Direct Minimization for Ensemble Electronic Structure Calculations

K. Baarman · V. Havu · T. Eirola

Received: 3 May 2012 / Revised: 3 June 2014 / Accepted: 16 June 2015 / Published online: 17 July 2015
© Springer Science+Business Media New York 2015

Abstract We consider a direct optimization approach for ensemble density functional theory electronic structure calculations. The update operator for the electronic orbitals takes the structure of the Stiefel manifold into account and we present an optimization scheme for the occupation numbers that ensures that the constraints remain satisfied. We also compare sequential and simultaneous quasi-Newton and nonlinear conjugate gradient optimization procedures, and demonstrate that simultaneous optimization of the electronic orbitals and occupation numbers improve performance compared to the sequential approach.

Keywords Quasi-Newton method · Nonlinear conjugate gradient · Ensemble optimization · Electronic structure · Stiefel manifold

1 Introduction

Advances in computer power and numerical methods during the past few decades has dramatically increased the scope of electronic structure problems that can be computationally studied. Kohn–Sham density functional theory (DFT) methods can be used to describe electronic properties of materials with high precision. Computationally most challenging system for DFT calculations are metallic systems. They lack a gap between occupied and unoccupied electronic states in the energy spectrum, which leads to slower convergence compared to insulators and semiconductors. Smearing of the Fermi surface is often used to enable convergence of metallic systems as well as insulators at positive temperatures. Ensemble DFT permits direct computation of the occupation numbers of the orbitals based on the entropic term in the Helmholtz free energy. We consider an optimization problem where the target
functional $A$ corresponds to the Helmholtz free energy and the variables $X$ and $f$ to the electronic orbitals and occupation numbers respectively.

The optimization problem is therefore minimize $A(X, f)$, where

$$A(X, f) = E(X, f) - T S(f),$$

and $E$ is the internal energy of the system, $T$ is the (electronic) temperature, and $S$ is the entropy. In this minimization problem we have the constraints

$$X \in \mathcal{M} = \{X \in \mathbb{R}^{m \times n} | X^T X = I\}.$$

$$f \in Q = \{f \in [0, 1]^n | \sum_{i=1}^{n} f_i = n_e\},$$

where $n_e$ is the number of electrons in the system so that $n_e \leq n$. The usual situation is that $\nabla_X A(X, f)$ and $\nabla_f A(X, f)$ are available, but expensive to compute. However, due to the form of $A(X, f)$ the price to compute $A$, $\nabla_X A$, and $\nabla_f A$ simultaneously is comparable to computing one of them separately. Furthermore we assume that $m \gg n$, and usually $m$ is too large to make storage of and operation with full $m \times m$ matrices prohibitively expensive.

The orthogonality constraint on $X$ means that $\mathcal{M} \subset \mathbb{R}^{m \times n}$ is a Stiefel manifold, and at $X$ we have the tangent space

$$T_X \mathcal{M} = \{Y = XB + Z | B^T = -B \text{ and } Z^T X = 0\},$$

where $Y, Z \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times n}$. We use the inner product for matrices

$$(X, Y) = \text{trace}(X^T Y).$$

Given an arbitrary matrix $V \in \mathbb{R}^{m \times n}$ its orthogonal projection onto $T_X \mathcal{M}$ is

$$Y = P_X(V) = (I - \frac{1}{2}X^T X) V - \frac{1}{2}XV^T X.$$

Minimization approaches to non-temperature dependent DFT do not in general permit fractional occupation of electronic orbitals [4,11,19–21,23]. In contrast, explicit minimization with regards to occupation numbers permits fractional occupation based on the entropy functional of the Helmholtz free energy and can improve convergence, especially for metallic systems [6,9,13,18]. It is also possible to transform the minimization of (1) into a nonlinear eigenvalue problem that can be solved through a self consistent field iteration. In this case several linear eigenvalue problems of the size $m \times m$ needs to be solved and it is not possible to optimize the objective function with respect to the occupation numbers [11,12,17,19]. The absence of well separated occupied and unoccupied orbitals make metallic systems challenging to compute, and broadening of the Fermi surface is used to facilitate convergence [5,22]. This broadening is often achieved by assigning the orbitals close to the Fermi level a fractional occupation number determined by the energy of the electronic orbital [10,14,15]. Direct minimization on the other hand does not require the orbital energies to be computed at every step, and these broadening schemes are therefore not well suited for minimization methods.

In Sect. 2 we first review the nonlinear conjugate gradient and the quasi-Newton methods adapted for the use on the Stiefel manifold. We then present an optimization procedure for the occupation numbers and end the section by presenting a simultaneous orbital-occupation optimization strategies. Then, in Sect. 3 we numerically demonstrate the method on a model problem that includes nonlinearities similar to a DFT problem. The conclusions are finally presented in Sect. 4.
The main contributions of this paper are the use of quasi-Newton and nonlinear conjugate gradient methods on a manifold and simple, effective stepsize selection strategies for the problem class arising from the electronic structure theory.

