Nonorthogonal localized molecular orbitals in electronic structure theory

Shubin Liu
Department of Chemistry, Duke University, Durham, North Carolina 27708-0346

José M. Pérez-Jordá
Department of Chemistry, Duke University, Durham, North Carolina 27708-0346 and Departamento de
Química Física, Universidad de Alicante, Apatado 99, E-03080, Spain

Weitao Yang
Department of Chemistry, Duke University, Durham, North Carolina 27708-0346

(Received 20 July 1999; accepted 26 October 1999)

The concept of nonorthogonal localized molecular orbital (NOLMO) is investigated in this paper. Given a set of the commonly used canonical molecular orbitals, a direct minimization algorithm is proposed to obtain both the orthogonal localized molecular orbitals (OLMO) and NOLMO by using the Boys criterion and conjugate gradient minimization. To avoid the multiple-minimum problem, the absolute energy minimization principle of Yang is employed to obtain initial guesses. Contrary to the early conclusion drawn by Lipscomb and co-workers who claimed that OLMOs and the corresponding NOLMOs are more or less the same, we found that NOLMOs are about 10%–30% more localized than OLMOs. More importantly, the so-called “delocalization tail” that plagues OLMOs is not present in NOLMOs, showing that NOLMOs are more compact and less oscillatory and capable of providing greater transferability in describing the electronic structure of molecules. We also found that main lobes of NOLMOs are slightly larger in size than those of OLMOs because of the normalization requirement. These features establish NOLMOs to be valuable as building blocks in electronic structure theory and for the understanding of chemical bonding. They show the promise for the utilization of NOLMOs—the most localized possible—in the linear scaling approaches of the electronic structure theory for molecules and solids. © 2000 American Institute of Physics. [S0021-9606(00)30204-5]

I. INTRODUCTION

Localized molecular orbitals (LMO) have been of considerable interest in quantum chemical studies for decades.1–3 The basic idea goes back to Fock’s observation4 that a single determinant many-electron wave function is invariant to unitary transformations. One thus has substantial freedom in the choice of the orbitals that describe a given system. There are several reasons for seeking LMOs: They connect with the classical concepts of chemical bonds, they are useful in isolating functional groups within chemically related molecules, and it is possible to transfer LMOs from one molecule to others with similar structure. Also, LMOs are especially suitable for a treatment of correlation because they are localized in different regions of the molecule, and presumably the correlation between electrons that are in LMOs separated by a long distance is much smaller than between those which are in LMOs near each other.5,6 Recent development of linear scaling methods7–19 of electronic structure theory for larger molecules and solids has led to renewed interest in LMOs. This is the precise reason that the present study was resulted.

There are two categories of LMOs in the literature. One is the orthogonal localized molecular orbital (OLMO)5 and the other is the nonorthogonal localized molecular orbital (NOLMO),20–29 sometimes also called strictly localized molecular orbital26 or extremely localized molecular orbital.29 The OLMO is obtained by a unitary transformation of the canonical molecular orbital, while NOLMO is obtained by a nonsingular transformation without the orthogonality condition. Since more degrees of freedom are available in the determination of NOLMO, it is generally anticipated that NOLMO should provide more localized description of molecular orbitals than OLMO.30,31

Early developments of the LMO concept focused on OLMOs.3 They could be determined either by introducing a localization potential22,23 into the integrodifferential equations that define the molecular orbitals (MOs) themselves or subjected to a unitary transformation that extremizes some functions of the MOs, such as Boys’ maximal spreads of orbitals (the sum of the squares of the distances between the centroids of the orbitals)1 and Edmiston-Ruedenberg’s maximal orbital self-repulsions.2 Recently, Vanderbilt et al. has applied the Boys method to determine the OLMOs for solids.34

The main problem with OLMOs is that they possess delocalization tails35 outside the localization center which complicates the transferability of descriptions of LMOs from one system to another. As pointed out by various authors,21–29 it is not necessary to use OLMOs, but instead one can use a single determinant of nonorthogonal, strictly local, molecular orbitals (NOLMOs). The concept of NOLMO was first introduced by Anderson20 and Diner et al.21 Most of the methods used in the literature to determine NOLMOs are procedures for variational optimization, consisting of partitioning a set
of basis functions into subsets corresponding to certain localization regions (atom, lone pair, bond). Each NOLMO is then expanded in the basis functions of the corresponding subset, which means that the delocalization tails are explicitly excluded. This type of approach produces only approximate NOLMOs because of the use of a subset of the basis sets for each NOLMO.

Another algorithm was introduced by Lipscomb and co-workers to obtain NOLMOs that are exact in the sense that the resulting wave function or density matrix is mathematically equivalent to that expressed from the canonical MOs. The Boys and Edmiston–Ruedenberg criteria were employed to determine NOLMOs, but the authors claimed that relaxation of the orthogonality condition leads to LMOs which are still fairly orthogonal to each other and which exhibit no spectacular increase in localization.

Very recently, OLMO and NOLMO have acquired renewed interests in linear scaling implementation for electronic structure calculations of solids and large molecular systems, e.g., proteins and DNAs. In last few decades, accurate calculation of electronic structures of molecules and solids has become possible with the rapid developments of computer software and hardware as well as robust algorithms from, for example, density functional theory. In conventional DFT algorithms involving full or iterative diagonalization, electronic structure is obtained in terms of canonical molecular orbitals that generally extend throughout the whole system under investigation. The numeric effort to compute such extended states scales at least as \( N^3 \) with \( N \) the number of occupied molecular orbitals. This \( N^3 \) unfavorable scaling behavior is one of the numerical bottlenecks that should ultimately be removed before applications of electronic structure calculations are routinely made for large systems.

