JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

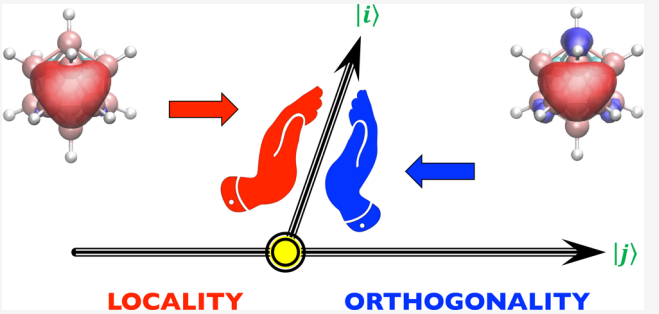
Variable-Metric Localization of Occupied and Virtual Orbitals

Ziling Luo and Rustam Z. Khaliullin*

Cite This: https://doi.org/10.1021/acs.jctc.1c00379 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The key idea of the variable-metric approach to orbital localization is to allow nonorthogonality between orbitals while, at the same time, preventing them from becoming linearly dependent. The variable-metric localization has been shown to improve the locality of occupied nonorthogonal orbitals relative to their orthogonal counterparts. In this work, numerous localization algorithms are designed and tested to exploit the conceptual simplicity of the variable-metric approach with the goal of creating a straightforward and reliable localization procedure for virtual orbitals. The implemented algorithms include the steepest descent, conjugate gradient (CG), limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS), and hybrid procedures as well as trust-region (TR) methods based on the CG and Cauchy-point subproblem solvers. Comparative analysis shows that the CG-based TR algorithm is the best overall method to obtain nonorthogonal localized molecular orbitals (NLMOs), occupied or virtual. The L-BFGS and CG algorithms can also be used to obtain NLMOs reliably but often at higher computational cost. Extensive tests demonstrate that the implemented methods allow us to obtain well-localized Boys–Foster (i.e., maximally localized Wannier functions) and Pipek–Mezey, orthogonal and nonorthogonal, and occupied and virtual orbitals for a variety of gas-phase molecules and periodic materials. The tests also show that virtual NLMOs, which have not been described before, are, on average, 13% (Boys–Foster) and 18% (Pipek–Mezey) more localized than their orthogonal counterparts.



I. INTRODUCTION

Spatially localized orbitals are widely used in the electronic structure theory. Occupied localized orbitals help to describe, visualize, and classify chemical bonding between atoms, thus facilitating our understanding of electronic-structure origins of observed properties of atomistic systems.[1−13] Occupied and virtual localized orbitals are crucial ingredients in all local electronic structure methods including Kohn–Sham density functional theory (DFT)[14−26] and wavefunction-based electron correlation methods.[23,27−37] In these methods, it is the locality of one-electron orbitals that allows us to dramatically reduce the computational cost of modeling of large systems.[24,38,39] Because of the importance of localized orbitals in electronic structure theory, such orbitals have been generated using many conceptually different approaches.[24,31,32,40] The focus of this work is on localized orbitals obtained through the optimization of a localization functional.[40]

The self-consistent field procedure at the heart of Kohn–Sham DFT and most single-reference electronic structure theories yield spatially delocalized one-electron states. Typically, these states are the eigenstates of the effective one-electron Hamiltonian and are known as canonical molecular orbitals (CMOs) in the case of molecules and Bloch orbitals in the case of periodic systems. Here, they will be collectively referred to as CMOs. Traditionally, spatially localized orbitals are constructed by finding the unitary transformation of CMOs that minimizes the spread of individual orbitals. The unitary transformation can be applied to the set of occupied orbitals, generating a localized description of chemical bonding, or to the set of virtual orbitals, producing local anti-bonding orbitals.

Multiple functionals have been proposed to measure orbital locality in molecular systems. The most known are Boys–Foster,[1] Edmiston–Ruedenberg,[2,41,42] Pipek–Mezey,[5] and Von Niessen[6] functionals. The Boys–Foster localization functional is perhaps the most popular because of the simplicity of its physical interpretation such as the orbital spread, low computational complexity, and ease of implementation. The Pipek–Mezey localization,[5] which maximizes atomic charges[43−45] of each orbital, is also widely used because it does not mix localized molecular orbitals (LMOs) representing σ and π bonds and thus gives a clear picture of bonding patterns. In the last decade, these well-established localization functionals have been modified to reduce orbital tails and produce more uniform localization across a set of orbitals,[46−50] especially virtual LMOs.[46,47]

Received: April 15, 2021

For periodic systems, maximally localized Wannier functions (MLWFs)[51,52] represent the solid-state equivalent of Boys–Foster orbitals. For large supercells of condensed phase-periodic systems, where electronic structure can be described with the Γ-point sampling of the Brillouin zone, generalized Pipek–Mezey Wannier functions have also been proposed.[53] In this work, Wannier functions and LMOs will be referred to as LMOs regardless of whether the system is isolated or treated with periodic boundary conditions.

Since CMOs are orthogonal and a unitary transformation applied to them during a localization procedure preserves the orbital metric, LMOs obtained in this way are orthogonal by construction. Because of the imposed orthogonality condition, orthogonal LMOs (OLMOs) exhibit small non-zero values even far away from their localization centers. These orthogonalization tails reduce orbital locality making orbital-based local correlation methods less computationally efficient. They also complicate the interpretation of chemically relevant electronic-structure information and make its transferability from one system to another more difficult. It is, however, known that the orthogonality of one-electron orbitals, while often mathematically convenient, is not a strict requirement in electronic structure theory where many methods have been reformulated in terms of nonorthogonal orbitals.[54]

To mitigate the undesirable orthogonality effects, metric-preserving unitary transformation has been applied to non-orthogonal orbitals[55] and, in a more dramatic procedure, the unitary localization transformations have been replaced with more general variable-metric nonsingular transformations.[56−63] In the latter procedure, the generalization lifts the orthogonality constraint imposed on LMOs and increases the number of degrees of freedom available to LMOs. It has been found that occupied nonorthogonal LMOs (NLMOs) obtained in a variable-metric procedure are indeed about 10−30% more localized than OLMOs if measured by the value of the Boys–Foster functional.[61,63,64]

In contrast to metric-preserving unitary transformations, the variable-metric localization must be formulated to avoid linear dependencies of orbitals during the minimization of the localization functional.[61,62,64] One approach to overcome the linear dependence problem is to fix the centers of NLMOs[61,62] at the positions guessed from the knowledge of bonding patterns in the system[62] or at the centers of OLMOs.[61] Another more recent approach is to augment the localization functional with a term that measures the deviation from the orthogonality and penalizes the states that are too close to linear dependence.[63] The latter reformulation of variable-metric localization allows not only to determine optimal positions of the NLMOs' centers in an unconstrained and straightforward minimization procedure but also to choose the desired balance between the orthogonality and locality of the orbitals.

A variety of iterative algorithms have been developed to minimize localization functionals and construct LMOs. They are briefly reviewed below, first for OLMOs and then for NLMOs.

In the case of OLMOs, the orthogonality can be preserved through the explicit parameterization of unitary transformations[65−67] or through implicit transformation of derivatives into the unitary manifold.[45,50]

The Jacobi rotations, a sequence of pairwise unitary mixings of orbitals, has been introduced in early works[2,65] and because of its simplicity, they have become the standard localization method for occupied OLMOs. Unfortunately, the Jacobi optimization fails to converge in the case of virtual orbitals with the exception of simple molecules and small basis sets. With the recognition of the utility of parametric representations of unitary transformations[50,65−67] and advent of iterative algorithms for unconstrained optimization of nonlinear functions, multidimensional line search methods that include steepest descent (SD), conjugate gradient (CG), and Newton and quasi-Newton algorithms[68] have been suggested for orbital optimization. It has been found that the SD approach is more efficient than Jacobi transformations for occupied OLMOs.[69] CG methods designed to outperform the SD algorithm have been reported not to improve the efficiency of the localization procedure significantly.[70] The limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm belonging to the family of quasi-Newton methods has also been implemented to localize occupied OLMOs.[67] It remains unclear whether L-BFGS is advantageous to simpler CG and SD algorithms. In the Newton–Raphson localization of occupied OLMOs, the incorporation of the second-derivative information results in the acceleration of the convergence in the vicinity of the minimum where the Hessian is positive-definite.[71] However, in the initial iterations of the Newton–Raphson localization, far away from the minimum, negative eigenvalues of the Hessian lead to poorly computed optimization directions and, thereby, to slow, erratic, or failed convergence.

The major shortcoming of the aforementioned line search methods is that they often fail to localize orbitals of the virtual subspace. This failure is attributed to the presence of negative Hessian eigenvalues that are more ubiquitous for more diffuse virtual orbitals. To resolve the indefinite Hessian problem, a trust-region (TR) optimization method employing the level-shifted Newton algorithm to solve the fixed-radius subproblem[68] has been investigated as an alternative to multidimensional line search methods. This variant of TR methods has been demonstrated to work reliably for both occupied and virtual Hartree–Fock OLMOs[40] with the standard and fourth-order Boys[46−48,72] and Pipek–Mezey localization functionals.[49,72]

There have been fewer studies of localization algorithms for nonorthogonal orbitals. The trust region method described above has been generalized and successfully applied to metric-preserving nonorthogonal occupied and virtual orbitals.[55] As to the more flexible variable-metric localization, it has been limited to occupied NLMOs and has successfully utilized multidimensional line search methods such as conjugated gradient,[63,64] the Gill–Murray variant of the Newton method,[61] and BFGS method.[62]

For periodic systems, the SD[51,52] and CG[53,63,66] algorithms have been used to construct occupied OLMOs[51−53,66] and NLNOs[63] using Boys (i.e., MLWFs)[51,63,66] and Pipek–Mezey[53,63,66] localization criteria.

This work focuses on NLMOs. The ability of the recently introduced variable-metric approach[63] to optimize NLMO mixing coefficients directly is exploited to design and compare numerous localization algorithms. It is demonstrated that several multidimensional line search and TR algorithms are robust and capable of extending variable-metric localization to virtual NLMOs, which have not been described before. Furthermore, the locality of occupied and virtual NLNOs obtained using Boys and Pipek–Mezey criteria for isolated and periodic systems is compared to that of their orthogonal counterparts.

https://doi.org/10.1021/acs.jctc.1c00379
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

II. THEORY

Conventional tensor notation is used to manipulate nonorthogonal orbitals.[54] Covariant quantities are denoted with subscripts, and contravariant quantities are denoted with superscripts. A summation is implied over the same covariant and contravariant orbital indices. Summation is not implied if two indices are both covariant or if indices do not refer to orbitals. Since the localization is performed separately for occupied and virtual orbitals, the implied summation runs over either occupied or virtual orbitals, not all orbitals of the system. All equations written in the standard matrix notation can be found in Supporting Information.