2 Optimization with Orthogonality Constraints

2.1 Update and Transport

At a given point \((X_k, f_k) \in \mathcal{M} \times Q\) we look for a next point \((X_{k+1}, f_{k+1})\) such that the value of \(A\) gets reduced. For this we need search directions which will be denoted \(Y_k \in T_{X_k} \mathcal{M}\) and \(y_k\) pointing in \(Q\) at \(f_k\).

We will consider several possibilities for choosing these search directions and step lengths. We also need operators that guarantee that our next point is still in \(\mathcal{M} \times Q\).

We ensure that \(X_{k+1}\) satisfies the orthogonality constraint by using a unitary update operator \(U\) which maps \(\mathcal{M} \rightarrow \mathcal{M}\), see [1]. A search direction \(Y_k \in T_{X_k} \mathcal{M}\) given by an optimization procedure can be written as

\[
Y_k = X_k B + QR, \tag{7}
\]

where \(Q \in \mathbb{R}^{m \times n}\), \(B \in \mathbb{R}^{n \times n}\), \(B^T = -B\), \(Q^T Q = I\), and \(Q^T X = 0\).

If we follow \(Y_k\) to update \(X_k\) along a Stiefel geodesic we obtain the update operator [7]

\[
U(\tau) = [X_k Q] \exp \left( \tau \begin{bmatrix} B & -R^T \\ R & 0 \end{bmatrix} \right) [X_k Q]^T, \tag{8}
\]

with step length parameter \(\tau\). Computation of the exponent has complexity \(O(n^3)\). It is not recommended to compute the full \(\mathbb{R}^{m \times m}\) product of these matrices, but just store matrices \(Q\) and the exponential and then use these for operations on matrices in \(\mathbb{R}^{m \times n}\). Such an operation has then complexity \(O(mn^2)\). The same applies to (9) below.

In order to use information gained from previous evaluations of \(A\) and \(\nabla A\) we must take \(\mathcal{M}\) into account. This requires us to transport vectors \(V \in T_{X_k} \mathcal{M}\) to \(T_{U(\tau)X_k} \mathcal{M}\) with the transport operator

\[
W(\tau) = I_m + [X_k Q] \left( \exp \left( \tau \begin{bmatrix} B & -R^T \\ R & 0 \end{bmatrix} \right) - I_{2n} \right) [X_k Q]^T. \tag{9}
\]

Here the subindices of \(I\) denote the dimensions. Notice that \(W\) does not modify matrices \(Z\) that satisfy \([X_k Q]^T Z = 0\).

2.2 Nonlinear Conjugate Gradients

The conjugate gradient (CG) method can be viewed as an optimization method for a quadratic problem. Several generalizations of the CG method have been presented to solve optimization problems that are not quadratic [16]. Below, we review a nonlinear CG (NLCG) method adapted to account for the curvature of the manifold [7].

Given \(X_0\) which satisfies \(X_0^T X_0 = I\), the gradient projected onto \(T_{X_0} \mathcal{M}\) is

\[
F_0 = P_{X_0}(\nabla_X A(X_0, f_0)), \tag{10}
\]

\footnote{For given \(V \in \mathbb{R}^{m \times n}\), \(X \in \mathcal{M}\), the orthogonal projection onto \(T_X \mathcal{M}\) is given by (6). Then, writing \(V = X C + Z\) with \(C = X^T V\), \(Z = V - X C\) we get (7) with \(B = \frac{1}{2} (C - C^T)\) and a QR-decomposition \(Z = QR\). These computations have complexity \(O(mn^2)\).}
and the initial search direction is the direction of steepest descent
\[ Y_0 = -F_0. \] (11)

On the manifold the NLCG method then proceeds by minimizing \( A \) along the path defined by the search direction \( Y_k \). In practice we evaluate \( A \) once along the search direction and construct a quadratic approximation that we minimize. The step length, \( \tau_k \), that minimizes the quadratic approximation along is then used to update \( X_k \) such that
\[ X_{k+1} = U(\tau_k)X_k, \] (12)
and the gradient and search directions are transported to \( TX_kM \) by \( W(\tau_k) \). The new projected gradient
\[ F_{k+1} = P_{X_{k+1}}(\nabla_X A(X_{k+1}, f_{k+1})), \] (13)
and search direction
\[ Y_{k+1} = -F_{k+1} + \gamma_k W(\tau_k)Y_k, \] (14)
are then computed, where
\[ \gamma_k = \frac{(F_{k+1} - W(\tau_k)F_k, F_{k+1})}{(F_k, F_k)}. \] (15)
This is directly analogical to the usual conjugate gradient recursion.

