

Realization of Quantum Chemistry without Wave Functions through First-Order Semidefinite Programming

David A. Mazziotti*

Department of Chemistry and the James Franck Institute, The University of Chicago, Chicago, Illinois 60637, USA

(Received 19 June 2004; published 15 November 2004)

Determining the energy and properties of an N -electron molecule through a two-electron variational optimization has been a dream for more than half a century. While optimizations, using two-electron reduced density matrices constrained to represent N electrons, have recently been achieved, the computational costs are prohibitive. In this report an efficient algorithm with an order-of-magnitude reduction in floating-point operations and memory usage is presented. Because the optimization occurs on the space of two electrons, this method automatically treats strong, multireference correlation. Application is made to N_2 and H_6 where the method yields consistent accuracy at all geometries.

DOI: 10.1103/PhysRevLett.93.213001

PACS numbers: 31.10.+z, 02.60.Pn, 31.25.-v, 31.50.Bc

At a conference in 1959 Coulson proposed “banishing” the wave function from electronic structure calculations in quantum chemistry [1]. Löwdin and Mayer [2] had shown earlier that the electronic energies and properties of atoms and molecules are computable simply from an effective two-electron density distribution known as the two-electron reduced density matrix (2-RDM). If electronic ground-state energies could be minimized with respect to 2-RDMs rather than wave functions, then many-electron calculations might be replaced by two-electron minimizations. Variational two-electron calculations, however, demonstrated that the two-electron density distribution in the 2-RDM must be constrained nontrivially to derive from the many-electron density distribution in the wave function; the necessary constraints became known as N -representability conditions [3,4]. Limitations in optimization software and computer resources prevented the practical development of 2-RDM methods.

Interest in 2-RDM calculations was rekindled in the 1990s through research on the contracted Schrödinger equation [5]. Variational 2-RDM methods with systematic, nontrivial N -representability conditions have recently been developed and applied to atomic and molecular electronic calculations [6–8]. The 2-RDM minimization problem may be treated by a special class of optimization techniques known as semidefinite programming. The recent 2-RDM methods employ an interior-point algorithm for semidefinite programming, developed for combinatorial optimization and control theory in the 1990s [9,10]. The utility of the 2-RDM techniques, however, is limited by the cost of the interior-point algorithm in both floating-point operations and storage [6]. In this letter we present a new optimization strategy and algorithm where the semidefinite program is converted into a nonlinear constrained optimization problem which, we demonstrate, may be efficiently solved by a first-order augmented Lagrange multiplier method. Application of the resulting 2-RDM

approach is made to describing the potential curves of the H_6 and N_2 molecules accurately, especially in the regions of the potential curve where it is difficult to treat multireference electron correlation. Through efficient first-order semidefinite programming, the variational 2-RDM method practically realizes the dream of a quantum chemistry without wave functions.

The energy of any atom or molecule may be written as a linear functional of only the 2-RDM. To demonstrate, consider the customary definition of the energy in quantum mechanics,

$$E = \int \Psi^*(1, 2, \dots, N) \hat{H} \Psi(1, 2, \dots, N) d1 d2 \dots dN, \quad (1)$$

where each number represents the spatial and spin coordinates for one of the N electrons. Because electrons interact with each other *pairwise* and are indistinguishable, the same energy may be procured by replacing the N -electron Hamiltonian in Eq. (1) with an effective two-electron Hamiltonian operator,

$${}^2\hat{K} = N \left(-\frac{1}{2} \nabla_1^2 - \sum_j \frac{Z_j}{r_{1j}} \right) + \frac{N(N-1)}{2} \frac{1}{r_{12}}. \quad (2)$$

Since the effective Hamiltonian depends only upon electrons 1 and 2, we can move the effective Hamiltonian outside the integration over electrons 3 to N ,

$$E = \int {}^2\hat{K} \left[\int \Psi(1, 2, \dots, N) \times \Psi^*(1', 2', \dots, N) d3 \dots dN \right] d1 d2. \quad (3)$$