In order to obtain a set of NLMOs, which tend to be more localized than OLMOs, the orthogonality constraint is replaced with the weaker normalization constraint. Thus, the coefficients in the linear expansion of NLMOs $|j\rangle$ in terms of the initial one-electron states $|i_0\rangle$

$$|j\rangle = |i_0\rangle A^i{}_j \tag{1}$$

are not required to form a unitary matrix. The normalization constraint is imposed on NLMOs by expressing the coefficients in terms of independent optimization parameters denoted with lowercase **a**

$$A^i{}_j = a^i{}_j (a^k{}_j \sigma^0_{kl} a^l{}_j)^{-\frac{1}{2}} \equiv a^i{}_j N_j \tag{2}$$

where N_j is the **a**-dependent normalization coefficient defined for brevity. Here and below, σ_0 and σ denote overlap matrices of the initial orbitals and NLMOs

$$\sigma^0_{ji} = \left\langle j_0 \middle| i_0 \right\rangle$$
$$\sigma_{kl} = \langle k|l\rangle = A^j{}_k \sigma^0_{ji} A^i{}_l \tag{3}$$

Note that the initial one-electron states are not required to be canonical or orthogonal but need to be normalized and linearly independent.

Without the orthogonality constraint, a conventional localization functional Ω_L must be augmented by a penalty functional Ω_P that prevents NLMOs from becoming linearly dependent during the optimization[63]

$$\Omega(\mathbf{A}) = \Omega_L(\mathbf{A}) + c_P \Omega_P(\mathbf{A}) \tag{4}$$

where $c_P > 0$ is the penalty strength. Although many penalty functionals can be envisioned, it has been found that

$$\Omega_P(\mathbf{A}) = -\ln \det[\sigma(\mathbf{A})] \tag{5}$$

works well with a variety of optimization algorithms. In eq 5, the dependence of the NLMO overlap matrix on the mixing coefficients is shown explicitly.

Since NLMOs remain normalized in the optimization procedure, the determinant of σ lies in the $(0, 1]$ interval. It is 1 for orthogonal NLMOs and 0 for linearly dependent NLMOs. Consequently, the penalty functional between 0 and $+\infty$ for these two extreme cases, respectively, makes the linearly dependent state inaccessible in the localization procedure with finite penalty strength c_P.

Without the orthogonality constraints, there are twice as many independent localization degrees of freedom in the variable-metric procedure as in the metric-preserving approach. These additional degrees of freedom are important for producing more localized occupied orbitals.[61,63,64] As has been shown previously,[63] these degrees of freedom can be handled efficiently during the unconstrained optimization procedure and the single-value penalty term plays the key role in avoiding linearly dependent states during the localization. It is this simplicity that makes the variable-metric approach described here straightforward to implement and use.

It is convenient to write the penalty strength c_P as a product of a dimensionless parameter α and the initial value of the localization functional

$$c_P = \alpha \Omega_L(\mathbf{I}) \tag{6}$$

This factorization makes units of the localization and penalty terms consistent. The value of α must be optimized to achieve a desirable compromise between orbital locality and linear independence. A simple strategy to determine α is to set it initially to a sufficiently large value and then gradually decrease it until the determinant of the overlap of NLMOs drops below a desired threshold $D_{tar} \in (0, 1]$. It has been found[63] that the initial value

$$\alpha^{init} = \left(\ln \frac{\det \sigma(\mathbf{I})}{D_{tar}} \right)^{-1} \tag{7}$$

is sufficiently large to ensure linear independence of NLMOs during the first crucial steps of the localization but, at the same time, is not significantly larger than its optimal value.

The localization functionals adopted in this work can be written as

$$\Omega_L(\mathbf{A}) = \sum_K \sum_i \omega_K (1 - |z_i^K|^2),$$
$$z_i^K = A^m{}_i \langle m_0 | \hat{B}^K | n_0 \rangle A^n{}_i = A^m{}_i B^K_{mn} A^n{}_i \tag{8}$$

with index i referring to NLMOs and other variables having different meaning depending on the localization method.

The Berghold[66] localization functional Ω_L^B is equivalent to the Boys–Foster localization scheme[66,73] for the gas-phase system. For periodic systems, the Berghold functional is a generalization of the Resta functional[73,74] to three dimensions and simulation cells of general shape and symmetry. It is suitable for the Γ-point-only description of electronic states.

$$\hat{B}^K = e^{i \mathbf{G}_K \cdot \hat{\mathbf{r}}} \tag{9}$$

where $\hat{\mathbf{r}}$ is the position operator in three dimensions, ω_K and $\mathbf{G}_K$ are suitable sets of weights and reciprocal lattice vectors, respectively, labeled by index K.[66,75]

In the Pipek–Mezey[5,45] localization functional Ω_L^{PM}

$$\hat{B}^K = \frac{1}{2} \sum_{\mu \in K} (|\chi_\mu\rangle\langle\chi^\mu| + |\chi^\mu\rangle\langle\chi_\mu|) \tag{10}$$

z_i^K is the contribution of orbital i to the Mulliken charge of atom K, $|\chi_\mu\rangle$ and $|\chi^\mu\rangle$ are atom-centered covariant and contravariant basis set functions, atomic weights ω_K are all set to one, and the imaginary part of z_i^K is zero.[66,75] The summation over μ is written explicitly to emphasize that it is restricted to the basis functions centered on atom K.

The unconstrained minimization of functional Ω can be carried out with a variety of algorithms, all of which require the first derivative with respect to the independent parameters (i.e., the gradient) and some of which can benefit from exact or approximate second derivatives (i.e., the Hessian).

https://doi.org/10.1021/acs.jctc.1c00379
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

The gradient $G_i^{\ j} \equiv \frac{\partial \Omega}{\partial a^i_{\ j}}$ is a sum of the localization $L_i^{\ j} \equiv \frac{\partial \Omega_L}{\partial a^i_{\ j}}$ and penalty $P_i^{\ j} \equiv \frac{\partial \Omega_P}{\partial a^i_{\ j}}$ components

$$G_i^{\ j} = L_i^{\ j} + c_P P_i^{\ j} \tag{11}$$

These components can be readily expressed in terms of the derivatives with respect to the transformation coefficients $\tilde{X}_k^{\ l} \equiv \frac{\partial \Omega_X}{\partial A^k_{\ l}}$

$$X_i^{\ j} = \tilde{X}_k^{\ l} \frac{\partial A^k_{\ l}}{\partial a^i_{\ j}} = \left[\tilde{X}_i^{\ j} - A_{ij}(A^m_{\ j} \tilde{X}_m^{\ j})\right] N_j \tag{12}$$

where X is either L or P and covariant coefficients arise from their contravariant-covariant counterparts $A_{ij} = \sigma^0_{in} A^n_{\ j}$. To clarify the tensor notation, the summation is not implied over j in the equation above because this index appears in the left-hand side of the equation (see also equations in the standard matrix notation in Supporting Information). The same convention is used in the equations below.

$$\tilde{L}_k^{\ l} = -\sum_K 4\omega_K \times \times [\mathrm{Re}(B^K_{kn}) A^n_{\ l}\, \mathrm{Re}(z_l^{\ K}) + \mathrm{Im}(B^K_{kn}) A^n_{\ l}\, \mathrm{Im}(z_l^{\ K})] \tag{13}$$

$$\tilde{P}_k^{\ l} = -2\sigma^0_{km} A^m_{\ n} \sigma^{nl} = -2A_k^{\ l} \tag{14}$$

The equation above defines covariant-contravariant coefficients $A_k^{\ l}$. Note that the penalty term ensures that the overlap matrix remains not only invertible but also well-conditioned. Thus, elements of its inverse σ^{nl} can be computed efficiently.

The Hessian $H_{is}^{\ jt} \equiv \frac{\partial^2 \Omega}{\partial a^i_{\ j} \partial a^s_{\ t}}$ is also a sum of the localization $\overline{L}_{is}^{\ jt} \equiv \frac{\partial^2 \Omega_L}{\partial a^i_{\ j} \partial a^s_{\ t}}$ and penalty $\overline{P}_{is}^{\ jt} \equiv \frac{\partial^2 \Omega_P}{\partial a^i_{\ j} \partial a^s_{\ t}}$ components

$$H_{is}^{\ jt} = \overline{L}_{is}^{\ jt} + c_P \overline{P}_{is}^{\ jt} \tag{15}$$

These components can be expressed in terms of the second derivatives with respect to the transformation coefficients $\tilde{\overline{X}}_{kx}^{\ ly} \equiv \frac{\partial^2 \Omega_X}{\partial A^k_{\ l} \partial A^x_{\ y}}$ and in terms of the first-derivative $\tilde{X}_k^{\ l}$ defined above. Here, X stands for either L or P.

$$\begin{aligned}\overline{X}_{is}^{\ jt} &= \tilde{\overline{X}}_{kx}^{\ ly} \frac{\partial A^k_{\ l}}{\partial a^i_{\ j}} \frac{\partial A^x_{\ y}}{\partial a^s_{\ t}} + \tilde{X}_k^{\ l} \frac{\partial^2 A^k_{\ l}}{\partial a^i_{\ j} \partial a^s_{\ t}} \\ &= [\tilde{\overline{X}}_{is}^{\ jt} - (\tilde{\overline{X}}_{ix}^{\ jt} A^x_{\ t}) A_{st} - (\tilde{\overline{X}}_{xs}^{\ jt} A^x_{\ j}) A_{ij} \\ &\quad + (A^k_{\ j} \tilde{\overline{X}}_{kx}^{\ jt} A^x_{\ t}) A_{ij} A_{st}] N_j N_t \\ &\quad + \delta_{jt}[(\tilde{X}_k^{\ t} A^k_{\ t})(3A_{it} A_{st} - \sigma^0_{is}) \\ &\quad - \tilde{X}_i^{\ t} A_{st} - \tilde{X}s^t A_{it}] N_t^{\ 2}\end{aligned} \tag{16}$$

$$\tilde{\overline{L}}_{kx}^{\ jt} = -\sum_K 4\omega_K \delta_{jt} \times [\mathrm{Re}(B^K_{kx}) \mathrm{Re}(z_t^{\ K}) + 2\mathrm{Re}(B^K_{kn}) A^n_{\ t}\, \mathrm{Re}(B^K_{xq}) A^q_{\ t} + \mathrm{Im}(B^K_{kx}) \mathrm{Im}(z_t^{\ K}) + 2\mathrm{Im}(B^K_{kn}) A^n_{\ t}\, \mathrm{Im}(B^K_{xq}) A^q_{\ t}] \tag{17}$$

$$\tilde{\overline{P}}_{kx}^{\ jt} = 2A_k^{\ t} A_x^{\ j} \tag{18}$$

III. COMPUTATIONAL METHODS

III.A. Multidimensional Line Search Methods. Multidimensional line search algorithms are the most popular methods to minimize a function of D independent variables. This large family of algorithms includes the SD, CG, and quasi-Newton methods. In these algorithms, the search direction $\mathbf{d}$ in D dimensions is determined first. The objective function, also known as the loss function, is then minimized along the chosen direction using a simple one-dimensional search algorithm to obtain the optimal step size γ^*. Using the variables of the previous section, an iteration in a multidimensional line search localization procedure is defined by

$$\begin{aligned}\gamma^* &= \mathrm{argmin}_\gamma \Omega(\mathbf{a}_k + \gamma \mathbf{d}_k) \\ \mathbf{a}_{k+1} &= \mathbf{a}_k + \gamma^* \mathbf{d}_k\end{aligned} \tag{19}$$

where boldface $\mathbf{a}$ and $\mathbf{d}$ denote *vectorized* forms of the independently varied parameters and search direction, respectively. The vectorized form is introduced to adapt the tensor notation of the previous section to the vector notation commonly employed in the field of numerical optimization. In the latter notation, the independent variables and gradients are denoted as vectors, whereas the second derivatives are treated as matrices. In the previous section of this work, the independent variable and gradient are two-index tensors, and the second derivative is a four-index tensor. Vectorization simply replaces double indices with a single collective index

$$\begin{aligned}(\mathbf{a})_Z &= a_i^{\ j} \\ (\mathbf{H})_{ZY} &= H_{is}^{\ jt}\end{aligned} \tag{20}$$

In the equation above, the range of the capitalized indices is equal to the number of independently varied parameters, that is, the square of the number of the orbitals being localized.

