

Towards a potential-based conjugate gradient algorithm for order- N self-consistent total energy calculations

Xavier Gonze

Unité de Physico—Chimie et de Physique des Matériaux, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

(Received 23 February 1996)

The determination of the total energy within density-functional theory can be formulated as a minimization problem in a space of trial self-consistent potentials. In order to apply a conjugate-gradient algorithm to this problem, a formula for the computation of the gradient of the energy with respect to the self-consistent potential is proposed. The second derivative of the energy with respect to potential changes is also analyzed, in order to obtain an efficient preconditioning operator. The wave functions do not appear explicitly in this approach, so that order- N algorithms could take advantage of it. The results of preliminary tests are reported. [S0163-1829(96)07031-2]

Recently, algorithms for the computation of the electronic energy of simple (non-self-consistent) tight-binding model, scaling linearly with the system size, have been proposed.^{1,2} Also for the much more accurate *ab initio* density-functional theory (DFT), in the case of *fixed* potentials, it was shown that the same scaling [O(N) where N is a measure of the size of the system] can be achieved.^{2,3} This is a considerable advance, since usual algorithms scale like O(N^3) in the large-size limit.⁴⁻⁶

Going one step further, Annett⁷ has investigated the scaling of *self-consistent* DFT algorithms, based on the minimization of the energy either with respect to trial localized orbitals, or with respect to trial self-consistent potentials, showing that the self-consistency requirement could destroy the O(N) scaling. In the present paper, I will concentrate on the minimization with respect to trial self-consistent potentials. The basic underlying ideas were already proposed by Bendt and Zunger more than ten years ago.⁸ Bendt and Zunger, as well as Annett, give formulas for calculating the gradient of the electronic energy versus the trial self-consistent potential, in which either the wave functions or the independent-particle polarizability are explicitly used. If Bendt and Zunger were not worried by this fact [they wanted to compute these wave functions anyhow, and were not looking after an O(N) method], Annett points out that the fast potential-based O(N) methods proposed until now do not allow an easy computation of wave functions or the independent-particle polarizability.

In the present paper, it is shown how to obtain the gradient of the energy with respect to the self-consistent potential by using ingredients that are produced by non-self-consistent O(N) methods. This opens the way for potential-based conjugate gradient (CG) algorithms for self-consistent total-energy calculations. Preconditioning operators are also examined, and preliminary results are presented.

Supposing that a system of noninteracting electrons is placed in the fixed potential $u(\mathbf{r})$, its energy, functional of the potential u , is given by

$$E^{\text{NSC}}[u] = \min_{\{\phi_i \text{ such that } \langle \phi_i | \phi_j \rangle = \delta_{ij}\}} \left\{ \sum_i s \langle \phi_i | T + u | \phi_i \rangle \right\}, \quad (1)$$

where the indices i and j run on the occupied orbitals (constrained to be orthonormal), the number of which is equal to half the number of electrons N_e . The spin degeneracy, s , is equal to 2. The corresponding noninteracting kinetic energy and density are evaluated for the orbitals that minimize Eq. (1):

$$T[u] = \sum_i s \langle \phi_i | T | \phi_i \rangle, \quad (2)$$

and

$$n[u](\mathbf{r}) = \sum_i s \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}). \quad (3)$$

Using the Lagrange-Euler equations associated with the minimization procedure in Eq. (1), one finds that $n[u](\mathbf{r})$ is the gradient of the energy $E^{\text{NSC}}[u]$ with respect to the potential $u(\mathbf{r})$:

$$n[u](\mathbf{r}) = \frac{\delta E^{\text{NSC}}[u]}{\delta u(\mathbf{r})}. \quad (4)$$

Thus, the kinetic energy can be written

$$T[u] = E^{\text{NSC}}[u] - \int \frac{\delta E^{\text{NSC}}[u]}{\delta u(\mathbf{r})} u(\mathbf{r}) d\mathbf{r}. \quad (5)$$

Equations (4) and (5) are orbital-free expressions giving the density and kinetic energy. In what follows, it is supposed that these quantities, as well as the energy $E^{\text{NSC}}[u]$, can be obtained with a fast O(N) algorithm. No knowledge of the orbitals will be assumed.