The inner integration of the wave function over electrons 3 through N , however, defines the 2-RDM,

$$E = \int {}^2\hat{K} {}^2D(1, 2; 1', 2') d1 d2. \quad (4)$$

Equation (4) suggests the tantalizing idea that the ground-state energy might be computed directly through a variational calculation with the 2-RDM. In 1932, von Neumann [11] gave the conditions for a general matrix to

be a density matrix; the matrix must be (i) Hermitian, (ii) normalized (fixed trace), (iii) antisymmetric (fermions) or symmetric (bosons) under particle exchange, and (iv) positive semidefinite to yield non-negative probabilities for finding electrons (a matrix is positive semidefinite if and only if its eigenvalues are non-negative). Variationally minimizing the energy in Eq. (4) with respect to the two-electron density matrix constrained by these von Neumann conditions, however, yields a result significantly below the energy from the ground-state wave function. A “violation” of the wave function variational principle occurs because not every two-electron density matrix (that is a matrix satisfying the von Neumann conditions) derives from the integration of an N -electron wave function. Boundary conditions in addition to the von Neumann constraints are necessary for the 2-RDM to ensure that it is N representable.

How might additional N -representability restrictions on the 2-RDM be obtained without invoking the N -electron wave function? There are three distinct matrix representations of the 2-RDM on the space of two particles—the D , Q , and G matrices with the D matrix appearing in Eq. (4). Each of these matrices contains the same information as the other two matrices, and linear mappings exist to convert between these matrices [6–8]. Constraining all three matrices to be positive semidefinite, however, furnishes three distinct N -representability conditions, known as the D , Q , and G conditions or, more simply, the 2-positivity conditions [4,6]. The positivity approach to N representability has been extended to higher RDMs to form a hierarchy of N -representability conditions with the restrictions on the p -RDM being called the p -positivity conditions [6].

The variational 2-RDM method with 2-positivity conditions raises the optimization problem of minimizing the energy with respect to the 2-RDM while the D , Q , and G matrices, connected by linear mappings, are restricted to be positive semidefinite. This class of optimization, known as positive semidefinite programming, arises in a variety of scientific areas including combinatorial optimization, control theory, and finance [9,10]. In 1993 Nesterov and Nemirovskii demonstrated how to extend a class of algorithms known as interior-point methods from linear programming to semidefinite programming [10]. This work led to significant research on semidefinite programming including the appearance of computer packages implementing interior-point algorithms.

Recent variational 2-RDM calculations with 2-positivity have solved the semidefinite program with one or more of packages, each of which employs a variant of the primal-dual interior-point algorithm [6–8]. The primal-dual interior-point algorithm solves two problems, known as the primal and dual programs, simultaneously to yield upper and lower bounds on the energy, respectively. Under weak assumptions minimizing the energy difference between the two programs to zero

ensures the convergence of the algorithm to the global minimum. While exhibiting a robust global convergence, the second-order primal-dual interior-point algorithm has an expensive scaling with respect to floating-point operations (r^{16}) and computer memory (r^8) where r is the number of one-electron basis functions.

For the variational 2-RDM approach to be practical requires a dramatic reduction in the scaling of semidefinite programming. Recently, large-scale problems in combinatorial optimization have fueled research interest in the development of alternative algorithms for semidefinite programming including spectral bundle methods, iterative interior-point techniques, and nonlinear programming methods. For the 2-RDM problem we have developed a nonlinear formulation of semidefinite programming which achieves a significant reduction in both computational operations and storage. We express the 2-RDM as a matrix factorization in terms of a Hermitian matrix R ,

$${}^2D = R^*R, \quad (5)$$

whose elements become the independent variables. Equation (5) automatically constrains the 2-RDM to be positive semidefinite with the eigenvalues of the 2-RDM being equal to the square of the R -matrix eigenvalues. This factorization was employed with the RDMs of the contracted Schrödinger equation by Mazziotti in 1998 [5]; Burer and Monteiro have recently used the factorization to solve semidefinite programs in which the matrix R has a low rank [12]. Analogous factorizations are introduced for the Q -matrix and G -matrix representations of the 2-RDM:

$${}^2Q = S^*S \quad \text{and} \quad {}^2G = T^*T. \quad (6)$$

With the D , Q , and G matrices positive semidefinite, the constraints assume the form of equalities enforcing the conversions between pairs of the D , Q , and G matrices. Note that the linear mappings connecting the D , Q , and G matrices produce *nonlinear* equalities in terms of the variables R , S , and T .