Multidimensional line search algorithms differ by how the search direction is computed. The SD, CG, L-BFGS, and CG-BFGS hybrid[76] are considered in this work.

For SD and CG, the search direction takes the following form

$$\begin{aligned}\mathbf{d}_{k+1} &= -\mathbf{G}_{k+1} + \beta_k \mathbf{d}_k \\ \beta_k^{\mathrm{SD}} &= 0, \\ \beta_k^{\mathrm{CG(FR)}} &= \frac{\langle \mathbf{G}_{k+1}, \mathbf{G}_{k+1} \rangle}{\langle \mathbf{G}_k, \mathbf{G}_k \rangle}, \qquad \beta_0^{\mathrm{CG(FR)}} = 0\end{aligned} \tag{21}$$

where $\mathbf{G}$ is the vectorized form of the gradient matrix defined in eq 11 and CG(FR) stands for the Fletcher–Reeves choice[77] for the update factor β. Note that a tensorially consistent inner product is defined for vectorized tensors

$$\langle \mathbf{K}, \mathbf{T} \rangle = \sigma_{km} K_n^{\ m} \sigma_0^{nl} T_l^{\ k} = K_n^{\ m} T^n_{\ m} = \sigma^{km} K^n_{\ m} \sigma^0_{nl} T^l_{\ k} \tag{22}$$

In the L-BFGS method, the search direction is determined by contracting an approximate inverted Hessian $\mathbf{B}^{-1}$ with the gradient

$$\begin{aligned}\mathbf{d}_{k+1}^{\mathrm{BFGS}} &= -\mathbf{B}_{k+1}^{\ -1} \mathbf{G}_{k+1} \\ \mathbf{B}_{k+1}^{-1} &= (\mathbf{I} - \rho_k \mathbf{s}_k \mathbf{y}_k^{\ T}) \mathbf{B}_k^{\ -1} (\mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^{\ T}) + \rho_k \mathbf{s}_k \mathbf{s}_k^{\ T} \\ \rho_k &= \frac{1}{\langle \mathbf{y}_k, \mathbf{s}_k \rangle}\end{aligned} \tag{23}$$

https://doi.org/10.1021/acs.jctc.1c00379
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

where $\mathbf{y}_k = \mathbf{G}_{k+1} - \mathbf{G}_k$ and $\mathbf{s}_k = \mathbf{a}_{k+1} - \mathbf{a}_k$. Note that the L-BFGS algorithm does not explicitly compute the Hessian nor the inverted Hessian. It only computes the contraction of the approximate inverted Hessian with the gradient. Obviating the need in the Hessian inversion makes L-BFGS a promising candidate for efficient orbital localization.

The hybrid CG-BFGS method[76] has been designed to combine advantages of BFGS and CG methods. It has been shown to reduce the total number of iterations to convergence[76] and is considered here as an alternative to the L-BFGS algorithm.

$$\mathbf{d}_{k+1}^{\text{CG-BFGS}} = -\mathbf{B}_{k+1}{}^{-1}\mathbf{G}_{k+1} - l_{k+1}\mathbf{G}_{k+1} + \beta_{k+1}^{\text{LSCD}}\mathbf{d}_k^{\text{CG-BFGS}},$$

$$\beta_{k+1}^{\text{LS}} = -\frac{\langle \mathbf{y}_k, \mathbf{G}_{k+1}\rangle}{\langle \mathbf{G}_k, \mathbf{d}_k\rangle}, \qquad \beta_{k+1}^{\text{CD}} = -\frac{\langle \mathbf{G}_{k+1}, \mathbf{G}_{k+1}\rangle}{\langle \mathbf{G}_k, \mathbf{d}_k\rangle}$$

$$\beta_{k+1}^{\text{LSCD}} = \max\{0, \min\{\beta_{k+1}^{\text{LS}}, \beta_{k+1}^{\text{CD}}\}\}$$

$$l_{k+1} = 1 + \beta_{k+1}^{\text{LSCD}}\frac{\langle \mathbf{G}_{k+1}, \mathbf{d}_k\rangle}{\langle \mathbf{G}_{k+1}, \mathbf{G}_{k+1}\rangle} \tag{24}$$

III.B. TR Methods. In contrast to the line search methods described above, the TR methods determine the search direction and one-dimensional step size simultaneously. At each TR iteration, the subproblem

$$\min_p m_k(\mathbf{p}) = \Omega(\mathbf{a}_k) + \langle \mathbf{G}_k, \mathbf{p}\rangle + \frac{1}{2}\langle \mathbf{p}, \mathbf{H}_k\mathbf{p}\rangle$$

$$\text{subject to } \langle \mathbf{p}, \mathbf{p}\rangle \le \Delta_k{}^2 \tag{25}$$

is solved to suggest an update $\mathbf{p}_k$ to the current values of independent variables $\mathbf{a}_k$. Here, Δ_k denotes the trust radius. If the update is judged acceptable, $\mathbf{a}_k$ is updated

$$\mathbf{a}_{k+1} = \mathbf{a}_k + \mathbf{p}_k \tag{26}$$

Otherwise, $\mathbf{a}_k$ values remain unchanged and the trust radius Δ_k is reduced to produce a new subproblem that is solved in the next iteration. The TR algorithm for updating $\mathbf{a}_k$ and Δ_k is based on comparing the actual reduction in the loss function to the reduction predicted by the model quadratic function $m_k(\mathbf{p})$. This algorithm is well-established[68] and various TR methods differ by how the model function is defined and how the subproblem is solved.

In this work, the model quadratic function in eq 25 is expressed in terms of the exact analytical gradient and Hessian given by eqs 11 and 15, respectively. The performance of two subproblem solvers is investigated.

The Cauchy-point subproblem solver minimizes the model function along the gradient within the TR bounds[68]

$$\mathbf{p}_k^{\text{TR(Cauchy)}} = -\tau_k\Delta_k\frac{\mathbf{G}_k}{\sqrt{\langle \mathbf{G}_k, \mathbf{G}_k\rangle}} \tag{27}$$

$$\tau_k = \begin{cases} 1, & \text{if } \langle \mathbf{G}_k, \mathbf{H}_k\Gamma_k\rangle \le 0 \\ \min\left(1, \dfrac{\langle \mathbf{G}_k, \mathbf{G}_k\rangle^{3/2}}{\Delta_k\langle \mathbf{G}_k, \mathbf{H}_k\Gamma_k\rangle}\right), & \text{otherwise} \end{cases} \tag{28}$$

where $\Gamma^i_j = \sigma_0^{in}G_n{}^m\sigma_{mj}$ is the contravariant-covariant version of the gradient. The first line in eq 28 shows that if the gradient lies in a direction of a negative curvature, the algorithm steps to the border of the trust region.

The second subproblem solver relies on the CG algorithm and is denoted as TR(CG). It has been shown that a sufficiently precise solution to the quadratic model function can be found using preconditioned CG algorithm if it is properly modified to monitor the satisfaction of the TR constraint.[78] Its main advantage over the dogleg method—perhaps the most popular subproblem solver—is the efficient utilization of the second-derivative information without the need to invert the Hessian. This enables the application of the TR(CG) method to problems with indefinite Hessian, which are often encountered during the localization of virtual orbitals.

It should be emphasized that the computational cost of the Hessian−vector contraction, required in both TR methods, grows cubically $O(N^3)$ with the number of NLMOs, not quartically $O(N^4)$ as expected from the straightforward counting of the elements. The scaling is naturally reduced because the localization function is a sum of single-orbital terms and, as a result, all the second derivatives with respect to the mixing coefficients of two different orbitals are zero. This drastic reduction in the number of non-zero elements in the localization part of the Hessian matrix is indicated by the Kronecker delta in eq 17. The penalty part of the Hessian in eq 18 is represented by an outer product of two vectors, which again results in the cubically scaling cost of the contraction.

III.C. Computational Details. All localization algorithms were implemented in the CP2K software package.[79] Key implementation details were described in the previous work.[63] NLMOs for multiple systems ranging from a simple water molecule to complex molecules with non-trivial bonding patterns and to large periodic systems were constructed. For all systems, the initial CMOs were obtained using the conventional diagonalization-based SCF procedure implemented in the electronic structure module of CP2K. The Becke−Lee−Yang−Parr generalized gradient approximation[80,81] was used as the exchange−correlation functional. Goedecker−Teter−Hutter pseudopotentials[82] were used together with a triple-ζ atom-centered Gaussian basis set with two sets of polarization functions (TZV2P) for all atoms. The energy cutoff was set at 600 Ry to define the auxiliary plane-wave basis set in the construction of the effective Hamiltonian. The integration over the Brillouin zone was performed using the Γ-point approximation. CMOs were used as the initial set of orbitals, and the initial transformation matrix $\mathbf{a}$ was set to be the identity matrix. It has been verified that the energy of all systems remains the same within the numerical accuracy of the calculations (10^{-12} a.u.) before and after the NLMO optimization indicating that there is no spurious mixing of occupied and virtual orbitals during their localization. All calculations were performed on 40-core nodes containing two Intel Xeon Gold 6148 Skylake CPUs running at 2.4 GHz.

For graphene, an 8×8 supercell containing 128 carbon atoms was used to demonstrate the localization of virtual orbitals and the 10×10 supercell containing 200 carbon atoms was used to construct occupied localized orbitals.

IV. RESULTS AND DISCUSSION

IV.A. Convergence and Computational Efficiency. The convergence and computational efficiency of the localization algorithms were compared for occupied and virtual orbitals in molecules and periodic solid-state systems. Unless stated otherwise, the Berghold localization functional was used in most tests. As a reminder, this functional corresponds to Boys

https://doi.org/10.1021/acs.jctc.1c00379
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

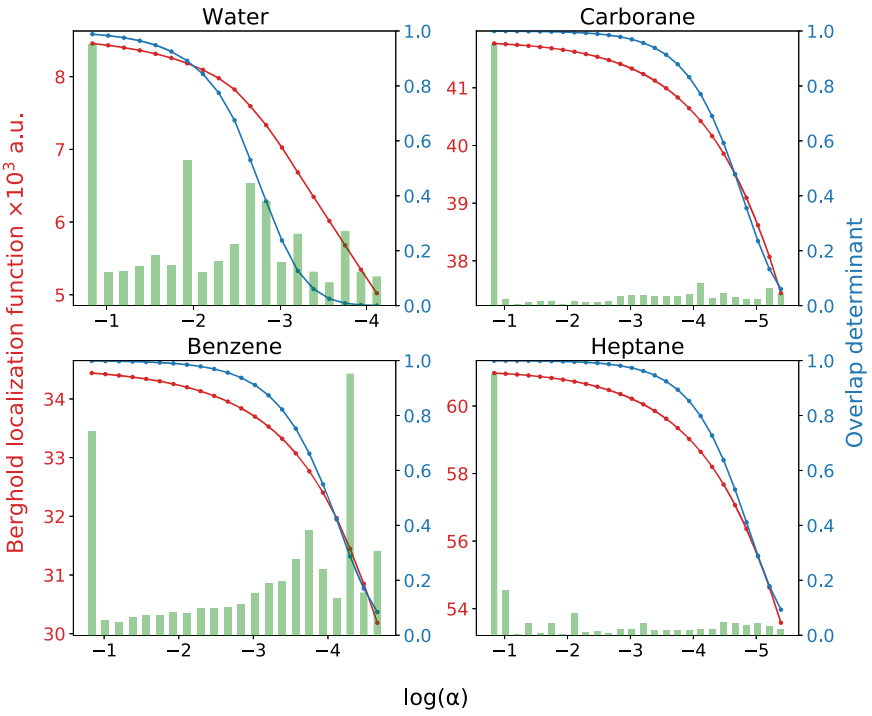


Figure 1. Dependence of the optimal localization functional and determinant of the virtual NLMO overlap on α—the adjustable part of the penalty strength. The first point on the left is $\alpha = (\ln 10)^{-1} \approx 0.434$. The green bars show the relative number of CG-BFGS iterations to localize orbitals for each value of α.

