The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Maximization of linear independence of basis function products

Cite as: J. Chem. Phys. 160, 234106 (2024); doi: 10.1063/5.0210971
Submitted: 28 March 2024 • Accepted: 3 June 2024 •
Published Online: 18 June 2024

View Online Export Citation CrossMark

Georgii N. Sizov, Vincent Lazeran, Llorenç Balada Gaggioli, and Viktor N. Staroverov[a)]

AFFILIATIONS

Department of Chemistry, The University of Western Ontario, London Ontario N6A 5B7, Canada

[a)]Author to whom correspondence should be addressed: vstarove@uwo.ca

ABSTRACT

Basis sets consisting of functions that form linearly independent products (LIPs) have remarkable applications in quantum chemistry but are scarce because of mathematical limitations. We show how to linearly transform a given set of basis functions to maximize the linear independence of their products by maximizing the determinant of the appropriate Gram matrix. The proposed method enhances the utility of the LIP basis set technology and clarifies why canonical molecular orbitals form LIPs more readily than atomic orbitals. The same approach can also be used to orthogonalize basis functions themselves, which means that various orthogonalization techniques may be viewed as special cases of a certain nonlinear optimization problem.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0210971

18 June 2024 15:18:30

I. INTRODUCTION

Suppose we have a basis set $\{f_i\}$ of n real square-integrable functions. Unlike the basis functions themselves, the products $f_i f_j$ $(i \le j)$ are not necessarily linearly independent. If they are, $\{f_i\}$ is said to be a *linearly independent product* (LIP) basis set. As pointed out by Harriman,[1–3] such basis sets provide remarkable ways to perform electronic structure calculations that would be impossible otherwise. In particular, within a LIP basis set, an electron density uniquely determines the corresponding one-electron reduced density matrix (1-RDM) and any nonlocal operator is equivalent to a certain multiplicative potential operator. This means that, by using a LIP basis set, one can reconstruct 1-RDMs from electron densities and represent the kinetic energy and exchange operators *exactly* by effective local potentials. Practical applications of LIP basis sets for such purposes have been reported not only for model systems[4–7] but also for real atoms and molecules.[8–11]

The main obstacle preventing LIP basis sets from wider use is their limited availability. In order for a given set $\{f_i\}$ to be a LIP basis set in practice, it is not enough that the products $f_i f_j$ be linearly independent in the mathematical sense; they must be also free from approximate linear dependencies that are indistinguishable from true linear dependencies in calculations with finite precision. Almost no standard basis of quantum chemistry has this property.[6,12–14] Previous attempts to devise general-purpose LIP basis sets have run into a wall, because the products $f_i f_j$ develop effective linear dependence rapidly and inexorably as the size of the underlying basis set $\{f_i\}$ increases beyond a few functions.[4–6] Of course, the onset of numerical linear dependence of the basis function products can be delayed by switching from double to quadruple or higher numerical precision, but it is desirable to have LIP basis sets within the standard double precision. Another way to unlock access to the LIP basis-set technology is by using canonical Hartree–Fock or Kohn–Sham molecular orbitals as basis functions, which are more likely to form LIP basis sets than nucleus-centered orbitals.[9–12] However, this remedy is not sufficiently effective, because canonical orbitals are system-specific and not transferable.

A basis set with exact linear dependencies in the products cannot be linearly transformed into a LIP basis set.[10] However, this does not preclude the possibility that a given basis set with *approximate* linear dependencies in the products can be turned into a numerically robust LIP basis set. In this work, we show how the linear independence of basis function products $f_i f_j$ can be enhanced through a specific linear transformation of the underlying basis set $\{f_i\}$. The suggested approach is of intrinsic mathematical interest and has the potential to make the LIP basis-set technology more practical than it is at present. Our findings also shed light on the observation that molecular orbitals are more likely to form LIPs than equivalent sets of atomic orbitals.

Published under an exclusive license by AIP Publishing

II. METHODOLOGY

A. Measure of linear independence

The extent of linear independence of a given set of functions $\{\phi_i\}$ can be characterized[15] by the magnitude of the Gramian, i.e., the determinant of the Gram matrix,

$$\mathbf{G} = \begin{pmatrix} \langle\phi_1|\phi_1\rangle & \langle\phi_1|\phi_2\rangle & \dots & \langle\phi_1|\phi_n\rangle \\ \langle\phi_2|\phi_1\rangle & \langle\phi_2|\phi_2\rangle & \dots & \langle\phi_2|\phi_n\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\phi_n|\phi_1\rangle & \langle\phi_n|\phi_2\rangle & \dots & \langle\phi_n|\phi_n\rangle \end{pmatrix}. \tag{1}$$

In quantum chemistry, $\mathbf{G}$ goes by the name of overlap matrix. If the functions ϕ_i are normalized to unity, which can always be effected, then the value of the Gramian is confined to the interval[16]

$$0 \le \det \mathbf{G} \le 1. \tag{2}$$

The value $\det \mathbf{G} = 0$ means that the set $\{\phi_i\}$ has at least one linear dependence. The maximum value, $\det \mathbf{G} = 1$, is attained if and only if all the normalized functions are orthogonal, i.e., "maximally linearly independent." If the functions $\{\phi_i\}$ have at least one approximate linear dependence, the Gramian will be small but positive. Thus, the magnitude of $\det \mathbf{G}$ meets all the requirements for a rigorous measure of the extent of linear independence for $\{\phi_i\}$, provided that each function is normalized to unity.

The smallest eigenvalue $\lambda_{\min}$ of $\mathbf{G}$ is another, perhaps better known, indicator[15,17,18] of linear independence of functions. Previously, Morrison[8] optimized LIP basis sets of orthonormal orbitals by unitary transformations that maximize the lowest eigenvalue of the overlap matrix of the product functions. The criterion proposed here $(\det \mathbf{G})$ measures the extent of linear independence of all n functions simultaneously, whereas the magnitude of $\lambda_{\min}$ reflects just one (the weakest) linear independence in $\{\phi_i\}$.

B. Application to basis functions

Before attacking the problem of linear independence of basis function products, we will show how one can employ the determinant of the Gram matrix to orthogonalize ordinary basis sets themselves. We will distinguish the new (transformed) basis functions by attaching a prime (e.g., f_i') and denote normalization to unity by a tilde, as in $\tilde{f}_i, \tilde{f}_i'$, etc.

Suppose we have an ordered set of linearly independent functions written as a column vector

$$\mathbf{f} = (f_1\ f_2\ \dots\ f_n)^T, \tag{3}$$

where T stands for transposition. The Gram matrix for these functions will be denoted by $\mathbf{S}$ and its elements by

$$S_{i,j} = \langle f_i|f_j\rangle. \tag{4}$$

Here and below, the row and column indices of two-dimensional arrays are separated by commas to distinguish them from collective subscripts such as ij (as in g_{ij}).