To solve the nonlinear formulation of the semidefinite program, we employ the augmented Lagrange multiplier method for constrained nonlinear optimization [12,13]. We introduce the augmented Lagrangian function

$$L(x) = E(x) - \sum_i \lambda_i c_i(x) + \frac{1}{2\mu} \sum_i c_i(x)^2, \quad (7)$$

where x is a vector containing the variables in the matrices R , T , and S , $E(x)$ is the ground-state energy as a function of x , $\{c_i(x)\}$ is the set of equality constraints, $\{\lambda_i(x)\}$ is the set of Lagrange multipliers, and μ is the penalty parameter. The augmented Lagrange multiplier method consists of a series of unconstrained minimizations of the Lagrangian function with updates of $\{\lambda_i(x)\}$ and μ which lead to the minimization of the energy $E(x)$ subject to the constraints. After each unconstrained mini-

mization of the Lagrangian function, if the maximum absolute error in the constraints is below a chosen threshold, the Lagrange multipliers are updated by a first-order correction $\lambda_i^{(n+1)} = \lambda_i^{(n)} - c_i(x)/\mu$; otherwise, the penalty parameter is decreased to enforce better the constraints, that is $\mu^{(n+1)} = 0.1\mu^{(n)}$. For the unconstrained optimization we employ a gradient-only approach in the form of a limited-memory quasi-Newton method. The resulting algorithm for the variational 2-RDM method requires r^6 floating-point operations for the matrix multiplications in Eqs. (5) and (6) and computer storage scaling with the number of 2-RDM elements (r^4), where r is the number of spatial one-electron basis functions.

To demonstrate the new algorithm for the variational 2-RDM method, we compute the potential curves for H_6 and N_2 . In all calculations we choose the initial 2-RDM within the nonlinear algorithm to be the Hartree-Fock 2-RDM. Calculations of electron integrals were performed with the electronic structure package GAMESS and its compiled implementation PC GAMESS [14]. For both the nonlinear formulation, employing an augmented Lagrange multiplier method, and the primal-dual interior-point algorithm within SEDUMI [9], Table I shows the amount of memory required by the program as a function of the number r of spatial one-particle basis functions. For $r = 8$ the SEDUMI program requires more than two gigabytes while the nonlinear Lagrange multiplier algorithm employs only four megabytes. Furthermore, the calculation of a molecule with $r = 7$ by the present method, taking less than 100 s, is approximately 30 times faster than the calculation with SEDUMI. The present algorithm is not fully optimized. Treatment of either H_6 or N_2 in nonminimal basis sets would not have been possible with the interior-point algorithms for semidefinite programming.

Metallic hydrogen is an infinite chain of equally spaced hydrogen atoms [15]. It can serve as a simple model for polymers and crystals. Despite the simplicity of the hydrogen atoms, the properties of metallic hydrogen as well as finite hydrogen rings and chains are significantly influenced by electron correlation. We consider the equally spaced, finite chain H_6 where the hydrogen atoms are described by the valence triple zeta basis set in GAMESS [14]. A potential energy curve may be generated by equally stretching the five bonds in H_6 . Ground-state energies from the variational 2-RDM method and a vari-

ety of wave function techniques are presented in Fig. 1(a) as functions of the distance R between adjacent hydrogen atoms. The 2-RDM method yields consistent energies with a maximum error of -10.8 mH at 1.5 Å. While the coupled cluster methods are accurate near the equilibrium geometry with errors of approximately 1 mH, their performance rapidly degrades as the bonds are stretched. At 3.5 Å each of the coupled cluster methods has an energy error of at least -160 mH while the 2-RDM method has an error of only -0.4 mH.

One of the most challenging correlation problems in electronic structure is the accurate description of the stretching and dissociation of the triple bond in nitrogen [16]. In Fig. 1(b) we examine the 2-RDM method and several wave function techniques. As in [16] a valence double-zeta basis set is employed with the two lowest core orbitals and the two highest virtual orbitals frozen. At the bond distance $R_c = 1.094\ 363\ 636$ Å the coupled cluster methods, CCSD and CCSD(T), yield energy errors of 6.2 mH and 1.8 mH respectively; however, at $2.5R_c$ they produce errors of -153 and -425 mH. In contrast, the 2-RDM method yields an error of -28.6 mH at R_c and an error of -24.0 mH at $2.5R_c$. Although the 2-RDM method overestimates the correlation energy in N_2 by 20–40 mH, there are several reasons why this is not as significant as it

TABLE I. The memory required as a function of the number r of one-particle spatial orbitals is shown for the present and previous (SEDUMI) 2-RDM methods.

