SIAM J. Sci. Comput.
Vol. 29, No. 5, pp. 1854–1875

A TRUST REGION DIRECT CONSTRAINED MINIMIZATION ALGORITHM FOR THE KOHN–SHAM EQUATION*

CHAO YANG[†], JUAN C. MEZA[†], AND LIN-WANG WANG[†]

Abstract. The self-consistent field (SCF) iteration, widely used for computing the ground state energy and the corresponding single particle wave functions associated with a many-electron atomistic system, is viewed in this paper as an optimization procedure that minimizes the Kohn–Sham (KS) total energy indirectly by minimizing a sequence of quadratic surrogate functions. We point out the similarity and difference between the total energy and the surrogate, and show how the SCF iteration can fail when the minimizer of the surrogate produces an increase in the KS total energy. A trust region technique is introduced as a way to restrict the update of the wave functions within a small neighborhood of an approximate solution at which the gradient of the total energy agrees with that of the surrogate. The use of trust regions in SCF is not new. However, it has been observed that directly applying a trust region-based SCF (TRSCF) to the KS total energy often leads to slow convergence. We propose to use TRSCF within a direct constrained minimization (DCM) algorithm we developed in [*J. Comput. Phys.*, 217 (2006), pp. 709–721]. The key ingredients of the DCM algorithm involve projecting the total energy function into a sequence of subspaces of small dimensions and seeking the minimizer of the total energy function within each subspace. The minimizer of a subspace energy function, which is computed by the TRSCF, not only provides a search direction along which the KS total energy function decreases, but also gives an optimal "step length" that yields a sufficient decrease in total energy. A numerical example is provided to demonstrate that the combination of TRSCF and DCM is more efficient than SCF.

Key words. nonlinear eigenvalue problem, Kohn–Sham total energy, constrained optimization, trust region

AMS subject classifications. 15A18, 65K10, 65F15

DOI. 10.1137/060661442

1. Introduction.

The *self-consistent field* (SCF) iteration is the most widely used procedure for computing the ground state energy and the corresponding single particle wave functions associated with a many-electron atomistic system. The procedure is often viewed as a fixed point iteration in which a sequence of linear eigenvalue problems is solved approximately. The approximate eigenvectors computed in the jth iteration are used to construct the potential component of the Kohn–Sham (KS) matrix Hamiltonian required in the $(j+1)$th iteration. Convergence is reached when the difference between Hamiltonians formed in two consecutive iterations becomes negligible. At this point, the eigenvalues and eigenvectors of the Hamiltonian become self-consistent and the KS total energy function associated with the system reaches the global minimum.

It has long been observed that the simplest form of SCF iteration often fails to converge. As such, a number of heuristics have been developed to prevent the

[*]Received by the editors May 30, 2006; accepted for publication (in revised form) October 24, 2006; published electronically September 28, 2007. This work was supported by the Director, Office of Science, Division of Mathematical, Information, and Computational Sciences of the U.S. Department of Energy under contract DE-AC03-76SF00098. This work was performed by an employee of the U.S. Government or under U.S. Government contract. The U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. Copyright is owned by SIAM to the extent not limited by these rights.

http://www.siam.org/journals/sisc/29-5/66144.html

[†]Computational Research Division, Lawerence Berkeley National Laboratory, Berkeley, CA 94720 (CYang@lbl.gov, JCMeza@lbl.gov, LWWang@lbl.gov).

SCF from diverging. These heuristics often involve combining either the potential or the charge densities computed in the previous SCF iterations to construct a new Hamiltonian so that the lack of self-consistency can be minimized. These techniques are known either as *charge mixing* [13, 15] in the material sciences community or as the direct inversion of iterative subspace (DIIS) extrapolation technique [24, 25] in the quantum chemistry community. Although these heuristics work remarkably well in stabilizing the SCF iteration for many problems, they can still fail in others. Even when they work, the reduction in the KS total energy is often not monotonic. Furthermore, no satisfactory theoretical foundation has been established to explain why charge mixing and DIIS work and under what circumstances they can fail. As a result, the convergence of SCF is often unpredictable for large atomistic systems with small (valence/conducting) band gaps.

In this paper, we will examine SCF from an optimization point of view. We view the SCF iteration as an indirect way to minimize the KS total energy function through the minimization of a sequence of quadratic surrogate functions. We point out the similarity and difference between the surrogate and the true objective function and introduce the concept of a *trust region* that can be used to restrict the update of the wave functions within a region in which the gradients associated with the true object and the surrogate differ very little. The trust region technique is a widely used methodology for promoting global convergence in numerical optimization procedures [5, 20]. Trust region-based SCF (TRSCF) iteration has been used in the past in the quantum chemistry community [28, 31], where it is sometimes called *level-shifted* SCF iteration [28].

By imposing a trust region with an appropriate radius at each SCF iteration, we can show that a monotonic reduction of the KS total energy can be achieved in the SCF procedure. However, it has been observed that for large systems, the use of TRSCF often leads to extremely slow convergence. In [31], the trust region technique is combined with DIIS to accelerate the convergence of SCF. In this paper, we propose a different approach. Instead of applying the trust region technique directly to an SCF iteration, we use it within a direct constrained minimization (DCM) algorithm developed in [36] to generate an effective search direction and step length simultaneously. This scheme is applicable to density functional theory (DFT) calculations in which the number of degrees of freedom in the discretized wave function is much larger than the number of electrons in the atomistic system.

The paper is organized as follows. In section 2, we establish some basic notation required for the discussion of SCF and DCM. In section 3, we examine the SCF iteration from an optimization point of view and provide the motivation for applying the trust region technique in SCF. A TRSCF iteration is presented in section 4. We discuss the use of TRSCF within the DCM algorithm in section 5. A numerical example is provided in section 6 to demonstrate the value of the trust region technique in KS total energy minimization.

2. Background and notation.

In this section, we establish the mathematical notation required to describe the SCF and DCM algorithms. We skip the description of the continuous formulation of the KS total energy optimization problem, which has been presented in [36] and many other references [15, 16, 23]. Instead, we will focus on the finite-dimensional version of the problem.

In the following discussion, we will use A^T to denote the transpose of a matrix A, and A^* to denote the complex conjugate of A. A submatrix of A consisting of rows i through j and columns p through q will be denoted by $A(i:j, p:q)$. If the submatrix contains all rows (columns) of A, it will be denoted by $A(:,p:q)$ (resp., $A(i:j,:)$). The Frobenius norm of a matrix A (defined as the square root of the sum of the absolute squares of all elements in A) is denoted by $\|A\|_F$.

With an appropriate discretization scheme, a single electron wave function can be approximated by a vector $x_i \in \mathbb{C}^n$, where n is the spatial degrees of freedom, i.e., the number of real space grid points. These vectors satisfy the orthonormality constraints

$$x_i^* x_j = \delta_{i,j}, \quad i,j = 1,2,\ldots,k,$$

where k is the number of occupied states. If we let $X = (x_1, x_2, \ldots, x_k)$, the matrix

$$D(X) = XX^* \tag{2.1}$$

is often known as the density matrix, and the *charge density* associated with the k occupied states can be expressed by

$$\rho(X) = \mathrm{diag}(XX^*), \tag{2.2}$$

where diag(A) denotes a column vector consisting of diagonal entries of the matrix A.

The KS total energy function consists of several components [23], i.e.,

$$E_{total}(X) = E_{kinetic}(X) + E_{ion}(X) + E_H(X) + E_{XC}(X),$$

where $E_{kinetic}$ is the kinetic energy and E_{ion}, E_H, and E_{XC} are potential energies induced by the electron-ion interaction (ionic potential), the electron-electron interaction (Hartree potential), and the exchange correlation potential, respectively.

Let $L \in \mathbb{C}^{n\times n}$ be a Hermitian matrix representing a discretized Laplacian operator. The kinetic energy is then defined by [23]

$$E_{kinetic}(X) = \frac{1}{2}\mathrm{trace}(X^* L X).$$

The ionic potential energy consists of a local and a nonlocal term. If we let D_{ion} be a real diagonal matrix representing a discretized local ionic potential function, then the local ionic potential energy is defined by [23]

$$E_{ion(local)}(X) = \mathrm{trace}(X^* D_{ion} X).$$

The contribution from the nonlocal ionic potential is defined by [23]

$$E_{ion(nonlocal)}(X) = \sum_i \sum_\ell \left| x_i^* w_\ell \right|^2,$$

where w_ℓ represents a discretized pseudopotential reference projection function.