We want to find a linear transformation matrix $\mathbf{A}$ such that the transformed basis set

$$\mathbf{f}' = \mathbf{A}\mathbf{f} \tag{5}$$

is maximally linearly independent, that is, has the greatest possible Gramian. The requisite matrix $\mathbf{A}$ is not unique, but this only simplifies the search, because any particular solution will do the job.

The Gram matrix for the new unnormalized basis $\mathbf{f}'$ is

$$\mathbf{S}' = \mathbf{A}\mathbf{S}\mathbf{A}^T, \tag{6}$$

so that

$$\det \mathbf{S}' = |\det \mathbf{A}|^2 \det \mathbf{S}. \tag{7}$$

The Gram determinant we need is not $\det \mathbf{S}'$ but that of the closely related matrix $\tilde{\mathbf{S}}'$, which involves the normalized new functions. The elements of $\tilde{\mathbf{S}}'$ are given by

$$\tilde{S}'_{i,j} = \frac{S'_{i,j}}{\sqrt{S'_{i,i}S'_{j,j}}}. \tag{8}$$

Examining the common factors appearing in each row and each column of $\tilde{\mathbf{S}}'$, one can see that the determinants of $\mathbf{S}'$ and $\tilde{\mathbf{S}}'$ are related by

$$\det \tilde{\mathbf{S}}' = \frac{\det \mathbf{S}'}{\prod_{i=1}^{n} \langle f_i'|f_i'\rangle}, \tag{9}$$

where f_i' are the raw (unnormalized) transformed functions. Finally, substituting Eq. (7) into Eq. (9) and denoting the result by Q_S, we obtain

$$Q_S \equiv \det \tilde{\mathbf{S}}' = \frac{|\det \mathbf{A}|^2}{\prod_{i=1}^{n} \langle f_i'|f_i'\rangle} \det \mathbf{S}. \tag{10}$$

We treat Q_S as a function of the elements of $\mathbf{A}$, which enter Eq. (10) through the numerator and the denominator but not through $\det \mathbf{S}$. As explained in Sec. II A, Q_S $(0 \le Q_S \le 1)$ is the measure of linear independence of the transformed basis functions $\{f_i'\}$.

Thus, the problem of orthogonalizing a given basis set $\{f_i\}$ is equivalent to maximizing the function Q_S, which is the ratio of two polynomials in the elements of $\mathbf{A}$. Global optimization of rational functions is generally complicated,[19–21] but in this case the problem is simplified by two facts: (i) the global maximum of Q_S is known *a priori* to be 1, and (ii) the problem has infinitely many solutions (orthogonalizing transformations $\mathbf{A}$, which differ from another by proper or improper rotations), any one of which is acceptable. As soon as one finds a point with $Q_S = 1$, starting from some initial guess, one can stop.

If applied to a set of non-orthogonal linearly independent functions, maximization of Q_S amounts to a peculiar orthogonalization procedure in which the transformation $\mathbf{A}$ is chosen without user control. This is to be contrasted with conventional orthogonalization methods, such as Löwdin's symmetric orthonormalization[22] and other techniques,[18] where the matrix $\mathbf{A}$ is chosen at the outset. This ability of the Gramian-maximization approach to cover all possible transformations of $\mathbf{f}$ proves crucial in what comes next.

C. Application to products of basis functions

Suppose we have a formally LIP basis set $\{f_i\}$ consisting of n real functions and consider all distinct basis function products

$$g_j = f_i f_j \quad (i \le j). \tag{11}$$

18 June 2024 15:18:30

Published under an exclusive license by AIP Publishing

We treat ij as a collective index with a numerical value of $i + j(j-1)/2$. The total number of products g_{ij} is

$$M = \frac{n(n+1)}{2}. \tag{12}$$

The $M \times M$ Gram matrix for these products is

$$\mathbf{W} = \begin{pmatrix} W_{11,11} & W_{11,12} & \dots & W_{11,nn} \\ W_{12,11} & W_{12,12} & \dots & W_{12,nn} \\ \vdots & \vdots & \ddots & \vdots \\ W_{nn,11} & W_{nn,12} & \dots & W_{nn,nn} \end{pmatrix}, \tag{13}$$

where

$$W_{ij,kl} = \langle g_{ij} | g_{kl} \rangle = \langle f_i f_j | f_k f_l \rangle \quad (i \le j, k \le l). \tag{14}$$

Following the convention of Sec. II B, we distinguish the products of the transformed functions by attaching a prime (as in $g'_{ij} = f'_i f'_j$) and denote the normalized products by a tilde. Note that $\tilde{g}_{ij} \ne \tilde{f}_i \tilde{f}_j$.

We want to find a linearly transformed basis $\{f_i'\}$ such that the products $f_i' f_j'$ are maximally close to forming an orthogonal basis. The purpose of enhancing the linear independence of the products $f_i f_j$ is to facilitate the numerical solution of ill-conditioned linear systems of equations such as

$$\mathbf{W}\mathbf{x} = \mathbf{b}, \tag{15}$$

where $\mathbf{b}$ is some given column vector. Such linear equations arise and must be solved accurately in Harriman's method.[2,8–11,23]

Since $\mathbf{W}$ tends to be nearly singular for atom-centered basis sets, the problem expressed by Eq. (15) is typically ill-conditioned. The relative error of the solution $\mathbf{x}$ is determined[24] by the appropriate condition number $\kappa = \lambda_{\max}/\lambda_{\min}$, where $\lambda_{\max}$ and $\lambda_{\min}$ are the greatest and the smallest eigenvalues of $\mathbf{W}$, respectively. It can be shown[24] that if Eq. (15) is solved using floating-point arithmetic with D decimal digits, and if $\kappa(\mathbf{W}) \approx 10^K$, then the solution $\mathbf{x}$ can be guaranteed to be correct up to $D - K - 1$ digits. Thus, when $\mathbf{W}$ is ill-conditioned and the linear transformation of the underlying basis set can reduce $\kappa(\mathbf{W})$ by m digits, then the accuracy of $\mathbf{x}$ is boosted by m digits.

The best possible condition number for any matrix is $\kappa = 1$. For $\mathbf{W}$, reaching this value would require the products g_{ij} to be orthonormal, which is impossible because

$$\langle g_{ii} | g_{jj} \rangle = \langle g_{ij} | g_{ij} \rangle \ne 0 \tag{16}$$

even if $i \ne j$. Instead of minimizing κ directly (which is too difficult), we propose to enhance the linear independence of the functions g_{ij} by maximizing the Gramian of the normalized products $\tilde{g}_{ij}$. This can be done as follows by analogy with Sec. II C.