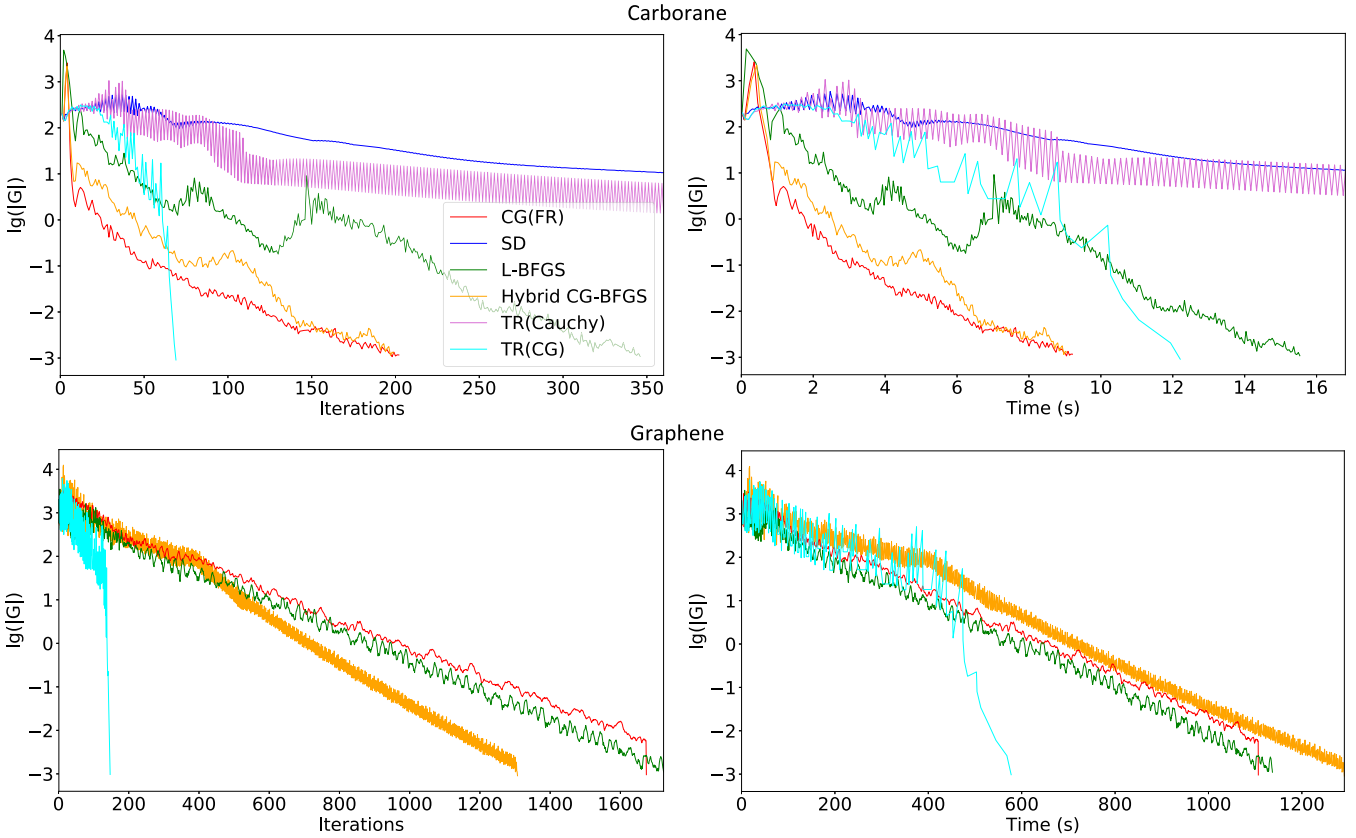


Figure 2. Maximum norm of the gradient in the Berghold localization of the occupied orbitals of carborane and graphene. Calculations were performed using 1 compute core and 200 compute cores, respectively.

localization for isolated systems and produces MLWFs for periodic systems with large supercells.

As shown in Figure 1, most computational effort of the localization procedure is devoted to converging the minimiza-

https://doi.org/10.1021/acs.jctc.1c00379
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

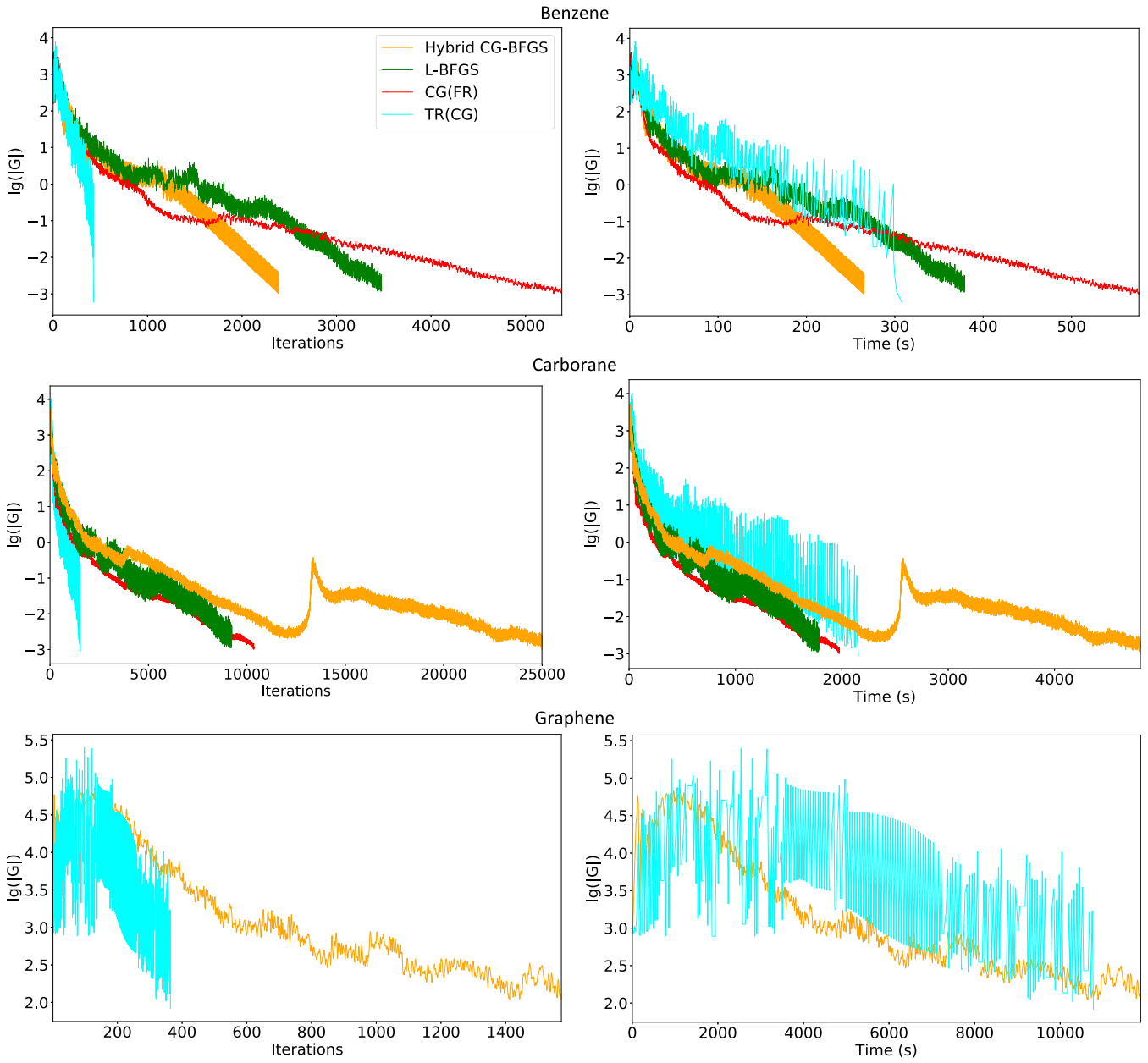


Figure 3. Maximum norm of the gradient in the Berghold localization of the virtual orbitals of benzene, carborane, and graphene. Calculations were performed using a single compute core for benzene and carborane. 120 compute cores were used in the case of graphene.

tion procedure for the initial value of the penalty strength α. After the first set of LMOs is obtained, the re-localization for weaker penalty strength α is fast in most cases. Taking this fact into account, the performance of the algorithms is compared only for the initial value of α.

Figure 2 compares the convergence of the localization procedure for occupied orbitals of carborane (25 occupied orbitals) and graphene (400 occupied orbitals)—representatives of molecular and solid-state systems. The maximum norm of the gradient was used as a convergence criterion. The convergence threshold was set to 10^{-3} a.u.—an excessively tight value that allows us to check the stability of the algorithms. The convergence of the simple steepest descend algorithm and the Cauchy-point TR algorithm—the TR equivalent of SD—is significantly lower than that of more advanced methods even for the small carborane molecule. For the CG, L-BFGS, and hybrid CG-BFGS algorithms, the rate of convergence is higher and mostly constant throughout the optimization. An attractive feature of the TR algorithm based on the iterative conjugate-gradient subproblem solver [TR(CG)] is that it tends to accelerate in the vicinity of the minimum. Unfortunately, the greatly reduced number of iterations in the TR(CG) algorithm does not result in the same reduction in computational time because each TR iteration (i.e., finding the solution to a subproblem) requires multiple CG iterations. Based on the data presented in Figure 2 for carborane, CG, BFGS, and hybrid CG-BFGS and TR(CG) algorithms appear to be reliable and efficient localization methods for occupied orbitals. Only these algorithms are considered for the localization of the occupied orbitals of graphene and all virtual orbitals. The SD and TR(Cauchy) algorithms are eliminated because of their low convergence rate.

Figure 3 compares the convergence of the most promising localization algorithms for virtual orbitals of benzene (172

https://doi.org/10.1021/acs.jctc.1c00379
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

virtual orbitals), carborane (257 virtual orbitals), and graphene (2560 virtual orbitals). The convergence of virtual orbitals is significantly slower compared with occupied orbitals for all algorithms. Apart from the slower convergence, the relative performance of the algorithms for virtual NLMOs is similar to that for occupied NLMOs. The L-BFGS algorithm appears to be better than the CG algorithm, and the hybrid CG-BFGS methods can produce benefits by combining the features of both approaches. However, the CG-BFGS algorithm fails to maintain its convergence rate near the solution in the case of carborane, casting doubt on its general superiority. As in the case of occupied NLMOs, the TR(CG) method requires significantly fewer iterations to converge but these iterations tend to take more time. Nevertheless, the TR(CG) algorithm is as efficient as or even more efficient as the line search methods.