In practice, appropriate boundary conditions are imposed so that L is nonsingular. If we use S to denote the inverse of the discrete Laplacian operator, then the Hartree potential energy, which is used to model the classical electrostatic average interaction between electrons, can be expressed [23] by

$$E_H(X) = \frac{1}{2}\rho(X)^T S\, \rho(X).$$

The exchange correlation function ϵ_{xc} is used to model the nonclassical interaction between electrons. The potential energy induced by this function is defined by [23]

$$E_{XC}(X) = e^T\left(\epsilon_{xc}[\rho(X)]\right),$$

where e is a column vector of ones.

Using the notation established above, we can state the KS total energy minimization problem as

$$\begin{aligned} &\min \quad E_{total}(X) \\ &\text{s.t.} \quad X^*X = I_k, \end{aligned} \tag{2.3}$$

where I_k denotes a $k \times k$ identity matrix.

The Lagrangian associated with (2.3) is

$$\mathcal{L}(X) = E_{total}(X) - \text{trace}\big[\Lambda^T (X^*X - I_k)\big], \tag{2.4}$$

where Λ is a $k \times k$ matrix containing the Lagrange multipliers associated with the constraints specified by $X^*X = I_k$ [20].

The solution to (2.3) must satisfy the first order necessary conditions

$$\begin{aligned} \nabla_X \mathcal{L}(X) &= 0, \\ X^*X &= I_k. \end{aligned} \tag{2.5}$$

Here, $\nabla_X \mathcal{L}$ represents an $n \times k$ matrix whose (i, j)th entry is the partial derivative of $\mathcal{L}$ with respect to the (i, j)th entry of X.

It is easy to verify that

$$\nabla_X E_{kinetic} = \frac{1}{2} LX, \tag{2.6}$$

$$\nabla_X E_{ion(local)} = D_{ion} X, \tag{2.7}$$

$$\nabla_X E_{ion(nonlocal)} = \sum_\ell (w_\ell w_\ell^*) X, \tag{2.8}$$

$$\nabla_X E_H = \text{Diag}(S\rho(X))X, \tag{2.9}$$

$$\nabla_X E_{XC} = \text{Diag}(\mu_{xc}(\rho))X, \tag{2.10}$$

where

$$\mu_{xc}(\omega) \equiv \frac{d\epsilon_{xc}(\omega)}{d\omega}$$

is the derivative of the exchange-correlation function. Here the notation $\text{Diag}(\rho)$ represents a diagonal matrix whose diagonal is determined by the vector ρ, and we have scaled (2.6)–(2.10) by 1/2 to be consistent with the convention used in the electronic structure community.

Substituting (2.6)–(2.10) into (2.5), we obtain the KS equation

$$H(X)X = X\Lambda_k, \qquad X^*X = I_k, \tag{2.11}$$

where

$$H(X) = \left[\frac{1}{2} L + D_{ion} + \sum_\ell w_\ell w_\ell^* + \text{Diag}(S\rho) + \text{Diag}(\mu_{xc}(\rho))\right]. \tag{2.12}$$

Because the vector ρ in (2.12) depends on X, the eigenvalue problem defined by (2.11) is nonlinear. Also note that the solution to (2.3) is not unique. If X is a solution, then XQ is also a solution for any $Q \in \mathbb{C}^{k\times k}$ such that $Q^*Q = I_k$. That is, the solution to the constrained minimization problem, or, equivalently, the nonlinear equations defined by (2.11), is a k-dimensional invariant subspace in $\mathbb{C}^n$ rather than a specific matrix. In particular, Q can be chosen such that Λ_k is diagonal. In this case, X consists of k KS eigenvectors associated with the k smallest eigenvalues of $H(X)$.

3. The optimization view of SCF.

Although the nonlinear eigenvalue problem defined by (2.11) may appear easier to solve than (2.3) because of its close connection to a linear eigenvalue problem, there is yet no robust and efficient general-purpose solver for this type of problem with guaranteed convergence. The most widely used technique for solving (2.11) is to reduce it to a sequence of linear eigenvalue problems that can be solved by many numerical linear algebra software packages such as LAPACK [1] or ARPACK [18]. This approach is often known as the SCF iteration. For completeness, we outline the major steps of the basic version of the SCF iteration procedure in Figure 3.1.

```
SCF iteration
Input:   The matrices $L$, $D_{ion}$, $S$, and the vectors $w_\ell$, $\ell = 1, 2, \ldots$. The
         derivative of the exchange-correlation function $\mu_{xc}(x)$; an initial
         guess $X^{(0)}$ for the optimal wave function $X \in \mathbb{C}^{n\times m}$;
Output:  $X \in \mathbb{C}^{n\times m}$ such that $X^*X = I_m$ and $E_{tot}(X)$ is minimized, where
         $E_{tot}(X)$ is defined by (2.6)–(2.10).

for $i = 1, 2, \ldots$ until convergence
   Form $H^{(i)} = H(X^{(i-1)})$;
   Compute $X^{(i)}$ such that $H^{(i)}X^{(i)} = X^{(i)}\Lambda^{(i)}$, and $\Lambda^{(i)}$
   contains the $k$ smallest eigenvalues of $H^{(i)}$;
end for
```

Fig. 3.1. *The SCF iteration.*

Depending on the discretization scheme used, it may not be necessary or possible to form the Hamiltonian $H^{(i)}$ explicitly in the SCF calculation. This is particularly true when the continuous problem is discretized by a spectral method using a plane wave basis. In that case, $H^{(i)}$ exists only in the form of a matrix vector multiplication procedure, and it is not feasible to solve the linear eigenvalue problem $H^{(i)}X^{(i)} = X^{(i)}\Lambda^{(i)}$ by using subroutines provided in LAPACK [1]. Iterative methods such as the Lanczos [17, 18, 35], preconditioned conjugate gradient [14, 12, 27], the Jacobi–Davidson type of method [6, 21, 26, 29], or multigrid accelerated Rayleigh-quotient iterations [3, 8, 11] are often used in this setting.

Because computing the k smallest eigenpairs of the discrete Hamiltonian $H^{(i)}$ is equivalent to solving the trace minimization problem

$$\begin{array}{ll} \min & q(X) = \frac{1}{2}\text{trace}(X^*H^{(i)}X) \\ \text{s.t.} & X^*X = I_k, \end{array} \tag{3.1}$$

the SCF iteration can be viewed as an iterative procedure that minimizes the total energy E_{total} indirectly by minimizing a sequence of quadratic *surrogate* functions of the form (3.1). Note that the discrete Hamiltonian $H^{(i)}$ is formed by using a fixed set of wavefunctions, $X^{(i-1)}$, computed in the previous SCF iteration.

At $X = X^{(i-1)}$, the gradient of the quadratic surrogate agrees with that of E_{total}, i.e.,

$$\nabla E_{total}(X)_{|X=X^{(i-1)}} = \nabla q(X)_{|X=X^{(i-1)}} = H^{(i)}X^{(i-1)}, \tag{3.2}$$

even though $E_{total}(X^{(i-1)})$ may be completely different from $q(X^{(i-1)})$ in general.

The gradient match between $E_{total}(X)$ and $q(X)$ at $X = X^{(i-1)}$ should come as no surprise because such a match is achieved by construction. A direct consequence of (3.2) is that, at least within a small neighborhood of $X^{(i-1)}$, a reduction of $q(X)$ is likely to result in a reduction in $E_{total}(X)$. However, if one moves too far away from $X^{(i-1)}$, $E_{total}(X)$ may actually increase because gradient matching does not hold in general when X is away from $X^{(i-1)}$.