The step length is determined by the minimizer of a quadratic approximation of \( A \) along the search direction. This is constructed by taking a trial step length \( \tau_e = \frac{1}{10} \max(\tau_{\min}, \tau_{k-1}) \), where \( \tau_{\min} \) is a predefined minimum trial step length and computing
\[ p(0) = A(X_k, f_k), \]
\[ p(\tau_e) = A(U(\tau_e)X_k, f_k), \]
\[ p'(0) = (Y_k, \nabla_X A(X_k, f_k)). \] (16)
Then \( \tau_k \) is solved by minimizing \( p \), limited by \( 2\tau_{k-1} \), and the update \( U(\tau_k)X_k \) is constructed. This approximate line search requires one extra evaluation of \( A \) per step, which is the minimum.

### 2.3 Quasi-Newton Method

The quasi-Newton (QN) method is similar to Newton’s method, but replaces the inverse Hessian with an approximation. This is often handy when the Hessian is too expensive to compute, but it can still be used to improve performance for a badly conditioned minimization problem.

We base the QN method on Broyden’s second or bad generalized update to construct an approximation \( G_k \) to the inverse of the Hessian of \( A \) at \( X_k \). While Broyden’s second update does not construct a symmetric approximation, or ensure that the approximation is positive definite it is a robust update choice for electronic structure calculations \[2, 12\]. Furthermore, \( X_k \) and \( \nabla_X A \) are \( \mathbb{R}^{m \times n} \) matrices, which we take into account when constructing the generalized Broyden update. The secant condition is then
\[ G_k \delta \Phi_k = \delta \Xi_k, \] (17)
where \( \delta \Phi_k \) and \( \delta \Xi_k \) are the collected orbital gradient and position differences projected onto the tangent space and transported to \( TX_kM \). That is
\[ \delta \Phi_k = [\delta F_{k-1} W(\tau_{k-1}) \delta F_{k-2} \ldots W(\tau_{k-l+1}) \delta F_{k-l-1}], \] (18)
and

$$\delta \Xi_k = [\delta X_{k-1} W(\tau_{k-1}) \delta X_{k-2} \cdots W(\tau_{k-\ell}) \cdots W(\tau_{k-l+1}) \delta X_{k-l}]$$

for history of length $l$. Here the gradient differences projected onto $T_{X_{i+1}}$ are

$$\delta F_i = F_{i+1} - W(\tau_i) F_i,$$

and $F_i$ is as in (13),

$$F_i = P X_i (\nabla X A(X_i, f_i)).$$

The projected occupation weighted orbital differences are

$$\delta X_i = P X_{i+1} (X_{i+1} \text{ diag}(f_{i+1}) - X_i \text{ diag}(f_i)).$$

Notice that we can update $\delta \Phi$ of (18) by

$$\delta \Phi_{k+1} = [\delta F_k W(\tau_k) \delta \Phi_k],$$

where $\delta \Phi_k$ is obtained from $\delta \Phi$ by deleting the last $n$ columns. We update the matrices $\delta \Xi_k$ similarly. These operations have complexity $O(lmn^2)$.

The motivation for including the weight is that the unoccupied electronic orbitals do not contribute to the energy of the system. The no change condition is

$$Z = G_k Z \quad \forall Z \text{ such that } Z^T \delta F_k = 0.$$  \hspace{1cm} (23)

The secant and no change conditions together correspond to the generalized Broyden’s second update where all single orbital secant conditions are simultaneously enforced for the entire history. We can therefore use the generalized update formula [8]

$$G_k = \mu I + (\delta \Xi_k - \mu \delta \Phi_k) (\delta \Phi_k^T \delta \Phi_k)^{-1} \delta \Phi_k^T,$$  \hspace{1cm} (24)

where dropping the empty orbitals ensure that $\delta \Phi^T \delta \Phi$ is nonsingular in practice. The search direction given by the QN method is

$$Y_k = -G_k F_k,$$  \hspace{1cm} (25)

and the next point $X_{k+1} \in \mathcal{M}$ is

$$X_{k+1} = U(\tau_k) X_k,$$  \hspace{1cm} (26)

where $Y_k$ determines $U$ as in Sect. 2.1. The line search is identical to the one described for the NLCG method in Sect. 2.2 with the addition of a constant underrelaxation $\beta X \in (0, 1]$ that we have included in the step length $\tau_k$.

In practice only the last few history steps contribute significantly to the rate of convergence. Consequently, we discard the oldest trial solutions and gradient information when a predetermined history length is reached. In case the inner product of the search direction $Y_k$ with the projected gradient $F_k$ is not negative we restart the QN iteration from the current point.