The first \( O(N) \) solution was developed by Yang in the divide-and-conquer approaches. Subsequently, there has been a surge of interests in the development of \( O(N) \) methods; for a review, see Ref. 8. There are two main categories of methods for localization developed so far. The local Hamiltonian approaches obtain the local properties, like the electron density and the kinetic energy density, from the one-electron potential locally, bypassing solution to the Hamiltonian approaches obtain the local properties, like the electron density and the kinetic energy density, from the one-electron potential locally, bypassing solution to the

The variational principle of Kim et al. who pointed out the need to allow the number of the localized, orthogonal orbitals to exceed the number of occupied states. The variational principle of Kim et al. is limited in that the ground state energy is only a local minimum of the variational functional, it is only established for non-interacting electrons and it uses OLMOs. Use of NOLMOs in the \( O(N) \) approaches of electronic structure calculations was advocated by Stechel et al. Two absolute minimum principles for the linear scaling determination of NOLMO was developed by Yang: In the variational principles, the ground state energy of density functional theory or the Hartree–Fock theory is expressed as the global minimum of the functionals of NOLMOs and the number of NOLMOs are allowed to exceed the number of occupied orbitals which alleviates the problem of multiple local minima.

In this paper, we intend to answer the following two questions. (i) How much more localized will an NOLMO be compared with an OLMO? (ii) What do OLMOs and NOLMOs look like for typical simple molecules? These questions need to be answered before any application of NOLMOs is seriously pursued in the \( O(N) \) method. Negative answers to these questions were seemingly available from Lipscomb et al., who concluded that OLMOs and NOLMOs are more or less the same, without apparent increase of localization in the latter. We employ a direct minimization algorithm to determine NOLMOs in this investigation and will show that substantial difference is observable between the two kinds of localized orbitals.

The rest of the present paper is organized as follows: In Sec. II, the methodological framework of this study is outlined. This includes the orthogonal and nonorthogonal representations of density matrix and the algorithm we employed to obtain both the OLMO and NOLMO. Section III describes the numerical techniques used in the implementation of the above algorithm, including conjugate gradient minimization, initial guesses, and procedures. In Sec. IV, results and discussion of OLMOs and NOLMOs are presented for a few simple molecules. Finally, a few concluding remarks are given in Sec. V.

II. METHODOLOGY

Theoretical basis of our investigation is based on the Hohenberg–Kohn–Sham density functional theory for molecules and solids. But the extension to other non-interacting electron methods is straightforward. To facilitate
The density operator \( \hat{\rho} \) obeys the following three required conditions: Hermitian \( \hat{\rho} = \hat{\rho}^\dagger \), idempotency \( \hat{\rho}^2 = \hat{\rho} \), and normalization \( 2 \text{Tr}(\hat{\rho}) = N \).

The conventional approach amounts to representing \( \hat{\rho} \) in terms of a set of \( N/2 \) orthogonal and normalized orbitals \( \{ \psi_i \} \)

\[
\hat{\rho} = \sum_i^{N/2} |\psi_i\rangle \langle \psi_i|.
\]

(1)

where \( S \) is the overlap matrix. Full rank of \( S \) or existence of \( S^{-1} \) is required.

We start with a set canonical MOs \( \{ \psi_i, i=1,\ldots,N/2 \} \), where \( N \) is the number of electrons. \( \psi_i \) are orthonormal

\[
\langle \psi_i | \psi_j \rangle = \delta_{ij}.
\]

(3)

Choose a \( (N/2 \times N/2) \) nonsingular nonsymmetric transformation matrix \( A \) such that the resulting nonorthogonal MOs, \( |\phi_k\rangle = \sum_{i=1}^{N/2} |\psi_i\rangle A_{ik} \), are most localized according to some criterion. No “localized region” is introduced. Except for the normalization requirement,

\[
\langle \phi_k | \phi_k \rangle = 1,
\]

(4)

or in terms of the transformation matrix \( A \)

\[
\sum_{i}^{N/2} (A_{ik})^2 = 1,
\]

(5)

all the elements of \( A \) are allowed to change freely during minimization. We use the Boys localization criterion\(^\dagger\) as the object functional, as used by others.\(^{36,34} \) For normalized non-orthogonal orbitals \( \phi_i \), we define the spread functional

\[
\Omega[A] = \sum_{i}^{N/2} \langle \phi_i | (\hat{r}^2 - r_i^2) | \phi_i \rangle
\]

\[
= \sum_{i}^{N/2} [\langle \phi_i | \hat{r} | \phi_i \rangle^2 - (\langle \phi_i | \hat{r} \hat{r} | \phi_i \rangle)^2].
\]

(6)

where

\[
r_i = \langle \phi_i | \hat{r} | \phi_i \rangle.
\]

(7)

Notice that both \( \hat{r} \) and \( r_i \) are vectors with \( x,y,z \) components. The minimum of \( \Omega[A] \) over all the nonsingular matrix \( A \) will give a set of nonorthogonal MOs which have minimum spread in the real space (i.e., most localized). The rank is just the number of independent nonorthogonal MOs \( \phi_i \).