This optimization view of the SCF iteration suggests the danger of minimizing the quadratic surrogate (3.1) by computing the k smallest eigenpairs of $H^{(i)}$. To illustrate this possibility, we will now give a concrete two-dimensional example below to demonstrate how SCF may fail to converge. Consider a simplified total energy function of the form

$$E_{total}(x) = \frac{1}{2}x^TLx + \frac{\alpha}{4}\rho(x)^TL^{-1}\rho(x), \tag{3.3}$$

where

$$L = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \text{and} \quad \rho(x) = \begin{pmatrix} x_1^2 \\ x_2^2 \end{pmatrix}$$

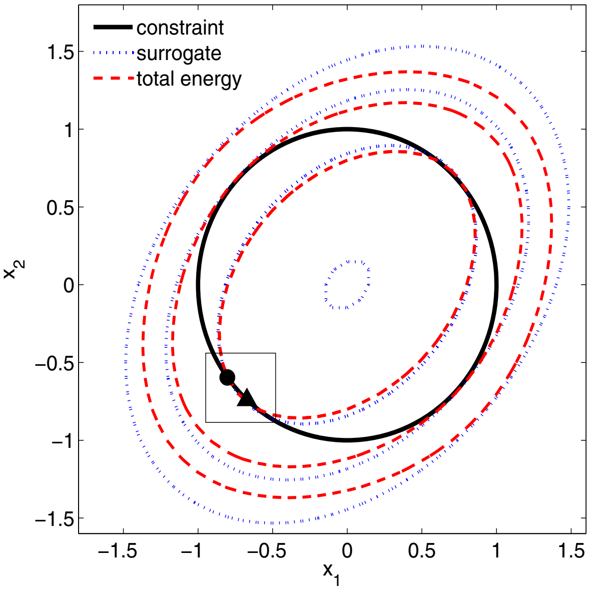
is used as our objective function, to be minimized subject to the constraint

$$x_1^2 + x_2^2 = 1. \tag{3.4}$$

Note that we deliberately ignored the ionic and exchange-correlation potentials to make this example simple enough for testing. As a result, both the total energy and the quadratic surrogate functions are convex, which may not necessarily be the case in general. Nonetheless, these examples exhibit the local convergence behavior of SCF.

When we set $\alpha = 2$ in (3.3), the simplest version of SCF algorithm shown in Figure 3.1 converges in 10 iterations. In Figure 3.2(a), we plot the contour of (3.3) (dashed contour) as well as that of the quadratic surrogate (dotted contour) constructed at the current approximation $\hat{x} = (\hat{x}_1, \hat{x}_2)^T = (-0.8033, -0.5956)^T$. The position of $\hat{x}$ is marked by a small solid circle in the figure. The minimizer of the quadratic surrogate $\hat{x}_q$ is marked by a small solid triangle. Both $\hat{x}$ and $\hat{x}_q$ lie on the large circle which corresponds to the constraint (3.4).

If we zoom into the box shown in Figure 3.2(a), we can see how the total energy changes as we move from $\hat{x}$ (the small solid circle, where the gradient of $E_{total}(x)$ matches the gradient of the quadratic surrogate) to $\hat{x}_q$ (the solid triangle, at which the dotted elliptical contour becomes tangent to the large circle). Figure 3.2(b) shows that the minimizer of the quadratic surrogate, $\hat{x}_q$, lies on the total energy contour (dashed) line that is above and to the right of the dashed contour line that passes $\hat{x}$. Because the total energy in this example is convex, this implies that moving from the current approximation $\hat{x}$ to $\hat{x}_q$ results in a reduction of the total energy. For reference purposes, we also plot the true minimizer of (3.3) as a small solid square in Figure 3.2(b).



(a)

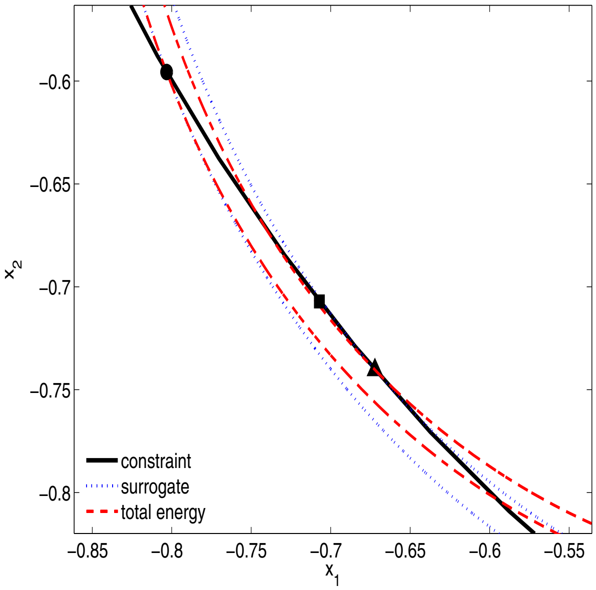
Fig. 3.2. (a) *When $\alpha = 2$ is set in* (3.3), *the SCF iteration converges. This figure shows the change of $E_{total}(x)$ in a single SCF iteration. The contours of the total energy defined in* (3.3) *and the surrogate $q(x) = \frac{1}{2}x^T(L + \alpha \mathrm{Diag}(L^{-1}\rho(x)))x$ for $\alpha = 2$. The current iterate $\hat{x}$ is marked by the small solid circle at which $\nabla E_{total}(\hat{x}) = \nabla q(\hat{x})$. The minimizer of $q(x)$ is marked by the solid triangle. Both of these points lie on the large circle which defines the orthonormality constraint* (3.4).

When we increase α to 12, the contour of (3.3) becomes less elliptical, as we can see in Figure 3.3(a). Figure 3.3(a) also shows the quadratic surrogate constructed from the current approximation (marked by a small solid circle) $\hat{x} = (\hat{x}_1, \hat{x}_2)^T = (-0.8904,\ -0.4551)^T$, as well as the location of $\hat{x}_q$ (marked by the solid triangle). The small solid square indicates the location of the true minimizer of (3.3) for $\alpha = 12$. All of these three points lie on the orthonormality constraint shown as the large circle in this figure.

As we zoom into the rectangular box shown in Figure 3.3(a), we can see more clearly why a single SCF iteration from $\hat{x}$ leads to an increase in (3.3). In Figure 3.3(b), which provides a zoom-in view, we can clearly see that the dashed contour line that passes through $\hat{x}_q$ lies to the lower left of the curve that passes through $\hat{x}$. This implies an increase in the total energy (3.3) as we move from the current approximation $\hat{x}$ to the minimizer of the surrogate $\hat{x}_q$.

Fortunately, the optimization view of the SCF iteration also suggests at least two ways to improve the convergence of a SCF:

1. develop a better surrogate function,
2. restrict the wave function update in a small neighborhood of the current approximation.

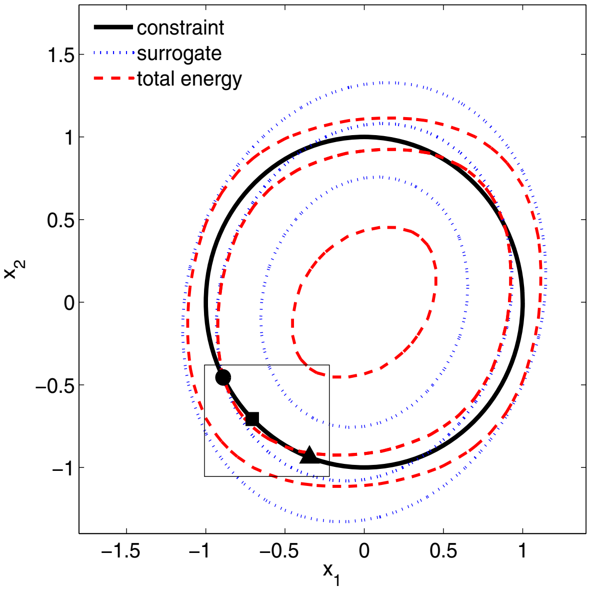


(b)

FIG. 3.2(*continued*). (b) *When* $\alpha = 2$ *is set in* (3.3), *the SCF iteration converges. This figure shows the change of* $E_{total}(x)$ *in a single SCF iteration. A zoom-in view of the boxed area in Figure* 3.2(a). *The minimizer of* $q(x)$ *(the small solid triangle) lies on the inner dashed contour line, indicating a decrease in the total energy. The solid square marks the true minimizer of* E_{total}.

One typical approach is to use the second order Taylor approximation to total energy as the surrogate model. The iterative method based on this quadratic model is the familiar Newton's method. However, due to the high cost associated with Hessian computation, this approach is not practical. Although it is possible to apply a limited memory version of a quasi-Newton method to minimize $E_{total}(X)$ subject to the orthonormality constraint, such an approach, in which only an approximate Hessian is used and updated by taking into account the gradient information computed in previous iterations, is generally not effective because of the large dimensionality of the minimization problem; hence the difficulty of obtaining a good approximation to the Hessian [33].

