Optimization of quantum Monte Carlo wave functions by energy minimization

Julien Toulouse, Cyrus J. Umrigar

Cornell Theory Center, Cornell University, Ithaca, New York, USA.
Email: toulouse@tc.cornell.edu

Web page: www.lct.jussieu.fr/pagesperso/toulouse/

January 2007
To fully benefit of the considerable flexibility in the form of the many-body wave functions used in QMC, it is crucial to be able to efficiently optimize their parameters.

So far, variance minimization in correlated sampling [1] has been the most frequently used optimization method because it is far more efficient than straightforward energy minimization.

Today, clever energy minimization techniques [2-10] are however superseding variance minimization.

**We study here three such efficient wave function optimization methods based on energy minimization in VMC : the Newton, Linear and Perturbative methods.**
We use standard Jastrow-Slater wave functions:

$$\Psi(p) = J(\alpha) \sum_{i=1}^{N_{CSF}} c_i C_i(\lambda)$$

where $J(\alpha)$ is a Jastrow factor and $C_i(\lambda)$ is a configuration state function (CSF), itself consisting of a symmetry-adapted linear combination of Slater determinants.

The Slater determinants are constructed from orbitals expanded in a (localized) one-electron basis:

$$\phi_{\mu}(r) = \sum_{\nu=1}^{N_{basis}} \lambda_{\mu\nu} \chi_{\nu}(r)$$

The parameters $p$ to be optimized are the Jastrow parameters $\alpha$, the CSF coefficients $c$ and the orbital coefficients $\lambda$. (For the orbitals, in practice, we optimize a set of non-redundant orbital rotation parameters.)
Newton optimization method (Refs. [2, 6, 7, 10])

- Expansion of the energy around $p^0$ to second order in the parameter variations $\Delta p = p - p^0$:

  \[
  E^{[2]}(p) = E_0 + \sum_i g_i \Delta p_i + \frac{1}{2} \sum_{i,j} h_{ij} \Delta p_i \Delta p_j
  \]

  where $E_0 = E(p^0)$, $g_i = \partial E(p^0) / \partial p_i$ and $h_{ij} = \partial^2 E(p^0) / \partial p_i \partial p_j$.

- Minimization of the energy gives parameter variations:

  \[
  \min_p E^{[2]}(p) \implies \Delta p = -h^{-1} \cdot g
  \]

- Monte Carlo estimators with low statistical fluctuations for gradient $g$ and Hessian $h$.

- Stabilization: $h_{ij} \rightarrow h_{ij} + a \delta_{ij}$ where $a \geq 0$. 
Expansion of the wave function around $p^0$ to first order in $\Delta p$:

$$\psi^{[1]}(p) = \psi_0 + \sum_i \Delta p_i \psi_i$$

where $\psi_0 = \psi(p^0)$ and $\psi_i = \partial \psi(p^0)/\partial p_i$.

Normalization of wave function chosen so that the derivatives $\psi_i$ are orthogonal to $\psi_0$.

Minimization of the energy $\rightarrow$ generalized eigenvalue equation:

$$\min_p \frac{\langle \psi^{[1]}(p)|\hat{H}|\psi^{[1]}(p) \rangle}{\langle \psi^{[1]}(p)|\psi^{[1]}(p) \rangle} \Rightarrow H \cdot \Delta p = E S \cdot \Delta p$$

where $H_{ij} = \langle \psi_i|\hat{H}|\psi_j \rangle$ and $S_{ij} = \langle \psi_i|\psi_j \rangle$.

Monte Carlo estimators for overlap matrix $S_{ij}$ and non-symmetric Hamiltonian matrix $H_{ij} \rightarrow$ strong zero-variance principle

**Stabilization**: $H_{ij} \rightarrow H_{ij} + a \delta_{ij} (1 - \delta_{i0})$ where $a \geq 0$. 

Linear optimization method (Refs. [4, 9, 10])
Perturbative optimization method (Refs. [3, 8, 9])

- Generalized eigenvalue equation of the linear method solved approximately by a **non-orthogonal first-order perturbation theory**:

\[
\Delta p_i^{(1)} = -\frac{1}{E_i - E_0} \sum_{j \geq 1} (S^{-1})_{ij} H_{j0}
\]

where \( E_i = H_{ii} / S_{ii} \).

- Energy denominators \( E_i - E_0 \) do not need to be calculated in QMC at each step but can be approximated.

- Monte Carlo evaluation of only overlap matrix \( S_{ij} \) and first row of Hamiltonian \( H_{j0} (=\) gradient) \( \Rightarrow \) **low computational cost**.

- **Stabilization**: \( E_i - E_0 \rightarrow E_i - E_0 + a \) where \( a \geq 0 \).
Optimization of the Jastrow factor

For $C_2$ molecule (24 parameters):

$\begin{align*}
-75.8 \\
-75.7 \\
-75.6 \\
-75.5 \\
-75.4 \\
-75.3 \\
-75.2 \\
\end{align*}$

$\begin{array}{cccccccc}
0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\hline
\text{Newton method (UF Hessian)} \\
\text{Newton method (TU Hessian)} \\
\text{Linear method} \\
\text{Perturbative method} \\
\end{array}$

$\Rightarrow$ Newton method and Linear method are very efficient.
Optimization of CSF coefficients

For $C_2$ molecule (49 parameters):

\[
\begin{align*}
-75.855 \\
-75.854 \\
-75.853 \\
-75.852 \\
-75.851 \\
-75.850 \\
-75.849 \\
-75.848 \\
-75.847 \\
-75.846 \\
-75.845 \\
\end{align*}
\]

\textbf{Newton method} \quad \textbf{Linear method} \quad \textbf{Perturbative method}

\[\Rightarrow\] \textbf{Linear method} converges in one iteration
Optimization of the orbitals

For $C_2$ molecule (44 parameters):

$$
\begin{align*}
-75.786 \\
-75.784 \\
-75.782 \\
-75.780 \\
-75.778 \\
-75.776 \\
-75.774 \\
-75.772 \\
-75.770 \\
-75.768 \\
-75.766 \\
-75.764
\end{align*}
$$

Newton method  
Linear method  
Perturbative method

$\Rightarrow$ The 3 methods are very efficient
Simultaneous optimization of all the parameters

For $\text{C}_2$ molecule (137 parameters) using the **Linear method**:

The energy converges with an accuracy of about $10^{-3}$ Hartree in 4 iterations.
Systematic improvement in quantum Monte Carlo

For C$_2$ molecule: VMC and DMC total energies for several wave functions of increasing “sizes” (Jastrow, CSF coefficients and orbitals optimized):

-75.94
-75.92
-75.90
-75.88
-75.86
-75.84
-75.82
-75.80

J*RAS(8,26)
J*CAS(8,8)
J*CAS(8,7)
J*CAS(8,5)
J*HF

Energy (Hartree)

Wave function

⇒ Systematic improvement in VMC and DMC!
Conclusions

Summary

- **Efficient wave function optimization methods by energy minimization** in VMC are now available.
- Achievement of **systematic improvement** of the total energy in VMC and DMC as the number of variational parameter increases thanks to proper optimization.

Perspectives

- Optimization of the exponents of the basis functions.
- Development of a hybrid optimization method taking the best of each method.
- Direct minimization of the DMC energy.
- Geometry optimization.