In the DFT, the potential that the noninteracting electrons feel has to be determined self-consistently. It includes the bare external potential and contributions from the Hartree

and the exchange-correlation terms, which are functionals of the density. In an orbital-based formulation, one defines the self-consistent energy by the following minimization procedure:

$$E^{\text{SC}}[v_{\text{ext}}] = \min_{\{\phi_i \text{ such that } \langle \phi_i | \phi_j \rangle = \delta_{ij}\}} \left\{ \sum_i s \langle \phi_i | T | \phi_i \rangle + \int n[\{\phi\}](\mathbf{r}) v_{\text{ext}}(\mathbf{r}) d\mathbf{r} + E_{\text{Hxc}}[n[\{\phi\}]] \right\}, \quad (6)$$

where the density is given by

$$n[\{\phi\}](\mathbf{r}) = \sum_i s \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}). \quad (7)$$

and $E_{\text{Hxc}}[n]$ is the sum of Hartree and exchange-correlation energies. The latter part must be approximated, since the exact exchange-correlation functional is not known.⁵

Instead of this formulation, relying on orbitals, one can obtain the same value of the energy and density by an unconstrained minimization procedure in the space of trial self-consistent potentials $u(\mathbf{r})$, as follows:^{5,7,8}

$$E^{\text{SC}}[v_{\text{ext}}] = \min_u E^{\text{trial}}[v_{\text{ext}}; u], \quad (8)$$

where the trial energy $E^{\text{trial}}[v_{\text{ext}}; u]$, a functional of the external potential v_{ext} and the trial self-consistent potential u , can be expressed as⁹

$$E^{\text{trial}}[v_{\text{ext}}; u] = T[u] + \int n[u](\mathbf{r}) v_{\text{ext}}(\mathbf{r}) d\mathbf{r} + E_{\text{Hxc}}[n[u]], \quad (9)$$

or equivalently as

$$E^{\text{trial}}[v_{\text{ext}}; u] = E^{\text{NSC}}[u] + \int n[u](\mathbf{r}) [v_{\text{ext}}(\mathbf{r}) - u(\mathbf{r})] d\mathbf{r} + E_{\text{Hxc}}[n[u]]. \quad (10)$$

The gradient of $E^{\text{trial}}[v_{\text{ext}}; u]$ with respect to $u(\mathbf{r})$, obtained by the differentiation of Eq. (10), and using Eq. (4), is

$$\frac{\delta E^{\text{trial}}[v_{\text{ext}}; u]}{\delta u(\mathbf{r})} = \int \frac{\delta n[u](\mathbf{r}')}{\delta u(\mathbf{r})} \left(v_{\text{ext}}(\mathbf{r}') + \frac{\delta E_{\text{Hxc}}[n]}{\delta n(\mathbf{r}')} \Big|_{n[u]} - u(\mathbf{r}') \right) d\mathbf{r}'. \quad (11)$$

At the minimum of the trial energy, the gradient Eq. (11) vanishes, and the trial self-consistent potential u becomes equal to the Kohn-Sham potential:

$$u_{\text{min}}(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + \frac{\delta E_{\text{Hxc}}[n]}{\delta n(\mathbf{r})} \Big|_{n[u_{\text{min}}]} = v_{\text{KS}}(\mathbf{r}). \quad (12)$$

In search for a practical CG algorithm¹⁰ to be applied to Eq. (8), one needs an efficient way to compute the gradient of $E^{\text{trial}}[v_{\text{ext}}; u]$ with respect to $u(\mathbf{r})$. Equation (11) is not the answer to this requirement, because it is not clear how to compute efficiently $\delta n[u](\mathbf{r}')/\delta u(\mathbf{r})$. Also, the gradient formula of Bendt and Zunger [Eq. (9) of Ref. 8] makes use of orbitals (valence as well as conduction orbitals).

Actually, $\delta n[u](\mathbf{r}')/\delta u(\mathbf{r})$ is not explicitly needed. This quantity, that is equal to the independent-particle polarizability of the system,¹¹ $\chi_0(\mathbf{r}, \mathbf{r}')$, is also the second derivative of the non-self-consistent energy with respect to change of potentials in \mathbf{r} and \mathbf{r}' , or the derivative of the density n at \mathbf{r} with respect to the potential u at \mathbf{r}' ,

$$\frac{\delta n[u](\mathbf{r}')}{\delta u(\mathbf{r})} = \frac{\delta^2 E^{\text{NSC}}[u]}{\delta u(\mathbf{r}) \delta u(\mathbf{r}')} = \chi_0(\mathbf{r}, \mathbf{r}') = \frac{\delta n[u](\mathbf{r}')}{\delta u(\mathbf{r}')}. \quad (13)$$