Using

\[
|\phi_k\rangle = \sum_{i=1}^{N/2} |\psi_i\rangle A_{ik},
\]

and

\[
|\phi_k\rangle = \sum_{j=1}^{N/2} A_{jk} |\psi_j\rangle.
\]

(8)

Eq. (6) becomes

\[
\Omega[A] = \sum_{i=1}^{N/2} \left( (A^\dagger R^2 A)_{ik} - ((A^\dagger R A)_{ik})^2 \right),
\]

(9)

where we have used Eq. (3).

\[
(R^2)_{ji} = \langle \psi_j | \hat{r}^2 | \psi_i \rangle = \langle \psi_j | \hat{x}^2 + \hat{y}^2 + \hat{z}^2 | \psi_i \rangle
\]

(10)

and

\[
(R)_{ji} = \{ \langle \psi_j | \hat{x} | \psi_i \rangle, \langle \psi_j | \hat{y} | \psi_i \rangle, \langle \psi_j | \hat{z} | \psi_i \rangle \}.
\]

(11)

In terms of matrix elements

\[
\Omega[A] = \sum_{i=1}^{N/2} \left( \sum_{j=1}^{N/2} \sum_{l=1}^{N/2} A_{jk}^* A_{lj} \langle \psi_j | \hat{r}^2 | \psi_l \rangle \right) - \left[ \sum_{i=1}^{N/2} \sum_{j=1}^{N/2} \sum_{l=1}^{N/2} A_{jk}^* \langle \psi_j | \hat{r} \hat{r} | \psi_l \rangle A_{lj} \right]^2 \left[ |y| - |z| \right],
\]

(12)

where \( [y] \) and \( [z] \) are short-hand notations for terms resulting from replacing \( x \) by \( y \) and \( x \) by \( z \) in the same expression for \( x \) within the bracket (this notation will also be used elsewhere).

The condition of \( A \) being nonsingular is equivalent to that \( A \) is a full rank matrix, that is, \( \text{rank}(A) = N/2 \). This is imposed in the following way. The rank of \( A \) is calculated as \( \text{rank}(A) = \text{Tr}(A A^{-1}) \). To convert to a positive definite matrix from the nonsymmetric \( A \), we use

\[
\text{rank}(A) = \text{rank}(A^\dagger A) = \text{Tr}((A^\dagger A)(A^\dagger A)^{-1}).
\]

(13)

Then, we can use the variational principle developed in Ref. 19; namely

\[
\text{rank}(A) = - \min_{x \sim x^1} \text{Tr}((-A^\dagger A)(2X - X A^\dagger A X A^\dagger)),
\]

(14)

where \( X \) is an auxiliary symmetric matrix which becomes \( (A^\dagger A)^{-1} \) at the minimum. In above formula, we have used the fact that \( A^\dagger A \) is positive semidefinite. To require \( A \) to be full rank in the minimization, we define following constraint functional:

\[
\Omega_a[A,X] = \{ 2N/2 - \text{Tr}( (A^\dagger A)(2X - X A^\dagger A X A^\dagger) ) \}^2.
\]

(15)

Note that

\[
\min_{x \sim x^1} \Omega_a[A,X] = \{ N/2 - \text{rank}(A^\dagger A) \}^2,
\]

(16)

and that any full rank \( A \) would give the following minimum:

\[
\min_{A} \Omega_a[A,X] = \min_{A} (N/2 - \text{rank}(A^\dagger A))^2 = 0.
\]

(17)

Therefore, we can formulate the final solution of finding NOLMOs as
with a constant $C_a$ to be adjusted.

In the above variational principle, if one does not vary $X$ and sets $X = I$, the $(N/2 \times N/2)$ unity matrix, then the present algorithm gives OLMOs, because orbitals become orthonormal when minimization is achieved. This can be seen from the fact $\{N/2 - \text{Tr}(\{A^\dagger A\}(2I - A^\dagger A))\}^2$ is minimum for orthonormal orbitals in which $A^\dagger A = I$. Thus for determining OLMO, we can use the following minimization procedure:

$$\min \{ \Omega[A] + C_a \Omega_a[A,I] \},$$

where $C_a$ to be determined. \textit{In practice}, we choose a more efficient way and simply force the condition that $A$ is an orthogonal transformation, $A^\dagger A = I$, or $\sum_i N/2 A_{i\ell} A_{i\ell'} = \delta_{\ell\ell'}$ at each step of minimization. Thus, for OLMOs, we simply use the following minimization procedure:

$$\min \{ \Omega[A] + C_a \Omega_a[A,I] \},$$

with a constant $C_a$ to be determined. \textit{In practice}, we choose a more efficient way and simply force the condition that $A$ is an orthogonal transformation, $A^\dagger A = I$, or $\sum_i N/2 A_{i\ell} A_{i\ell'} = \delta_{\ell\ell'}$ at each step of minimization. Thus, for OLMOs, we simply use the following minimization procedure:

$$\min \{ \Omega[A] + C_a \Omega_a[A,I] \},$$

Equations (18) and (20) are our main working procedures for determining NOLMOS and OLMOS, respectively.

Lipscomb and co-workers also used the Boys criterion to obtain NOLMOS. Two points distinguish our algorithm from theirs. First, instead of minimizing the sum of the spread of each orbital as in Eq. (6), making each orbital as compact as possible, Lipscomb and co-workers maximized the sum of the squares of the distances between the orbital centroids with the following Boys functional

$$B = \sum_{k > l} (\langle \phi_k | \mathbf{r} | \phi_k \rangle - \langle \phi_l | \mathbf{r} | \phi_l \rangle)^2,$$

making orbitals to stay as far away from each other as possible. The criteria of minimizing Eq. (6) and maximizing Eq. (21) become identical only when the LMO are orthogonal, $\langle \phi_k | \phi_l \rangle = \delta_{kl}$. In general, they are not equivalent in the case of NOLMO. Second, a factor of $|S|^2$, the square of determinant of the overlap matrix $S$ of the NOLMO, was put in front of the Boys criterion. Therefore, maximization in their approach was carried out for the product of $|S|^2$ and the Boys functional, Eq. (21). The second point is more severe, because we observe in all our calculations that $|S|^2$ is maximal only when all orbitals are orthonormal, giving $|S|^2 = 1$. For NOLMOS, as we will show below in the next section, $|S|^2$ is far less than the optimal value. Therefore, when the $|S|^2$ factor is maximized together with the Boys criterion, NOLMOS obtained will tend to be orthogonal. This explains why little difference was observed between OLMOs and NOLMOS in their implementation.  