A technique that is widely used in the current practice of electronic structure calculation is to modify the Hessian of the quadratic surrogate (3.1) so that the *lack of self-consistency* in KS Hamiltonian $H(X)$ is minimized. The lack of self-consistency in X can be measured in a number of ways. If $X^{(i+1)}$ is the approximate solution to the quadratic minimization problem (3.1), in which H is defined in terms of $X^{(i)}$, then the lack of self-consistency can be directly measured by computing $\delta_H = \|H(X^{(i+1)}) - H(X^{(i)})\|_F$, or simply $\delta_\rho = \|\rho(X^{(i+1)}) - \rho(X^{(i)})\|$, because the change in $H(X)$ is contributed by the last two terms of (2.12), which are both functions of the charge density $\rho(X)$. The minimization of δ_ρ is often achieved approximately by choosing $\rho^{(i+1)}$ to be a linear combination of the charge densities computed in

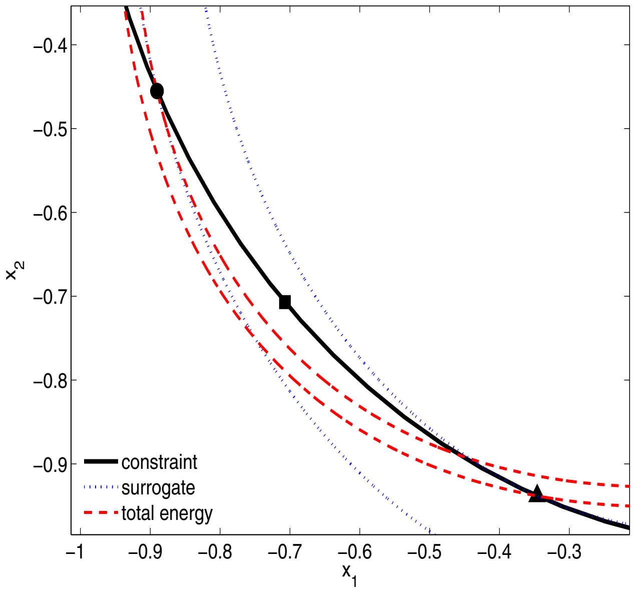


(a)

FIG. 3.3. (a) *When* $\alpha = 12$ *is set in* (3.3), *the SCF iteration fails to converge. This figure shows the change of* $E_{total}(x)$ *in a single SCF iteration. The contours of the total energy defined in* (3.3) *and the surrogate* $q(x) = \frac{1}{2}x^T(L + \alpha \mathrm{Diag}(L^{-1}\rho(x)))x$ *for* $\alpha = 12$. *The current iterate* $\hat{x}$ *is marked by the small solid circle at which* $\nabla E_{total}(\hat{x}) = \nabla q(\hat{x})$. *The minimizer of* $q(x)$ *is marked by the solid triangle. Both of these points lie on the large circle which defines the orthonormality constraint* (3.4).

the previous SCF iterations and solving an equality constrained quadratic program which returns the "optimal" linear combination. This approach is often referred to as the *Pulay mixing* scheme or the DIIS scheme [15, 24, 25]. In the material sciences community, the Pulay mixing scheme is often followed by another procedure in which selected components of the new charge density are modified or scaled. This procedure is sometimes called the *Kerker mixing* scheme [13, 15]. These mixing schemes work remarkably well for many problems but can fail for others. To our knowledge, no theoretical analysis of these schemes is yet available. For that reason, we will not go into the details of these mixing schemes but rather refer readers to [15] for more discussions.

The second way to improve the convergence of SCF, which we will discuss in the next section, is to restrict the minimization of the quadratic surrogate in (3.1) (hence the update of the single particle wave functions) to a small neighborhood of the current approximation $X^{(i)}$. This technique is known as the *trust region* technique in the numerical optimization community [5, 20]. The gradient matching (between the total energy and the surrogate) property of SCF implies that a reduction of the total energy in the SCF iteration can be guaranteed if one imposes a trust region with a sufficiently small radius to (3.1).



(b)

FIG. 3.3(*continued*). (b) *When* $\alpha = 12$ *is set in* (3.3), *the SCF iteration fails to converge. This figure shows the change of* $E_{total}(x)$ *in a single SCF iteration. A zoom-in view of the boxed area in Figure* 3.2(a). *The minimizer of* $q(x)$ *(the small solid triangle) lies on the outer dashed contour line, indicating an increase in the total energy. The solid diamond marks the true minimizer of* E_{total}.

4. Trust region SCF.

The simplest type of trust region one may consider for (3.1) is

$$\|X - X^{(0)}\|_F \leq \Delta, \tag{4.1}$$

where Δ is a trust region radius that may be reduced or enlarged depending on the ratio of the actual reduction of $E_{total}(X)$ over the predicted reduction measured in terms of the change in $q(X)$. The constrained optimization problem

$$\begin{gathered} \min q(X) \\ X^*X = I, \\ \|X - X^{(0)}\|_F \leq \Delta, \end{gathered} \tag{4.2}$$

associated with a particular choice of Δ is known as a *trust region subproblem*.

However, there are two serious drawbacks associated with this type of constraint. First of all, (4.1) is not rotationally invariant; i.e., the inequality (4.1) is not equivalent to

$$\|XQ - X^{(0)}\|_F \leq \Delta$$

for all $Q \in \mathbb{C}^{k\times k}$ such that $Q^*Q = I_k$. As a result, the solution to the trust region subproblem (4.2) is not rotation invariant, whereas the solution to the original total energy minimization problem is.

Secondly, adding (4.1) as a constraint makes the constrained quadratic minimization problem much more difficult to solve. If we introduce the constraint (4.1) as a penalty function and solve

$$\begin{aligned} &\min \hat{q}(X;\sigma) \equiv q(X) + \sigma \|X - X^{(0)}\|_F^2, \\ &X^* X = I_k \end{aligned} \tag{4.3}$$

for an appropriately chosen penalty parameter σ, the first order necessary condition of (4.3) cannot be expressed as a linear eigenvalue problem or other simple form. Thus it cannot be solved easily.

To preserve the rotational invariance property of the solution to (2.3) in a trust region subproblem, we must define the trust region in terms of quantities that are rotationally invariant. Both the charge density $\rho(X)$ defined in (2.2) and the density matrix $D(X)$ defined in (2.1) satisfy this desirable property. However, we will show in the following that a trust region defined in terms of $D(X)$ allows us to reduce the minimizing of the constrained quadratic surrogate function to a linear eigenvalue problem.

Note that $D(X)$ is an orthogonal projector associated with the subspace spanned by columns of X when $X^*X = I_k$. It is well known [10] that

$$\|D(X) - D(X^{(0)})\|_F \tag{4.4}$$

measures the distance between the subspaces defined by columns of X and $X^{(0)}$, where $\|A\|_F$ denotes the Frobenius norm of A. Therefore, we impose the constraint

$$\|D(X) - D(X^{(0)})\|_F \le \Delta$$

on the solution to the quadratic minimization problem. Because $X^*X = {X^{(0)}}^* X^{(0)} = I_k$, it is easy to verify that

$$\begin{aligned} \|D(X) - D(X^{(0)})\|_F^2 &= \|D(X)\|_F^2 + \|D(X^{(0)})\|_F^2 + 2\text{trace}\big[D(X)^* D(X^{(0)})\big] \\ &= 2k - 2\text{trace}\big[X^* X^{(0)} {X^{(0)}}^* X\big]. \end{aligned}$$

If we solve the trust region subproblem by introducing (4.4) as a penalty function in the quadratic objective $q(x)$, i.e., we solve

$$\begin{aligned} &\min \hat{q}(X;\sigma) \equiv \frac{1}{2}\text{trace}\big[X^* H(X^{(0)}) X\big] - \frac{\sigma}{2}\text{trace}\big[X^* X^{(0)} {X^{(0)}}^* X\big], \\ &\text{s.t. } X^*X = I_k, \end{aligned} \tag{4.5}$$

where σ is an appropriately chosen penalty parameter, then the first order necessary condition associated with (4.5) becomes

$$\begin{aligned} \big[H(X^{(0)}) - \sigma X^{(0)} {X^{(0)}}^*\big] X &= X\Lambda, \\ X^*X &= I_k, \end{aligned}$$

where Λ (which can be diagonalized by postmultiplying X by an unitary transformation) is a matrix of Lagrange multipliers.

Therefore, when a trust region is defined with respect to $D(X)$, solving the corresponding trust region subproblem is equivalent to computing the eigenvectors associated with the smallest k eigenvalues of the *level shifted* Hamiltonian $H(X^{(0)}) -$

$\sigma X^{(0)} X^{(0)^*}$, a problem that we generally know how to solve efficiently, at least when the dimension of $H(X^{(0)})$ is small.