Inserting this result in Eq. (11) gives

$$\frac{\delta E^{\text{trial}}[v_{\text{ext}}; u]}{\delta u(\mathbf{r})} = \int \frac{\delta n[u](\mathbf{r}')}{\delta u(\mathbf{r}')} \left(v_{\text{ext}}(\mathbf{r}') + \frac{\delta E_{\text{Hxc}}[n]}{\delta n(\mathbf{r}')} \Big|_{n[u]} - u(\mathbf{r}') \right) d\mathbf{r}', \quad (14)$$

which is nothing more than the linear rate of change of the density due to a change in trial potential u by $\{v_{\text{ext}}(\mathbf{r}') + [\delta E_{\text{Hxc}}[n]/\delta n(\mathbf{r}')]|_{n[u]} - u(\mathbf{r}')\}$. In a finite-difference approach, for λ small enough,

$$\frac{\delta E^{\text{trial}}[v_{\text{ext}}; u]}{\delta u(\mathbf{r})} \approx \frac{1}{\lambda} \left\{ n \left[u + \lambda \left(v_{\text{ext}} + \frac{\delta E_{\text{Hxc}}[n]}{\delta n} \Big|_{n[u]} - u \right) \right](\mathbf{r}) - n[u](\mathbf{r}) \right\}. \quad (15)$$

Neither the wave functions, nor the independent-particle polarizability appear in Eq. (15). Since the density is one of the output of the fast $O(N)$ non-self-consistent calculation, while the computation of this finite-difference formula relies only on the differences of densities, the $O(N)$ character of the underlying non-self-consistent algorithm is maintained. So, this formula provides an efficient way to evaluate the gradient of the energy with respect to change of self-consistent potential.

However, the knowledge of the gradient is not the only ingredient needed for an efficient CG algorithm. A good preconditioning operator can prove crucial.^{6,7} Indeed, the rate of convergence of a CG algorithm depends mainly on the ratio of extremal eigenvalues of the Hessian matrix H : When this ratio is close to 1, the convergence is extremely fast, while a larger ratio leads to a slower convergence. In this last case, applying a preconditioner operator K that mimicks as closely as possible the inverse of the Hessian matrix will improve the rate of convergence, since it is then governed by the eigenvalues of KH .

In our case, the Hessian matrix at the minimum $u = u_{\text{min}}$ is exactly

$$\begin{aligned}
\frac{\delta^2 E^{\text{trial}}[v_{\text{ext}}; u_{\text{min}}]}{\delta u(\mathbf{r}) \delta u(\mathbf{r}')} &= -\frac{\delta^2 E^{\text{NSC}}[u_{\text{min}}]}{\delta u(\mathbf{r}) \delta u(\mathbf{r}')} + \int \int \frac{\delta^2 E^{\text{NSC}}[u_{\text{min}}]}{\delta u(\mathbf{r}) \delta u(\mathbf{r}'')} \frac{\delta^2 E_{\text{Hxc}}[n]}{\delta n(\mathbf{r}'') \delta n(\mathbf{r}''')} \Big|_{n=n[u_{\text{min}}]} \frac{\delta^2 E^{\text{NSC}}[u_{\text{min}}]}{\delta u(\mathbf{r}''') \delta u(\mathbf{r}')} d\mathbf{r}'' d\mathbf{r}''' \\
&= -\int \frac{\delta^2 E^{\text{NSC}}[u_{\text{min}}]}{\delta u(\mathbf{r}) \delta u(\mathbf{r}'')} \left(\delta(\mathbf{r}'' - \mathbf{r}') - \int \frac{\delta^2 E_{\text{Hxc}}[n]}{\delta n(\mathbf{r}'') \delta n(\mathbf{r}''')} \Big|_{n=n[u_{\text{min}}]} \frac{\delta^2 E^{\text{NSC}}[u_{\text{min}}]}{\delta u(\mathbf{r}''') \delta u(\mathbf{r}')} d\mathbf{r}''' \right) d\mathbf{r}'' \\
&= -\int \chi_0(\mathbf{r}, \mathbf{r}'') \epsilon_e(\mathbf{r}'', \mathbf{r}') d\mathbf{r}'', \tag{16}
\end{aligned}$$

where ϵ_e is the electron dielectric response function.¹¹ The ratio of extremal eigenvalues of this Hessian unfortunately increases with the system size.⁷

If we take $K = H^{-1} = -(\chi_0 \epsilon_e)^{-1}$, and apply this (best) preconditioner operator to minus the gradient direction Eq. (14), the independent-particle polarizability χ_0 *disappears*. The preconditioned change of potential is

$$\Delta u(\mathbf{r}) = \int \epsilon_e^{-1}(\mathbf{r}, \mathbf{r}') \left(v_{\text{ext}}(\mathbf{r}') + \frac{\delta E_{\text{Hxc}}[n[u]]}{\delta n(\mathbf{r}')} - u(\mathbf{r}') \right) d\mathbf{r}'. \tag{17}$$

Supposing that we are close to the minimum (and thus in the quadratic region of the energy functional), this formula would give the change of potential needed to reach the self-consistency in one step only, independently of the size of the system.