III. NUMERICAL IMPLEMENTATION

A. Gradients of the functionals

The above formalism, Eqs. (18) and (20), can be minimized using standard off-the-shelf optimization algorithms. We have chosen the standard conjugate gradient (CG) approach, since it only requires the value of the functional and its gradient. The CG minimization generates a series of “noninterfering” search directions along which the object functional is minimized. In our case, CG minimization should be performed twice, one for $X$, the real and symmetric auxiliary matrix which converges to $(A^\dagger A)^{-1}$ at the minimum, and the other for $A$. For each of the minimizations, an initial guess for $A$ or $X$ is obtained, as is the corresponding search direction. Minimization proceeds as a series of one-dimensional line searches along the CG search direction. The line search step is chosen to correspond to a minimum of the object functional along the search direction. The search direction of the next step is chosen such that it is conjugate to the previous step.

The gradient of $\Omega[A]$ needed in CG minimization is as follows:

$$\frac{\partial \Omega[A]}{\partial A_{kl}} = 2 \left( \sum_{i=1}^{N/2} \langle \psi_{hi} | \mathbf{r} | \psi_{hi} \rangle A_{il} \right) - 4 \left( \sum_{j=1}^{N/2} \sum_{i=1}^{N/2} A_{ij} \langle \psi_{hi} | \mathbf{r} | \psi_{hi} \rangle A_{il} \right) \times \left( \sum_{k} \langle \psi_{hi} | \mathbf{r} | \psi_{hi} \rangle A_{kl} \right) - [y] - [z].$$

For the penalty functional

$$\frac{\partial \Omega[A,X]}{\partial A_{kl}} = \left\{ 8 \{1/2 - \text{Tr}[A^\dagger A(2X - A^\dagger A X)] \} ight\}$$

$$\times [A(-X + A^\dagger A X)]_{kl}$$

and

$$\frac{\partial \Omega[A,X]}{\partial X_{kl}} = \left\{ 4 \{1/2 - \text{Tr}[A^\dagger A(2X - A^\dagger A X)] \} \right\}$$

$$\times [-A^\dagger A + A^\dagger A A X A^\dagger A]_{kl}.$$
The five occupied localized molecular orbitals are in columns. The symbol "$r$" stands for a nonzero number to be generated randomly.

$$\phi_i(\mathbf{r}) = \sum_{\gamma} C_{\gamma i} \chi_{\gamma}(\mathbf{r}),$$

(25)

Eq. (8) thus becomes

$$\mathbf{C} = \mathbf{C}_0 \mathbf{A}_0,$$

(26)

where $\mathbf{C}_0$ is the $(N_{\text{basis}} \times N/2)$ coefficients matrix of canonical orthonormal molecular orbital and $\mathbf{C}$ is that for the localized orbitals. We obtain $\mathbf{C}_0$ by performing conventional Kohn–Sham calculations implemented in our laboratory\cite{66,67} and a trial NOLMO $\mathbf{C}$ by employing Yang's absolute energy minimization principle.\cite{19} According to this principle, the total energy functional could be minimized in the following way:

$$E(N) = \min_{\{\phi_i\}} \left\{ \min_{\mathbf{x}} \Omega[(2\mathbf{X} - \mathbf{X}\mathbf{X}), \{\phi_i\}] + \eta N \right\},$$

(27)

where $\mathbf{X}$ is an auxiliary symmetric matrix which at the minimum becomes $\mathbf{S}^{-1}$

$$\Omega[\mathbf{A}, \{\phi_i\}] = E[\hat{\rho}_\mathbf{A}] - 2\eta \text{ Tr}(\hat{\rho}_\mathbf{A}),$$

(28)

$$\hat{\rho}_\mathbf{A} = \sum_{ij} |\phi_i\rangle \langle \phi_j| \mathbf{A}_{ij},$$

(29)

and

$$E[\hat{\rho}_\mathbf{A}] = \text{ Tr}[\hat{\rho}_\mathbf{A} (\mathbf{h} - \eta \mathbf{s})],$$

(30)

where the $(N_{\text{basis}} \times N_{\text{basis}})$ Kohn–Sham matrix $\mathbf{h}$ is defined as $h_{\gamma\delta} = \langle \chi_{\gamma} | \hat{h} | \chi_{\delta} \rangle$, the $(N_{\text{basis}} \times N_{\text{basis}})$ basis function overlap matrix $\mathbf{s}$ is defined by $s_{\gamma\delta} = \langle \chi_{\gamma} | \chi_{\delta} \rangle$, with the Kohn–Sham Hamiltonian operator defined as $\hat{h} = -\frac{1}{2} \nabla^2 + \mathbf{v}_{\text{eff}}(\mathbf{r})$, and $\mathbf{v}_{\text{eff}}(\mathbf{r})$ the Kohn–Sham effective potential.

To obtain a trial guess for $\mathbf{A}$, an $N_{\text{basis}} \times N/2$ normalized (column-wise) coefficient matrix $\mathbf{C}$ is formed which corresponds to a set of approximate NOLMOS. The elements of $\mathbf{C}$ are 0 if atomic orbitals do not contribute to LMOs. Otherwise, they are generated by random numbers. To determine $\mathbf{C}$, chemical information from the Lewis structure of molecules is required. An example is shown in Table I for the CH$_4$ molecule. The first LMO comes from the orbital of carbon. So, for the first column of $\mathbf{C}$, nonzero coefficients come from the carbon atom and all contributions from hydrogen atoms are zero. The rest four occupied LMOs represent four equivalent C–H bonds. Using the second LMO (the second column of the $\mathbf{C}$ matrix) as an example, as shown in Table I, nonzero coefficients are for atomic orbitals from the carbon and first hydrogen atoms and the rest of the column all vanish. The other three LMOs can be built similarly. By doing that, the orbitals to be optimized possess localization patterns similar to that described in the Lewis structure of chemical bonding and thus become nonorthogonal localized molecular orbitals. These patterns are preserved until minimization is finished. Note that the energy from this process will be higher than without these constraints because degrees of freedom is limited to preserve the bonding patterns. Since matrix $\mathbf{C}$ still has a number of elements that can be freely adjusted, the increase in the total energy is anticipated to be small.