Given an arbitrary basis set $\mathbf{f}$, we denote the corresponding product basis by an M-component column vector,

$$\mathbf{g} = (g_{11}\ g_{12}\ g_{22}\ g_{13}\ g_{23}\ g_{33}\ \dots\ g_{nn})^T. \tag{17}$$

After applying a linear transformation $\mathbf{A}$ to $\mathbf{f}$, the product basis $\mathbf{g}$ becomes

$$\mathbf{g}' = \mathbf{C}\mathbf{g}. \tag{18}$$

Here, $\mathbf{C}$ is an M-by-M matrix with the elements

$$C_{j,kl} = \begin{cases} A_{i,k}A_{j,l} & \text{if } k = l, \\ A_{i,k}A_{j,l} + A_{i,l}A_{j,k} & \text{if } k < l, \end{cases} \tag{19}$$

where $i \le j$ and $k \le l$. Symbolically, we write $\mathbf{C} = \mathbf{A} \otimes_s \mathbf{A}$, where the operation $\otimes_s$ is the symmetric Kronecker product (Appendix A). In Appendix B, it is shown that

$$\det \mathbf{C} = \det(\mathbf{A} \otimes_s \mathbf{A}) = (\det \mathbf{A})^{n+1}. \tag{20}$$

Under the linear transformation $\mathbf{A}$ of $\mathbf{f}$, the Gram matrix of the basis function products transforms as

$$\mathbf{W}' = \mathbf{C}\mathbf{W}\mathbf{C}^T. \tag{21}$$

We denote the Gram matrix of the normalized transformed products $\{\tilde{g}'_{ij}\}$ by $\tilde{\mathbf{W}}'$. The elements of this matrix are

$$\tilde{W}'_{ij,kl} = \frac{W'_{ij,kl}}{\sqrt{W'_{ij,ij} W'_{kl,kl}}}. \tag{22}$$

From Eqs. (20)–(22), the Gramian for $\{\tilde{g}'_{ij}\}$ is given by

$$Q_W \equiv \det \tilde{\mathbf{W}}' = \frac{|\det \mathbf{A}|^{2(n+1)}}{\prod_{i \le j}^{M} \langle g'_{ij} | g'_{ij} \rangle} \det \mathbf{W}, \tag{23}$$

where the functions g'_{ij} in the denominator are unnormalized.

Thus, to maximize the linear independence of the basis function products $\{g'_{ij}\}$, one needs to maximize Q_W with respect to the elements of $\mathbf{A}$. This is analogous to how the basis functions themselves were orthogonalized by maximizing the function Q_S in Sec. II B. Because of Eq. (16), the simultaneous orthogonality of all M products is impossible, so the optimal value of Q_W must be necessarily less than 1.

There are other essential differences between the objective functions Q_S and Q_W. Unlike Q_S, the function Q_W is *not* invariant under orthogonal transformations of the underlying basis $\mathbf{f}$. This is because, when $\mathbf{A}$ is orthogonal, the matrix $\mathbf{C} = \mathbf{A} \otimes_s \mathbf{A}$ is not (cf. Ref. 8). As a consequence, Q_W has a more complicated topography with many local maxima and an unknown global maximum. The problem of maximizing Q_W is, therefore, akin to searching for the global minimum on a molecular potential energy surface.[25] Global optimization of rational functions such as Q_S and Q_W is an active area of mathematical research.[19–21] In this proof-of-concept project, we rely on the *Maple* software[26] to maximize both Q_S and Q_W.

D. Practical considerations

In the most general case, the linear transformation matrix $\mathbf{A}$ has n^2 independent elements (degrees of freedom),

$$\mathbf{A} = \begin{pmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,n} \\ A_{2,1} & A_{2,2} & \dots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \dots & A_{n,n} \end{pmatrix}. \tag{24}$$

18 June 2024 15:18:30

Published under an exclusive license by AIP Publishing

Of these, n degrees of freedom scale the transformed functions f_i' and do not affect the "angles" between them. Thus, without loss of generality, one can consider only transformations with $n(n-1)$ variables, e.g., matrices of the form

$$\mathbf{A} = \begin{pmatrix} 1 & A_{1,2} & \dots & A_{1,n} \\ A_{2,1} & 1 & \dots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \dots & 1 \end{pmatrix}. \tag{25}$$

Furthermore, $n(n-1)/2$ of these $n(n-1)$ variables describe orthogonal transformations (proper or improper rotations), which do not affect the value of Q_S. The linear transformation of Eq. (25) can thus be thought of as a product of an orthogonal matrix and an upper triangular matrix (QR decomposition) effecting a rotation and an orthogonalization, respectively. For the purpose of maximizing Q_S, the orthogonal matrix is irrelevant and it is sufficient to consider only upper triangular transformation matrices such as

$$\mathbf{A} = \begin{pmatrix} 1 & A_{1,2} & \dots & A_{1,n} \\ 0 & 1 & \dots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}. \tag{26}$$

A matrix of this form realizes the Gram–Schmidt orthogonalization.

To maximize Q_W, one must not constrain $\mathbf{A}$ to any form less general than Eq. (25). This is because, in contrast with Q_S, the quantity Q_W is not invariant under orthogonal transformations of $\mathbf{f}$ and Eq. (26) does not cover all possible linear transformations of $\mathbf{f}$.

III. ILLUSTRATIVE EXAMPLES

The following examples illustrate the proposed method in application to simple basis functions of one variable. This is not a limitation, because the number of variables of Q_S and Q_W depends only on the number of basis functions and not on their form. All calculations were carried out with *Maple*[26] using 16 digits for floating-point numbers. The results were rounded to the number of decimals shown.

A. Two basis functions centered at the same point

Consider a set of two unnormalized, non-orthogonal functions

$$\begin{aligned} f_1(x) &= e^{-x^2}, \\ f_2(x) &= e^{-3x^2}. \end{aligned} \tag{27}$$

The most general linear transformation of this basis is specified by a four-parameter matrix of the type of Eq. (24). As explained in Sec. II D, for the purpose of optimizing the functions Q_S and Q_W, there is no loss of generality in considering only linear transformations of the form

$$\mathbf{A} = \begin{pmatrix} 1 & a \\ b & 1 \end{pmatrix}. \tag{28}$$

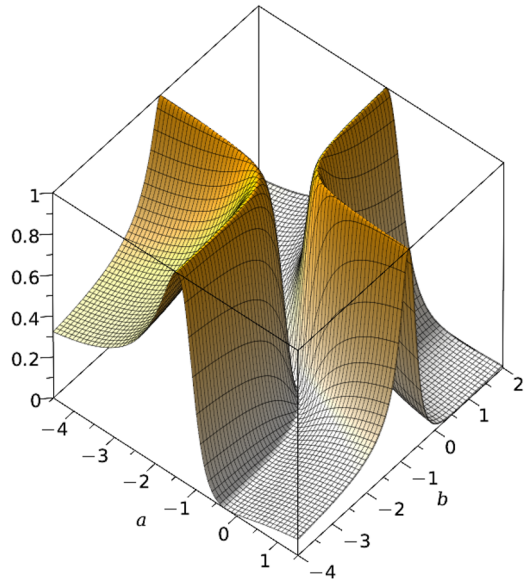


FIG. 1. Objective function $Q_S(a,b)$ of Eq. (29). The function reaches its maximum values along the tops of two infinite ridges.