However, the inverse dielectric function ϵ_e^{-1} needed in Eq. (17) is not known *a priori*. In order to bypass this lack of knowledge, one can either use a model (approximate) dielectric function, or decide that the preconditioning brought by the inverse of the independent-particle polarizability, $K = -\chi_0^{-1}$, is good enough. In this case, the preconditioned change of potential is given by

$$\Delta u(\mathbf{r}) = v_{\text{ext}}(\mathbf{r}) + \frac{\delta E_{\text{Hxc}}[n[u]]}{\delta n(\mathbf{r})} - u(\mathbf{r}). \tag{18}$$

Neither the wave functions, nor the independent-particle polarizability χ_0 appear in Eq. (18).

With $K = -\chi_0^{-1}$, giving Eq. (18), the eigenvalues of the modified (preconditioned) Hessian are those of the electronic dielectric response function ϵ_e . In a cell calculation, the only fluctuations of density that are allowed are those commensurate with the cell geometry, so that not all eigenvalues of ϵ_e will influence the rate of convergence.¹² In the limit of very large supercells, however, nearly all the eigenvalues will be sampled. For a system with a gap (molecule or insulator), it is expected that the eigenvalues are distributed between a value close to 1 and a finite positive value [on the order of the macroscopic dielectric constant for an insulator, see the analysis of the dielectric band structure in Ref. 13, and Fig. 1(b) of Ref. 11]. For such a system, when the size of the cell is increased, the ratio of extremal eigenvalues will reach a plateau after an eventual initial growth, due to the increase of the allowed fluctuations wavelength: The *charge sloshing* problem will affect the rate of convergence, but not the O(N) scaling.¹⁴ So, a better preconditioning is not mandatory to keep the same O(N) scaling of the overall algorithm in that case. For metals, this is not true, and the O(N)

behavior of the non-self-consistent algorithms cannot be maintained without a better preconditioning method: one obtains an O(N^{4/3}) scaling.⁷

A preconditioned CG algorithm based on Eqs. (15) and (18), was implemented and tested on crystalline silicon (with two atoms per unit cell).¹⁰ Instead of using an O(N) algorithm, the non-self-consistent energy and density were obtained by simple diagonalization of a plane wave pseudopotential Hamiltonian,⁵ with a six Hartree kinetic energy cutoff, and two special points in the irreducible part of the Brillouin zone. The LDA exchange-correlation functional of Ceperley and Alder⁵ was used. First, the validity of Eq. (15) was carefully checked. Then a preconditioned CG minimization run was performed. The starting potential was the bare electron potential, without self-consistent screening. The initial error on the energy, compared with a fully converged value, was 10⁻¹ Ha/cell. After four line minimizations, this error was decreased to 10⁻¹⁰ Ha/cell. Each line minimization needed three non-self-consistent calculations at most.

Actually there are different ways to implement a CG algorithm, and the purpose of this preliminary test was to demonstrate the feasibility of the idea, not to obtain the best possible efficiency. For this reason, the details of the implementation are not described in this paper.

In this test case, the ratio of extremal eigenvalues of the preconditioned Hessian matrix, for the fluctuations of density compatible with the periodic boundary conditions and the symmetry, was rather small, about 1.5, so that the replacement of Eq. (17) by Eq. (18) does not spoil the efficiency. For larger silicon cells, the ratio will raise to about 13, in which case a better preconditioning would be advantageous.

I now discuss briefly improved preconditioning schemes. Since the preconditioning operator should be easy to apply, one aims at an approximation of the inverse dielectric response function that is either diagonal in reciprocal space, or diagonal in real space, or separable, or a multiplicative (or additive) combination of such forms. In the limit of long wave length fluctuations, the inverse dielectric matrix tends to be diagonal in reciprocal space: The local field effects are usually corrections on the order of a few percents to the macroscopic dielectric constant of silicon and other insulators.¹¹ Levine and Louie¹⁵ have proposed a model dielectric function that is diagonal in reciprocal space, and that relies only on two adjustable parameters. These could be fixed by sum-rule requirements, or by performing their adjustment in the course of the computer run. Other models are discussed in Ref. 13. So, I tend to believe that all-purpose improved preconditioners can be designed.