After both $\mathbf{C}_0$ and $\mathbf{C}$ are determined, an initial guess for $\mathbf{A}$, $\mathbf{A}_0$, is obtained by multiplying both sides of Eq. (26) by $\mathbf{C}_0$ and using the fact that $\mathbf{C}_0 \mathbf{s} \mathbf{C}_0 = \mathbf{I}$. Thus, one has

$$\mathbf{A}_0 = \mathbf{C}_0 \mathbf{s} \mathbf{C}.$$  

(31)

It should be mentioned that the form of the initial $\mathbf{C}$ matrix is not unique. Even for simple molecules like CH$_4$, we still can construct 120 (or 5!) different forms for it by permuting the

<table>
<thead>
<tr>
<th>Basis set$^a$</th>
<th>Canonical Spread</th>
<th>%b</th>
<th>NOLMO Spread</th>
<th>%c</th>
</tr>
</thead>
<tbody>
<tr>
<td>2s1p1s</td>
<td>21.38</td>
<td>14.88</td>
<td>30.4</td>
<td>12.50</td>
</tr>
<tr>
<td>3s2p1d/2s1p</td>
<td>17.50</td>
<td>10.58</td>
<td>39.5</td>
<td>9.54</td>
</tr>
<tr>
<td>4s4p2d/2s1p</td>
<td>17.06</td>
<td>10.64</td>
<td>37.6</td>
<td>9.63</td>
</tr>
<tr>
<td>4s4p2d/3s2p</td>
<td>17.04</td>
<td>10.57</td>
<td>37.7</td>
<td>9.56</td>
</tr>
</tbody>
</table>

$^a$The basis set in the numerator is for carbon and denominator for hydrogen. For example, 2s1p1s means that the basis set for C is 2s1p and for H is 1s, i.e., the minimal basis set for both C and H.

$^b$Compared with the canonical (delocalized) MOs.

$^c$Compared with the corresponding OLMOs.

| $C_a$ | Spread $\Omega[\mathbf{A}]$ | Rank of $\mathbf{A}$ | Eigenvectors of NOLMO overlap matrix | Value of $|\mathbf{S}|^2$ |
|-------|--------------------------|---------------------|-------------------------------------|------------------------|
| $10^{-2}$ | 0.492                   | 3.995               | $3 \times 10^{-9}$, $6 \times 10^{-6}$, $7 \times 10^{-6}$, $6 \times 10^{-7}$, 5.000 | $2 \times 10^{-45}$ |
| $10^{-1}$ | 9.542                   | 5.000               | 0.58, 0.70, 0.71, 1.16, 1.84         | 0.387 |
| $10^0$    | 9.539                   | 5.000               | 0.59, 0.70, 0.71, 1.16, 1.84         | 0.385 |
| $10^1$    | 9.536                   | 5.000               | 0.57, 0.69, 0.70, 1.19, 1.85         | 0.378 |
| $10^2$    | 9.535                   | 5.000               | 0.57, 0.69, 0.70, 1.18, 1.86         | 0.367 |
| $10^3$    | 9.535                   | 5.000               | 0.57, 0.69, 0.70, 1.18, 1.86         | 0.367 |

$^a$The basis set in the numerator is for carbon and denominator for hydrogen. For example, 2s1p1s means that the basis set for C is 2s1p and for H is 1s, i.e., the minimal basis set for both C and H.
columns of initial $C$ matrix in Table I. Note that each such permutation equally represents the Lewis structure. A particular chosen pattern of $C$ and the associated $A_0$ may still result in orbital linear dependency during the NOLMO minimization. This can be resolved by taking a different permutation for the initial $C$ matrix.

With the initial guess just described, we were able to obtain the lowest minimum of the spreads, Eqs. (18) and (20), compared to any other initial guess generated randomly.

C. Minimization procedures

Kohn–Sham calculations \cite{66,67} are performed first to optimize the molecular structure and to output the canonical MOs coefficient matrix $C_0$ and matrices needed in LMO optimization and initial guess determination. These matrices include the second-order moment matrix $\langle \psi_j | \vec{r}^2 | \psi_i \rangle$, the electron position matrices $\langle \psi_j | \hat{x} | \psi_i \rangle$, $\langle \psi_j | \hat{y} | \psi_i \rangle$, and $\langle \psi_j | \hat{z} | \psi_i \rangle$, the Kohn–Sham matrix $h$, and the basis function overlap matrix $s$. Then an absolute energy minimization is conducted to obtain the matrices $C$ in Eq. (26), with which, using Eq. (31), one obtains an initial guess for $A_0$. After that, the matrix $A$ for both OLMOs and NOLMOs are determined from the search of the minimum by using conjugate gradient minimization.