To show how the Gramian maximization works, let us first orthogonalize the basis set $\{f_i\}$ by optimizing Q_S. For $n = 2$, the objective function Q_S of Eq. (10) assumes the form

$$Q_S(a,b) = \frac{|\det \mathbf{A}|^2}{\langle f_1'|f_1'\rangle\langle f_2'|f_2'\rangle} \det \mathbf{S}, \tag{29}$$

where f_i' are the transformed functions and $\det \mathbf{S}$ is the Gramian for the original functions. The plot of Q_S as a function of a and b (Fig. 1) shows that there are infinitely many (a,b) pairs that maximize Q_S and orthogonalize f_1 and f_2. All such pairs form two infinite ridges of constant height $Q_S = 1$, as seen in Fig. 1.

To maximize the linear independence of the basis function products, we maximize the objective function Q_W of Eq. (23). For $n = 2$, Q_W is given by

$$Q_W(a,b) = \frac{|\det \mathbf{A}|^6}{\langle g_{11}'|g_{11}'\rangle\langle g_{12}'|g_{12}'\rangle\langle g_{22}'|g_{22}'\rangle} \det \mathbf{W}, \tag{30}$$

where $\mathbf{A}$ is the same as in Eq. (29) and $\det \mathbf{W}$ is the Gramian for the products of the original basis functions.

Figure 2 shows that Q_W has only two equal maxima indicated in the legend. The first maximum corresponds to the optimized basis

$$\begin{aligned} \tilde{f}_1' &= 2.3160 f_1 - 2.4659 f_2, \\ \tilde{f}_2' &= -0.7142 f_1 + 1.9988 f_2, \end{aligned} \tag{31}$$

while the second one corresponds to

$$\begin{aligned} \tilde{f}_1' &= 0.7142 f_1 - 1.9988 f_2, \\ \tilde{f}_2' &= -2.3160 f_1 + 2.4659 f_2. \end{aligned} \tag{32}$$

These two basis sets are clearly equivalent, which explains why the two maxima seen in Fig. 2 are equal. The transformation of $\mathbf{f}$ to

18 June 2024 15:18:30

Published under an exclusive license by AIP Publishing

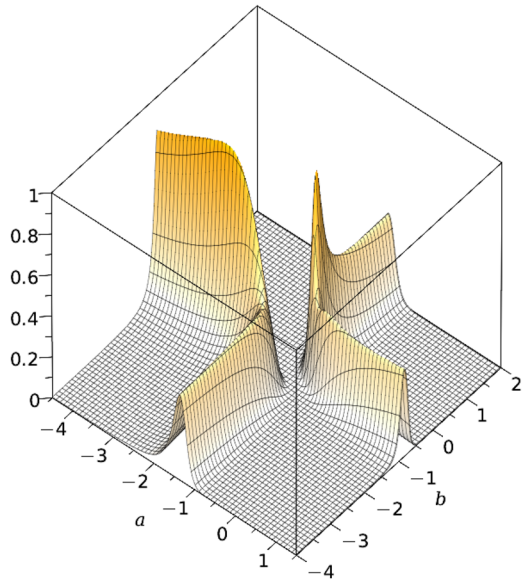


FIG. 2. Objective function $Q_W(a,b)$ of Eq. (30). The function reaches its maximum value $Q_W^{\max} = 0.9945$ at two different points: $(-1.0648, -0.3573)$ and $(-2.7987, -0.9392)$.

the basis set of Eqs. (31) and (32) reduces the condition number of the Gram matrix for the normalized products from $\kappa(\tilde{\mathbf{W}}) = 3.57 \times 10^3$ to $\kappa(\tilde{\mathbf{W}}') = 1.07$. The smallest eigenvalue increased by three orders of magnitude from $\lambda_{\min}(\tilde{\mathbf{W}}) = 1.08 \times 10^{-3}$ to $\lambda_{\min}(\tilde{\mathbf{W}}') = 0.926$, which shows that the transformation is beneficial.

The normalized new basis functions are shown in Fig. 3 and their normalized products in Fig. 4. It is noteworthy that the functions of Eqs. (31) and (32) are almost, but not quite, orthogonal,

$$\tilde{\mathbf{S}}' = \begin{pmatrix} 1 & 0.023\,67 \\ 0.023\,67 & 1 \end{pmatrix}, \tag{33}$$

and are similar to the Löwdin-orthonormalized functions (Fig. 3). The lack of orthogonality of $\tilde{f}_1'$ and $\tilde{f}_2'$ is consistent with the fact that the two ridges of $Q_W(a,b)$ are positioned similarly, but not identically, to the ridges of $Q_S(a,b)$ (compare Figs. 1 and 2).

B. Three basis functions centered at different points

Consider the following non-orthogonal basis set consisting of three unnormalized Gaussians:

$$\begin{aligned} f_1(x) &= e^{-\left(x+\frac{1}{5}\right)^2}, \\ f_2(x) &= e^{-x^2}, \\ f_3(x) &= e^{-\left(x-\frac{3}{5}\right)^2}. \end{aligned} \tag{34}$$

Because of the spatial proximity of f_1 and f_2, the products $f_i f_j$ are almost linearly dependent: their Gram matrix has a small determinant $\det \tilde{\mathbf{W}} = 5.43 \times 10^{-17}$ and a large condition number $\kappa(\tilde{\mathbf{W}}) = 1.39 \times 10^8$.

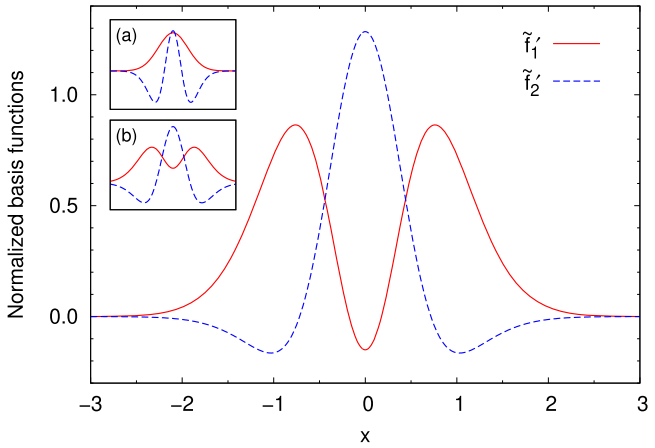


FIG. 3. Normalized basis functions of Eq. (31), which are not exactly orthogonal. The insets show the original functions orthonormalized by the (a) Gram–Schmidt and (b) Löwdin methods.

