Active-Space *N*-Representability Constraints for Variational Two-Particle
Reduced Density Matrix Calculations

Neil Shenvi[1,*] and Artur F. Izmaylov[2]

[1]*Department of Chemistry, Duke University, Durham, North Carolina 27708, USA*
[2]*Department of Chemistry, Yale University, New Haven, Connecticut 06520, USA*
(Received 3 August 2010; published 18 November 2010)

The ground-state energy of a system of fermions can be calculated by minimizing a linear functional of the two-particle reduced density matrix (2-RDM) if an accurate set of *N*-representability conditions is applied. In this Letter we introduce a class of linear *N*-representability conditions based on exact calculations on a reduced active space. Unlike wave-function-based approaches, the 2-RDM methodology allows us to combine information from calculations on different active spaces. By adding active-space constraints, we can iteratively improve our estimate for the ground-state energy. Applying our methodology to a 1D Hubbard model yields a significant improvement over traditional 2-positivity constraints with the same computational scaling.

The first-principles prediction of the electronic structure of atoms, molecules, and solids has been a major focus of theoretical chemistry since the advent of quantum mechanics in the early 1900s. The difficulty of electronic structure calculation is a consequence of the exponential growth of the fermionic Hilbert space with particle number [1]. For a system with N electrons in L spin orbitals, the ground-state N-particle wave function will be some linear combination of $\binom{L}{N} = L!/N!(L-N)!$ possible electronic configurations.

A non-wave-function-based approach to electronic structure calculations was proposed in the 1950s by Löwdin, Coleman, and others [2–5]. This approach treats the two-particle reduced density matrix (2-RDM) rather than the N-particle wave function as the fundamental object of interest [5–9]. A Hamiltonian of N fermions with at most two-body interactions can be written in second-quantized form as

$$\hat{H} = \sum_{ik} {}^1H_k^i c_i^\dagger c_k + \sum_{ijkl} {}^2H_{kl}^{ij} c_i^\dagger c_j^\dagger c_l c_k, \qquad (1)$$

where $c_i^\dagger$ and c_i are, respectively, the creation and annihilation operators for a fermion in state i. The tensors ${}^1H$ and ${}^2H$ are the one- and two-particle integrals. If we evaluate the two-particle reduced density matrix corresponding to the state $|\Psi\rangle$,

$$ {}^2D_{kl}^{ij} = \langle \Psi | c_i^\dagger c_j^\dagger c_l c_k | \Psi \rangle, \qquad (2)$$

then we can write the energy as a linear functional of only the 2-RDM [2,10,11]

$$E_\Psi = \langle \Psi | \hat{H} | \Psi \rangle \qquad (3)$$

$$= \sum_{ijkl} {}^2K_{kl}^{ij} {}^2D_{kl}^{ij} \qquad (4)$$

$$= \mathrm{Tr}[{}^2K\,{}^2D], \qquad (5)$$

where we have defined the two-particle reduced Hamiltonian

$$ {}^2K_{kl}^{ij} = \frac{1}{N-1}\,{}^1H_k^i \delta_l^j + {}^2H_{kl}^{ij}. \qquad (6)$$

Because not all 2-RDMs are derived from physical, fermionic N-particle systems, minimization over the space of 2-RDMs in Eq. (5) to find the ground-state energy must be *constrained* to include only physical 2-RDMs. The problem of determining which 2-RDMs are derived from fermionic N-particle systems is known as the N-representability problem [4]. We will denote the set of N-representable 2-RDMs as ${}^2\mathcal{D}_\mathcal{N}$ so that the true ground-state energy can be obtained by performing a constrained minimization over this set:

$$E_g = \min_{{}^2D \in {}^2\mathcal{D}_\mathcal{N}} \mathrm{Tr}[{}^2K\,{}^2D]. \qquad (7)$$

Although no simple, complete set of N-representability constraints is known to exist, many sets of partial constraints provide approximate N representability. The most well known of these are the D, G, and Q constraints, known collectively as the 2-positivity constraints, developed by Coleman, Garrod, and Percus [4,12]. Just as ${}^2D$ is the two-particle RDM, there exists a two-hole RDM ${}^2Q$ and a particle-hole RDM ${}^2G$ which are linearly interrelated [9] according to the equations

$$ {}^2G_{kl}^{ij} = {}^2D_{jk}^{il} + {}^1D_k^i \delta_l^j, \qquad (8)$$

$$ {}^2Q_{kl}^{ij} = {}^2D_{kl}^{ij} - {}^1D_k^i \delta_l^j - \delta_k^i {}^1D_l^j + \delta_k^i \delta_l^j. \qquad (9)$$

The 2-positivity constraints require that all three of the matrices ${}^2D$, ${}^2G$, and ${}^2Q$ be positive semidefinite, having no negative eigenvalues.