A few points should be addressed. (i) Choice of the constant $C_n$ in Eq. (18) for the penalty function. Since the penalty function vanishes at the solution point, the value of $C_n$ should have no impact on the final result. However, it influences the minimization process. This constant should be large enough to guide the minimization toward the right direction. We normally choose it to be 1000. (ii) The rank of the matrix $A$ should always be an integer. However, its numerical calculation by Eq. (14) may lead to noninteger values, because of numerical limits of the machine. Therefore, it is possible that the matrix is not full rank and the orbitals have linear dependence, even though it may numerically be very close to full rank. To check if the full rank condition is satisfied, we always calculate the eigenvalues of the overlap matrix $S$ of the LMOs. If they have one or more very small eigenvalues, less than $10^{-5}$, we regard the LMOs are linearly dependent, and must be discarded. In all the cases we studied so far, the failure, if exits, is invariably due to the initial input matrix $A_0$. We solve this problem simply by readjusting the initial $C$ matrix by permutation. (iii) Our algorithm is designed for normalized NOLMO orbitals. After

<table>
<thead>
<tr>
<th>MO</th>
<th>$\langle x \rangle$</th>
<th>$\langle y \rangle$</th>
<th>$\langle z \rangle$</th>
<th>$\Omega_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>12.00</td>
</tr>
<tr>
<td>2</td>
<td>0.79</td>
<td>0.79</td>
<td>2.60</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>-0.79</td>
<td>-0.79</td>
<td>2.61</td>
<td>0.77</td>
</tr>
<tr>
<td>4</td>
<td>-0.79</td>
<td>-0.79</td>
<td>2.62</td>
<td>0.77</td>
</tr>
<tr>
<td>5</td>
<td>0.79</td>
<td>-0.79</td>
<td>2.63</td>
<td>0.77</td>
</tr>
<tr>
<td>Total</td>
<td>10.58</td>
<td>9.54</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE IV. Coordinates of the atoms and orbital contributions to the total spread in the methane molecule calculated with the BLYP exchange-correlation energy functional, $C_n = 1000$, and the polarized 3s2p1d2s1p basis set (a.u.).

<table>
<thead>
<tr>
<th>Basis Set</th>
<th>$\epsilon_{\text{canonical}}$ (a.u.)</th>
<th>$\epsilon_{\text{Levvi}}$ (a.u.)</th>
<th>Canonical Spread (a.u.)</th>
<th>% OLMO</th>
<th>% NOLMO</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH</td>
<td>-23.0931</td>
<td>-23.0930</td>
<td>17.50</td>
<td>10.58(39.5)</td>
<td>9.54(9.8)</td>
</tr>
<tr>
<td>C$_2$</td>
<td>-43.4069</td>
<td>-43.4039</td>
<td>22.88</td>
<td>11.83(48.3)</td>
<td>9.47(19.9)</td>
</tr>
<tr>
<td>H$_2$O</td>
<td>-41.6119</td>
<td>-41.6083</td>
<td>10.11</td>
<td>8.18(19.1)</td>
<td>6.35(22.1)</td>
</tr>
<tr>
<td>C$_2$H$_2$</td>
<td>-44.0475</td>
<td>-44.0437</td>
<td>30.87</td>
<td>15.47(49.9)</td>
<td>13.24(14.4)</td>
</tr>
<tr>
<td>C$_2$H$_4$</td>
<td>-44.7604</td>
<td>-44.7422</td>
<td>41.94</td>
<td>18.42(56.1)</td>
<td>16.63(9.8)</td>
</tr>
<tr>
<td>C$_2$H$_6$</td>
<td>-45.3939</td>
<td>-45.3743</td>
<td>55.44</td>
<td>19.55(64.7)</td>
<td>16.70(14.6)</td>
</tr>
<tr>
<td>NH$_3$</td>
<td>-62.2723</td>
<td>-62.2426</td>
<td>29.98</td>
<td>15.18(49.4)</td>
<td>13.10(13.7)</td>
</tr>
<tr>
<td>CH$_2$OH</td>
<td>-64.0485</td>
<td>-64.0241</td>
<td>34.94</td>
<td>16.40(53.1)</td>
<td>11.38(30.5)</td>
</tr>
<tr>
<td>CH$_3$NH$_2$</td>
<td>-54.0167</td>
<td>-53.9896</td>
<td>38.63</td>
<td>17.95(53.5)</td>
<td>14.36(20.0)</td>
</tr>
<tr>
<td>C$_2$H$_6$</td>
<td>-132.200</td>
<td>-131.986</td>
<td>231.4</td>
<td>54.58(76.4)</td>
<td>41.33(24.3)</td>
</tr>
<tr>
<td>CO</td>
<td>-63.3556</td>
<td>-63.3540</td>
<td>14.02</td>
<td>12.41(11.5)</td>
<td>9.62(22.5)</td>
</tr>
<tr>
<td>CO$_2$</td>
<td>-104.914</td>
<td>-104.746</td>
<td>57.21</td>
<td>18.45(67.8)</td>
<td>14.90(19.2)</td>
</tr>
<tr>
<td>LiF</td>
<td>-55.0655</td>
<td>-55.0516</td>
<td>8.74</td>
<td>7.69(12.0)</td>
<td>6.57(14.6)</td>
</tr>
<tr>
<td>N$_2$</td>
<td>-61.9513</td>
<td>-61.9496</td>
<td>19.71</td>
<td>11.73(40.5)</td>
<td>9.32(20.5)</td>
</tr>
<tr>
<td>B$_2$H$_6$</td>
<td>-30.9332</td>
<td>-30.8665</td>
<td>59.76</td>
<td>19.14(68.0)</td>
<td>17.59(8.9)</td>
</tr>
<tr>
<td>C$_3$H$_5$(cycle)</td>
<td>-67.1013</td>
<td>-67.0565</td>
<td>73.13</td>
<td>25.88(64.6)</td>
<td>22.25(14.0)</td>
</tr>
<tr>
<td>SF$_6$</td>
<td>-552.899</td>
<td>-552.776</td>
<td>275.47</td>
<td>46.12(83.3)</td>
<td>40.31(12.6)</td>
</tr>
<tr>
<td>CH$_3$CH$_3$</td>
<td>-67.7133</td>
<td>-67.6725</td>
<td>109.23</td>
<td>29.04(73.4)</td>
<td>24.26(16.5)</td>
</tr>
<tr>
<td>CH$_3$CH$_2$CH$_3$</td>
<td>-90.0236</td>
<td>-89.7110</td>
<td>182.83</td>
<td>39.59(78.3)</td>
<td>33.20(16.1)</td>
</tr>
<tr>
<td>CH$_3$CONHCH$_2$</td>
<td>-139.334</td>
<td>-139.194</td>
<td>168.56</td>
<td>40.30(76.1)</td>
<td>35.34(12.6)</td>
</tr>
<tr>
<td>CH$_2$CHCH$_2$</td>
<td>-88.8651</td>
<td>-88.7990</td>
<td>167.43</td>
<td>34.90(79.2)</td>
<td>30.48(12.7)</td>
</tr>
<tr>
<td>Tetra-glycine</td>
<td>-508.885</td>
<td>-508.397</td>
<td>1745.29</td>
<td>127.33(92.7)</td>
<td>105.81(16.9)</td>
</tr>
<tr>
<td>Average %</td>
<td>57.1</td>
<td>17.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
each cycle of minimization, updated NOLMOs should be renormalized. For OLMO orbitals, Gram–Schmidt orthonormalization is carried out to ensure that \( A^\dagger A = I \). (iv) For the density functional calculations with our programs, we used the Becke–Lee–Yang–Parr \(^{68,69}\) functional for the exchange-correlation energy. All numerical integrations were performed with an error tolerance of \( 10^{-5} \).