The results for occupied and virtual orbitals show that, among the line search methods, the hybrid CG-BFGS algorithm has the best overall performance. TR(CG) is as robust and efficient. In addition, it offers benefits of accelerated convergence at higher computational cost per iteration. It should be noted that the L-BFGS algorithm and especially the CG algorithm should not be discarded as very simple methods to obtain NLMOs reliably.

In order to speed up the construction of virtual NLMOs in difficult cases, the localization procedure for the initial values of α can be converged using a loose convergence criterion because NLMOs will be re-localized in subsequent α-iterations anyway. It is only the final α iteration that should be converged tightly. This approach is used for the challenging case of virtual NLMOs of graphene. For the initial value of α, the localization convergence threshold was set to 2×10^2 a.u. (Figure 3), whereas the final convergence threshold was set to 1×10^{-1} a.u. (Figure 4). The localization procedure of 2560 virtual orbitals takes almost 7800 hybrid CG-BFGS iterations and ∼21 h on 120 compute cores. The TR(CG) localization for the same system takes almost 1500 iterations and ∼32 h on the same number of compute cores.

The performance of the best localization algorithms was also compared for the Pipek−Mezey functional. Pipek−Mezey occupied and virtual orbitals of carborane were constructed (Figure 5). The ranking of the algorithms in terms of the number of iterations and time to convergence is different from that for the Berghold localization. For example, the hybrid CG-BFGS is now slower than the CG algorithm. TR(CG), on the other hand, demonstrates the same high performance as before, especially in the case of costly virtual localization. It should be noted that the cost of the evaluation of the Pipek−Mezey localization functional and its derivatives grows more rapidly with the number of atoms in the system than that of the Berghold (i.e., Boys and MLWFs) functional. This is because the calculation of charges is required for each atom in the Pipek−Mezey scheme. Even for a relatively small carborane molecule, Pipek−Mezey localization requires more time per iteration, making the TR(CG) scheme with its fewer iterations more advantageous than the line search algorithms. This effect can be seen by comparing Figure 5 and the carborane panels in Figures 2 and 3. The advantage of the TR(CG) algorithm is expected to become more pronounced in larger systems.

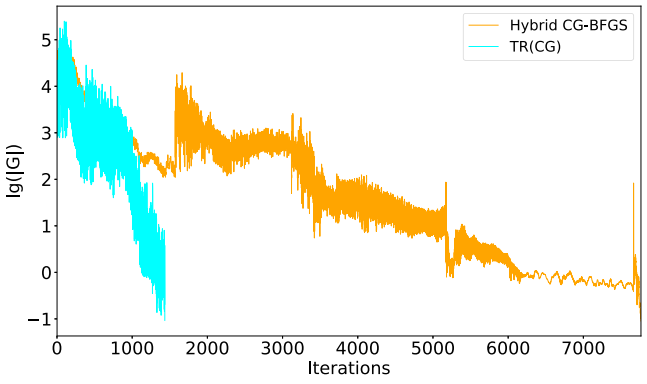


Figure 4. Maximum norm of the gradient in the Berghold localization of the virtual orbitals of graphene. The convergence threshold was set to 2×10^2 a.u. for the first α-iteration and then gradually decreased by 2 to the final value of 1×10^{-1} a.u.

To conclude, TR(CG) is the preferable algorithm for the variable-metric localization of one-electron orbitals as it demonstrates accelerated convergence near a minimum, requires fewer iterations, and works well for computationally demanding cases such as virtual orbitals and/or Pipek−Mezey localization functional. The main disadvantage of TR(CG) is that it is harder to implement than the straightforward CG algorithm,[63] which is also shown to be robust but often more costly.

IV.B. Comparison of OLMOs and NLMOs. The localization of occupied and virtual orbitals was performed using the CG-BFGS algorithm for a variety of test systems. The minimum allowed determinant of the NLMO overlap D_{tar} was set to 10^{-1} as suggested in the previous work on occupied NLMOs.[63] The initial value of α was set to $(\ln 10)^{-1}$ according to eq 7. The value of α was decreased iteratively by dividing it by a factor of 1.2 until the determinant of the NLMO overlap drops below D_{tar} or until the optimal value of the loss function stopped changing appreciably.

Figure 1 demonstrates the relation between the penalty strength, orbital localization, and determinant of the overlap matrix. It shows that within a wide range, spanning several orders of magnitude, the penalty strength can be tuned to produce nonorthogonal but linearly independent virtual LMOs. What is even more important is that, within this range, it is possible to achieve substantial reduction in orbital spread.

As in the case of occupied orbitals,[63] the chosen initial α is sufficiently large to generate nearly orthogonal localized orbitals in all test systems but, at the same time, is sufficiently small to permit deviation from orthogonality after a few α-iterations. It should be noted that, if α is allowed to decrease further, the numerical precision of the implemented code is sufficient to obtain NLMOs with $\det(\sigma)$ as low as 10^{-8}. As demonstrated for water molecules (Figure 1), if the final determinant of the virtual NLMOs overlap is allowed to drop from the accepted value of 10^{-1} to 10^{-4}, the localization functional value decreases to 5100 a.u., which is noticeably lower than 8400 a.u. for OLMOs and 6300 a.u. for NLMOs with $\det(\sigma) = 0.06$. However, NLMOs with $\det(\sigma) \approx 10^{-4}$ cannot be assigned orthogonal counterparts unambiguously and it is unclear whether they are sufficiently distinct to be practically useful in local electron correlation methods. Until further investigation of such nearly-dependent NLMOs is performed, it is recommended to use safer 10^{-1} threshold as the target determinant value.

The spatial extent of OLMOS and NLMOs relative to that of CMOs is shown in Figure 6 for the Berghold functional and in Figure 7 for the Pipek−Mezey functional. The NLMO overlap determinants shown on the right vertical axis in these figures are slightly lower than D_{tar}, because the penalty strength adjustment is designed to stop when $\det(\sigma)$ decreases below, and does not

https://doi.org/10.1021/acs.jctc.1c00379
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

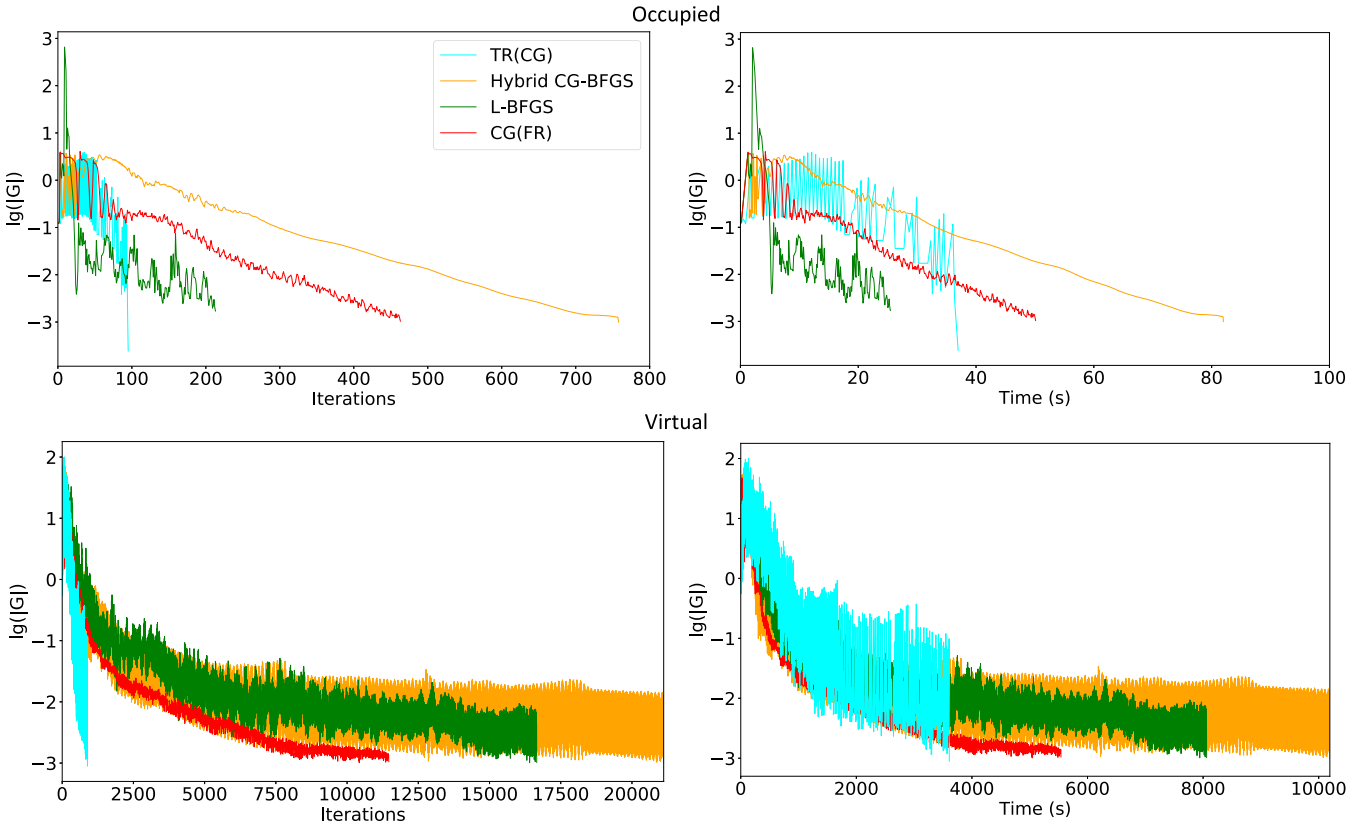


Figure 5. Maximum norm of the gradient in the Pipek−Mezey localization of the occupied and virtual orbitals of carborane. Calculations were performed using 1 compute core.

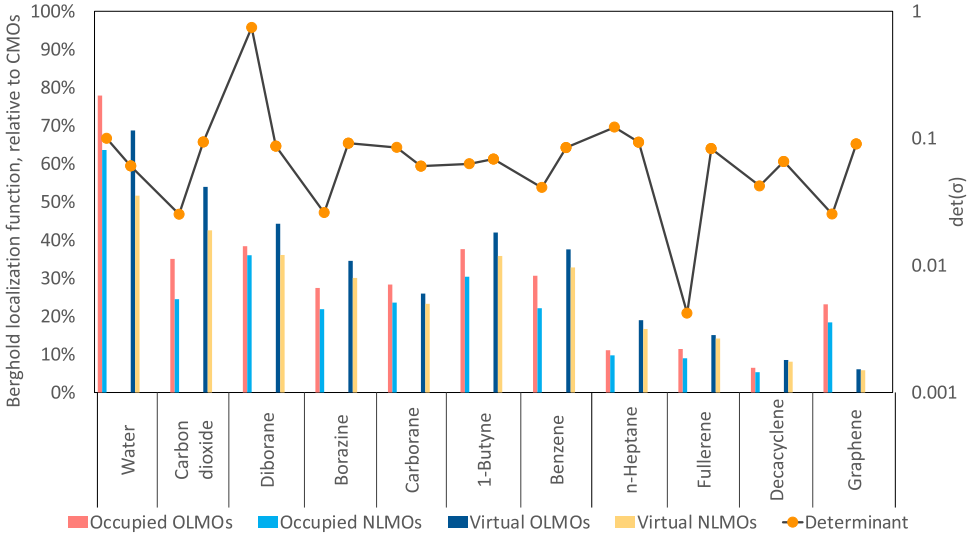


Figure 6. Berghold localization functional relative to that of CMOs. The final determinant of the NLMO overlap is shown on the right vertical axis.

equal to D_{tar}. It is worth mentioning that the algorithm can be modified to treat the penalty strength as a Lagrange multiplier that imposes the det $\sigma(\mathbf{A}) = D_{tar}$ constraint rigorously.

$$\Omega(\mathbf{A}) = \Omega_L(\mathbf{A}) + c_P(\Omega_P(\mathbf{A}) + \ln D_{tar}) \quad (29)$$

While this reformulation needs a constrained optimization algorithm, it would avoid the non-zero penalty values at the solution.