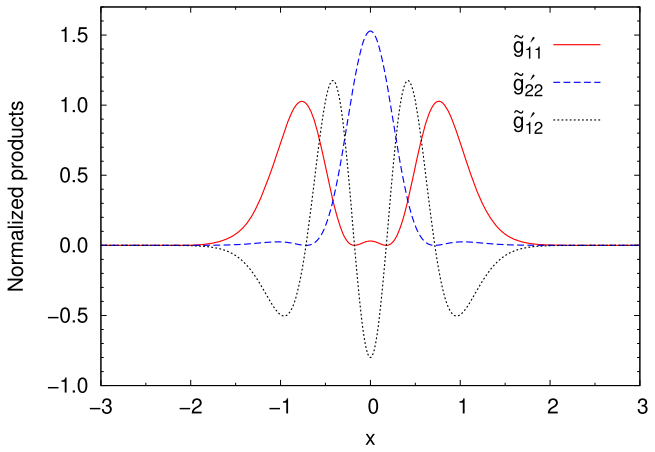


FIG. 4. Normalized products of the transformed basis functions shown in Fig. 3.

To optimize this LIP basis set, we employed the linear transformation matrix $\mathbf{A}$ given by Eq. (25). Since $n = 3$, Q_W is a function of $n(n-1) = 6$ variables. The optimized linear combinations of the basis functions f_i were found to be (after normalization)

$$\begin{aligned} \tilde{f}_1' &= 4.5966 f_1 - 4.3994 f_2 + 0.6045 f_3, \\ \tilde{f}_2' &= -7.2976 f_1 + 9.6397 f_2 - 2.4311 f_3, \\ \tilde{f}_3' &= 1.9705 f_1 - 3.1284 f_2 + 1.9635 f_3. \end{aligned} \tag{35}$$

As in the previous example, these functions are almost, but not quite, orthogonal,

$$\tilde{\mathbf{S}}' = \begin{pmatrix} 1 & -0.047\,53 & 0.048\,81 \\ -0.047\,53 & 1 & -0.043\,56 \\ 0.048\,81 & -0.043\,56 & 1 \end{pmatrix}. \tag{36}$$

This time, however, the products $f_i' f_j'$ are less linearly independent overall than in the previous example. No off-diagonal element of $\tilde{\mathbf{W}}'$

18 June 2024 15:18:30

Published under an exclusive license by AIP Publishing

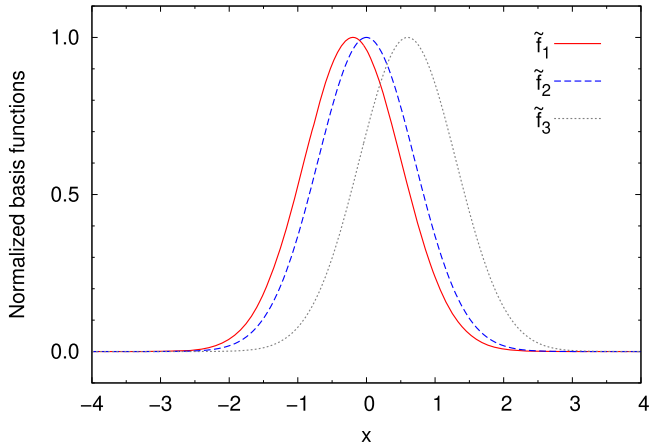


FIG. 5. Original basis functions of Eq. (34) after normalization.

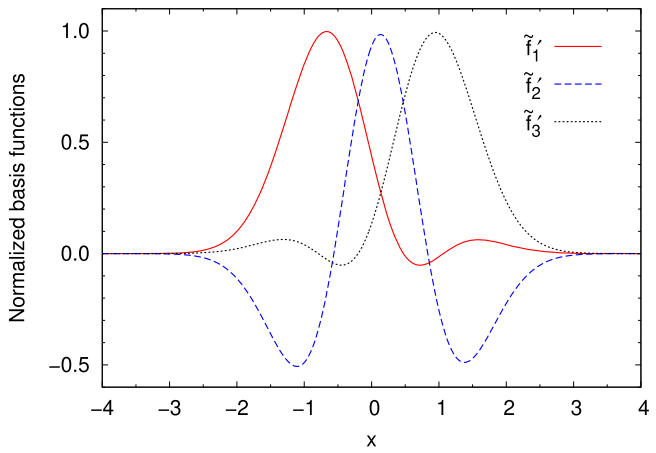


FIG. 6. Transformed normalized basis functions of Eq. (35).

is exactly zero, the smallest ones being of the order of 10^{-3}. Nevertheless, the improvement over the original basis set is substantial: the condition number of the Gram matrix has dropped to $\kappa(\tilde{\mathbf{W}}') = 7.39 \times 10^4$, which is an improvement by four orders of magnitude, and the smallest eigenvalue increased by three orders of magnitude from $\lambda_{\min}(\tilde{\mathbf{W}}) = 5.17 \times 10^{-8}$ to $\lambda_{\min}(\tilde{\mathbf{W}}') = 6.16 \times 10^{-5}$. For comparison, the optimization of LIP basis sets by maximizing $\lambda_{\min}$ has not been observed to increase the magnitude of $\lambda_{\min}$ by more than one order of magnitude.[8]

A comparison of the original and the transformed functions (Figs. 5 and 6) shows that the linear transformation enhancing the independence of the products pushes the maxima of f_i^2 apart, as should be expected.

IV. DISCUSSION

The optimization of the function Q_S has an appealing geometric interpretation in terms of concepts of linear algebra. The quantity $\det \mathbf{S}$ may be regarded as the square of the volume of an n-dimensional "parallelepiped" spanned by "vectors" $\{f_i\}$,[27] whereas $|\det \mathbf{A}|$ is the volume expansion factor associated with the linear transformation $\mathbf{A}$.[28] Thus, $|\det \mathbf{A}|^2 \det \mathbf{S}$ is the square of the volume of the transformed n-parallelepiped spanned by $\{f_i'\}$. The denominator $\prod_{i=1}^n \langle f_i'|f_i'\rangle$ is the square of the product of all sides of the transformed n-parallelepiped and as such is equal to the square of the maximum possible volume. This makes it clear why the quotient Q_S varies between 0 and 1. Observe that the Löwdin transformation[18] $\mathbf{A} = \mathbf{S}^{-1/2}$ makes all functions f_i' normalized and automatically gives $Q_S = 1$, which proves that this transformation is an orthonormalization. The objective functions Q_W can be interpreted similarly, except that Q_W can never be equal to 1.

A conceivable alternative to maximizing Q_W is to try to set up and solve a system of nonlinear equations $W'_{ii,ii} = 1$, $W'_{ii,jj} = W'_{ij,ij}$, $W'_{ij,kl} = 0$ (if permitted by the subscripts), etc., where the coefficients $A_{i,j}$ are the unknowns. This would not work for several reasons. First, such a system is in general inconsistent, because the number of equations is of the order of n^4, whereas the number of unknowns is of the order of n^2. Second, as Eq. (36) shows, it may happen that no matrix element of $\mathbf{W}'$ is exactly zero, contrary to the assumption that at least some of them should vanish. Finally, systems of nonlinear equations, even if consistent, are more difficult to solve than optimization problems. Formulating the search for $\mathbf{A}$ as an unconstrained optimization problem resolves all these complications.