Recently, it has been shown that the imposition of these 2-positivity constraints allows the calculation of atomic and molecular ground-state energies to a high degree of

accuracy [6,8,9,13]. To improve such calculations even further, additional constraints must be added. Again, one obvious approach is to construct three- and four-particle analogues of the two-particle positivity constraints [6,7,9,14,15]. However, implementing these higher-order positivity constraints is extremely costly; the operational complexity of the 3-positivity constraints scales at minimum as L^9, where L is the number of one-electron basis functions used in the calculation [16,17]. Consequently, it is highly desirable to develop new N-representability constraints which can be used to improve the accuracy of existing RDM calculations without adding substantial computational overhead.

In this Letter, we introduce linear inequality constraints based on the expectation value of two-body operators. Our constraints take the form

$$\mathrm{Tr}[A\,^2D] \leq a_{\max}, \qquad (10)$$

where A is a Hermitian matrix of the same dimension as ^2D and where $a_{\max}$ is the maximum value of $\mathrm{Tr}[A\,^2D]$ that can be achieved for an N-representable RDM ^2D. In other words,

$$a_{\max} \equiv \max_{^2D \in \,^2\mathcal{D}_{\mathcal{N}}} \mathrm{Tr}[A\,^2D] \qquad (11)$$

$$= \max_{\Psi} \langle \Psi | \hat{A} | \Psi \rangle, \qquad (12)$$

where

$$\hat{A} = \sum_{ijkl}^{L} A_{kl}^{ij} c_i^\dagger c_j^\dagger c_l c_k, \qquad (13)$$

such that the elements A_{kl}^{ij} constitute a Hermitian matrix. Therefore, in order to obtain $a_{\max}$ in Eq. (11) we simply need to find the maximum eigenvalue of the operator $\hat{A}$.

For a general two-body operator $\hat{A}$, this task is as difficult as the original many-body problem. However, the central observation of this Letter is that if the operator $\hat{A}$ in Eq. (12) acts only on some small active space S of $L_S < L$ one-particle states [i.e., if the sum in Eq. (13) runs up to L_S instead of L], then the problem of calculating $a_{\max}$ becomes tractable. To further simplify the calculation of $a_{\max}$, we will assume that $\hat{A}$ is also a spinless and charge-conserving operator. In this case, $\hat{A}$ commutes with the electron number operator n_S on the subspace S and all spin operators on the subspace S. This assumption results in a block-diagonal matrix representation for $\hat{A}$ in the eigenstate basis of the n_S and S_z operators. Thus, we can calculate the maximum eigenvalue of $\hat{A}$ far more efficiently (i.e., by diagonalizing a block-diagonal 2^{L_S}-dimensional matrix) than we could if $\hat{A}$ were a general operator, which would require the diagonalization of a $\binom{L}{N}$-dimensional matrix.

Next, we must consider how to select the operator $\hat{A}$ to produce the most useful constraints for our ground-state calculations. There are an infinite number of operators $\hat{A}$ which act on only L_S orbitals. Then which operators $\hat{A}$

should we select as constraints? Similarly, which subspace S will give us the constraints which are most effective in implementing partial N-representability for a given problem? We have chosen to take an iterative approach to constraint generation similar to that taken in [18].

First, the ground-state energy in Eq. (7) is minimized with respect to ^2D subject to the standard closed-shell constraints defined in [9] and the 2-positivity constraints. Next, we search for the matrix $A(\mathbf{c}, \mathbf{x})$ which minimizes the objective function $J(\mathbf{c}, \mathbf{x}) = (a_{\max} - \mathrm{Tr}[A(\mathbf{c}, \mathbf{x})\,^2D])/a_{\max}$, the relative difference between the exact upper bound and the expectation value calculated from our approximate 2-RDM, ^2D. To reduce the size of the search space, we parametrize $A(\mathbf{c}, \mathbf{x})$ as