Algorithm	Memory usage (GigaBytes)						
	Number r of spatial one-electron basis functions						
	5	6	7	8	12	15	18
SEDUMI	0.075	0.224	0.731	2.100
Present	0.002	0.002	0.003	0.004	0.011	0.026	0.052

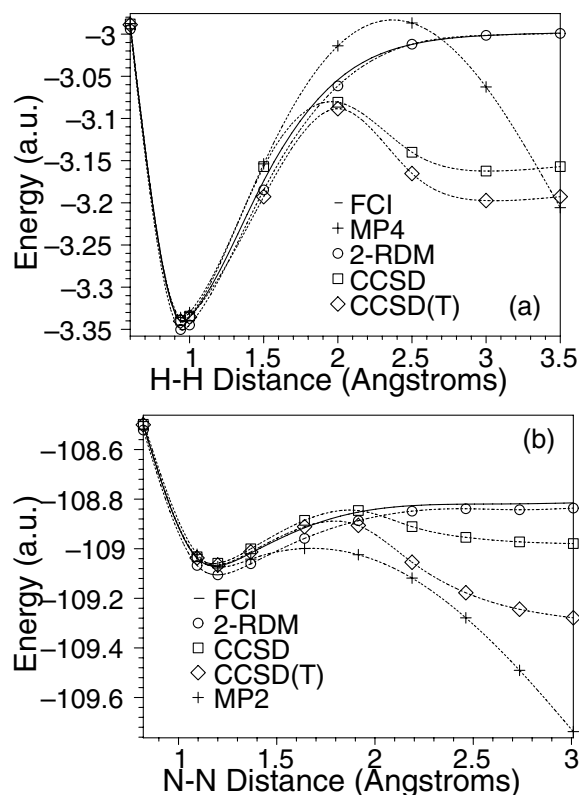


FIG. 1. Ground-state potential energy curves of (a) H_6 and (b) N_2 are shown from 2-RDM and wave function methods, second-order and fourth-order many-body perturbation theories (MP2 and MP4) and coupled cluster singles-doubles (CCSD) and CCSD with triples (CCSD(T)).

may first appear: (i) the potential curve has the correct shape because all of the energies are shifted by at least 20 mH; (ii) calculations on a large data set of organic molecules show that the percentage of the correlation energy errors may be systematized by functional groups [17], and (iii) an additional order of magnitude of accuracy appears achievable by using a subset of the 3-positivity conditions [6,8]. The dissociation of nitrogen may also be treated accurately with *multireference* wave function methods which use a linear combination of determinants for the reference rather than a single Slater determinant [18,19]. Using a second-order multi-reference perturbation theory, Andersson, Malmqvist, and Roos have computed nitrogen at a variety of geometries with an accuracy of ± 5 mH [19]. Within these methods the reference determinants must be carefully selected to balance accuracy and computational cost.

The purpose of the present Letter is not to give a detailed comparison of 2-RDM methods with multireference techniques. However, two features of the 2-RDM method which distinguish it from multireference approaches should be highlighted: (i) because the set of 2-RDMs satisfying the 2-positivity conditions contains the 2-RDMs from all correlated N -particle wave functions, the variational 2-RDM method automatically captures on the two-particle space many-particle excitations which are difficult to treat systematically in wave function methods; (ii) as a consequence of (i), treating multireference effects in the 2-RDM method requires only an initial guess for the 2-RDM from a single-determinant, Hartree-Fock wave function rather than a carefully constructed, multideterminant reference space.