Numerical illustrations of the proposed method suggest, in agreement with previous reports,[4–6,12] that atomic-orbital-type (localized) basis functions are poor LIP basis sets. Linear independence of basis function products is enhanced when the largest maxima of $|f_i|$ are spatially separated (as in Figs. 3 and 6). Since canonical molecular orbitals tend to have their maxima/minima at different locations, the products of molecular orbitals tend to be more linearly independent than the products of atomic orbitals.[9–12] This also signals that the problem of LIP basis set construction is related in some sense to the processes of orbital orthogonalization and localization.[29–34]

In practice, the extent to which the products f_if_j can be made linearly independent is limited by two interrelated factors: (i) the products f_if_j cannot be all simultaneously orthogonal (for reasons explained in Sec. II C), and (ii) the number of strictly non-negative products such as f_i^2, which especially hinder orthogonality, increases with the size of the basis set. This suggests that making Q_W close to 1 should be harder for a larger basis set than for a smaller one. The examples presented in Sec. III are consistent with this conclusion.

V. CONCLUSION

The LIP basis set technology has a great appeal in principle but severe limitations in practice. This is because the products of basis functions tend to exhibit approximate linear dependencies whose extent rapidly increases with the size of the basis set. We have shown here that one can always find linear transformations of a given basis set $\{f_i\}$ which maximize the extent of linear independence of the basis function products f_if_j in the sense of making their Gramian (after normalization) as close to 1 as possible. In this manner, a basis set that is not LIP in a numerical sense can be sometimes turned into a numerically robust LIP basis set using the technique developed in this work. Such basis set transformations enhance the practical utility of the LIP basis set technology.

18 June 2024 15:18:30

Published under an exclusive license by AIP Publishing

A by-product of the proposed method is a reformulation of the orthogonalization process for a given set of functions as an unconstrained nonlinear optimization problem. This perspective may be conceptually useful because it subsumes all particular orthogonalization techniques.

ACKNOWLEDGMENTS

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Discovery Grants Program (Application No. RGPIN-2020-06420). L.B.G. also acknowledges the support from the Mitacs Globalink Research Internship program.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Georgii N. Sizov: Conceptualization (equal); Methodology (equal); Writing – original draft (supporting); Writing – review & editing (supporting). **Vincent Lazeran**: Conceptualization (supporting); Methodology (equal); Writing – original draft (equal); Writing – review & editing (supporting). **Llorenç Balada Gaggioli**: Conceptualization (supporting); Methodology (equal); Writing – original draft (equal); Writing – review & editing (supporting). **Viktor N. Staroverov**: Conceptualization (equal); Funding acquisition (lead); Methodology (equal); Writing – original draft (lead); Writing – review & editing (lead).

DATA AVAILABILITY

The data that support the findings of this study are available within the article.

APPENDIX A: SYMMETRIC KRONECKER PRODUCT

Let $\mathbf{A}$ be an $n \times n$ matrix and let $M = n(n+1)/2$. The symmetric Kronecker product of $\mathbf{A}$ with itself, denoted by $\mathbf{A} \otimes_s \mathbf{A}$, is an M-by-M matrix with the elements

$$(\mathbf{A}\otimes_s\mathbf{A})_{j,kl} = \frac{1}{1+\delta_{k,l}}\left(A_{i,k}A_{j,l} + A_{i,l}A_{j,k}\right), \tag{A1}$$

where $i \le j$ and $k \le l$. This definition is motivated as follows.

Consider a LIP set of n functions $\{f_i\}$ subjected to a linear transformation specified by $\mathbf{A}$ according to

$$f_i' = \sum_{k=1}^{n} A_{i,k} f_k. \tag{A2}$$

The product of two transformed functions is

$$f_i' f_j' = \left(\sum_{k=1}^{n} A_{i,k} f_k\right)\left(\sum_{l=1}^{n} A_{j,l} f_l\right) = \sum_{k=1}^{n}\sum_{l=1}^{n} A_{i,k}A_{j,l} f_k f_l. \tag{A3}$$

There are n^2 such products, but only M of them are distinct due to the symmetry $f_k f_l = f_l f_k$. We denote the distinct products by $g_{kl} = f_k f_l$ $(k \le l)$. In terms of these, Eq. (A3) becomes

$$\begin{aligned} g_{ij}' &= \sum_{k=1}^{n} A_{i,k}A_{j,k} g_{kk} + \sum_{k<l} \left(A_{i,k}A_{j,l} + A_{i,l}A_{j,k}\right) g_{kl} \\ &= \sum_{k\le l} \frac{1}{1+\delta_{k,l}} \left(A_{i,k}A_{j,l} + A_{i,l}A_{j,k}\right) g_{kl}. \end{aligned} \tag{A4}$$

If we think of g_{kl} and g_{ij}' as components of column vectors $\mathbf{g}$ and $\mathbf{g}'$, respectively, then Eq. (A4) may be rewritten in matrix form as

$$\mathbf{g}' = \mathbf{C}\mathbf{g}, \tag{A5}$$

where $\mathbf{C} = \mathbf{A} \otimes_s \mathbf{A}$ as defined by Eq. (A1).

The operation $\otimes_s$ defined by Eq. (A1) can be generalized to products of different matrices,

$$(\mathbf{A}\otimes_s\mathbf{B})_{j,kl} = \frac{1}{2(1+\delta_{k,l})}\left(A_{i,k}B_{j,l} + A_{i,l}B_{j,k} + B_{i,k}A_{j,l} + B_{i,l}A_{j,k}\right), \tag{A6}$$

where again $i \le j$ and $k \le l$. Neither $\mathbf{A} \otimes_s \mathbf{B}$ nor even $\mathbf{A} \otimes_s \mathbf{A}$ is a symmetric matrix itself. The name "symmetric Kronecker product" refers to the fact that $\mathbf{A} \otimes_s \mathbf{B} = \mathbf{B} \otimes_s \mathbf{A}$, in contrast to the conventional Kronecker product, which is not commutative.

Our definition of the symmetric Kronecker product $\mathbf{A} \otimes_s \mathbf{A}$ by Eq. (A1) differs from the one found in the literature,[35–38] where the function products are defined as $g_{ij} = \sqrt{2} f_i f_j$ for $i < j$ and $g_{ij} = f_i f_j$ for $i = j$, whereas we set $g_{ij} = f_i f_j$ for all i and j. The factor of $\sqrt{2}$ alters certain rows and columns of the matrix $\mathbf{A} \otimes_s \mathbf{A}$. However, all the properties of Eq. (A1) deduced in Appendix B also hold for the symmetric Kronecker product of Refs. 35–38. One notable exception concerns unitary matrices: in Refs. 35–38, the matrix $\mathbf{A} \otimes_s \mathbf{A}$ is unitary if $\mathbf{A}$ is unitary, whereas our definition of $\mathbf{A} \otimes_s \mathbf{A}$ by Eq. (A1) lacks this property. This is consistent with the fact that the Gram determinant of the matrix $\mathbf{W}'$ of Eq. (21) is not invariant under unitary transformations of the functions $\{f_i\}$.