D. Keeping a full rank

Given as an example, Table II shows the dependence of the optimized spread [Eq. (9)] on the constant \( C_\alpha \) for the methane molecule. When \( C_\alpha \) is too small, say \( 10^{-2} \), though one can get very small spread after minimization, the rank of the NOLMO is incorrect. Moreover, analysis of the eigenvalues of the NOLMO overlap matrix shows that four eigenvalues are less than \( 10^{-3} \), indicating that the NOLMOs obtained are linearly dependent. As \( C_\alpha \) increases, the full rank condition is satisfied and the optimized spread decreases monotonically with the increase of \( C_\alpha \). Beyond some point, \( 10^1 \), the minimized spread keeps unchanged and no longer depends on the value of the constant \( C_\alpha \). This constant, differing from the Lagrange multiplier, serves merely as the penalty and provides guidance for the object functional, Eq. (9), toward the minimum. The net contribution of its product with the penalty function, Eq. (15), should vanish at the solution point when the full rank condition is satisfied. Table II shows that when the penalty is too small, it does not provide robust guidance, leading to incorrect rank and interdependence of the NOLMOs. The constant \( C_\alpha \) should not be too large either because convergence of the CG minimization will be hindered. Based on what we have observed from methane, the value of \( C_\alpha \) used in the present work is fixed to be \( 10^3 \) for all systems we investigated.

In addition, shown in the last column of Table II are the values of \( |S|^2 \), the square of determinant of the overlap matrix \( S \) for NOLMOs (note that \( S = A^\dagger A \)). They are far less than the optimal value, which corresponds to the orthonormal orbital case, i.e., for OLMOs, \( |S|^2 = 1 \). This implies that if the optimization procedure has \( |S|^2 \) as a part of the object function, as did by Lipscomb and co-workers,\(^{36}\) the NOLMO
solution will possess tendency of being orthonormal. Therefore, the NOLMOs derived in this way are not true nonorthogonal LMOs.

IV. RESULTS AND DISCUSSION

A. Basis set effects

Basis set plays a role in NOLMOs calculations. Shown in Table III is a test for methane. The minimal basis and three different polarized basis sets are investigated. One finds that the minimal basis set gives the smallest percentage (30.4%) of OLMO spread decrease and the largest percentage (16.0%) of NOLMO spread decrease. Among three polarized basis sets, however, not much difference exists, indicating that stable results can be obtained with the polarized basis set. We use 2s1p for hydrogen, 3s2p1d for first-row elements, and 4s3p2d for second-row elements thereafter. The notion 2s1p stands for a $s$-type and 1 $p$-type numerical basis functions. 66,67

B. Comparing the degree of localization

To see whether OLMOs and NOLMOs obtained are reasonable, one convenient way is to check the orbital contributions of the total spread and coordinates of the localized orbital centers. In Table IV, we tabulated the coordinates of five OLMOs and NOLMOs of methane and their orbital contributions. The first LMO sits on the origin and corresponds to the inner 1$s$ molecular orbital stemmed from the 1$s$ atomic orbital of carbon. The other four LMOs are equivalent and their centers locate on the four different vortices of a tetrahedral, representing four identical C–H bonds of CH$_4$.

Now, let us turn to the first question raised in the Sec. I, namely, how much more localized NOLMOs are, compared with corresponding OLMOs? Lipscomb et al. 36 have previously shown that, based on their algorithm, the two kinds of LMOs are more or less the same and no substantial increase of localization was observed from NOLMOs. Table V summarizes the results we obtained from our algorithms for a number of simple molecules. Columns 4–6 tabulate the spreads of canonical MOs, OLMOs, and NOLMOs, respec-
tively. We calculated the relative percentages of localization increase (i.e., spread decrease) for both OLMOs and NOLMOs as well. For the systems we investigated, the relative percentage of OLMOs ranges from 11% to 83%, while that of NOLMOs from 9% to 30%. The average percentages of OLMO and NOLMO spread decreases are 52.9% and 17.0%, respectively. From these data, we observe that significant localization gain can be obtained from NOLMOs over OLMOs. For organic species (see, for example, the last five rows of Table V), about 75% of localization gain from OLMOs over canonical MOs and about 15% of localization gain from NOLMOs over OLMOs are typical. For small organic systems, such as $\text{C}_2\text{H}_4$ and $\text{C}_2\text{H}_6$, these numbers change to about 50% for OLMOs and about 15% for NOLMOs. There are a few inorganic systems investigated, such as NH$_3$, CO, H$_2$O, and LiF, where NOLMOs give even larger percentage than OLMOs.