2.4 Optimization of Occupation Numbers

Given a set of electronic orbitals $X_k$ it is possible to further reduce $A$ by optimizing with respect to the occupation numbers $f$. Forcing occupation towards a uniform distribution increases contributions to $A$ from higher energy states, while simultaneously increasing the entropy which contributes to a reduction of $A$ at nonzero temperatures. The relative strength of both of these effects determine the ground state of the system, and can lead to nonzero occupation of higher energy states at positive temperatures.
Therefore, given $X_k$, we want to find $f$ that minimizes $A$. We also have an approximation $f_k$ to the minimizer. To keep the number of electrons constant we determine the search direction $y_k$ which is the vector closest to $-\nabla_f A(X_k, f_k)$ and such that it points into $Q$. To this end we solve
\begin{equation}
\text{minimize } \|y + \nabla_f A(X_k, f_k)\|,
\end{equation}
with the constraints $\sum_{i=1}^n y_i = 0$, $y_i \leq 0$ if $f_i = 1$, and $y_i \geq 0$ if $f_i = 0$. The first constraint on $y$ ensures that the minimization step conserves the number of electrons while the second and third conditions prohibit unphysical occupation numbers. Note that the problem (27) is quite simple, and we use the quadprog routine available in MATLAB to solve it. Given the search direction $y$ we minimize $A$ by constructing a quadratic approximation similar to (16).

After we have solved $y_k$ the occupation step length $\sigma_k$ is determined like in Sect. 2.2 with the addition of the constant underrelaxation $\beta_l \in (0, 1]$ included in $\sigma_k$. In addition, we ensure that the occupation remains physical by limiting $\sigma_k$ with $\sigma_M$ such that $0 \leq f_i + \sigma_M y_i \leq 1$ for all $i$. It is possible to take a longer step than $\sigma_M$ by recomputing $y_k$ from Eq. (27) with the updated boundary information when an entry in $f$ reaches the boundary of physical occupation (0 or 1). However, convergence of occupation numbers is faster than orbital convergence, and the numbers of steps needed for convergence is therefore determined by the orbital convergence. Furthermore, if the occupation number of the least populated orbital has been less than $10^{-12}$ on two consecutive iterations we drop the associated orbitals.

### 2.5 Simultaneous Step Size Selection

Typically an ensemble DFT problem is solved by sequentially optimizing the orbitals with fixed occupation numbers and then fixing the orbitals and optimizing the occupation numbers. This process is then repeated until a satisfactory solution is obtained.

The cost of evaluating $A$, $\nabla_X A$, and $\nabla_f A$ is comparable to evaluating one of them separately, and simultaneous optimization of $A$ with respect to $X$ and $f$ can for this reason potentially reduce computational effort.

Given a pair of search directions $(Y_k, y_k)$ for the orbitals and occupation numbers respectively and starting guesses for step lengths, $\tau_{k-1}$ and $\sigma_{k-1}$ we evaluate $A$ and its gradients with the following trial step lengths
\begin{equation}
\tau_e = \frac{1}{10} \max(\tau_{\text{min}}, \tau_{k-1}) \quad \text{and} \quad \sigma_e = \min(\sigma_M, \frac{1}{10} \max(\sigma_{\text{min}}, \sigma_{k-1})).
\end{equation}

Here $\tau_{\text{min}}$ and $\sigma_{\text{min}}$ are minimum trial step lengths. With this we construct a quadratic surface approximation
\begin{equation}
p(\tau, \sigma) = c_1 \tau^2 + c_2 \sigma^2 + c_3 \tau + c_4 \sigma + c_5,
\end{equation}
that we use to simultaneously update both $X$ and $f$ by evaluation in one trial point. This surface is determined by the system of equations
\begin{equation}
\begin{align*}
p(0, 0) &= A(X_k, f_k), \\
p_\tau(0, 0) &= (Y_k, \nabla_X A(X_k, f_k)), \\
p_\sigma(0, 0) &= (y_k, \nabla_f A(X_k, f_k)), \\
p_\tau(\tau_e, \sigma_e) &= (Y_k, \nabla_X A(U(\tau_e)X_k, f_k + \sigma_e y_k)), \\
p_\sigma(\tau_e, \sigma_e) &= (y_{\tau_e}, \nabla_f A(U(\tau_e)X_k, f_k + \sigma_e y_k)).
\end{align*}
\end{equation}
Solving this system and finding the minimum gives the optimal step lengths \( \hat{\tau}_k \) and \( \hat{\sigma}_k \) for the quadratic approximation of the search directions. For the simultaneous NLCG method the step lengths are then \( \tau_k = \hat{\tau}_k \) and \( \sigma_k = \hat{\sigma}_k \) while the QN method uses \( \tau_k = \beta_X \hat{\tau}_k \) and \( \sigma_k = \beta_f \hat{\sigma}_k \), where \( \beta_X, \beta_f \in (0, 1] \) are constant underrelaxation parameters. We then simultaneously update \( X \) and \( f \) with \( X_{k+1} = U(\tau_k)X_k \) and \( f_{k+1} = f_k + \min(\sigma_M, \sigma_k) y_k \), respectively.