APPENDIX B: DETERMINANT OF THE SYMMETRIC KRONECKER PRODUCT

The following property of the symmetric Kronecker product is shared by all of its variants.

Theorem 1. If $\mathbf{A}$ is an $n \times n$ matrix, then

$$\det\left(\mathbf{A}\otimes_s\mathbf{A}\right) = \left(\det\,\mathbf{A}\right)^{n+1}. \tag{B1}$$

The proof presented below is more general than the one given in Ref. 10, where we assumed that the matrix $\mathbf{A}$ has n eigenvectors. We begin by establishing three preliminary results.

Lemma 1. If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices, then

$$(\mathbf{AB})\otimes_s(\mathbf{AB}) = (\mathbf{A}\otimes_s\mathbf{A})(\mathbf{B}\otimes_s\mathbf{B}). \tag{B2}$$

Proof. The proof is by direct manipulation. According to Eq. (A1), the element (ij, kl) of the $M \times M$ matrix on the left-hand

18 June 2024 15:18:30

Published under an exclusive license by AIP Publishing

side of Eq. (B2) is

$$\begin{aligned}
&[(\mathbf{AB})\otimes_s(\mathbf{AB})]_{ij,kl}\\
&= \frac{1}{1+\delta_{k,l}}\left[(\mathbf{AB})_{i,k}(\mathbf{AB})_{j,l} + (\mathbf{AB})_{i,l}(\mathbf{AB})_{j,k}\right]\\
&= \frac{1}{1+\delta_{k,l}}\sum_{p=1}^{n}\sum_{q=1}^{n}\left(A_{i,p}B_{p,k}A_{j,q}B_{q,l} + A_{i,q}B_{q,l}A_{j,p}B_{p,k}\right)\\
&= \frac{1}{1+\delta_{k,l}}\sum_{p=1}^{n}\sum_{q=1}^{n}\left(A_{i,p}A_{j,q} + A_{i,q}A_{j,p}\right)B_{p,k}B_{q,l}\\
&= \sum_{p\le q}\frac{\left(A_{i,p}A_{j,q} + A_{i,q}A_{j,p}\right)}{1+\delta_{p,q}}\frac{\left(B_{p,k}B_{q,l} + B_{p,l}B_{q,k}\right)}{1+\delta_{k,l}}\\
&= \sum_{pq}(\mathbf{A}\otimes_s\mathbf{A})_{ij,pq}(\mathbf{B}\otimes_s\mathbf{B})_{pq,kl}\\
&= [(\mathbf{A}\otimes_s\mathbf{A})(\mathbf{B}\otimes_s\mathbf{B})]_{ij,kl}. \qquad \text{(B3)}
\end{aligned}$$

This is true for every matrix element of $(\mathbf{AB})\otimes_s(\mathbf{AB})$ and hence proves Eq. (B2).

Lemma 2. If $\mathbf{A}$ is non-singular, then

$$(\mathbf{A}\otimes_s\mathbf{A})^{-1} = \mathbf{A}^{-1}\otimes_s\mathbf{A}^{-1}. \qquad \text{(B4)}$$

Proof. This follows from Eq. (B2) if we set $\mathbf{B} = \mathbf{A}^{-1}$ and observe that

$$(\mathbf{A}\otimes_s\mathbf{A})(\mathbf{A}^{-1}\otimes_s\mathbf{A}^{-1}) = \mathbf{I}_n\otimes_s\mathbf{I}_n = \mathbf{I}_M, \qquad \text{(B5)}$$

where $\mathbf{I}_M$ is an identity matrix of order M.

Lemma 3. If $\mathbf{A}$ is an upper triangular matrix, then $\mathbf{C} = \mathbf{A}\otimes_s\mathbf{A}$ is also upper triangular. Moreover, the diagonal elements of $\mathbf{C}$ are given by

$$C_{ij,ij} = A_{i,i}A_{j,j}. \qquad \text{(B6)}$$

Proof. All elements $C_{ij,kl}$ below the main diagonal have either $i > k$ or $j > l$ and, simultaneously, either $i > l$ or $j > k$. By Eq. (A1), all such elements are zero if the matrix $\mathbf{A}$ is upper triangular. The second part of the lemma, Eq. (B6), follows from Eq. (A1) and the fact that in an upper triangular matrix, all $A_{j,i} = 0$ for $j > i$.

Now, we are ready to prove Theorem 1. The argument is similar to the proof of Stephanos's theorem for eigenvalues of ordinary Kronecker products.[39]

Proof of Theorem 1. By Schur's unitary triangularization theorem,[40] for every square matrix $\mathbf{A}$ there exists a unitary matrix $\mathbf{U}$, such that

$$\mathbf{T} = \mathbf{U}^{-1}\mathbf{AU} \qquad \text{(B7)}$$

is upper triangular. Neither $\mathbf{U}$ nor $\mathbf{T}$ is unique, but this is immaterial. Clearly,

$$\det\mathbf{T} = \prod_{i=1}^{n} T_{i,i}. \qquad \text{(B8)}$$

Now, by Lemma 1 and Lemma 2,

$$\begin{aligned}
\mathbf{T}\otimes_s\mathbf{T} &= (\mathbf{U}^{-1}\mathbf{AU})\otimes_s(\mathbf{U}^{-1}\mathbf{AU})\\
&= (\mathbf{U}^{-1}\otimes_s\mathbf{U}^{-1})(\mathbf{A}\otimes_s\mathbf{A})(\mathbf{U}\otimes_s\mathbf{U})\\
&= (\mathbf{U}\otimes_s\mathbf{U})^{-1}(\mathbf{A}\otimes_s\mathbf{A})(\mathbf{U}\otimes_s\mathbf{U}). \qquad \text{(B9)}
\end{aligned}$$

This shows that the matrices $\mathbf{A}\otimes_s\mathbf{A}$ and $\mathbf{T}\otimes_s\mathbf{T}$ are similar, and, therefore,

$$\det(\mathbf{A}\otimes_s\mathbf{A}) = \det(\mathbf{T}\otimes_s\mathbf{T}). \qquad \text{(B10)}$$

By Lemma 3, the matrix $\mathbf{T}\otimes_s\mathbf{T}$ is upper triangular, so

$$\begin{aligned}
\det(\mathbf{T}\otimes_s\mathbf{T}) &= \prod_{i\le j}^{M}(\mathbf{T}\otimes_s\mathbf{T})_{ij,ij} = \prod_{i=1}^{n}\prod_{j\ge i}^{n} T_{i,i}T_{j,j}\\
&= \left(\prod_{i=1}^{n} T_{i,i}\right)^{n+1} = (\det\mathbf{T})^{n+1}. \qquad \text{(B11)}
\end{aligned}$$

Since $\det\mathbf{T} = \det\mathbf{A}$, Eqs. (B10) and (B11) imply Eq. (B1). Note that instead of Schur's decomposition of $\mathbf{A}$, one could have relied on the Jordan decomposition, which has the same form as Eq. (B7) except that the matrix $\mathbf{U}$ is not unitary.