Also shown in Table V, in the second and third columns, are the sum of eigenvalues from the absolute energy minimizations. The first minimization obtains the canonical single-particle wave function without the localization constraints superimposed on $\Phi$, and the second uses the Lewis structure information of the system to get a trial localized wave function. The second minimization gives an initial guess for the CG minimization of the matrix $A$. The value of the third column is always larger than that of the second as we anticipated because localization patterns are imposed in the former. But the difference between the two is often very small, indicating the good quality of description by the Lewis structure.

C. Compare the shape of localization

Although there exists substantial difference between the spreads of OLMOs and NOLMOs, as shown above, it is still unclear how much different the respective orbitals look like. From the literature, it is well known that OLMOs attain delocalization tails whose presence is necessary for OLMOs to satisfy the orthogonal condition. These tails destroy the description transferability of OLMOs from one system to another. NOLMOs from other previous studies$^{21-29}$ do not possess such “tails,” but localized regions for confining NOLMOs have to be introduced artificially. Our method is different in the fact that it is a direct minimization algorithm and no artificial localized region is introduced. What we use in our algorithm is to directly minimize the Boys functional, the sum of spreads of all occupied orbitals, with an arbitrary nonsingular nonsymmetric matrix transformation subject to the condition that the transformation matrix is full rank. The resulting NOLMOs and OLMOs from our procedure are completely equivalent to each other and to the canonical MOs in their representation of the electronic structure.

Figures 1–6 are contour plots of a few OLMOs and NOLMOs obtained from the present procedure. Figures 1 and 2 are two OLMOs (top) and NOLMOs (bottom), respectively, for $\text{C}_2\text{H}_4$, one for $\text{C}–\text{C}$ bond (Fig. 1) and the other for a $\text{C}–\text{H}$ bond (Fig. 2). Carbon atoms are represented by dark dots linked by a solid line. Four hydrogen atoms sit on the page but not shown here. It is seen from the top plot of Figs. 1 and 2 that while C–C and C–H bonds can be well reproduced by an OLMO, delocalization tails show up elsewhere, near nuclei of both carbon and hydrogen atoms. These tails are not present for the corresponding NOLMOs, as shown in the bottom plots of Figs. 1 and 2. As a compensation, however, we find that the size of the main lobe is larger than that of OLMO. This fact is readily discernible from Figs. 1 and 2 by comparing respective OLMOs (top) and NOLMOs (bottom) of the C–C and C–H bonds. The reason for larger main lobes of NOLMOs is simply the normalization condition. Both OLMO and NOLMO are normalized. OLMOs are more oscillatory and occupy more space, thus the main lobes can be smaller. For NOLMOs, no tails exit and less space is occupied. Hence, the main lobes will have larger sizes.

These two features of NOLMOs, i.e., no tails and larger main lobes, are confirmed by Figs. 3 and 4 for propane, $\text{C}_3\text{H}_8$. Figure 3 shows the OLMO (top) and NOLMO (bottom) for one C–C $\sigma$ bond and Fig. 4 shows the OLMO (top) and NOLMO (bottom) for one C–H bond (bottom). In both
OLMO cases (top) in Figs. 3 and 4, delocalization tails are present, but no tails are observed in the NOLMO (bottom) plots of the same bonds of the same compound. Main lobes of NOLMOs are again larger than the corresponding OLMOs in both figures. As a further illustration, the benzene molecule has been investigated. Figures 5 and 6 each show two contour maps of OLMOs (top) and NOLMOs (bottom) of one C–C bond (Fig. 5) and one C–H bond (Fig. 6) for the molecule. The same characteristics are confirmed.

In summary, we found that NOLMOs are significantly more localized and less oscillatory than OLMOs. They do not have delocalization tails extending elsewhere, but because of normalization requirement, the main lobe of NOLMOs is a little larger in size than OLMOs.

V. CONCLUDING REMARKS

The questions of whether NOLMOs are substantially different from OLMOs and what NOLMOs generally look like for simple molecular systems are addressed in this paper. Contrary to early conclusions drawn by Lipscomb et al.,

our results from a direct minimization algorithm using conjugate gradient minimization and Boys functional criterion show that much localization gain over OLMOs can be achieved by releasing the orthogonality condition. Depending on the systems under investigation, about 10%–30% decrease of spreads (increase of localization) can be obtained. Moreover, without explicitly introducing localized regions, our results show that delocalization tails that plague traditional OLMOs do not show up in NOLMOs. We also found that because of the normalization requirement, the main lobe of NOLMOs tends to occupy more space and is a little larger in size than that of corresponding OLMOs. These properties imply that NOLMOs can be more transferable than OLMOs in the description of the electronic structure of different but structurally similar molecules.

The conclusions drawn from this work may have impact on the development of O(N) approaches of electronic structure theory for large molecules and solids. The essence of an O(N) molecular orbital approach is to employ LMOs in the Kohn–Sham equation, leading to sparse Hamiltonian and overlap matrices that could be solved in an O(N) manner. Our present results support the use NOLMOs, in place of OLMOs, in the O(N) implementation. They have better transferability because they do not possess delocalization tails. They are more localized and less oscillatory than OLMOs, so less mesh points or smaller localized regions are needed to simulate NOLMOs, leading to efficiency improvements of algorithms.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation and the National Institute of Health through the program project at the University of North Carolina–Chapel Hill and at Duke University. We thank Professor David Vanderbilt for providing with us a program to draw contour plots presented in this study.

1S. F. Boys, Rev. Mod. Phys. 32, 296 (1960).
4V. Fock, Z. Phys. 61, 126 (1930).
64 M. C. Payne et al., Rev. Mod. Phys. 64, 1045 (1992).