Augmented Lagrangian method for order-N electronic structure

Satrajit Adhikari and Roi Baer

Citation: **115**, 11 (2001); doi: 10.1063/1.1383590

JOURNAL OF CHEMICAL PHYSICS VOLUME 115, NUMBER 1 1 JULY 2001

Augmented Lagrangian method for order-*N* electronic structure

Satrajit Adhikari and Roi Baer[a)]
Department of Physical Chemistry and the Lise Meitner Minerva-Center for Quantum Chemistry,
The Hebrew University of Jerusalem, Jerusalem 91904, Israel

(Received 6 December 2000; accepted 15 May 2001)

Molecular electronic ground-state theories, whether *ab initio*, or semiempirical are most often formulated as a variational principle, where the electronic ground-state energy, considered a linear or nonlinear functional of a reduced density matrix, obtains a constrained minimum. In this communication, we present a Lagrangian analysis of the self-consistent-field electronic structure problem, which does not resort to the concept of orthogonal molecular orbitals. We also develop a method of constrained minimization efficiently applicable to nonlinear energy functional minimization, as well as to linear models such as tight-binding. The method is able to treat large molecules with an effort that scales linearly with the system size. It has built-in robustness and leads directly to the desired minimal solution. Performance is demonstrated on linear alkane and polyene chains. [DOI: 10.1063/1.1383590]

I. INTRODUCTION

Ab initio and semiempirical electronic structure methods for studying systems containing hundreds or more atoms are useful for diverse fields such as chemistry, materials science, biology, and condensed matter physics. Developing methods for accomplishing this feat includes adapting for computation suitable formulations of electronic structure theories. Recently, such theories and methods have drawn considerable attention.[1,2] One of the widely used approaches is based on minimizing an energy functional dependent on the idempotent density matrix.[3–10] These methods utilize the fact that the density matrix (DM) is sparse.[11,12]

This communication develops a new linear scaling method based on the theory of constrained optimization. The functionals may be linear as in extended Huckel or tight binding approximations[13] or nonlinear, as in self-consistent field (SCF) theories, whether semiempirical or *ab initio*. The method has built-in robustness and flexibility to enhance convergence rate based on the combined use of penalty and Lagrange multiplier terms. We now outline the theory and then demonstrate the algorithmic complexity properties on linear polyenes.

II. THEORY

Given $E[\rho]$, an energy functional of the one-electron density matrix $\rho(r,r')$, the ground-state density and energy are obtained by solving the doubly constrained optimization problem:

$$E_{gs}=\min_{\rho} E[\rho] \ \text{ under: } \ tr\rho=N_e; \ \ \rho^2=\rho, \tag{1.1}$$

where the following shorthand matrix-multiply notation is used

$$\rho^2(\mathbf{r},\mathbf{r}')\equiv\int\rho(\mathbf{r},\mathbf{r}'')\rho(\mathbf{r}'',\mathbf{r}')d^3r''. \tag{1.2}$$

The first constraint fixes the number of electrons N_e. The second, DM idempotency, results from Pauli's principle.

A. Lagrangian saddle point

The constrained minimum problem can be handled using the method of Lagrange multipliers, which we discuss now. Define the Lagrangian

$$L_{\mu,A}[\rho]=E[\rho]-\mu(tr\rho-N_e)-trA\rho(1-\rho), \tag{1.3}$$

where the number μ and the symmetric matrix elements $A(r,r')$ are Lagrange multipliers. Next, define the following functional of the Lagrange multipliers:

$$\Lambda[\mu,A]=\min_{\rho} L_{\mu,A}[\rho], \tag{1.4}$$

and assume now that Λ obtains a maximum at μ^* and A^*. For any ρ under the constraints of Eq. (1.1), the following holds:

$$\Lambda[\mu^*,A^*]\leqslant L_{\mu^*,A^*}[\rho]=E[\rho]. \tag{1.5}$$

In particular, ρ_{gs} obeys the constraints so

$$\Lambda[\mu^*,A^*]\leqslant E_{gs}. \tag{1.6}$$

Thus, the maximum of Λ is a lower bound to E_{gs}, which is tight under conditions of convexity.[14] The constrained minimization problem is thus transformed to a *saddle point problem*: The Lagrangian is minimum with respect to the density, when $\mu=\mu^*$ and $A=A^*$, and maximum with respect to μ and A when the density $\rho=\rho_{gs}$ is held fixed. At the saddle-point, the following relations hold:

$$0=\left.\frac{\delta L}{\delta\rho}\right|_{\rho_{gs},A^*,\mu^*}=H'-A^*+\rho_{gs}A^*+A^*\rho_{gs}, \tag{1.7}$$

where $H'=\delta E/\delta\rho|_{\rho=\rho_{gs}}-\mu^*$, and

[a)] Electronic mail: Roi.Baer@huji.ac.il

0021-9606/2001/115(1)/11/4/$18.00

$$0=\left.\frac{\partial\Lambda}{\partial\mu}\right|_{*}=N_e-tr\rho_{gs}\,,\quad 0=\left.\frac{\delta\Lambda}{\delta A}\right|_{*}=\rho_{gs}(\rho_{gs}-1). \quad (1.8)$$