The spatial spread of occupied Berghold OLMOs is between 78 and 6.5% of that of CMOs, showing that the localization procedure is highly effective at reducing the size of the orbitals (Table 1 and Figure 6). As expected, the relative spread is lower for larger systems. The average relative spread over all tested systems is 30%, same as that calculated previously for occupied Berghold orbitals of a slightly different set of systems.[63] Removal of the orthogonality constraint makes occupied orbitals even more localized. The spread of occupied NLMOs relative to that of OLMOs ranges from 70 to 94% with the average being 81%. The same quantitative analysis is applicable to virtual OLMOs and NLMOs obtained using the Berghold functional (Figure 6).

https://doi.org/10.1021/acs.jctc.1c00379
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

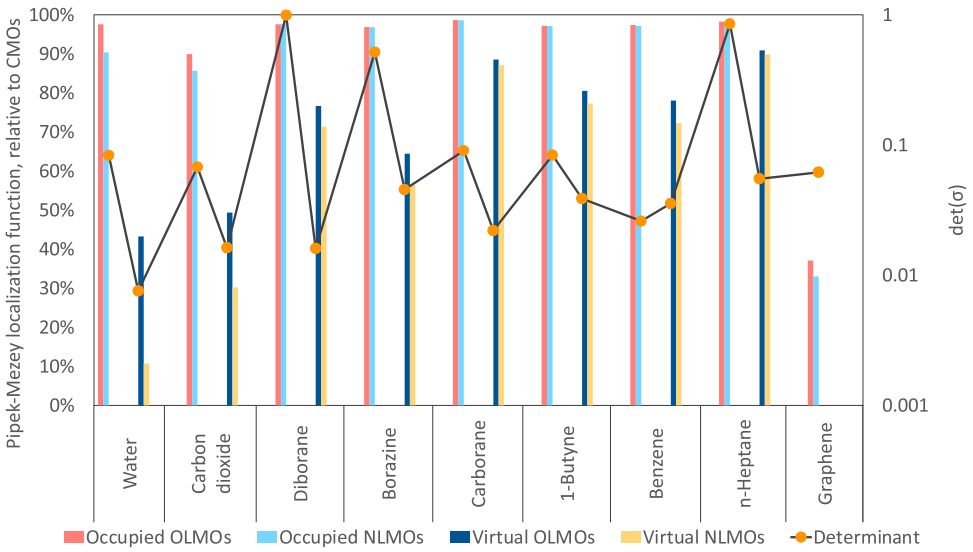


Figure 7. Pipek–Mezey localization functional relative to that of CMOs. The final determinant of the NLMO overlap is shown on the right vertical axis.

Table 1. Relative Localization Functional Values for the Test Systems Shown in Figures 6 and 7

		min–max, %	mean, %
Ω_L^B, occ.	OLMOs/CMOs	6.5–78	30
	NLMOs/OLMOs	70–94	81
Ω_L^B, virt.	OLMOs/CMOs	6.1–69	32
	NLMOs/OLMOs	75–95	87
Ω_L^{PM}, occ.	OLMOs/CMOs	37–99	90
	NLMOs/OLMOs	89–100	97
Ω_L^{PM}, virt.	OLMOs/CMOs	43–91	72
	NLMOs/OLMOs	25–99	82

Although the average spatial extent of virtual NLMOs relative to that of OLMOs is slightly larger than of the occupied orbitals (87%), the additional gain in the localization is still noticeable and can substantially reduce the number of retained occupied virtual amplitudes if employed in local electron correlation methods. It should be mentioned that most local electron correlation methods have been formulated to deal with nonorthogonal localized orbitals[38,54,83,84] and, therefore, can be readily used with NLMOs. It has been shown that, for large systems, the computational cost of dealing with nonorthogonal localized orbitals is fully compensated by the locality-enabled computational savings stemming from the dramatic reduction of significant double-, triple-, and higher-order substitution amplitudes.[38]

The same qualitative trends are observed for occupied and virtual OLMOs and NLMOs constructed using the Pipek–Mezey functional. However, because the Pipek–Mezey functional measures orbital locality differently—atomic charges instead of the spatial spread—the quantitative reduction in Ω_L upon orbital localization is lower especially for occupied orbitals (Table 1). The locality of occupied OLMOs, as measured by Ω_L^{PM}, is between 99 and 37% of that of CMOs, with the average being 90%. Lifting the orthogonality constraint makes occupied Pipek–Mezey orbitals only slightly more localized. The quantitative descriptors of the virtual Pipek–Mezey localization are noticeably better. The conventional metric-preserving localization reduces Ω_L^{PM} to 72% of the corresponding CMO value. Allowing for nonorthogonality lowers this number substantially to 59%.

Figures 8, 9, 10, and 11 compare virtual Berghold OLMOs and NLMOs for representative gas-phase molecules and periodic systems. Visual examination of these and other systems reveals that there is always a clear one-to-one correspondence between virtual OLMOs and NLMOs. For example, both NLMOs and OLMOs of carborane $C_2B_{10}H_{12}$ (Figure 8) correctly represent the σ^* orbital of a B–B bond and B–H bond. The figures show that main lobes of NLMOs tend to be slightly larger than those of OLMOs, and that peripheral tails of NLMOs are visibly reduced compared to OLMOs—the effects described previously for occupied orbitals.[63,64] The tail reduction in virtual NLMOs is visible when orbitals are shown with small isosurface values: 0.02 a.u. for fullerene in Figure 9 and 0.001 a.u. for graphene in Figure 10. The results of visual comparison are consistent with the numerical characterization reported in Figure 6.

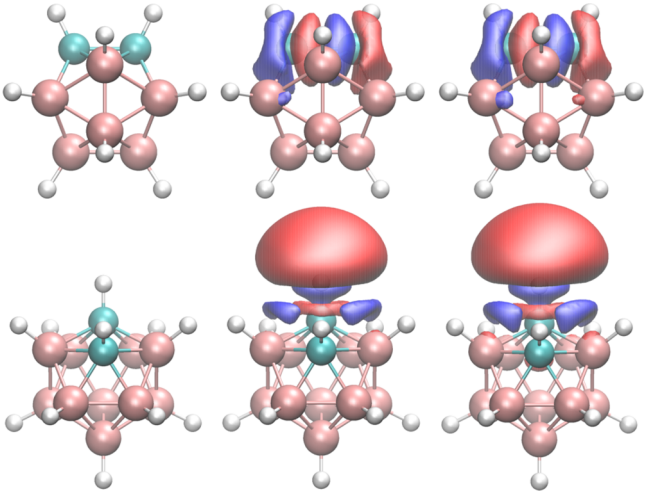


Figure 8. Nonorthogonal (middle) and orthogonal (right) LMOs of one virtual orbital of carborane $C_2B_{10}H_{12}$. The isosurface value is 0.04 a.u.

https://doi.org/10.1021/acs.jctc.1c00379
J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

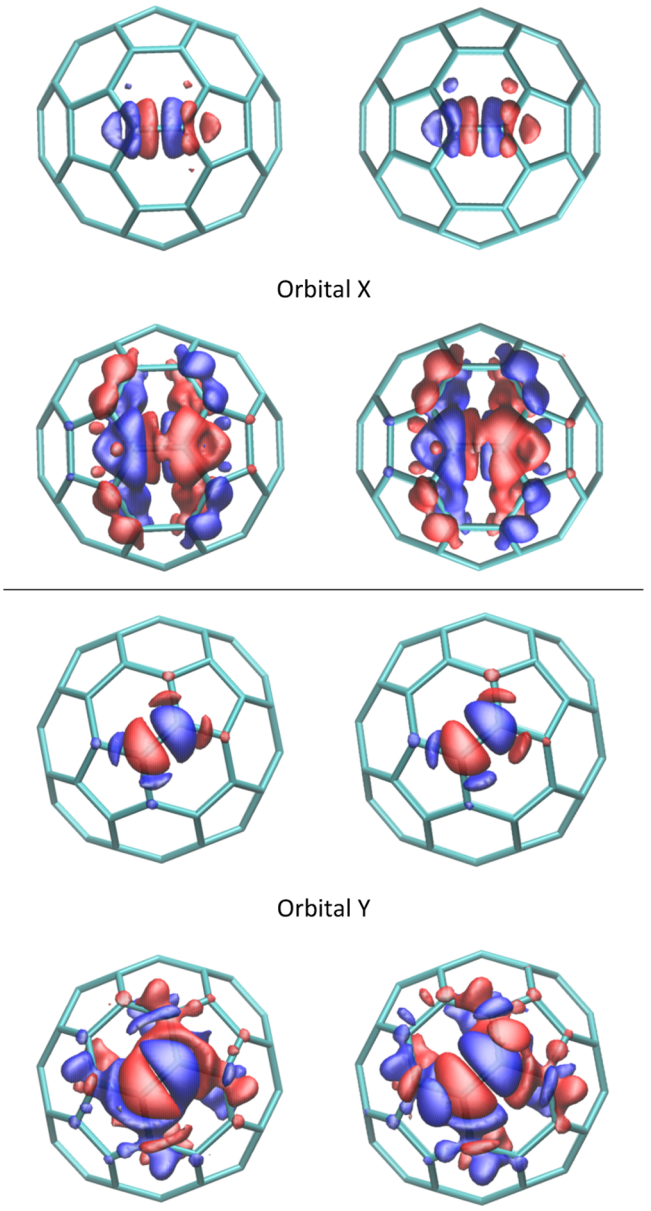


Figure 9. Nonorthogonal (left) and orthogonal (right) LMOs of two virtual orbitals of C_{60} fullerene. The isosurface value are 0.06 a.u. (top) and 0.02 a.u. (bottom). The low isosurface value emphasizes orbital tails.

Finally, the L-BFGS algorithm was used to localize virtual Berghold orbitals of pentacene using a triple-zeta basis set now augmented with diffuse functions (aug-TZV2P). In contrast to some traditional methods that struggle to localize virtual LMOs for molecules with delocalized electrons when diffuse basis set functions are included,[72] the variable-metric approach produces well localized aug-TZV2P orbitals of pentacene as efficiently as it localizes TZV2P orbitals (Figure S1 in Supporting Information). The final locality of an average aug-TZV2P NLMO of pentacene is 92% of that of an OLMO, very close to the 91% value of an average TZV2P orbital and within the 75−95% range reported for the other molecules (TZV2P) in Table 1. The distribution of LMOs according to the value of their Berghold localization functional is shown in Figure S2 in Supporting Information.