In conclusion, in the present paper, some elements leading to a self-consistent potential-based conjugate-gradient $O(N)$ algorithm were proposed and analyzed. Neither the wave functions, nor the independent-particle polarizability appear in Eqs. (15), (17), or (18). Preliminary results showing the feasibility of these ideas have been presented. A conjugate gradient algorithm has the advantage of robustness and reduced memory needs over most other algorithms, while its efficiency is excellent. With the simple preconditioning operator $K = -\chi_0^{-1}$ used here, no parameter that ought to be tuned for each different problem is present. A better precon-

ditioning scheme is certainly possible, but should keep the same desirable features. Finally, a potential-based CG algorithm could also be combined with non-self-consistent $O(N^3)$ schemes that are suited for small or medium size calculations. Its efficiency should be compared with the band-by-band conjugate-gradient algorithm,⁶ as well as Broyden or Anderson methods.¹⁶

I acknowledge interesting discussions with S. Goedecker, and I thank Ph. Ghosez and G.-M. Rignanese for a careful reading of the manuscript. I am grateful to FNRS-Belgium for financial support.

-
- ¹R. Haydock, *Solid State Phys.* **35**, 215 (1980); D.A. Drabold and O.F. Sankey, *ibid.* **70**, 3631 (1993); M. Aoki, *ibid.* **71**, 3842 (1993); S. Goedecker and L. Colombo, *Phys. Rev. Lett.* **73**, 122 (1994); S. Goedecker and M. Teter, *Phys. Rev. B* **51**, 9455 (1995); L.-W. Wang and A. Zunger, *Comput. Mat. Sci.* **2**, 326 (1994).
- ²W. Yang, *Phys. Rev. Lett.* **66**, 1438 (1991); S. Baroni and P. Giannozzi, *Europhys. Lett.* **17**, 547 (1992); G. Galli and M. Parrinello, *Phys. Rev. Lett.* **69**, 3547 (1992); X.-P. Li, R.W. Nunes, and D. Vanderbilt, *Phys. Rev. B* **47**, 10 891 (1993); M.S. Daw, *ibid.* **47**, 10 895 (1993).
- ³F. Mauri, G. Galli, and R. Car, *Phys. Rev. B* **47**, 9973 (1993); J. Kim, F. Mauri, and G. Galli, *ibid.* **52**, 1640 (1995); P. Ordejon, D. Drabold, R.M. Martin, and M.P. Grumbach, *ibid.* **51**, 1456 (1995); W. Kohn, *Chem. Phys. Lett.* **208**, 167 (1993); E.B. Stechel, A.R. Williams, and P.J. Feibelman, *Phys. Rev. B* **49**, 10 088 (1994); W. Hierse and E.B. Stechel, *ibid.* **50**, 17 811 (1994); E. Hernandez and M.J. Gillan, *ibid.* **51**, 10 157 (1995); Y. Wang *et al.*, *Phys. Rev. Lett.* **75**, 2867 (1995).
- ⁴R. Car and M. Parinello, *Phys. Rev. Lett.* **55**, 2471 (1985).
- ⁵W.E. Pickett, *Comput. Phys. Rep.* **9**, 115 (1989).
- ⁶M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias, and J.D. Joannopoulos, *Rev. Mod. Phys.* **64**, 1045 (1992).
- ⁷J.F. Annett, *Comput. Mater. Sci.* **4**, 23 (1995).
- ⁸P. Bendt and A. Zunger, *Phys. Rev. B* **26**, 3114 (1982).
- ⁹ $T[u]$ and $n[u](\mathbf{r})$ are defined by Eqs. (4) and (5).
- ¹⁰For the basics of conjugate-gradient algorithms, see Refs. 6,7, and references therein.
- ¹¹M.S. Hybertsen and S.G. Louie, *Phys. Rev. B* **35**, 5585 (1987).
- ¹²Also, symmetry constraints can prohibit some eigenvalues to influence the rate of convergence.
- ¹³R. Car, E. Tosatti, S. Baroni, and S. Leelapruete, *Phys. Rev. B* **24**, 985 (1981).
- ¹⁴F. Tassone, F. Mauri, and R. Car, *Phys. Rev. B* **50**, 10 561 (1994).
- ¹⁵Z.H. Levine and S.G. Louie, *Phys. Rev. B* **25**, 6310 (1982).
- ¹⁶See D.D. Johnson, *Phys. Rev. B* **38**, 12 807 (1988), and references therein.