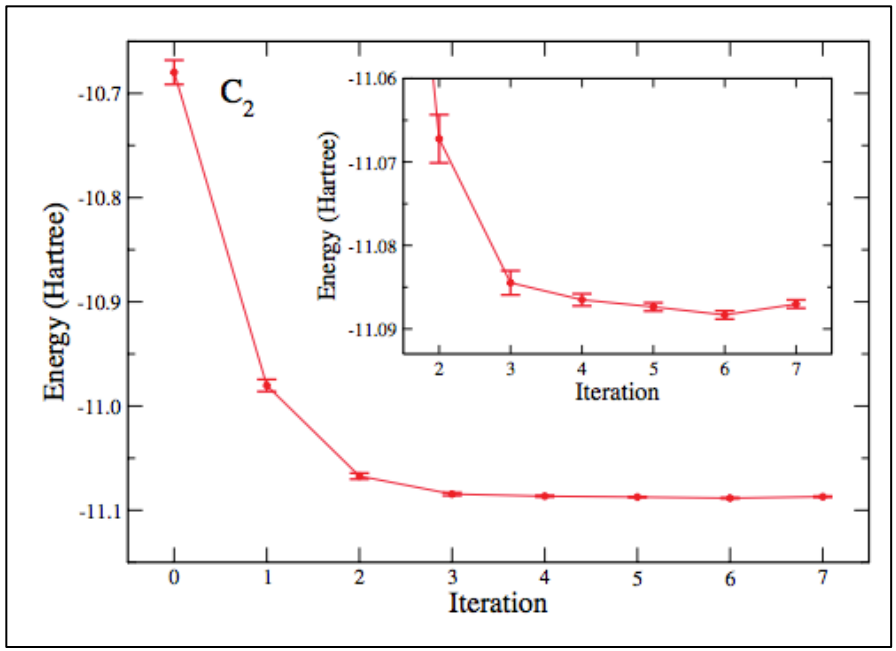
- Important parameters:
  - ***bigchange***: (default 50.0) largest parameter change allowed
  - ***usebuffer***: (default "no") Save useful information during VMC
  - ***nonlocalpp***: (default "no") Include non-local energy on 1-D min
  - ***MinMethod***: (default "quartic") Method to calculate magnitude of parameter change
    - quartic: fit quartic polynomial to 4 values of the cost function obtained using reweighting along chosen direction
    - linemin: direct line minimization using reweighting
    - rescale: no 1-D minimization. Uses Umrigar's suggestions.
  - ***stepsize***: (default 0.25) step size in either quartic or linemin methods.
  - ***alloweddifference***: (default 1e-4) Allowed increased in energy
  - ***exp0***: (default -16.0) Initial value for stabilizer (shift to diagonal of H)
    - Actual value of stabilizer is 10^exp0
  - ***nstabilizers***: (default 3) Number of stabilizers to try
  - ***stabilizaterScale***: (default 2.0) Increase in value of exp0 between iterations.
  - ***max_its***: (default 1) number of inner loops with same sample

QMCPACK Specifics

- Important parameters:
  - ***minwalkers***: (default 0.3) minimum value allowed for the ratio of effective samples to actual number of walkers in a reweighting step. The optimization will stop if the effective number of walkers in any reweighting calculation drops below this value. Last set of acceptable parameters are kept.
  - **maxWeight:** (defaul 1e6) Maximum weight allowed in reweighting. Any weight above this value will be reset to this value.

- Recommendations:
  - Set ***samples*** to equal to (#***threads***)*****blocks***.
  - Set ***steps*** to 1. Use ***substeps*** to control correlation between samples.
  - For cases where equilibration is slow, increase both **substeps** and **warmupsteps**.
  - For hard cases (e.g. simultaneous optimization of long MSD and 3-Body J), set ***exp0*** to 0 and do a single inner iteration (***max_its***=1) per sample of configurations.

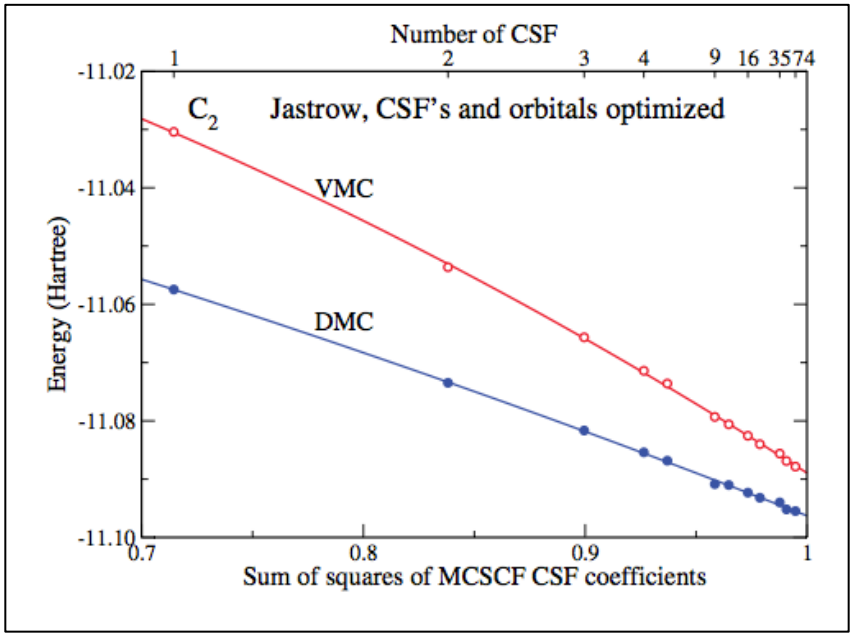
Examples from literature



Simultaneous optimization of Jastrow, CSF and orbital parameters.

- Robust and automatic
- Little need for human intervention

- Convergence to mHa accuracy with less than 10 iterations
- Can recover energy in large MSD expansions!



C.Umrigar, et al., "*Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions*" PRL **98**, 110201 (2007).

Important References

- X. Lin, et al., "*Optimization of quantum Monte Carlo wave functions using analytical energy derivatives*", J. Chem. Phys. **112**, 2650 (2000).
- C. Umrigar and C. Filippi, "*Energy and Variance Optimization of Many-Body Wave Functions",* PRL **94**, 150201 (2005).
- S. Sorella, "Wave function optimization in the variational Monte Carlo method", PRB **71**, 241103 (R) (2005).
- A. Scemama, et al., "*Simple and efficient approach to the optimization of correlated wave functions*", J. Chem. Phys. **73**, 241101 (R) (2006).
-J. Toulouse and C. Umrigar, "*Optimization of quantum Monte Carlo wave functions by energy minimization* ",J. Chem. Phys. **126**, 084102 (2007).
- C.Umrigar, et al., "*Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions*" PRL **98**, 110201 (2007).
- J. Toulouse and C. Umrigar, "*Full optimization of Slater-Jastrow wavefunctions with application to the first-row atoms and homonuclear dimers*", J. Chem. Phys. **128**, 174101 (2008).