What remains to be determined now is the penalty or trust region parameter σ in (4.5). Choosing a large σ value has the same effect as setting a small trust region radius Δ in (4.2). When σ is sufficiently large, the convergence of the SCF iteration can be guaranteed, although the rate of convergence may be very slow. When σ is too small, the solution to (4.6) may lead to an increase in the total energy. Unfortunately, the optimal choice of σ cannot be obtained analytically in general. The standard recipe for choosing such a parameter is usually dynamic [20]. In an unconstrained optimization problem, one starts with an arbitrary guess Δ_0 bounded by the maximum step length allowed. After the trust region subproblem is solved, the trust region radius may be reduced or increased based on the ratio of reduction in the true objective E_{true} over the predicted reduction E_{pred} measured from the surrogate function. A reduction of the trust region radius implies that the trust region subproblem must be resolved.

Two special features of the SCF iteration require the selection of penalty or trust region parameters to be made in a slightly different manner. First of all, the quadratic surrogate function in (3.1) does not match the true objective, i.e., the KS total energy function, at the current iterate $X^{(i)}$. Recall that the only thing that matches between $E_{total}(X)$ and $q(X)$ at $X^{(i)}$ is their gradients. Hence, the ratio between the changes in $E_{total}(X)$ and $q(X)$ can be difficult to predict. Consequently, an adjustment of the penalty parameter based on this ratio is unlikely to be effective. Secondly, the evaluation of $E_{total}(X)$ tends to be costly. Therefore, the selection of the penalty parameter should avoid repeated evaluation of the total energy.

A heuristic for estimating the penalty parameter is developed in [31] by expressing the change in E_{total} as a function of a matrix built from a linear combination of previous density matrices. The optimal σ is estimated by applying an inexact line search to the approximate model. Such a scheme requires saving wave functions or density matrices obtained in previous SCF iterations; thus is not practical for large problems.

We propose to use a simpler heuristic in this paper. Our heuristic is based on the following observation. As $X^{(i)}$ converges to X in a trust region enabled SCF iteration, the eigenvalues of the level-shifted Hamiltonian $H(X) - \sigma X X^*$ converge to

$$\lambda_1 - \sigma, \lambda_2 - \sigma, \ldots, \lambda_k - \sigma, \lambda_{k+1}, \lambda_{k+2}, \ldots, \lambda_n, \tag{4.6}$$

where

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$$

are eigenvalues of the KS Hamiltonian $H(X)$ as defined in (2.12). Thus, adding a trust region of the form (4.4) has the effect of increasing the gap between the kth and the $(k+1)$th eigenvalues of the shifted Hamiltonian. Even though (4.6) does not hold in general for eigenvalues of the shifted Hamiltonian $H(X^{(i)})$ before $X^{(i)}$ converges to the solution of (2.3), our numerical experiments and those presented in [31] show an increased gap between the kth and $(k+1)$th eigenvalues of the shifted $H(X^{(i)})$ when σ is sufficiently large.

It is well known [22] that a larger gap between the eigenvalues associated with the desired invariant subspace and the rest of the spectrum generally makes it easier to compute the desired invariant subspace. Thus our heuristic tries to enlarge such a gap when total energy increases in an SCF iteration. To be specific, we set σ to zero

```
TRSCF
Input:  The matrices $L$, $D_{ion}$, $S$, and the vectors $w_\ell$, $\ell = 1, 2, \ldots$. The
        derivative of the exchange-correlation function $\mu_{xc}(x)$; an initial
        guess $X^{(0)}$ for the optimal wave function $X \in \mathbb{C}^{n \times m}$;
Output: $X \in \mathbb{C}^{n \times m}$ such that $X^* X = I_m$ and $E_{tot}(X)$ is minimized, where
        $E_{tot}(X)$ is defined by (2.6)–(2.10).

Set $\sigma = 0.0$;
for $i = 1, 2, \ldots$ until convergence
  if ($\sigma = 0.0$)
    Set $H^{(i)} = H(X^{(i-1)})$;
  else
    Set $H^{(i)} = H(X^{(i-1)}) - \sigma X^{(i-1)} X^{(i-1)^*}$;
  endif
  Compute $X^{(i)}$ such that $H^{(i)} X^{(i)} = X^{(i)} \Lambda^{(i)}$ and $\Lambda^{(i)}$
  contains the $k$ smallest eigenvalues of $H^{(i)}$;
  if ($E_{tot}(X^{(i)}) > E_{tot}(X^{(i-1)})$), set $\sigma \leftarrow \gamma(\lambda_{k+1}^{(i)} - \lambda_k^{(i)})$
  for some constant $\gamma > 1$;
end for
```

FIG. 4.1. *A trust region-based SCF iteration.*

initially. If the minimizer of $q(X)$ yields an increase in $E_{total}(X)$, we increase σ by setting it to $\gamma\eta_{\max}$, where $\eta_{\max}$ is the maximum gap defined as

$$\eta_{\max} = \max_{\ell \in \{1,2,\ldots,n-1\}} \lambda_{\ell+1}^{(i)} - \lambda_\ell^{(i)},$$

where $\lambda_\ell^{(i)}$ is the ℓth eigenvalue of $H(X^{(i)})$ and γ is a small constant. Empirically, we found a good choice of γ to be around 2 to 5. This particular strategy for choosing the penalty parameter is somewhat conservative in the sense that σ is never decreased in subsequent SCF iterations to allow a larger reduction in total energy. A more sophisticated and efficient scheme will be described in a separate paper.

The major steps of a TRSCF iteration is summarized in Figure 4.1. We now return to the simple example (3.3) and show the effect of applying the trust region technique in the SCF iteration. The eigenvalues of the Hamiltonian constructed at $\hat{x}$ are

$$\hat{\lambda}_1 = 6.4597 \quad \text{and} \quad \hat{\lambda}_2 = 9.5403.$$

We set $\sigma = \hat{\lambda}_2 - \hat{\lambda}_1$ to increase the gap between the eigenvalues of the shifted Hamiltonian by a factor of two. Figure 4.2 is almost identical to Figure 3.3(b). The only difference is that we plotted the solution to the trust region subproblem (4.5), which is marked by a solid diamond, in addition to the solution to the unpenalized surrogate (3.1), which is marked by a solid triangle. As we can see in this figure, the solution to the trust region subproblem lies inside the inner dashed contour, indicating a reduction in total energy as we move from the small solid circle to the solid diamond. Furthermore, the solid diamond is clearly closer to the minimizer of the total energy

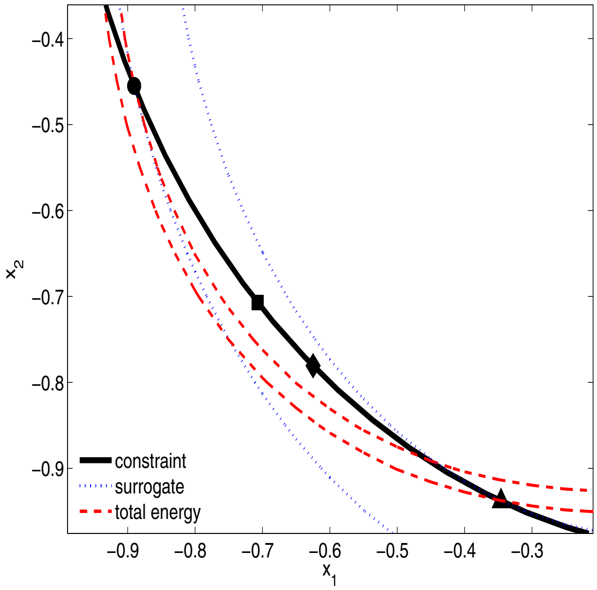


FIG. 4.2. *This is the same contour plot as shown in Figure* 3.3(b), *except that the figure also shows the effect of applying the trust region technique in SCF. The solution to the trust region subproblem or the penalized problem* (4.5) *is marked by the solid diamond, which lies to the upper right of the inner dashed contour. Clearly, it is closer to the true minimizer (the small solid square) than $\hat{x}$ (the small solid circle), indicating a reduction in the total energy in the SCF iteration.*

(the solid square) than either the current approximation $\hat{x}$ (the small solid circle) or the minimizer of the unpenalized surrogate $\hat{x}_q$ (the solid triangle).

5. Direct minimization of the total energy. A monotonic reduction of the total energy can be guaranteed in TRSCF if the penalty parameter σ is sufficiently large. However, choosing a large penalty parameter may lead to a slow convergence rate for large atomistic systems [4] because one is forced to take a shorter step in each TRSCF iteration. In [31], TRSCF is combined with a DIIS-like acceleration scheme to improve the convergence rate. In this paper, we proposed to use TRSCF within a DCM that we developed in [36].

Instead of minimizing $E_{total}(X)$ by minimizing a sequence of quadratic surrogate functions, the DCM algorithm minimizes the $E_{total}(X)$ directly.

This general approach has been discussed in a number of papers [2, 9, 15, 23, 30, 32, 33]. In [23, 30], a conjugate gradient (CG) type of algorithm is used to minimize the total energy. The minimization is carried out "band-by-band"; i.e., the total energy is minimized with respect to one wave function at a time. For the jth band (wavefunction), a search direction $p_j^{(i)}$ is generated from a linear combination of the wavefunction $x_j^{(i)} = X^{(i)}e_j$ and the residual

$$r_j = H^{(i)}x_j^{(i)} - x_j^{(i)}\lambda_j,$$

where λ_j is the jth eigenvalue of the projected Hamiltonian ${X^{(i)}}^* H^{(i)} X^{(i)}$. Note that r_j is simply the jth column of the gradient matrix $\nabla_X \mathcal{L}(X^{(i)})$. Similar to the standard CG algorithm, the linear combination of $x_j^{(i)}$ and r_j is chosen so that $p_j^{(i)}$ is

$H^{(i)}$-conjugate to the previous search direction $p_j^{(i-1)}$. The new wave function $x_j^{(i+1)}$ is then computed by minimizing the KS total energy in the subspace spanned by $x_j^{(i)}$ and $p_j^{(i)}$. To simplify this minimization problem, $p_j^{(i)}$ is first orthogonalized against $x_j^{(i)}$ and normalized so that $\|p_j^{(i)}\| = 1$. The new wave function is then parameterized by

$$x_j^{(i+1)} = x_j^{(i)} \cos\theta + p_j^{(i)} \sin\theta,$$

where the optimal θ is obtained by a standard line search procedure. Instead of using the KS function to perform the line search, Teter, Payne, and Allan [30] proposed using a surrogate function that is cheaper to evaluate. However, this approach was shown in [15] to be less efficient than the SCF iteration. We believe this is primarily due to the "band-by-band" nature of the algorithm.

The methods presented in [9, 2, 33, 32] were designed to minimize the total energy with respect to all wave functions (associated with the occupied state) simultaneously. The method developed in [9] modifies the unconstrained CG search direction so that the orthonormality constraint $X^*X = I_k$ can be satisfied. The approaches taken in [2, 32] reparameterize the search direction so that standard unconstrained minimization can be used directly. The algorithm developed in [33] first computes the search direction via a limited-memory BFGS (Broyden–Fletcher–Goldfarb–Shannon) scheme [19], and the search direction is then modified through a parallel transport technique [7] to ensure that the orthonormality constraint $X^*X = I_k$ is satisfied in the line search procedure. In all of these methods, the search direction is computed first, and a step length is then determined to reduce the total energy along the computed search direction.

The direct minimization algorithm that we presented in [36] also seeks the optimal wave functions associated with all occupied states simultaneously. However, we choose the search direction and the step length simultaneously from a subspace that consists of the existing wave functions $X^{(i)}$, the gradient of the Lagrangian (2.4), and the search direction produced in the previous iteration. A special strategy is developed to minimize the total energy within the search space while maintaining the orthonormality constraint required for $X^{(i+1)}$. This strategy requires us to solve a projected nonlinear eigenvalue problem, as we will illustrate below.

Let $R^{(i)}$ be the preconditioned gradient of the Lagrangian (2.4) with respect to X evaluated at $X^{(i)}$, and let $P^{(i-1)}$ be the search direction obtained in the $(i-1)$th iteration. In our algorithm, the wave function update is performed within the $3k$-dimensional subspace spanned by $X^{(i)}$, $R^{(i)}$, and $P^{(i-1)}$. This is in the same spirit as the locally optimal block preconditioned conjugate gradient (LOBPCG) algorithm proposed in [14] for solving large-scale linear eigenvalue problems. Note that the inclusion of $P^{(i-1)}$ is important. It prevents the search direction constructed at the ith step from being parallel to the steepest descent direction which often results in a tiny change between $X^{(i)}$ and $X^{(i+1)}$ (hence a small reduction in E_{total} from $X^{(i)}$ to $X^{(i+1)}$).

If we let

$$Y = (X^{(i)},\ R^{(i)},\ P^{(i-1)}),$$

we can then express the new approximation, $X^{(i+1)}$, by

$$X^{(i+1)} = YG, \tag{5.1}$$

where $G \in \mathbb{C}^{3k\times k}$ is chosen to minimize $\hat{E}(G) \equiv E_{total}(YG)$; i.e., we must solve

$$\begin{aligned} &\min_G E_{total}(YG) \\ &\text{s.t. } \; G^* Y^T Y G = I_k. \end{aligned} \tag{5.2}$$

The first order necessary condition of (5.2) can be derived by examining the gradient of the Lagrangian associated with $\hat{E}(G)$ (with respect to G). It is easy to verify [36] that

$$\nabla_G \hat{E}(G) = \hat{H}(G)G, \tag{5.3}$$

where

$$\hat{H}(G) = Y^* \left[\frac{1}{2}L + D_{ion} + \sum_\ell w_\ell w_\ell^* + \mathrm{Diag}\Big(S\rho(YG)\Big) + \mathrm{Diag}\Big(\mu_{xc}(\rho(YG))\Big)\right] Y. \tag{5.4}$$

(Note that (5.4) has been scaled by 1/2 to be consistent with the convention used in the electronic structure community.)

Consequently, solving (5.2) is equivalent to solving

$$\hat{H}(G)G = BG\Omega_k, \qquad G^*BG = I_k, \tag{5.5}$$

where $B = Y^*Y$ and the $k \times k$ diagonal matrix Ω_k contains the k smallest eigenvalues of (5.5).

Note that the projected nonlinear eigenvalue problem defined by (5.5) is much smaller than the nonlinear eigenvalue solved in an SCF iteration. The reduction in size provides us with more flexibility in choosing appropriate algorithms to solve the nonlinear eigenvalue problem. In particular, we can apply the TRSCF iteration introduced in section 4 to compute the desired eigenpairs of (5.5). The presence of the mass matrix B does not pose any difficulty in defining a trust region. Through a change of variables (i.e., let $\tilde{G} = RG$, where R is the Cholesky factor of B, i.e., $G = R^H R$), we can easily show that the first order necessary condition of the penalized quadratic surrogate defined at $G^{(i)}$ can be expressed by

$$\left[\hat{H}(G^{(i)}) - \sigma B G^{(i)} {G^{(i)}}^* B\right] G = BG\Omega_k, \qquad G^*BG = I_k. \tag{5.6}$$

The generalized linear eigenvalue problem defined by (5.6) can be solved by calling an appropriate LAPACK [1] subroutine. Choosing a sufficiently large penalty parameter will guarantee that $E_{total}(YG)$ decreases monotonically. Furthermore, it should be noted that it is not necessary to solve (5.5) to full accuracy in the early stage of the direct minimization process, because all we need is a G that yields sufficient decrease in the objective function within the subspace spanned by columns of Y.

Once G is computed, we can update the wave function following (5.1). In addition, we can compute the search direction associated with this update using [14]

$$P^{(i)} \equiv X^{(i+1)} - X^{(i)}G(1:k,:) = Y(:,k+1:3k)G(k+1:3k,:).$$

Because the solution to (5.5) ensures that columns of $X^{(i+1)}$ are orthonormal, there is no need to explicitly orthogonalize $P^{(i)}$ against $X^{(i)}$ in our algorithm.

A complete description of the constrained minimization algorithm is shown in Figure 5.1. We should point out that solving the projected optimization problem

```
ALGORITHM: A CONSTRAINED MINIMIZATION ALGORITHM FOR TOTAL ENERGY MINIMIZATION

Input:   An initial set of wave functions $X^{(0)} \in \mathbb{C}^{n\times k}$, where $k$ is the number
         of occupied states; the matrices $L$, $D_{ion}$, $S$; the vectors $w_\ell$,
         $\ell = 1, 2, \ldots$. The derivative of the exchange-correlation function
         $\mu_{xc}(x)$; a preconditioner $K$;
Output:  $X \in \mathbb{C}^{n\times k}$ such that the KS total energy function $E_{total}(X)$ is
         minimized and $X^*X = I_k$.
1.       Orthonormalize $X^{(0)}$ such that ${X^{(0)}}^* X^{(0)} = I_k$;
2.       for $i = 0, 1, 2, \ldots$ until convergence
3.          Compute $\Theta = {X^{(i)}}^* H^{(i)} X^{(i)}$;
4.          Compute $R = K^{-1}\left[H^{(i)}X^{(i)} - X^{(i)}\Theta\right]$,
5.          if ($i > 1$) then
               $Y \leftarrow (X^{(i)}, R, P^{(i-1)})$
            else
               $Y \leftarrow (X^{(i)}, R)$;
            endif
6.          $B \leftarrow Y^*Y$;
7.          Find $G \in \mathbb{C}^{2k\times 2k}$ or $\mathbb{C}^{3k\times 3k}$ that minimizes $E_{total}(YG)$
            subject to the constraint $G^*BG = I$;
8.          Set $X^{(i+1)} = YG$;
9.          if ($i > 1$) then
               $P^{(i)} \leftarrow Y(:, k+1:3k)G(k+1:3k, :)$;
            else
               $P^{(i)} \leftarrow Y(:, k+1:2k)G(k+1:2k, :)$;
            endif
10.      end for
```

FIG. 5.1. *A direct constrained minimization algorithm for total energy minimization.*

in step 7 of the algorithm requires us to evaluate the projected Hamiltonian (5.4) repeatedly as we search for the best G. However, since the first three terms of $\hat{H}$ do not depend on G, they can be computed and stored in advance. Only the last two terms of (5.4) need to be updated. These updates require that the charge density, the Hartree, and the exchange-correlation potentials be recomputed.

6. Numerical examples.

In this section, we compare the performance of the DCM algorithm presented in the previous section with that of the SCF iteration implemented in the software package PEtot [34] through two numerical examples. In PEtot, single particle wave functions are discretized by a spectral method using plane waves as the basis. These basis functions are eigenfunction of the Laplacian operator (L) associated with the kinetic energy of the atomistic system. Thus, PEtot stores only the Fourier coefficients of each wave function $x_j = Xe_j$ instead of x_j itself so that $y \leftarrow Lx_j$ can be carried out in $\mathcal{O}(n)$ floating point operations (flops) in the frequency space. However, because the potential terms of the Hamiltonian (with the exception of the nonlocal ionic potential) are diagonal in the spatial domain, PEtot converts the

Fourier space representation of x_j into the real space representation before operations involving these potential terms are performed. The complexity of this conversion is $\mathcal{O}(n \log n)$ when it is carried out by a fast Fourier transform (FFT). We measure the convergence of both algorithms by examining the relative reduction of the total energy computed in each outer iteration. The relative reduction is evaluated by

$$\Delta E_i = E_{total}(X^{(i)}) - E_{min},$$

where E_{min} is a lower bound of the total energy.

In a PEtot SCF iteration, the minimization of the surrogate (3.1) is accomplished by applying a preconditioned conjugate gradient (PCG) algorithm to minimize the Rayleigh quotient $x^* H^{(i)} x / x^* x$. Explicit deflation is put in place to accelerate the convergence of the smallest k eigenpairs. Each PCG iteration requires a single matrix vector (MATVEC) multiplication followed by a preconditioning operation. When n is sufficiently large, the complexity of each MATVEC is dominated by the cost of the FFT calculation used to convert the Fourier space representation of x_j to the real space representation. The Laplacian operator L is used as the preconditioner. Because it is diagonal in the frequency space, the cost of preconditioning is relatively small compared to a MATVEC. If m PCG iterations are taken on average to compute an approximate eigenpair of $H^{(i)}$, then the total number of MATVECs used per SCF iteration is $m \times k$.

In the DCM algorithm, k MATVECs are performed in each outer iteration to obtain the gradient. When TRSCF is used to solve the projected problem (5.5), each outer DCM iteration contains a number of inner TRSCF iterations in which the projected Hamiltonian (5.4) must be updated repeatedly. The update of the projected Hartree potential requires us to compute $S\rho(YG)$. Because S is the inverse of L, this calculation is typically carried out by a fast Poisson solver. The complexity of this computation is approximately $\mathcal{O}(n \log n)$, which is equivalent to a single MATVEC used in the SCF iteration asymptotically. Thus, if p inner TRSCF iterations are taken in the DCM algorithm, the total number of MATVECs used per DCM iteration is $k + p$.

6.1. The PtNiO system. We applied both algorithms to a relatively large system consisting of 9 atoms and 86 valence electrons. It represents a thin PtNi slab with one O atom attached to the surface. The system is used for catalysis to dissociate O_2 molecules. The wave function is defined on a $96 \times 48 \times 48$ real space grid. The number of plane wave basis functions used in the Fourier representation is 15181, and the number of occupied states for this molecule is $k = 43$.

In the SCF calculation, we set the convergence tolerance of each PCG run to $\tau = 10^{-12}$, and the maximum number of PCG iterations allowed to 10. That is, we terminate the PCG iteration when

$$\|H^{(i)} x_j^{(i)} - \lambda_j^{(i)} x_j^{(i)}\| \leq 10^{-12},$$

or when the number of PCG iterations taken reaches 10. In our experiment, the PCG convergence tolerance was never reached before the maximum number of iterations were taken. Thus each outer SCF iteration consumed 43 × 10 MATVECs. Both Pulay [25] (DIIS) and Kerker [13] charge-mixing schemes were used in the outer SCF iteration to accelerate the convergence.

In the DCM calculation, the projected minimization problem was solved by applying TRSCF iteration to (5.5). We set the number of inner TRSCF iterations to 5.

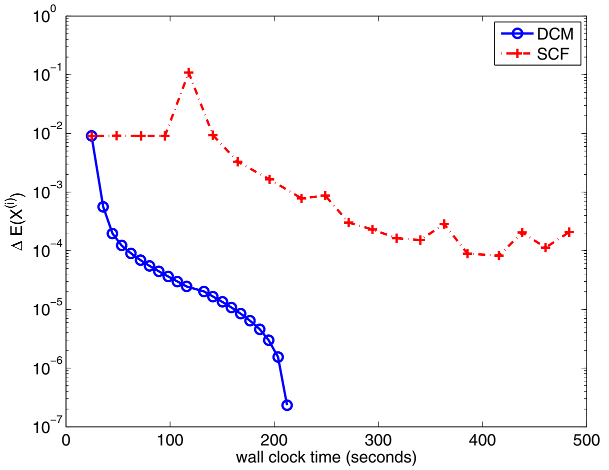


FIG. 6.1. *Comparing the convergence of SCF and DCM when they are applied to the PtNiO system.*

Thus, the number of MATVECs used in each DCM iteration is roughly $5 + 43$, which is significantly smaller than that used in SCF.

Both SCF and DCM have been parallelized using MPI (message passing interface). For the PtNiO system, we ran both codes on the IBM SP maintained at the National Energy Research Scientific Computing (NERSC) Center using 64 CPUs. Each IBM SP node contains 16 Power3 CPUs and 16 GB memory. Each Power3 CPU runs at a 375 MHz clock speed, and has 2 MB L2 cache. Figure 6.1 shows that DCM exhibits monotonic convergence whereas the total energy does not decrease monotonically in SCF even when both Pulay and Kerker charge-mixing schemes are used to stabilize the algorithm. After 20 iterations, DCM was able to reach a much lower KS total energy level.

6.2. The $\mathbf{Si_{29}H_{36}}$ cluster. In the second example, we show that the trust region DCM algorithm is more efficient than the SCF iteration (accelerated by charge mixing) for atomistic systems that are "well behaved" (in the sense that the total energy decreases monotonically in SCF). A smaller example was presented in our earlier work [36] in which both DCM and SCF were applied to the SiH_4 system. We showed that DCM was almost four times faster than SCF when both were executed on 16 Power3 CPUs. No trust region was used in that experiment.

In this example, we apply both trust region DCM and SCF (accelerated by charge mixing) to a larger silicon cluster that contains 29 silicon atoms and 36 hydrogen atoms. The total number of valence electrons in the system is 152; i.e., the number of occupied states is $k = 76$. The system is discretized by plane waves on a real space grid of $96 \times 96 \times 96$. The number of plane wave basis functions used is 35585. We ran both DCM and SCF on 128 Power3 CPUs at NERSC. The PCG tolerance and iteration limits used in this example are the same as those set in the PtNiO example. We also set the number of inner TRSCF iterations used in DCM to 5.

Figure 6.2 shows that the total energy decreases monotonically in both the DCM and SCF runs for this problem. However, the reduction of the total energy is much faster in DCM than it is in SCF.

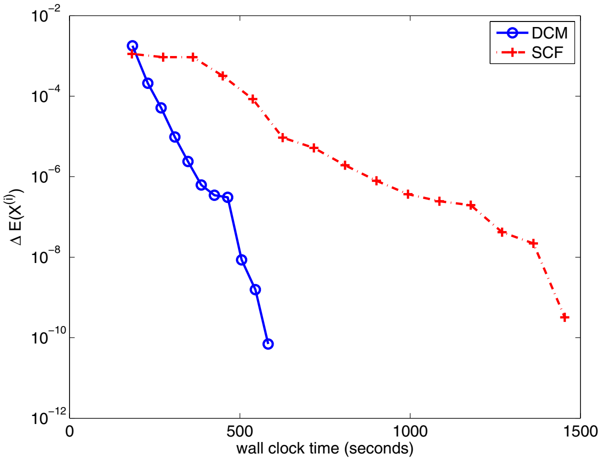


FIG. 6.2. *Comparing the convergence of SCF and DCM when they are applied to the $Si_{29}H_{36}$ cluster.*

7. Concluding remarks.

We viewed the SCF iteration, commonly used for solving the KS equation, as an optimization procedure that minimizes the KS total energy indirectly by minimizing a sequence of quadratic surrogate. Such a viewpoint allows us to easily explain when SCF works and how it can fail. It also allows us to devise techniques that can either stabilize or accelerate the SCF iteration. We showed that the convergence of SCF can be stabilized by introducing a quadratic constraint into the surrogate minimization problem. Such a constraint restricts the wave function update to a small neighborhood of the current iterate at which the gradients of the KS total energy and the surrogate match, hence defining a "trust region." However, applying a trust region-based SCF iteration directly to the KS equation may lead to slow convergence. We proposed using the trust region technique within the DCM algorithm developed in [36] to compute optimal search directions and step lengths simultaneously. We demonstrated through two numerical examples that such a scheme outperforms the SCF iteration combined with charge mixing in terms of both efficiency and reliability. We should point out that our numerical results are still somewhat preliminary. For atomistic systems that are "well behaved," the performance difference between SCF and DCM may be less dramatic if the single vector PCG algorithm used in PEtot to solve the surrogate minimization problem (3.1) is replaced by a block algorithm. Also, it will be interesting to see how these two methods differ in performance when the number of electrons becomes a larger fraction of n. In that case, dense matrix operations will constitute a significant portion of the computational cost in both DCM and SCF. We will undertake a more systematic performance analysis in a future study that will also take into account different discretization schemes.

REFERENCES

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen, *LAPACK Users' Guide*, 2nd ed., SIAM, Philadelphia, 1995.

[2] T. A. Arias, M. C. Payne, and J. D. Joannopoulos, *Ab initio molecular dynamics: Analytically continued energy functionals and insights into iterative solutions*, Phys. Rev. Lett., 69 (1992), pp. 1077–1080.

[3] E. L. Briggs, D. J. Sullivan, and J. Berholc, *Real-space multigrid-based approach to large-scale electronic structure calculations*, Phys. Rev. B, 54 (2001), pp. 14362–14375.

[4] C. Le Bris, *Computational chemistry from the perspective of numerical analysis*, Acta Numer., 14 (2005), pp. 363–444.

[5] A. R. Conn, N. I. M. Gould, and Ph. L. Toint, *Trust-Region Methods*, MPS-SIAM Ser. Optim. 1, SIAM, Philadelphia, 2000.

[6] E. R. Davidson, *The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real symmetric matrices*, J. Comput. Phys., 17 (1975), pp. 87–94.

[7] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 303–353.

[8] J. L. Fatteberg, *A block Rayleigh quotient iteration with local quadratic convergence*, Electron. Trans. Numer. Anal., 7 (1998), pp. 56–74.

[9] M. J. Gillan, *Calculation of the vacancy formation in aluminum*, J. Phys. Condens. Matter, 1 (1989), pp. 689–711.

[10] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins, Baltimore, MD, 1989.

[11] M. Heiskanen, T. Torsti, M. J. Puska, and R. M. Nieminen, *Multigrid method for electronic structure calculations*, Phys. Rev. B, 63 (2001), pp. 1–8.

[12] M. R. Hestenes and W. Karush, *A method of gradients for the calculation of the characteristic roots and vectors of a real symmetric matrix*, J. Res. National Bureau of Standards, 47 (1951), pp. 45–61.

[13] G. P. Kerker, *Efficient iteration scheme for self-consistent pseudopotential calculations*, Phys. Rev. B, 23 (1981), pp. 3082–3084.

[14] A. V. Knyazev, *Toward the optimal preconditioned eigensolver: Locally optimal block preconditioned conjugate gradient method*, SIAM J. Sci. Comput., 23 (2001), pp. 517–541.

[15] G. Kresse and J. Furthmüller, *Efficiency of ab initio total energy calculations for metals and semiconductors using a plane-wave basis set*, Comput. Materials Sci., 6 (1996), pp. 15–50.

[16] G. Kresse and J. Furtthmüller, *Efficient iterative schemes for* ab initio *total-energy calculations using a plane-wave basis set*, Phys. Rev. B, 54 (1996), pp. 11169–11185.

[17] C. Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, J. Res. National Bureau of Standards, 45 (1950), pp. 255–282.

[18] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, Software Environ. Tools 6, SIAM, Philadelphia, 1998.

[19] D. C. Liu and J. Nocedal, *On the limited memory method for large scale optimization*, Math. Programming B, 45 (1989), pp. 503–528.

[20] J. Nocedal and S. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.

[21] J. Olsen, P. Jorgensen, and J. Simons, *Passing the one-billion limit in full configuration-interaction (FCI) calculations*, Chem. Phys. Lett., 169 (1990), pp. 463–472.

[22] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice–Hall, Englewood Cliffs, NJ, 1980.

[23] M. C. Payne, M. P. Teter, D. C. Allen, T. A. Arias, and J. D. Joannopoulos, *Iterative minimization techniques for* ab initio *total energy calculation: Molecular dynamics and conjugate gradients*, Rev. Modern Phys., 64 (1992), pp. 1045–1097.

[24] P. Pulay, *Convergence acceleration of iterative sequences: The case of SCF iteration*, Chem. Phys. Lett., 73 (1980), pp. 393–398.

[25] P. Pulay, *Improved SCF convergence acceleration*, J. Comput. Chem., 3 (1982), pp. 556–560.

[26] Y. Saad, A. Sathopoulos, J. Chelikowsky, K. Wu, and S. Ogut, *Solution of large eigenvalue problems in electronic structure calculations*, BIT, 36 (1996), pp. 563–578.

[27] A. H. Sameh and J. A. Wisniewski, *A trace minimization algorithm for the generalized eigenvalue problem*, SIAM J. Numer. Anal., 19 (1982), pp. 1243–1259.

[28] V. R. Saunders and I. H. Hillier, *A level-shifting method for converging closed shell Hartree-Fock wavefunctions*, Int. J. Quantum Chem., 7 (1973), pp. 699–705.

[29] G. L. G. Sleijpen and H. A. Van der Vorst, *A Jacobi–Davidson iteration method for linear eigenvalue problems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 401–425.

[30] M. P. Teter, M. C. Payne, and D. C. Allan, *Solution of Schrödinger's equation for large systems*, Phys. Rev. B, 40 (1989), pp. 12255–12263.

[31] L. Thogersen, J. Olsen, D. Yeager, and P. Jorgensen, *The trust-region self-consistent field method: Towards a black-box optimization in Hartree-Fock and Kohn-Sham theories*, J. Chem. Phys., 121 (2004), pp. 16–27.

[32] J. VandeVondele and J. Hutter, *An efficient orbital transformation method for electronic structure calculations*, J. Chem. Phys., 118 (2003), pp. 4365–4369.

[33] T. Van Voorhis and M. Head-Gordon, *A geometric approach to direct minimization*, Molecular Phys., 100 (2002), pp. 1713–1721.

[34] L. Wang, *PETOT*, software package available online at http://hpcrd.lbl.gov/∼linwang/PEtot/PEtot.htm.

[35] K. Wu, A. Canning, H. D. Simon, and L.-W. Wang, *Thick-restart Lanczos method for electronic structure calculations*, J. Comput. Phys., 154 (1999), pp. 156–173.

[36] C. Yang, J. C. Meza, and L.-W. Wang, *A constrained optimization algorithm for total energy minimization in electronic structure calculation*, J. Comput. Phys., 217 (2006), pp. 709–721.