IV.C. Comparison of Berghold and Pipek−Mezey Virtual NLMOs. Figure 11 compares NLMOs and OLMOs

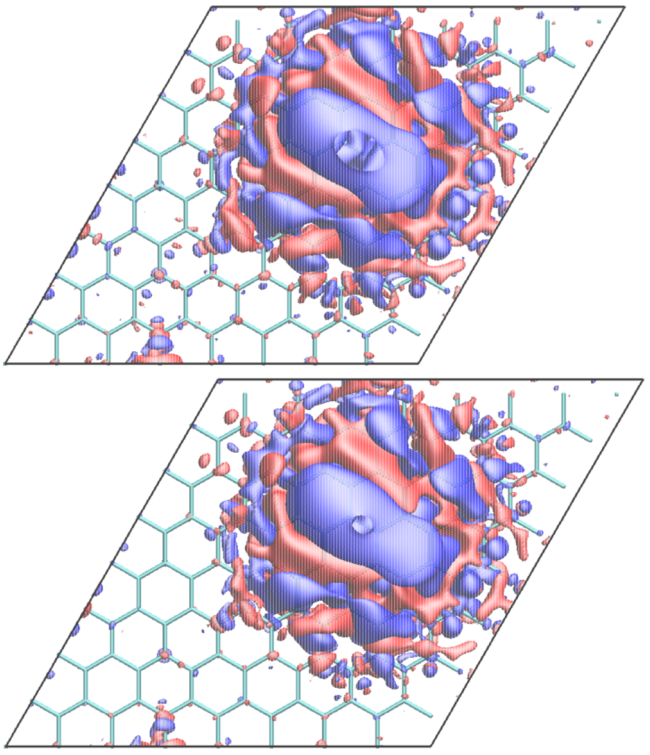


Figure 10. Orthogonal (top) and nonorthogonal (bottom) LMOs of the virtual orbital of graphene. The low isosurface value of 0.001 a.u emphasizes the tail reduction of the orbital.

obtained with the Berghold and Pipek−Mezey localization functionals for benzene. All C−H σ^* orbitals are well-reproduced by NLMOs and OLMOs with both localization schemes. In the case of OLMOs, it is widely known that the minimization of the Berghold functional produces τ^* anti-bonding orbitals centered on C−C bonds. These τ^* orbitals can be viewed as mixtures of σ^* and π^* anti-bonding orbitals. The minimization of the Pipek−Mezey functional, on the other hand, has been proven to guarantee to produce well-separated σ^* and π^* orbitals if orbital orthogonality is enforced. However, it has been shown that the minimization of the Pipek−Mezey functional is not guaranteed to preserve the $\sigma-\pi$ separation[63] if the orthogonality constraint is lifted. It has also been demonstrated that occupied Pipek−Mezey NLMOs of benzene indeed have the τ-character, unlike occupied Pipek−Mezey OLMOs of benzene.[63] Virtual Pipek−Mezey NLMOs of benzene obtained in this work are nevertheless well-separated σ^* and π^* orbitals. This is in agreement with the previously presented mathematical arguments[63] that state that Pipek−Mezey NLMOs can retain the $\sigma-\pi$ separation but this cannot be guaranteed.

V. CONCLUSIONS

It has been shown recently that a variable-metric approach to orbital localization allows us to obtain well-localized occupied orbitals by directly optimizing orbital mixing coefficients in a straightforward unconstrained minimization procedure. In this work, the simplicity of the variable-metric approach is exploited to design and compare numerous minimization algorithms with the goal of creating a straightforward and reliable method for localizing nonorthogonal *virtual* orbitals. The algorithms implemented in CP2K—a popular massively parallel freely distributed software package for modeling of materials and

https://doi.org/10.1021/acs.jctc.1c00379
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

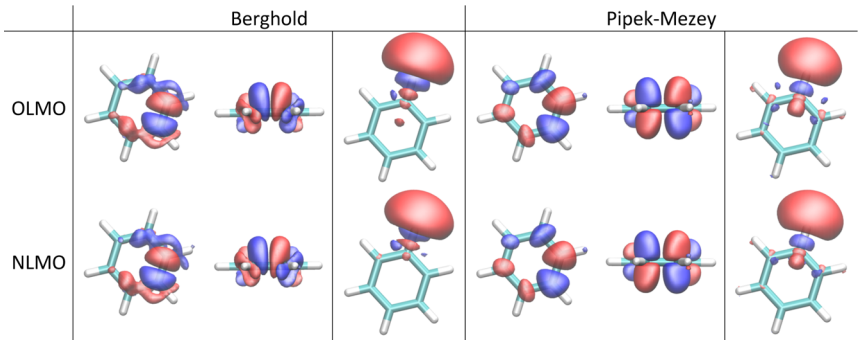


Figure 11. Comparison of OLMOs (top) and NLMOs (bottom) computed with the Berghold and Pipek−Mezey localization functionals for benzene. The isosurface value is 0.04 a.u.

molecules—belong to two families of optimization strategies: multidimensional line search and TR optimization. The list includes SD, CG, L-BFGS, and hybrid CG-BFGS algorithms as well as TR methods based on the Cauchy-point and CG subproblem solvers.

Based on the results of an extensive performance analysis, it can be argued that the TR(CG) algorithm is the best overall method to obtain NLMOs using the variable-metric approach. While the computational cost of a TR(CG) iteration is higher than that of all line search methods, its convergence rate often compensates for this drawback, especially in difficult cases. The tests show that the L-BFGS and CG algorithms can also be used to obtain virtual NLMOs reliably, disputing their poor reputation in metric-preserving localization of virtual orbitals. The simplicity of the CG algorithm and the ease of its implementation can outweigh its computational overhead in many cases. The convergence of SD and TR(Cauchy) is too slow to make them practical.

Comprehensive tests demonstrate that the implemented algorithms are capable of producing well-localized Berghold (i.e., Boys and MLWFs) and Pipek−Mezey, orthogonal and nonorthogonal, and occupied and virtual orbitals for molecules and periodic materials. Relaxing the orthogonality constraint produces noticeable gains in the locality of orbitals. If averaged across the tested systems, virtual NLMOs are 13% (Berghold) and 18% (Pipek−Mezey) more localized than their orthogonal counterparts. The reduction for the occupied orbitals is 19% (Berghold) and 3% (Pipek−Mezey). The ability of the variable-metric localization method to improve the locality of both occupied and virtual orbitals makes it a promising tool in designing more efficient local electron correlation methods.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00379.

All equations written in the standard matrix notation and localized molecular orbitals of pentacene (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Rustam Z. Khaliullin − *Department of Chemistry, McGill University, Montreal QC H3A 0B8, Canada;* orcid.org/0000-0002-9073-6753; Email: rustam.khaliullin@mcgill.ca

Author

Ziling Luo − *Department of Chemistry, McGill University, Montreal QC H3A 0B8, Canada*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c00379

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) through Discovery Grant (RGPIN-2016-0505) and by Triagency Institutional Programs Secretariat through New Frontiers in Research Fund (NFRFE-2018-00852). The authors are grateful to Compute Canada for computer time.

■ REFERENCES

(1) Boys, S. F. Construction of Some Molecular Orbitals to Be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys.* **1960**, *32*, 296−299.
(2) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457−464.
(3) Kitaura, K.; Morokuma, K. A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation. *Int. J. Quantum Chem.* **1976**, *10*, 325−340.
(4) Stevens, W. J.; Fink, W. H. Frozen fragment reduced variational space analysis of hydrogen bonding interactions. Application to the water dimer. *Chem. Phys. Lett.* **1987**, *139*, 15−22.
(5) Pipek, J.; Mezey, P. G. A Fast Intrinsic Localization Procedure Applicable for Ab Initio and Semiempirical Linear Combination of Atomic Orbital Wave Functions. *J. Chem. Phys.* **1989**, *90*, 4916−4926.
(6) von Niessen, W. Density Localization of Atomic and Molecular Orbitals. I. *J. Chem. Phys.* **1972**, *56*, 4290−4297.
(7) Staub, R.; Iannuzzi, M.; Khaliullin, R. Z.; Steinmann, S. N. Energy Decomposition Analysis for Metal Surface-Adsorbate Interactions by Block Localized Wave Functions. *J. Chem. Theory Comput.* **2019**, *15*, 265−275.
(8) Bagus, P. S.; Illas, F. Decomposition of the chemisorption bond by constrained variations: order of the variations and construction of the variational spaces. *J. Chem. Phys.* **1992**, *96*, 8962−8970.
(9) Glendening, E. D.; Streitwieser, A. Natural energy decomposition analysis: An energy partitioning procedure for molecular interactions with application to weak hydrogen bonding, strong ionic, and moderate donor-acceptor interactions. *J. Chem. Phys.* **1994**, *100*, 2900−2909.
(10) Mo, Y.; Gao, J.; Peyerimhoff, S. D. Energy decomposition analysis of intermolecular interactions using a block-localized wave function approach. *J. Chem. Phys.* **2000**, *112*, 5530−5538.
(11) Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M. Unravelling the origin of intermolecular interactions using