$$A(\mathbf{c}, \mathbf{x}) = P_S(\mathbf{x})(c_1^2 K + c_2^2 D_{\alpha\alpha} + c_3^2 D_{\alpha\beta})P_S(\mathbf{x}), \quad (14)$$

where $D_{\alpha\alpha}$ and $D_{\alpha\beta}$ are the spin-symmetrized diagonal blocks of the 2-RDM as defined in [9],

$$[D_{\alpha\alpha}]_{kl}^{ij} = \langle \Psi | c_{i\alpha}^\dagger c_{j\alpha}^\dagger c_{l\alpha} c_{k\alpha} | \Psi \rangle, \qquad (15)$$

$$[D_{\alpha\beta}]_{kl}^{ij} = \langle \Psi | c_{i\alpha}^\dagger c_{j\beta}^\dagger c_{l\beta} c_{k\alpha} | \Psi \rangle. \qquad (16)$$

The operator $P_S(\mathbf{x})$ projects the matrix A onto some set of L_S one-particle orbitals and can be written in terms of an orthogonal rotation of the canonical orbitals [i.e., $\mathbf{U} = \exp(\mathbf{x})$]. Given the definition in Eq. (14), we can search the parameter space of $\mathbf{c}$ and $\mathbf{x}$ to find minima of the objective function $J(\mathbf{c}, \mathbf{x})$. The many local minima of this objective function can be identified using any large-scale nonlinear optimization scheme; we made use of a Monte Carlo sampling scheme to optimize $\mathbf{c}$ along with a gradient descent method to find an optimal rotation $\mathbf{x}$. For all $(\mathbf{c}, \mathbf{x})$ such that $J(\mathbf{c}, \mathbf{x}) < 0$, $A(\mathbf{c}, \mathbf{x})$ defines a new constraint $\mathrm{Tr}[A(\mathbf{c}, \mathbf{x}) \cdot\,^2D] \leq a_{\max}$ which is violated by our current 2-RDM. This type of constraint is similar to that employed by Verstichel et al. in their study of molecular dissociation via RDM theory [19]. The major difference is that in their examples Verstichel and co-workers assume that the active-space and constraint operator are known in advance, whereas a crucial aspect of our work is the optimization of both the active-space and the constraint operator. Consequently, our methodology can be used to treat dissociation but can also be applied far more generally.

Having obtained one (or several) linear constraints which are violated by the approximate ground-state 2-RDM, we add these constraints to our constraint set and rerun the ground-state energy minimization algorithm. Constraints of the form in Eq. (10) are very easy to implement, since they are linear inequality conditions on the 2-RDM. Thus, they require considerably fewer resources than positivity constraints, which require the storage and manipulation of large matrices [13]. At each stage of the optimization procedure, constraints which are no longer active can be removed or "pruned" to improve

performance [18]. The constrained minimization of the ground-state energy and the addition of new constraints are repeated until some degree of convergence with respect to the ground-state energy is reached or until no new violated constraints (i.e., those with $J < 0$) can be found.

Our active-space constraints are similar in spirit to the complete active-space self-consistent field (CASSCF) method except for the crucial difference that the 2-RDM methodology allows us to simultaneously impose multiple constraints generated from different active spaces. Thus, we can effectively combine information which would be obtained from different CASSCF calculations, an approach that would not be possible apart from RDM methodology. The current approach should also be contrasted with other active-space RDM approaches which implement approximations to the 2-positivity constraints, but which are unrelated to the active-space constraints derived in this Letter [20,21].

To test the utility of these constraints, we performed calculations on the half filled, 1D Hubbard model which had previously been studied using RDM techniques in conjunction with the 2-positivity constraints [22]. The 1D Hubbard Hamiltonian is given by

$$H = -t \sum_{i=1}^{L} (c_i^\dagger c_{i+1} + c_{i+1}^\dagger c_i) + U \sum_{i=1}^{L} c_{i\alpha}^\dagger c_{i\alpha} c_{i\beta}^\dagger c_{i\beta}, \quad (17)$$

where L is the number of lattice sites and period boundary conditions are assumed. The constrained minimization of the ground-state energy was performed using the quasi-Newton L-BFGS algorithm following the matrix-factorization method for large-scale semidefinite programming discussed in [13,23,24]. Optimization proceeded until all constraints were satisfied to within a tolerance of 10^{-5}. The active-space size was chosen to be $L_S = 8$ (four spatial orbitals and two spin states) so that the maximum block size in the matrix representation of $\hat{A}$ was 36×36. Table I shows the results of the application of our constraints to the half filled, six-site Hubbard model for several values of the on-site repulsion term U, along with the number of active-space constraints n_A imposed at the final iteration. We ran our iterative procedure, adding five active-space

constraints at each iteration and pruning inactive constraints, until no additional violated constraints could be found or until the energy converged to within 0.1 mH (assuming atomic units). For small values of $U < 8$, we were unable to find constraints of the form specified in Eq. (14) which were violated by the approximate 2-RDM (although numerical results indicate that more complicated parametrizations of $\hat{A}$ did yield some violated constraints). However, for high values of U, Table I shows that new constraints were found which provided significant improvements to the approximate ground-state energy. For instance, at $U = 20$ the 2-positivity constraints recover $\sim 103\%$ of the correlation energy, but still give an excess correlation energy of 190 mH with respect to the exact ground-state energy [22]. Implementing the active-space constraints in addition to the 2-positivity constraints removes over half of this excess using only 17 extra linear constraints. Figure 1 shows the convergence of the algorithm with respect to the number of iterations, as constraints are added for $L_S = 6$ and $L_S = 8$. The algorithm successively adds constraints and prunes inactive constraints, slowly converging to some active-space limit given the size of the active space. In general, the larger the active space L_S, the more accurate the constraints will be.

With regard to the algorithm outlined above, several points merit attention. First, it must be stressed that the parametrization of the matrix A suggested in Eq. (14) is entirely arbitrary; there is no guarantee that the optimal choice of A can be parametrized as stated in Eq. (14). It would be extremely valuable to know in advance which matrices A and which subspaces S are likely to give the best improvement to ground-state energies. A second, related issue is the optimization procedure to select useful constraints. There is no guarantee that constraints which minimize the objective function $J(\mathbf{c}, \mathbf{x})$ will necessarily provide the most dramatic corrections to the ground-state energy. A more thoughtful procedure would not merely maximize the degree of violation of a given constraint, but would maximize the impact of such a constraint on the ground-state energy $\text{Tr}[^2K\,^2D]$. Third, we have already alluded to the fact that the active-space constraints, being

TABLE I. Exact energy and deviations for the $L = 6$ site Hubbard model with periodic boundary conditions and $t = 1$. E_{FCI} is the exact, full configuration interaction (FCI) ground-state energy. ΔE_{DGQ} and ΔE_A are, respectively, the energy differences between the exact energy and the approximate ground-state energies obtained by imposing the 2-positivity constraints and 2-positivity plus active-space constraints with $L_S = 8$. n_A gives the number of constraints imposed at the final iteration to obtain the approximate ground-state energy.

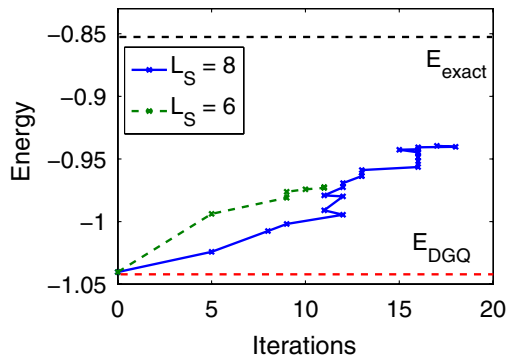| U	E_{FCI}	ΔE_{DGQ}	ΔE_A	$\Delta E_{\text{DGQ}} - \Delta E_A$	Excess recovered	n_A
8	−2.048	−0.300	−0.289	0.011	3.7%	11
20	−0.853	−0.187	−0.087	0.100	53.5%	17
40	−0.429	−0.102	−0.039	0.063	61.8%	16
80	−0.215	−0.053	−0.025	0.028	52.8%	13

FIG. 1 (color online). The approximate energy versus the number of active-space constraints for a $L = 6$ site Hubbard model with $U/t = 20$. Initially, no additional constraints are added, yielding E_{DGQ}. A significant fraction of the energy difference between the exact result and the DGQ answer can be recovered by imposing a small number of active-space constraints. The larger the active-space, the more accurate the final energy that is obtained.

linear, can be implemented far more efficiently than the 2-positivity constraints. Consequently, it would be quite interesting to examine whether we could abandon the 2-positivity constraints altogether for certain problems and simply perform 2-RDM calculations using only active-space constraints. Fourth, we have assumed that the operator $\hat{A}$ does not act on spin; this restriction could be dropped. All of these suggested modifications could potentially improve both the accuracy and efficiency of the active-space constraints suggested herein. Nonetheless, even as they are currently implemented, active-space constraints appear to be a valuable new tool in variational 2-RDM calculations.