Multiplying Eq. (1.7) by ρ_{gs} from one side and by $(1-\rho_{gs})$ from the other, we obtain

$$(1-\rho_{gs})H'\rho_{gs}=\rho_{gs}H'(1-\rho_{gs})=0, \quad (1.9)$$

which also leads to

$$[H',\rho_{gs}]=0. \quad (1.10)$$

Now multiplying Eq. (1.7) by ρ_{gs} from the right leads to

$$H'\rho_{gs}=-\rho_{gs}A^*\rho_{gs}\,, \quad (1.11)$$

likewise, multiplying by $(1-\rho_{gs})$ from the right

$$H'(1-\rho_{gs})=(1-\rho_{gs})A^*(1-\rho_{gs}). \quad (1.12)$$

These two conditions are fulfilled by the following choice of the Lagrange multiplier matrix:

$$A^*=H'(1-2\rho_{gs}). \quad (1.13)$$

The energy is minimized when the density matrix ρ_{gs} describes the occupation of the lowest eigenstates of H'. Now consider the Hessian of the Lagrangian. This is composed of the Hessian of the energy D^2E *and* the Hessian of the constraint

$$D^2L=D^2E+I\otimes A^*+A^*\otimes I. \quad (1.14)$$

The Hessian of the energy is problem specific. For example, in tight binding models it is zero. In most cases, the demand that the matrix A^* be positive definite is enough to ensure a stable minimum. We shall suppose this is so. Because ρ_{gs} is a projection operator which is commutative with H', the only way to ensure positive definiteness of A^* and compatibility with Eq. (1.13) is to take

$$A^*=|H'|. \quad (1.15)$$

(We define the absolute value of a matrix X as follows. Suppose $X=\eta x\eta^{-1}$ where x is a diagonal matrix of eigenvalues, then $|X|=\eta|x|\eta^{-1}$.) This leads to the condition: $1-2\rho_{gs}=Sign(-H')$, or the usual zero-temperature Fermi–Dirac distribution

$$\rho_{gs}=\theta(-H'), \quad (1.16)$$

where $\theta(x)$ is the Heaviside function. Consider the following Lagrangian, obtained by inserting in Eq. (1.3) the relation $A[\rho]=H'(1-2\rho)$:

$$\begin{aligned}L_{\mu^*}[\rho]&=L_{A[\rho],\mu^*}[\rho]\\&=E[\rho]-tr(H\rho)+tr(H'(3\rho^2-2\rho^3))+\mu^*N_e\,,\end{aligned} \quad (1.17)$$

where, $H=H'+\mu$. After proper selection of μ, this Lagrangian has a minimum at the ground-state density. Note the natural emergence of the McWeeny purified density[15] $3\rho^2-2\rho^3$. This Lagrangian is a generalization of the functional of Li *et al.*[3] that is obtained for linear functionals, where $E[\rho]=tr(H\rho)$, is obtained. It should be pointed out, that the use of Eq. (1.17) in an actual minimum search algorithm is not necessarily an efficient or robust method. Convergence is not guaranteed unless one starts from a close approximation to the ground-state density.

Another way to impose the constraints, suggested by Haynes and Payne,[9] is to minimize an analytical penalized functional

$$L_c[\rho]=E[\rho]+\tfrac{1}{2}c\{(tr\rho-N_e)^2+tr(\rho^2(1-\rho)^2)\}. \quad (1.18)$$

The second term here, weighted by the cost $c>0$ is a penalty for violating the constraints. Another possibility is the square root of the right trace in Eq. (1.18), proposed by Kohn,[6] but this leads to a nonanalytical functional.[16] Minimizing L_c leads to ill posedness because of the absolute necessity to closely approach the limit $c\rightarrow\infty$.

B. The augmented Lagrangian

Enhancing robustness and convergence without encountering ill posedness may be achieved by using the penalized function $L_c[\rho]$ as a *function to be minimized under constraints*. Such an approach is termed *augmented Lagrangian* approach[14]

$$L_{A,\mu,c}[\rho]=L_c[\rho]-\mu(tr\rho-N_e)-tr\{A\rho(1-\rho)\}. \quad (1.19)$$

At the saddle point, $L_c[\rho]$ and its first derivatives are identical to those of $E[\rho]$ and the Lagrange multipliers equal those of Eq. (1.3). It is possible to show that there exists a constant c^* such that for all $c>c^*$, the unconstrained minimum of $L_{A^*,\mu^*,c}[\rho]$ is the solution to the primary problem [Eq. (1.1)]. There is no need to take the limit $c\rightarrow\infty$ but one can increase c *to* accelerate convergence without hampering stability.