A practical variational calculation of the 2-RDM without the N -electron wave function is achieved. Three different representations of the 2-RDM are constrained via a quadratic factorization to be positive semidefinite, and the resulting nonlinear program is solved by an augmented Lagrange multiplier method. Unlike other recent 2-RDM methods which scale as r^{16} , the present formulation scales as r^6 with a concomitant reduction in memory requirements. The rate of energy convergence with basis set size r is similar to a full configuration interaction calculation. Improving the rate of convergence would require abandoning the orbital product basis set in favor of explicitly correlated basis functions. Because of its two-electron formulation, the 2-RDM method may be especially suited to explicitly correlated two-electron basis functions. We have applied for the first time the 2-RDM method to the calculation of potential energy curves for H_6 and N_2 in nonminimal basis sets. The bond stretchings in these molecules display strong multireference correlation which we show is described by the 2-RDM method without selecting a multideterminant reference wave function. Accurate treatment of potential surfaces is critical for understanding dynamics, kinetics, and spectroscopy. This nonlinear approach to semidefinite pro-

gramming may also have a broad impact on applications in combinatorial optimization, control theory, and finance. The present 2-RDM method realizes a quantum chemistry without wave functions and a new paradigm for describing strong electron correlations.

The author expresses his appreciation to Professors Dudley Herschbach and Stuart Rice and Dr. Alexander Mazziotti for reading the manuscript. The Camille & Henry Dreyfus Foundation and the NSF are gratefully acknowledged.

*Electronic address: damazz@uchicago.edu

- [1] A. J. Coleman and V. I. Yukalov, *Reduced Density Matrices: Coulson's Challenge* (Springer-Verlag, New York, 2000).
- [2] P. O. Löwdin, Phys. Rev. **97**, 1474 (1955); J. E. Mayer, Phys. Rev. **100**, 1579 (1955).
- [3] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963).
- [4] C. Garrod and J. Percus, J. Math. Phys. (N.Y.) **5**, 1756 (1964).
- [5] F. Colmenero and C. Valdemoro, Phys. Rev. A **47**, 979 (1993); H. Nakatsuji and K. Yasuda, Phys. Rev. Lett. **76**, 1039 (1996); D. A. Mazziotti, Phys. Rev. A **57**, 4219 (1998); D. A. Mazziotti, J. Chem. Phys. **116**, 1239 (2002).
- [6] D. A. Mazziotti and R. M. Erdahl, Phys. Rev. A **63**, 042113 (2001); D. A. Mazziotti, Phys. Rev. A **65**, 062511 (2002); D. A. Mazziotti, Phys. Rev. A **66**, 062503 (2002); G. Gidofalvi and D. A. Mazziotti, Phys. Rev. A **69**, 042511 (2004); T. Juhasz and D. A. Mazziotti, J. Chem. Phys. **121**, 1201 (2004).
- [7] M. Nakata, H. Nakatsuji, M. Ehara, M. Fukuda, K. Nakata, and K. Fujisawa, J. Chem. Phys. **114**, 8282 (2001); M. Nakata, M. Ehara, and H. Nakatsuji, J. Chem. Phys. **116**, 5432 (2002).
- [8] Z. Zhao, B. J. Braams, H. Fukuda, M. L. Overton, and J. K. Percus, J. Chem. Phys. **120**, 2095 (2004).
- [9] J. F. Sturm, Optimization Methods and Software **11**, 625 (1999).
- [10] Y. Nesterov and A. S. Nemirovskii, *Interior Point Polynomial Method in Convex Programming: Theory and Applications* (SIAM, Philadelphia, 1993).
- [11] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1995).
- [12] S. Burer and R. D. C. Monteiro, Math. program. **95**, 329 (2003).
- [13] R. Fletcher, *Practical Methods of Optimization* (John Wiley and Sons, New York, 1987).
- [14] M. W. Schmidt *et al.*, J. Comput. Chem. **14**, 1347 (1993).
- [15] S. Suhai, Phys. Rev. B **50**, 14791 (1994).
- [16] P. Piecuch, S. A. Kucharski, and K. Kowalski, Chem. Phys. Lett. **344**, 176 (2001).
- [17] G. Gidofalvi and D. A. Mazziotti (to be published).
- [18] C. W. Bauschlicher, Jr. and S. R. Langhoff, J. Chem. Phys. **86**, 5595 (1987).
- [19] K. Andersson, P. Malmqvist, and B. O. Roos, J. Chem. Phys. **96**, 1218 (1992); J. E. Stevens, R. K. Chaudhuri, and K. F. Freed, J. Chem. Phys. **105**, 8754 (1996).