Finally, it must be stressed that the idea suggested in this Letter is quite general. The key observation is that if $\hat{A}$ acts only on some reduced active space, then a_{max} can be calculated efficiently. However, there are many other situations in which an exact a_{max} can be calculated efficiently. For instance, if $\hat{A}$ is a single-particle operator, if it commutes with some symmetry operation like translation or rotation, or if it is a pairing Hamiltonian, then a_{max} can be calculated much more efficiently than for a general operator $\hat{A}$. In fact, in the specific case of a pairing Hamiltonian, Van Neck *et al.* recently showed that an iterative scheme very similar to our own could be used to obtain an improved ground-state energy [25]. Regardless of how the linear constraints are obtained, the possibility of applying multiple constraints to a single ground-state calculation would not be possible using a wave-function-based approach, and necessitates a RDM methodology. Another key point to note is that the Hamiltonian $\hat{H}$ of the system itself need not have the same symmetry as $\hat{A}$ in order for new, efficient constraints to be obtained. For instance, we could consider a system which is nearly translationally invariant (a Hubbard model with mild disorder) and yet could implement constraints based on exact translational invariance. When we consider that any form of $\hat{A}$ which allows for an efficient calculation of a_{max} will yield new constraints, it becomes clear that the approach presented in this Letter is extremely versatile and could potentially be quite powerful.

*Corresponding author.
Neil.Shenvi@duke.edu

[1] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (Dover, Mineola, NY, 1996).
[2] P.-O. Löwdin, Phys. Rev. **97**, 1474 (1955).
[3] R. McWeeny, Rev. Mod. Phys. **32**, 335 (1960).
[4] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963).
[5] A. J. Coleman and V. I. Yukalov, *Reduced Density Matrices: Coulson's Challenge* (Springer, Berlin, 2000).
[6] M. Nakata, H. Nakatsuji, M. Ehara, M. Fukuda, K. Nakata, and K. Fujisawa, J. Chem. Phys. **114**, 8282 (2001).
[7] D. A. Mazziotti and R. M. Erdahl, Phys. Rev. A **63**, 042113 (2001).
[8] M. Nakata, M. Ehara, and H. Nakatsuji, J. Chem. Phys. **116**, 5432 (2002).
[9] D. A. Mazziotti, Phys. Rev. A **65**, 062511 (2002).
[10] K. Husimi, Proc. Phys. Math. Soc. Jpn. **22**, 264 (1940).
[11] J. Mayer, Phys. Rev. **100**, 1579 (1955).
[12] C. Garrod and J. K. Percus, J. Math. Phys. (N.Y.) **5**, 1756 (1964).
[13] D. A. Mazziotti, Phys. Rev. Lett. **93**, 213001 (2004).
[14] R. M. Erdahl, Int. J. Quantum Chem. **13**, 697 (1978).
[15] Z. Zhao, B. J. Braams, M. Fukada, M. L. Overton, and J. K. Percus, J. Chem. Phys. **120**, 2095 (2004).
[16] D. A. Mazziotti, Phys. Rev. A **74**, 032501 (2006).
[17] G. Gidofalvi and D. A. Mazziotti, J. Chem. Phys. **126**, 024105 (2007).
[18] T. Juhász, N. Shenvi, and D. A. Mazziotti, Chem. Phys. Lett. **445**, 79 (2007).
[19] B. Verstichel, H. van Aggelen, D. Van Neck, P. W. Ayers, and P. Bultinck, J. Chem. Phys. **132**, 114113 (2010).
[20] G. Gidofalvi and D. A. Mazziotti, J. Chem. Phys. **127**, 244105 (2007).
[21] G. Gidofalvi and D. A. Mazziotti, J. Chem. Phys. **129**, 134108 (2008).
[22] J. R. Hammond and D. A. Mazziotti, Phys. Rev. A **73**, 062505 (2006).
[23] D. A. Mazziotti, J. Chem. Phys. **121**, 10 957 (2004).
[24] C. Zhu, R. Byrd, and J. Nocedal, ACM Trans. Math. Softw. **23**, 550 (1997).
[25] D. Van Neck and P. W. Ayers, Phys. Rev. A **75**, 032502 (2007).