Since A^* and μ^* are not known, an iterative algorithm needs to be set up. For a guess A_k and μ_k, the density ρ_k that minimizes $L_{A_k,\mu_k,c_k}[\rho]$, satisfies the following condition:

$$\begin{aligned}0&=\frac{\delta L_{A_k,\mu_k,c_k}}{\delta\rho}\\&=\left.\frac{\delta E}{\delta\rho}\right|_{\rho_k}+c_k(tr\rho_k-N_e)-\mu_k-A_k+\rho_kA_k\\&\quad+A_k\rho_k+c_k+c_k\rho_k(1-\rho_k)(1-2\rho_k).\end{aligned} \quad (1.20)$$

Let us define the ‘‘updated’’ Lagrange multipliers as

$$\begin{aligned}\mu_{k+1}&=\mu_k-c_k(tr\rho_k-N_e),\\A_{k+1}&=A_k-c_k\rho_k(1-\rho_k).\end{aligned} \quad (1.21)$$

To see why these updates are used, let us insert them into Eq. (1.20), obtaining after rearrangement

$$0=\left.\frac{\delta E}{\delta\rho}\right|_{\rho_k}-\mu_{k+1}-A_{k+1}+\rho_kA_{k+1}+A_{k+1}\rho_k\,, \quad (1.22)$$

which is exactly the minimum condition of Eq. (1.7).
The following algorithm for the constrained minimization of the energy is then obtained:

(1) Initialize: Set $k=0$. Guess A_0, μ_0, and c_0;

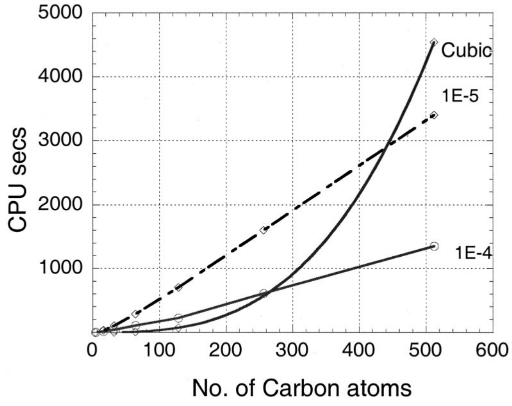

FIG. 1. The total CPU time (including search for μ^* and A^*) for the TB DM computation of an alkane chain as a function of chain length. Two truncation thresholds are shown, as well as the cubic CPU time based on diagonalization.

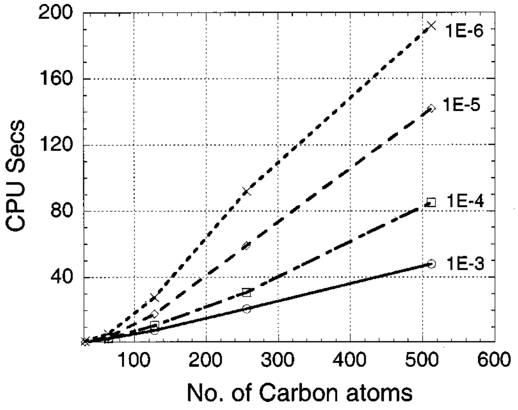

FIG. 2. The CPU time for the DM SCF computation of polyene chains as a function of chain length. The four curves refer to four values of the truncation threshold T. We use $c_0=1$ and $\lambda=2$.

(2) Minimize unconstrained, $L_{c_k,A_k,\mu_k}[\rho]$, for example, using a conjugate gradient method;
(3) Use Eq. (1.21) to update A_{k+1} and μ_{k+1} and set $c_{k+1}=\lambda c_k$ (a value of $\lambda>1$ leading to efficient convergence can be determined by some experimentation);
(4) Set $k\rightarrow k+1$ and go to step 2.

III. RESULTS

The method in the previous section was coded using SPARSKIT 2, sparse matrix library by Y. Saad. The sparsity is imposed by zeroing matrix elements of the DM, which are smaller than a threshold T during each iteration. During the computation the matrix of Lagrange multipliers tends to loose sparsity, however, since its exact value is immaterial, we truncate it using a larger threshold.