REFERENCES

[1] J. E. Harriman, "Geometry of density matrices. IV. The relationship between density matrices and densities," Phys. Rev. A **27**, 632–645 (1983).
[2] J. E. Harriman, "Densities, operators, and basis sets," Phys. Rev. A **34**, 29–39 (1986).
[3] J. E. Harriman, "Electron densities, momentum densities, and density matrices," Z. Naturforsch. A **48**, 203–210 (1993).
[4] W. H. E. Schwarz and B. Müller, "Density matrices from densities," Chem. Phys. Lett. **166**, 621–626 (1990).
[5] W. H. E. Schwarz, A. Langenbach, and L. Birlenbach, "Density matrices from position and momentum densities," Theor. Chim. Acta **88**, 437–445 (1994).
[6] G. T. Smith and H. L. Schmider, "Linearly dependent subspaces and the eigenvalue spectrum of the one-particle reduced density matrix," J. Mol. Struct.: THEOCHEM **527**, 181–191 (2000).
[7] G. N. Sizov and V. N. Staroverov, "Analytic construction of one-electron reduced density matrices from electron densities within finite basis dets," J. Chem. Theory Comput. (published online) (2024).
[8] R. C. Morrison, "Density and density matrix from optimized linearly independent product basis functions for Be," Int. J. Quantum Chem. Symp. **22**, 43–49 (1988).
[9] Y. Oueis and V. N. Staroverov, "Reconstruction of exchange-correlation potentials from their matrix representations," J. Chem. Theory Comput. **18**, 6092–6098 (2022).
[10] Y. Oueis and V. N. Staroverov, "Multiplicative potentials for kinetic energy and exact exchange," J. Chem. Phys. **157**, 204107 (2022).
[11] Y. Oueis, G. N. Sizov, and V. N. Staroverov, "Local potentials reconstructed within linearly independent product basis sets of increasing size," J. Phys. Chem. A **127**, 2664–2669 (2023).
[12] D. E. Hoch and J. E. Harriman, "Linear dependencies among basis set products and near locality of some nonlocal operators," J. Chem. Phys. **102**, 9590–9597 (1995).
[13] D. E. Hoch and J. E. Harriman, "Basis set dependence of the locality of the kinetic energy operator," J. Chem. Phys. **104**, 5898–5911 (1996).
[14] J. E. Harriman and D. E. Hoch, "Locality of exchange matrices for common Gaussian basis sets," Int. J. Quantum Chem. **63**, 111–119 (1997).

18 June 2024 15:18:30

Published under an exclusive license by AIP Publishing

[15]B. Klahn and W. A. Bingel, "Completeness and linear independence of basis sets used in quantum chemistry," Int. J. Quantum Chem. **11**, 943–957 (1977).
[16]G. Sansone, *Orthogonal Functions* (Dover, New York, 1991).
[17]R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1953), Vol. 1.
[18]P.-O. Löwdin, "On the nonorthogonality problem," Adv. Quantum Chem. **5**, 185–199 (1970).
[19]H. Tuy, *Convex Analysis and Global Optimization*, 2nd ed. (Springer, New York, 2016).
[20]D. Jibetean and E. de Klerk, "Global optimization of rational functions: A semidefinite programming approach," Math. Program. **106**, 93–109 (2006).
[21]J. B. Lasserre, V. Magron, S. Marx, and O. Zahm, "Minimizing rational functions: A hierarchy of approximations via pushforward measures," SIAM J. Optim. **31**, 2285–2306 (2021).
[22]P.-O. Löwdin, "Quantum theory of cohesive properties of solids," Adv. Phys. **5**, 1–171 (1956).
[23]V. N. Staroverov, "Noninteracting *v*-representable subspaces of orbitals in the Kohn–Sham method," J. Phys. Chem. Lett. **13**, 6839–6844 (2022).
[24]H. T. Schwarz and J. Waldvogel, *Numerical Analysis: A Comprehensive Introduction* (John Wiley & Sons, Chichester, 1989).
[25]T. Schlick, "Optimization methods in computational chemistry," in *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz and D. B. Boyd (VCH Publishers, New York, 1992), Vol. 3, pp. 1–71.
[26]Maple, *Maplesoft, a division of Waterloo* (Maple Inc., Waterloo, Ontario, 2022).
[27]W. H. Greub, *Linear Algebra*, 3rd ed. (Springer-Verlag, Berlin, 1967).
[28]M. A. Akivis and V. V. Goldberg, *An Introduction to Linear Algebra and Tensors* (Dover, Mineola, 1977).
[29]J. M. Foster and S. F. Boys, "Canonical configurational interaction procedure," Rev. Mod. Phys. **32**, 300–302 (1960).
[30]C. Edmiston and K. Ruedenberg, "Localized atomic and molecular orbitals," Rev. Mod. Phys. **35**, 457–465 (1963).
[31]W. von Niessen, "Density localization of atomic and molecular orbitals. I," J. Chem. Phys. **56**, 4290–4297 (1972).
[32]J. Pipek and P. G. Mezey, "A fast intrinsic localization procedure applicable for *ab initio* and semiempirical linear combination of atomic orbital wave functions," J. Chem. Phys. **90**, 4916–4926 (1989).
[33]S. Lehtola and H. Jónsson, "Unitary optimization of localized molecular orbitals," J. Chem. Theory Comput. **9**, 5365–5372 (2013).
[34]G. Del Re, "The non-orthogonality problem and orthogonalization procedures," in *Quantum Science Methods and Structure: A Tribute to Per-Olov Löwdin*, edited by J.-L. Calais, O. Goscinski, J. Linderberg and Y. Öhrn (Plenum Press, New York, 1976), pp. 53–75.
[35]F. D. Murnaghan, *The Theory of Group Representations* (The Johns Hopkins Press, Baltimore, 1938).
[36]E. de Klerk, *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications* (Kluwer, New York, 2002).
[37]M. J. Todd, K. C. Toh, and R. H. Tütüncü, "On the Nesterov–Todd direction in semidefinite programming," SIAM J. Optim. **8**, 769–796 (1998).
[38]F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton, "Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results," SIAM J. Optim. **8**, 746–768 (1998).
[39]P. Lancaster and M. Tismenetsky, *The Theory of Matrices with Applications*, 2nd ed. (Academic Press, San Diego, 1985).
[40]R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1985).

18 June 2024 15:18:30

Published under an exclusive license by AIP Publishing