Remark The surface (30) is determined by computing \( \nabla_X A \) and \( \nabla_f A \) at the trial step. The system of Eqs. (31) could alternatively be determined by computing both \( A \) and \( \nabla_X A \) or \( A \) and \( \nabla_f A \) at \((W(\tau_e)X, f + \sigma_e y)\).

3 Numerical Experiments

We use a two dimensional model problem to compare the sequential and simultaneous NLCG and QN methods. This model problem is inspired by ensemble DFT, and corresponds to a three dimensional system of electrons constrained to two dimensions without spin effects and exchange-correlation terms while taking entropy into account. The continuous model problem adapted from Reference [13] is

\[
\tilde{A}(\{\psi_i\}^{n_e}_{i=1}, f) = -\frac{1}{2} \sum_{i=1}^{n_e} f_i \int \psi_i(r) \Delta \psi_i(r) \, dr \\
+ \int \nabla_{ext}(r) \tilde{\rho}(r) \, dr + \frac{1}{2} \int \nabla_{int}(r) \tilde{\rho}(r) \, dr - TS(f).
\]

(32)

Here \( \nabla_{ext}(r) \) is the external potential of the nuclei

\[
\nabla_{ext}(r) = -\sum_{j=1}^{N} \frac{Z_j}{||r - R_j||}
\]

(33)

![Fig. 1 Orbital energy levels with occupation for \( Z_2 \) at varying temperatures. Fractional occupation numbers are indicated and the same data is also presented in Table 1](image-url)
Fig. 2 Orbital energy levels with occupation for \( \{Z_3, Z_2\} \) at varying temperatures with orbital energy shifted by +5. Fractional occupation numbers are indicated and the same data is also presented in Table 2.

Fig. 3 Orbital energy levels with occupation for \( \{Z_4, Z_3\} \) at varying temperatures with orbital energy shifted by +10. Fractional occupation numbers are indicated and the same data is also presented in Table 3.

Table 1 Orbital energy levels with occupation for \( Z_2 \) at varying temperatures

| E        | Occ. (T = 0) | Occ. (T = 1) | Occ. (T = 2) | Occ. (T = 3) |
|----------|--------------|--------------|--------------|--------------|
| 4.172259 | 1.000000     | 1.000000     | 1.000000     | 0.996380     |
| 21.328241| 0.500000     | 0.500000     | 0.499955     | 0.498751     |
| 21.328241| 0.500000     | 0.500000     | 0.499880     | 0.498751     |
| 36.836577| 0.000000     | 0.000000     | 0.000165     | 0.006117     |
| 43.225667| 0.000000     | 0.000000     | 0.000000     | 0.000000     |
| 46.034373| 0.000000     | 0.000000     | 0.000000     | 0.000000     |

The same data is graphically presented in Fig. 1.
Table 2 Orbital energy levels with occupation for \{Z_3, Z_2\} at varying temperatures with orbital energies shifted by +5

| E     | Occ. (T = 0) | Occ. (T = 1) | Occ. (T = 2) | Occ. (T = 3) |
|-------|--------------|--------------|--------------|--------------|
| 4.606322 | 1.000000     | 1.000000     | 1.000000     | 1.000000     |
| 17.773445 | 1.000000     | 1.000000     | 1.000000     | 1.000000     |
| 23.744218 | 1.000000     | 1.000000     | 0.999833     |              |
| 36.378253 | 1.000000     | 0.994738     | 0.508795     |              |
| 46.607469 | 0.554627     | 0.504757     | 0.508795     |              |
| 46.624356 | 0.445373     | 0.500505     | 0.506011     |              |
| 61.308726 | 0.000000     | 0.000000     | 0.008575     |              |
| 62.566830 | 0.000000     | 0.000000     | 0.005816     |              |
| 78.212923 | 0.000000     | 0.000000     | 0.000000     |              |
| 80.870921 | 0.000000     | 0.000000     | 0.000000     |              |
| 86.752316 | 0.000000     | 0.000000     | 0.000000     |              |
| 93.853450 | 0.000000     | 0.000000     | 0.000000     |              |
| 96.049712 | 0.000000     | 0.000000     | 0.000000     |              |

The same data is graphically presented in Fig. 2

Table 3 Orbital energy levels with occupation for \{Z_4, Z_3\} at varying temperatures with orbital energies shifted by +10

| E     | Occ. (T = 0) | Occ. (T = 1) | Occ. (T = 2) | Occ. (T = 3) |
|-------|--------------|--------------|--------------|--------------|
| 4.327524 | 1.000000     | 1.000000     | 1.000000     | 1.000000     |
| 17.873544 | 1.000000     | 1.000000     | 1.000000     | 1.000000     |
| 25.639021 | 1.000000     | 1.000000     | 1.000000     | 1.000000     |
| 38.541992 | 1.000000     | 1.000000     | 1.000000     | 1.000000     |
| 47.960214 | 1.000000     | 0.999983     | 0.994464     |              |
| 48.278074 | 1.000000     | 0.999841     | 0.993917     |              |
| 63.300484 | 1.000000     | 0.699980     | 0.564432     |              |
| 64.759294 | 0.000000     | 0.330020     | 0.439697     |              |
| 78.938961 | 0.000000     | 0.000000     | 0.005937     |              |
| 82.626842 | 0.000000     | 0.000000     | 0.001553     |              |
| 89.029063 | 0.000000     | 0.000000     | 0.000000     |              |
| 95.252731 | 0.000000     | 0.000000     | 0.000000     |              |
| 98.574017 | 0.000000     | 0.000000     | 0.000000     |              |

The same data is graphically presented in Fig. 3

and

$$V_{\text{int}}(r) = \int \frac{\tilde{\rho}(r')}{\|r - r'\|} \, dr'$$  \hspace{1cm} (34)$$

is the electrostatic (Hartree) potential of the electrons. The electron density is given by

$$\tilde{\rho}(r) = \sum_{i=1}^{n} f_i |\psi_i(r)|^2.$$  \hspace{1cm} (35)$$

Finally, $T$ is the temperature and $S$ the entropy.
For our calculations we consider a discretized version of the problem (32) given as

\[ A(X, f) = -\frac{1}{2} \text{tr}(X^T \Delta_h X \text{diag}(f)) + v_{\text{ext}}^T \rho + \frac{1}{2} v_{\text{int}}^T \rho - TS(f). \]  

(36)

Here \( \Delta_h \in \mathbb{R}^{m \times m} \) is the discretized Laplace operator, \( v_{\text{ext}} \in \mathbb{R}^m \) and \( v_{\text{int}} = V \rho \) the discretized potentials, and \( \rho \in \mathbb{R}^m \) the discretized electron density. The electronic orbitals are represented by \( X \) and the electron density is given by the orbitals and their occupations by

\[ \rho = (X \circ X)f, \]  

(37)

where \( \circ \) is the entrywise, or Hadamard, product. The entropy term is

\[ S(f) = -\sum_{i=1}^{n} f_i \ln(f_i + \delta(1 - f_i)) + (1 - f_i) \ln(1 - f_i + \delta f_i), \]  

(38)

where \( \delta > 0 \) is a small regularization parameter that ensures that the derivative of \( S \) remains finite.
To calculate the discretized potentials we use

$$(v_{\text{ext}})_i = -\sum_{j=1}^{N} \frac{Z_j}{\|r_i - R_j\| + \alpha},$$

where the sum is over the nuclei with charge $Z_j$ and position $R_j$. The position corresponding to the discretization point $i$ is $r_i$, and the parameter $\alpha$ is used to regularize the potential. $V \in \mathbb{R}^{m \times m}$ is similarly given by

$$V_{ij} = \frac{1}{\|r_i - r_j\| + \alpha}.$$

We solve the problem in the unit square with zero boundary conditions corresponding to an infinite potential well. We use a uniform finite difference discretization with $m$ inner points to obtain a system where $X \in \mathbb{R}^{m \times n}$. Here $n$ corresponds to the number of electronic orbitals in the calculation. As initial guess we use the solution of the quadratic problem using the first two terms of (36). The occupation numbers are initialized to

$$f_i = \frac{n_e}{n} + \frac{1}{2} \eta - 1 - \frac{2i}{n + 1},$$

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig5}
\caption{Energy convergence for $\{Z_3, Z_2\}$ at varying temperatures}
\end{figure}
where $\eta = \min(n_e/n, 1 - n_e/n)$ and $n_e \leq n$ is the number of electrons. This choice ensures that the initial occupation of all orbitals is nonzero and emphasizes lower energy orbitals.

We demonstrate the methods for three external potentials. For all models we use potential regularization $\alpha = 5 \times 10^{-2}$ and entropy regularization $\delta = 10^{-3}$.

The first model is a single nucleus with charge $Z = 2$ centered at the center of the unit square with two electrons. For this system the second and third orbitals are degenerate. We calculate the model with $n = 10$ electronic orbitals and a first order finite difference discretization with 25 interior points in one dimension resulting in $m = 625$ spatial degrees of freedom. We will refer to this system as $Z_2$.

The second model, which we name $\{Z_3, Z_2\}$, consists of two nuclei, with a nuclei of charge $Z = 3$ placed at $\left(\frac{1}{3}, \frac{1}{2}\right)$ and another with charge $Z = 2$ placed at $\left(\frac{2}{3}, \frac{2}{3}\right)$ and $n_e = 5$ electrons. This system has four well separated electronic orbitals, while the fifth and sixth are relatively close. The computation is initialized with $n = 10$ orbitals and 29 interior grid points in one dimension for $m = 841$.

The last model, $\{Z_4, Z_3\}$, consists of two nuclei, $Z = 4$ placed at $\left(\frac{1}{4}, \frac{1}{4}\right)$, and $Z = 3$ at the grid point closest to $\left(\frac{2}{3}, \frac{13}{24}\right)$ and $n_e = 7$ electrons. The off diagonal placement is chosen to break the symmetry of the system. This model initially has $n = 14$ orbitals and 29 interior grid points in one dimension ($m = 841$).
For the sequential QN orbital minimizer uses the parameters $\beta_X = 0.4$, $\mu = 5 \times 10^{-5}$, and history of length 6. The sequential QN and NLCG methods minimum trial step length is $\tau_0 = 10^{-3}$ and we perform 6 orbital optimization steps before engaging the occupation number minimizer. Both sequential optimization routines use an identical steepest descent (SD) routine with $\beta_f = 0.5$, $\mu = 10^{-4}$ and $\sigma_0 = 10^{-4}$ for occupation number optimization with two optimization steps. We have tried several different combinations of orbital and occupation optimization steps and observed that this combination offers a good compromise. For the SD method $\mu$ only serves to scale the approximate line search.

We measure convergence by the energy difference to a reference energy computed by running the simultaneous methods for 3000 steps and the sequential methods for 3000 optimization rounds. We then use the lowest energy obtained as the reference energy.

The change in occupation numbers with rising temperature is graphically presented in Figs. 1, 2, and 3. The same data is repeated in Tables 1, 2, and 3 and for clarity all the orbital energies are shifted to be positive. At $T = 0$ the lowest electronic orbitals are fully occupied for $\{Z_4, Z_3\}$, while one electron is split even between two degenerate orbitals for $Z_2$. Even though there is a small gap ($1.69 \times 10^{-2}$) between the fifth and sixth electron orbitals for $\{Z_3, Z_2\}$ the fifth electron is split (0.55 vs 0.45) between these orbitals. We successfully replicated this split with a 3000 round sequential SD orbital occupation number optimization. Furthermore, restarting the SD iteration with a five orbital initial guess based on the split orbital reference solution results in convergence to a higher energy state.
Figures 4, 5, and 6 illustrate energy convergence for the different methods. The simultaneous methods generally perform better than the sequential methods, and the simultaneous NLCG method is more robust than the simultaneous QN approach. In the energy convergence for the sequential optimization routines the switch between orbital and occupation optimization is readily seen in the steplike energy convergence. Furthermore, the performance of the sequential versions of QN and NLCG methods is nearly identical for all models. This might be due to the limited number of steps available for orbital optimization before occupation optimization is enabled.

The simultaneous NLCG method outperforms the QN method for the \( \{Z_4, Z_3\} \) system shown in Fig. 6. Increasing history generally improves the convergence rate of the QN method, but this did not significantly change the rate of convergence for this model. Frequent restarts limit history length and provide at least a partial explanation for this effect. Figure 7 presents \( \{Z_4, Z_3\} \) restarts for the simultaneous QN method. Restarts are frequent for this model at all temperatures compared to \( Z_2 \) and \( \{Z_3, Z_2\} \). However, for \( T > 0 \) there are generally sufficiently many steps between restarts for the history to grow to full length, and the rate of convergence does improve somewhat.

For the \( \{Z_4, Z_3\} \) model the energy difference between the highest occupied and lowest unoccupied orbital is \( 1.69 \times 10^{-2} \), see Fig. 3 and Table 3. This difference is comparatively
small and could explain the poor performance of the QN method, particularly for $T = 0$. In Fig. 8 the convergence rate of the optimization procedures for $\{Z_4, Z_3\}$ for $T = 0.3, 0.5, 0.7$, and the convergence rate for $T = 0$ is included for reference. At $T = 0.3$ the rate of convergence for the QN method is considerably improved and the convergence rate remains superior to $T = 0$ for $T = 0.5$ and $T = 0.7$. The elevated temperature broadens the Fermi surface and this could explain the improved convergence at $T = 0.3$, while the convergence of higher energy orbitals makes the problem more challenging at higher temperatures. This would also explain the decreasing performance of NLCG for higher temperatures.

4 Conclusion

We have presented two schemes for energy optimization of ensemble DFT computations. The updates take the problem constraints into account and permit us to use information obtained from previous evaluations of the target functional and gradients to improve rate of convergence. We have further demonstrated the methods numerically on a model problem inspired by the electronic structure theory and compared simultaneous and sequential schemes based on the QN and NLCG methods.

The ensemble model successfully concentrates occupation to low energy orbitals at low temperatures, and gradually increases occupation of higher energy orbitals at increasing temperature to increase the entropy of the system. Optimization of the occupation numbers also enables ensemble DFT calculations to automatically handle degenerate and near degenerate orbitals at $T = 0$, which are challenging for methods that construct the electron density by the Aufbau principle. Furthermore, it seems possible to broaden the Fermi surface by increasing temperature to accelerate convergence of small gap systems.

Simultaneous optimization schemes provide improved convergence compared to sequential approaches for both the NLCG and QN methods. While the NLCG and QN methods are often comparable in performance, the NLCG method is overall more robust. In contrast, Reference [3] found that QN method is more robust than the NLCG method. It is possible that the quadratic approximate line search gives a better result for NLCG in the model problem. As the NLCG approximate line search gives a better result for NLCG in the model problem. As the NLCG method depends heavily on a high quality line search this might provide a possible explanation. In the present case, the QN method performs poorly for problems with frequent restarts and while the restarts do not fully explain the lack of convergence they can be used as a problem indicator. It is typical for gradient based methods that convergence cannot be guaranteed in general. However, our tests show reasonable convergence in most of the cases.

References

1. Baarman, K., Eirola, T., Havu, V.: Direct minimization of electronic structure calculations with Householder reflections (preprint), arXiv:1204.1204
2. Baarman, K., Eirola, T., Havu, V.: Robust acceleration of self consistent field calculations in density functional theory, J. Chem. Phys. 134, 134109 (2011)
3. Baarman, K., VandeVondele, J.: A comparison of accelerators for direct energy minimization in electronic structure calculations, J. Chem. Phys. 134, 244104 (2011)
4. Bekas, C., Kokioleopoulou, E., Saad, Y.: Computation of large invariant subspaces using polynomial filtered Lanczos iterations with applications in density functional theory, SIAM J. Matrix Anal. Appl. 30, 397 (2008)
5. Blum, V., Hehrke, R., Havu, P., Havu, V., Ren, X., Reuter, K., Scheffler, M.: Ab initio molecular simulations with numeric atom-centered orbitals, Comput. Phys. Commun. 180, 2175 (2009)
6. Cancès, E.: Self-consistent field algorithms for Kohn–Sham models with fractional occupation numbers. J. Chem. Phys. 114, 244104 (2001)
7. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. 20, 303 (1998)
8. Fang, H., Saad, Y.: Two classes of multisecant methods for nonlinear acceleration. Numer. Linear Algebra Appl. 16, 197 (2009)
9. Freysoldt, C., Boeck, S., Neugebauer, J.: Direct minimization technique for metals in density functional theory. Phys. Rev. B 79, 241103 (2009)
10. Fu, C., Ho, K.: First-principles calculation of the equilibrium ground-state properties of transition metals: applications to Nb and Mo. Phys. Rev. B 28, 5480 (1983)
11. Kresse, G., Furthmüller, J.: Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. Comput. Mater. Sci. 6, 15 (1996)
12. Marks, L.D., Luke, D.R.: Robust mixing for ab initio quantum mechanical calculations. Phys. Rev. B 78, 075114 (2008)
13. Marzari, N., Vanderbilt, D., Payne, M.C.: Ensemble density-functional theory for ab initio molecular dynamics of metals and finite-temperature insulators. Phys. Rev. Lett. 79, 1337 (1997)
14. Mermin, N.D.: Thermal properties of the inhomogeneous electron gas. Phys. Rev. 137, A1441 (1965)
15. Methfessel, M., Paxton, A.T.: High precision sampling for Brillouin-zone integration in metals. Phys. Rev. B 40, 3616 (1989)
16. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, Berlin (2006)
17. Pulay, P.: Convergence acceleration in iterative sequences: the case of SCF iteration. Chem. Phys. Lett. 73, 393 (1980)
18. Ruiz-Serrano, A., Skylaris, C.-K.: A variational method for density functional theory calculations on metallic systems with thousands of atoms. J. Chem. Phys. 139, 054107 (2013)
19. Saad, Y., Chelikowsky, J.R., Shontz, S.M.: Numerical methods for electronic structure calculations of materials. SIAM Rev. 52, 3 (2010)
20. VandeVondele, J., Hutter, J.: An efficient orbital transformation method for electronic structure calculations. J. Chem. Phys. 118, 4365 (2003)
21. van Voorhis, T., Head-Gordon, M.: A geometric approach to direct minimization. Mol. Phys. 100, 1713 (2002)
22. Wagner, F., Lalloyaux, T., Scheffler, M.: Errors in Hellmann-Feynman forces due to occupation-number broadening and how they can be corrected. Phys. Rev. B 57, 2102 (1998)
23. Zhou, Y., Saad, Y., Tiago, M.L., Chelikowsky, J.R.: Parallel self-consistent-field calculations via Chebyshev-filtered subspace acceleration. Phys. Rev. E 74, 066704 (2006)