We show two applications for hydrocarbon chains. The first is a linear unbranched alkane C_nH_{2n+2} using a tight binding model of Wang *et al.*[17] The algorithm starts by setting $A_0=1$, and $\mu_0=0$ and ends when the Lagrangian corrections are smaller than a given threshold. CPU timing results are shown in Fig. 1. The algorithmic complexity approaches a linear scaling regime although differences in number of conjugate gradient iterations cause fluctuations. The CPU times, measured on a PENTIUM III 500 MHZ/LINUX, correspond to a total energy computation, including the search for μ^* and A^*. The search takes typically between 8 and 12 iterations. The number of iterations is almost system independent, increasing slightly as system grows.

Next, an SCF system is tested. Linear conjugated polyenes $H_2C{=}CH\text{–}(CH{=}CH)_{n-1}\text{–}CH{=}CH_2$ are treated with Pople's molecular orbital SCF[18] where the energy functional of the π-electron density matrix P_{ij} is:

$$E[P]=tr(hP)+\frac{\gamma}{2}\sum_{i=1}^{N}P_{ii}^2+\frac{1}{2}\sum_{i\neq j=1}^{N}\left[\frac{(2P_{ii}-Z_i)(2P_{jj}-Z_j)}{b_{ij}}-2\frac{P_{ij}^2}{b_{ij}}\right]. \tag{1.23}$$

Here h is a $N\times N$ matrix describing interaction of π-electrons with the σ-system. h has the following nonzero elements: $h_{ii}=\varepsilon$, $h_{i,i+1}=h_{i+1,i}=\beta_0$ and $h_{i,i-1}=h_{i-1,i}=\beta_1$ for i even. The second term is an on-site direct repulsion and the last term describes long-range interactions between the π-electrons, and the nuclei. In the equation Z_i is the residual charge on nucleus i, and $b_{ij}^{-1}=r_{ij}^{-1}e^{-\alpha r_{ij}}$ is a Yukawa interaction at a distance. Yukawa rather than Coulomb potential is used because of linear scaling Fock build considerations. The following parameters and definitions are used:

$$\beta_0=-0.08E_h,\quad \beta_1=-0.1E_h,\quad \gamma=0.8E_h,$$
$$\varepsilon_i=-0.4E_h,\quad Z_i=1,\quad r_{ij}=D_{CC}|i-j|,$$
$$b_{ij}=r_{ij}e^{\alpha r_{ij}}\quad \alpha=0.7a_0^{-1},\quad D_{cc}=2.6a_0.$$

In Fig. 2, it is seen that the algorithmic complexity approaches the linear scaling regime. CPU times are for the complete energy computation, starting from a density matrix of $P=0.5I$ and initial value of $A_0=I$ and $\mu_0=0$. The SCF iterations are converged in about 10 iterations. Work is under way to study applications on two- and three-dimensional systems and semiconductors.

ACKNOWLEDGMENTS

This research was supported by Grant No. 9800108 from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel. The authors gratefully acknowledge the use of computational resources at the Fritz Haber center in the Hebrew University of Jerusalem.

[1] S. Goedecker, Rev. Mod. Phys. **71**, 1085 (1999).
[2] D. R. Bowler, I. J. Bush, and M. J. Gillan, Int. J. Quantum Chem. **77**, 831 (2000).
[3] X. P. Li, W. Nunes, and D. Vanderbilt, Phys. Rev. B **47**, 10891 (1993).
[4] M. S. Daw, Phys. Rev. B **47**, 10895 (1993).
[5] E. Hernandez and M. J. Gillan, Phys. Rev. B **51**, 10157 (1995).
[6] W. Kohn, Phys. Rev. Lett. **76**, 3168 (1996).
[7] C. H. Xu and G. E. Scuseria, Chem. Phys. Lett. **262**, 219 (1996).
[8] A. H. R. Palser and D. E. Manolopoulos, Phys. Rev. B **58**, 12704 (1998).
[9] P. D. Haynes and M. C. Payne, Phys. Rev. B **59**, 12173 (1999).
[10] D. R. Bowler and M. J. Gillan, Comput. Phys. Commun. **120**, 95 (1999).
[11] R. Baer and M. Head-Gordon, Phys. Rev. Lett. **79**, 3962 (1997).

[12] R. Baer and M. Head-Gordon, J. Chem. Phys. **109**, 10159 (1998).
[13] C. M. Goringe, D. R. Bowler, and E. Hernandez, Rep. Prog. Phys. **60**, 1447 (1997).
[14] P. Gill, *Practical Optimization* (Academic, New York, 1982).
[15] R. McWeeny, Rev. Mod. Phys. **32**, 335 (1960).
[16] P. D. Haynes and M. C. Payne, Solid State Commun. **108**, 737 (1998).
[17] Y. Wang and C. H. Mak, Chem. Phys. Lett. **235**, 37 (1995).
[18] J. A. Pople, Trans. Faraday Soc. **49**, 1375 (1953).