absolutely localized molecular orbitals. *J. Phys. Chem. A* **2007**, *111*, 8753−8765.
(12) Weinhold, F. Natural Bond Orbital Analysis: A Critical Overview of Relationships to Alternative Bonding Perspectives. *J. Comput. Chem.* **2012**, *33*, 2363−2379.
(13) Shi, Y.; Scheiber, H.; Khaliullin, R. Z. Contribution of the Covalent Component of the Hydrogen-Bond Network to the Properties of Liquid Water. *J. Phys. Chem. A* **2018**, *122*, 7482−7490.
(14) Adams, W. H. On the Solution of the Hartree-Fock Equation in Terms of Localized Orbitals. *J. Chem. Phys.* **1961**, *34*, 89−102.
(15) Stoll, H.; Wagenblast, G.; Preu*β*, H. On the use of local basis sets for localized molecular orbitals. *Theor. Chim. Acta* **1980**, *57*, 169.
(16) Pulay, P.; Saebø, S. Orbital-Invariant Formulation and Second-Order Gradient Evaluation in Møller-Plesset Perturbation Theory. *Theor. Chim. Acta* **1986**, *69*, 357−368.
(17) Hampel, C.; Werner, H. J. Local Treatment of Electron Correlation in Coupled Cluster Theory. *J. Chem. Phys.* **1996**, *104*, 6286−6297.
(18) Förner, W.; Knab, R.; Čížek, J.; Ladik, J. Numerical Application of the Coupled Cluster Theory with Localized Orbitals to Polymers. IV. Band Structure Corrections in Model Systems and Polyacetylene. *J. Chem. Phys.* **1997**, *106*, 10248−10264.
(19) Goedecker, S. Linear Scaling Electronic Structure Methods. *Rev. Mod. Phys.* **1999**, *71*, 1085.
(20) Saebø, S.; Pulay, P. A Low-Scaling Method for Second Order Møller−Plesset Calculations. *J. Chem. Phys.* **2001**, *115*, 3975−3983.
(21) Pisani, C.; Busso, M.; Capecchi, G.; Casassa, S.; Dovesi, R.; Maschio, L.; Zicovich-Wilson, C.; Schütz, M. Local-MP2 Electron Correlation Method for Nonconducting Crystals. *J. Chem. Phys.* **2005**, *122*, 094113.
(22) Wesolowski, T. A. *Computational Chemistry: Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific: Singapore, 2006; Vol. *10*; pp 1−82.
(23) Zaleśny, R.; Papadopoulos, M.; Mezey, P.; Leszczynski, J. *Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications*; Springer Science & Business Media, 2011.
(24) Bowler, D. R.; Miyazaki, T. \mathcal{O}(N) methods in electronic structure calculations. *Rep. Prog. Phys.* **2012**, *75*, 036503.
(25) Watanabe, Y.; Matsuoka, O. Nonorthogonal molecular orbital method: Single-determinant theory. *J. Chem. Phys.* **2014**, *140*, 204111.
(26) Shi, Y.; Khaliullin, R. Z. Robust Linear-Scaling Optimization of Compact Localized Orbitals in Density Functional Theory. **2020**, arXiv:2004.05901 2020.
(27) Meyer, W. PNO-CI Studies of electron correlation effects. I. Configuration expansion by means of nonorthogonal orbitals, and application to the ground state and ionized states of methane. *J. Chem. Phys.* **1973**, *58*, 1017−1035.
(28) Saebo, S.; Pulay, P. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213−236.
(29) Schütz, M.; Hetzer, G.; Werner, H.-J. Low-Order Scaling Local Electron Correlation Methods. I. Linear Scaling Local MP2. *J. Chem. Phys.* **1999**, *111*, 5691−5705.
(30) Schütz, M.; Werner, H. Low-Order Scaling Local Electron Correlation Methods. IV. Linear Scaling Local Coupled-Cluster (LCCSD). *J. Chem. Phys.* **2001**, *114*, 661−681.
(31) Neese, F.; Wennmohs, F.; Hansen, A. Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method. *J. Chem. Phys.* **2009**, *130*, 114108.
(32) Yang, J.; Kurashige, Y.; Manby, F. R.; Chan, G. K. L. Tensor factorizations of local second-order Møller-Plesset theory. *J. Chem. Phys.* **2011**, *134*, 044123.
(33) Azar, R. J.; Head-Gordon, M. An energy decomposition analysis for intermolecular interactions from an absolutely localized molecular orbital reference at the coupled-cluster singles and doubles level. *J. Chem. Phys.* **2012**, *136*, 024103.
(34) Misquitta, A. J. Charge transfer from regularized symmetry-adapted perturbation theory. *J. Chem. Theory Comput.* **2013**, *9*, 5313−5326.
(35) Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.* **2013**, *138*, 034106.
(36) Schneider, W. B.; Bistoni, G.; Sparta, M.; Saitow, M.; Riplinger, C.; Auer, A. A.; Neese, F. Decomposition of intermolecular interaction energies within the local pair natural orbital coupled cluster framework. *J. Chem. Theory Comput.* **2016**, *12*, 4778−4792.
(37) Azar, R. J.; Head-Gordon, M. Similarity-transformed perturbation theory on top of truncated local coupled cluster solutions: Theory and applications to intermolecular interactions. *J. Chem. Phys.* **2015**, *142*, 204101.
(38) Korona, T.; Kats, D.; Schütz, M.; Adler, T. B.; Liu, Y.; Werner, H.-J. *Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications*; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: Netherlands, Dordrecht, 2011; pp 345−407.
(39) Scheiber, H.; Shi, Y.; Khaliullin, R. Z. Communication: Compact orbitals enable low-cost linear-scaling ab initio molecular dynamics for weakly-interacting systems. *J. Chem. Phys.* **2018**, *148*, 231103.
(40) Høyvik, I.-M.; Jørgensen, P. Characterization and Generation of Local Occupied and Virtual Hartree−Fock Orbitals. *Chem. Rev.* **2016**, *116*, 3306−3327.
(41) Bytautas, L.; Ruedenberg, K. Electron Pairs, Localized Orbitals and Electron Correlation. *Mol. Phys.* **2002**, *100*, 757−781.
(42) Bytautas, L.; Ivanic, J.; Ruedenberg, K. Split-Localized Orbitals Can Yield Stronger Configuration Interaction Convergence Than Natural Orbitals. *J. Chem. Phys.* **2003**, *119*, 8217−8224.
(43) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833−1840.
(44) Löwdin, P. O. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *J. Chem. Phys.* **1950**, *18*, 365−375.
(45) Lehtola, S.; Jónsson, H. Pipek−Mezey Orbital Localization Using Various Partial Charge Estimates. *J. Chem. Theory Comput.* **2014**, *10*, 642−649.
(46) Jansík, B.; Høst, S.; Kristensen, K.; Jørgensen, P. Local Orbitals by Minimizing Powers of the Orbital Variance. *J. Chem. Phys.* **2011**, *134*, 194104.
(47) Høyvik, I.-M.; Jansik, B.; Jørgensen, P. Orbital Localization Using Fourth Central Moment Minimization. *J. Chem. Phys.* **2012**, *137*, 224114.
(48) Høyvik, I.-M.; Jansik, B.; Jørgensen, P. Trust Region Minimization of Orbital Localization Functions. *J. Chem. Theory Comput.* **2012**, *8*, 3137−3146.
(49) Høyvik, I.-M.; Jansik, B.; Jørgensen, P. Pipek−Mezey Localization of Occupied and Virtual Orbitals. *J. Comput. Chem.* **2013**, *34*, 1456−1462.
(50) Lehtola, S.; Jónsson, H. Unitary Optimization of Localized Molecular Orbitals. *J. Chem. Theory Comput.* **2013**, *9*, 5365−5372.
(51) Marzari, N.; Mostofi, A. A.; Yates, J. R.; Souza, I.; Vanderbilt, D. Maximally Localized Wannier Functions: Theory and Applications. *Rev. Mod. Phys.* **2012**, *84*, 1419−1475.
(52) Marzari, N.; Vanderbilt, D. Maximally Localized Generalized Wannier Functions for Composite Energy Bands. *Phys. Rev. B* **1997**, *56*, 12847.
(53) Jónsson, E. Ö.; Lehtola, S.; Puska, M.; Jónsson, H. Theory and Applications of Generalized Pipek−Mezey Wannier Functions. *J. Chem. Theory Comput.* **2017**, *13*, 460−474.
(54) Head-Gordon, M.; Maslen, P. E.; White, C. A. A Tensor Formulation of Many-Electron Theory in a Nonorthogonal Single-Particle Basis. *J. Chem. Phys.* **1998**, *108*, 616−625.
(55) Høyvik, I.-M.; Olsen, J.; Jørgensen, P. Generalising Localisation Schemes of Orthogonal Orbitals to the Localisation of Non-Orthogonal Orbitals. *Mol. Phys.* **2017**, *115*, 16−25.
(56) Anderson, P. W. Self-Consistent Pseudopotentials and Ultra-localized Functions for Energy Bands. *Phys. Rev. Lett.* **1968**, *21*, 13−16.
(57) Diner, S.; Malrieu, J. P.; Claverie, P.; Jordan, F. Fully Localized Bond Orbitals and the Correlation Problem. *Chem. Phys. Lett.* **1968**, *2*, 319−323.

https://doi.org/10.1021/acs.jctc.1c00379
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(58) Magnasco, V.; Musso, G. F. Localized orbitals and short-range molecular interactions. I. Theory. *J. Chem. Phys.* **1974**, *60*, 3744−3748.
(59) Payne, P. W. The Hartree-Fock Theory of Local Regions in Molecules. *J. Am. Chem. Soc.* **1977**, *99*, 3787−3794.
(60) Mehler, E. L. Self-Consistent, Nonorthogonal Group Function Approximation for Polyatomic Systems. I. Closed Shells. *J. Chem. Phys.* **1977**, *67*, 2728−2739.
(61) Feng, H.; Bian, J.; Li, L.; Yang, W. An Efficient Method for Constructing Nonorthogonal Localized Molecular Orbitals. *J. Chem. Phys.* **2004**, *120*, 9458−9466.
(62) Cui, G.; Fang, W.; Yang, W. Efficient Construction of Nonorthogonal Localized Molecular Orbitals in Large Systems†. *J. Phys. Chem. A* **2010**, *114*, 8878−8883.
(63) Luo, Z.; Khaliullin, R. Z. Direct Unconstrained Variable-Metric Localization of One-Electron Orbitals. *J. Chem. Theory Comput.* **2020**, *16*, 3558−3566.
(64) Liu, S.; Pérez-Jordá, J. M.; Yang, W. Nonorthogonal Localized Molecular Orbitals in Electronic Structure Theory. *J. Chem. Phys.* **2000**, *112*, 1634−1644.
(65) Barr, R.; Basch, H. Improved Convergence in Orbital Localization Methods. *Chem. Phys. Lett.* **1975**, *32*, 537−540.
(66) Berghold, G.; Mundy, C. J.; Romero, A. H.; Hutter, J.; Parrinello, M. General and Efficient Algorithms for Obtaining Maximally Localized Wannier Functions. *Phys. Rev. B* **2000**, *61*, 10040−10048.
(67) Kari, R. Parametrization and comparative analysis of theBFGS optimization algorithm for the determination of optimum linear coefficients. *Int. J. Quantum Chem.* **1984**, *25*, 321−329.
(68) Nocedal, J.; Wright, S. *Numerical Optimization*; Springer Science & Business Media, 2006.
(69) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. II. *J. Chem. Phys.* **1965**, *43*, S97−S116.
(70) Ryback, W.; Poirier, R.; Kari, R. An Application of the Method of Conjugate Gradients to the Calculation of Minimal Basis Set Localized Orbitals. *Int. J. Quantum Chem.* **1978**, *13*, 1−16.
(71) Leonard, J. M.; Luken, W. L. Quadratically Convergent Calculation of Localized Molecular Orbitals. *Theor. Chim. Acta* **1982**, *62*, 107−132.
(72) Høyvik, I.-M.; Jørgensen, P. Localized Orbitals from Basis Sets Augmented with Diffuse Functions. *J. Chem. Phys.* **2013**, *138*, 204104.
(73) Resta, R.; Sorella, S. Electron Localization in the Insulating State. *Phys. Rev. Lett.* **1999**, *82*, 370−373.
(74) Resta, R. Quantum-Mechanical Position Operator in Extended Systems. *Phys. Rev. Lett.* **1998**, *80*, 1800−1803.
(75) Silvestrelli, P. L. Maximally Localized Wannier Functions for Simulations with Supercells of General Symmetry. *Phys. Rev. B* **1999**, *59*, 9703−9706.
(76) Stanimirović, P.; Ivanov, B.; Djordjević, S.; Brajević, I. New Hybrid Conjugate Gradient and Broyden−Fletcher−Goldfarb−Shanno Conjugate Gradient Methods. *J. Optimiz. Theory App.* **2018**, *178*, 860−884.
(77) Fletcher, R.; Reeves, C. M. Function Minimization by Conjugate Gradients. *Comput J.* **1964**, *7*, 149−154.
(78) Steihaug, T. The Conjugate Gradient Method and Trust Regions in Large Scale Optimization. *SIAM J. Numer. Anal.* **1983**, *20*, 626−637.
(79) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; et al. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, 194103.
(80) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(81) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785−789.
(82) Goedecker, S.; Teter, M.; Hutter, J. Separable Dual-Space Gaussian Pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703−1710.
(83) Lee, M. S.; Maslen, P. E.; Head-Gordon, M. Closely approximating second-order Mo/ller-Plesset perturbation theory with a local triatomics in molecules model. *J. Chem. Phys.* **2000**, *112*, 3592−3601.
(84) Saebø, S. *Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications*; Zalesny, R., Papadopoulos, M. G., Mezey, P. G., Leszczynski, J., Eds.; Springer: Netherlands, Dordrecht, 2011; pp 65−82.

https://doi.org/10.1021/acs.jctc.1c00379
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX