

5 June 2026

Is a bit over okay?—Overcomplete sets of localized molecular orbitals

Ágnes Szabados¹, János Daru², András Gombás¹, Péter R. Surján¹, Agnes Szabados

1. Institute of Chemistry, Faculty of Science, ELTE Eötvös Loránd University

2. Institute of Chemistry, Faculty of Science, ELTE Eötvös Loránd University

Abstract

Quantum chemical bases are mostly linearly independent, whether orthonormal or not. However, in certain applications, using an overcomplete set of functions, not forming a basis in the strict sense offers distinct advantages. Localized molecular orbitals (LMO) projected from atomic orbitals (PAO) are a primary example. Use of redundant LMOs is facilitated by the mathematical theory of frames. Along with a systematic presentation of the necessary background, it is shown that the trace of one-electron operators can be obtained by a formula analogous to the orthonormal case. For instance, the electronic dipole moment can be expressed as a sum of frame LMO centroids multiplied by their fractional populations as weights. An advantage of frame LMOs is demonstrated when following reaction coordinates. When a system passes through a high-symmetry point between lower-symmetry arrangements, orthogonal LMOs often exhibit abrupt changes. This occurs when a specific set of localized orbitals does not represent the molecular point group of the highsymmetry structure, resulting in equivalent sets of orthogonal LMOs that may induce discontinuities in LMO based quantities. It is shown that PAO related frame LMOs transition smoothly in these scenarios, without any extra algorithmic effort.

Is a bit over okay? – Overcomplete sets of localized molecular orbitals

Ágnes Szabados ^{*,†} János Daru ^{*,‡} András Gombás ^{*,†,¶} and Péter R. Surján ^{*,†}

[†]*Laboratory of Theoretical Chemistry, Institute of Chemistry, Faculty of Science, ELTE Eötvös Loránd University, H-1518 Budapest 112, P.O.B. 32, Hungary*

[‡]*Institute of Chemistry, Faculty of Science, ELTE Eötvös Loránd University, H-1518 Budapest 112, P.O.B. 32, Hungary*

[¶]*György Hevesy PhD School of Chemistry, Institute of Chemistry, Faculty of Science, ELTE Eötvös Loránd University, H-1518 Budapest 112, P.O.B. 32, Hungary*

E-mail: agnes.szabados@ttk.elte.hu; janos.daru@ttk.elte.hu; gombasandras@student.elte.hu; peter.surjan@ttk.elte.hu

Abstract

Quantum chemical bases are mostly linearly independent, whether orthonormal or not. However, in certain applications, using an overcomplete set of functions, not forming a basis in the strict sense offers distinct advantages. Localized molecular orbitals (LMO) projected from atomic orbitals (PAO) are a primary example.

Use of redundant LMOs is facilitated by the mathematical theory of frames. Along with a systematic presentation of the necessary background, it is shown that the trace of one-electron operators can be obtained by a formula analogous to the orthonormal case. For instance, the electronic dipole moment can be expressed as a sum of frame LMO centroids multiplied by their fractional populations as weights.

An advantage of frame LMOs is demonstrated when following reaction coordinates. When a system passes through a high-symmetry point between lower-symmetry arrangements, orthogonal LMOs often exhibit abrupt changes. This occurs when a specific set of localized orbitals does not represent the molecular point group of the high-symmetry structure, resulting in equivalent sets of orthogonal LMOs that may induce discontinuities in LMO based quantities. It is shown that PAO related frame LMOs transition smoothly in these scenarios, without any extra algorithmic effort.

1 Introduction

Apart from fully numerical approximation schemes, most quantum chemical (QC) approaches are based on a set of one-electron functions. Considering a set of N vectors $\{\varphi_\mu\}_{\mu=1}^N$ expanded on an auxiliary orthonormal basis $\{e_i\}_{i=1}^N$ as

$$|\varphi_\mu\rangle = \sum_{i=1}^N |e_i\rangle x_{i\mu} , \quad (1)$$

expansion coefficients $x_{i\mu}$ can be collected into the nonhermitian square matrix $\mathbf{X}$, possessing the singular value decomposition (SVD)

$$\mathbf{X} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\dagger , \quad (2)$$

where $\mathbf{\Sigma}$ is a diagonal matrix and matrices $\mathbf{U}$ and $\mathbf{V}$ are unitary. When vectors $\{\varphi_\mu\}_{\mu=1}^N$ are linearly independent, σ_μ entries in the diagonal of $\mathbf{\Sigma}$ are nonzero, and the set $\{\varphi_\mu\}_{\mu=1}^N$ is considered a basis.

The overlap matrix of $\{\varphi_\mu\}$ defined as

$$S_{\mu\nu} = \langle \varphi_\mu | \varphi_\nu \rangle$$

can be expressed as $\mathbf{S} = \mathbf{X}^\dagger \mathbf{X}$, admitting the eigenvalue decomposition

$$\mathbf{S} = \mathbf{V} \mathbf{\Sigma}^2 \mathbf{V}^\dagger . \quad (3)$$

Diagonal entries σ_μ^2 of $\mathbf{\Sigma}^2$ are also nonzero when $\{\varphi_\mu\}_{\mu=1}^N$ represents a basis.

Nonorthogonal bases, characterized by nonunitary $\mathbf{X}$ and nonunit $\mathbf{S}$, are routinely handled in QC practice, either by explicit treatment of the overlap or by applying orthonormalization. Of the several possibilities for the latter,¹⁻⁵ we here consider Löwdin's symmetrical scheme⁶

$$|\varphi_\mu^L\rangle = \sum_{\nu=1}^N |\varphi_\nu\rangle S_{\nu\mu}^{-1/2} \quad (4)$$

(with a shorthand for the elements of matrix power $\mathbf{S}^{-1/2}$), yielding $\langle\varphi_\mu^L|\varphi_\nu^L\rangle = \delta_{\mu\nu}$.

A central role of a spanning set in a vector space is the expansion of other vectors. This is most compactly achieved by formulating the identity operator of the N -dimensional space

$$\hat{I}_N = \sum_{i=1}^N |e_i\rangle\langle e_i| \quad (5)$$

with the vectors in question. Relying on the set $\{\varphi_\mu\}_{\mu=1}^N$, the identity can be composed as

$$\hat{I}_N = \sum_{\mu\nu=1}^N |\varphi_\mu\rangle S_{\mu\nu}^{-1} \langle\varphi_\nu| , \quad (6)$$

where $\{\varphi_\mu^L\}_{\mu=1}^N$ can be introduced as

$$\hat{I}_N = \sum_{\mu=1}^N |\varphi_\mu^L\rangle\langle\varphi_\mu^L| . \quad (7)$$

Applying e.g. (5), $\hat{I}_N|\varphi_\mu\rangle$ obviously yields $x_{i\mu} = \langle e_i|\varphi_\mu\rangle$ for the coefficients in (1).

In this work a generalization of the above issues for the case of linear dependence, i.e. one or more eigenvalues of $\mathbf{S}$ becoming zero, is explored. This situation is often accompanied by numerical instabilities, requiring algorithmic adjustments to address the pathological behavior.⁷⁻¹⁰ Beyond the field of QC, the use of overcomplete sets is

however common. Pioneered in signal processing, the application of overcomplete sets of functions has spread in diverse disciplines, including mathematics, physics, engineering, and computer science. Elaboration of the strict mathematical background, termed the theory of frames, surged in the 1980s.¹¹ Literature on the subject has grown extensive to date, we point to Refs.^{12–14} as instructive readings.

While the benefit of nonorthogonal bases has been stressed often,^{15–20} the application of redundant sets is less common in QC. Bringing robustness or mitigating data loss are obvious advantages of frames in signal analysis, parallel fruitful applications are however not immediate in our field. A couple of works employing frames in electron correlation theories make use of overcompleteness in preserving symmetry, boosting computational efficiency or retaining unitary invariance via avoiding any filtering of functions.^{21–25}

This paper focuses on a further practical use of overcompleteness, in connection with localized molecular orbitals (LMO). Centers of occupied LMOs are widely applied for introducing a coarse-grained representation of the electron density, e.g. calculating and analyzing dipole moments,²⁶ polarizabilities,²⁷ and the corresponding spectral properties via their time-autocorrelation functions. Furthermore, they have been utilized in the analysis of chemical bonding, estimating oxidation states,²⁸ and deriving dispersion²⁹ or electrostatic models for classical force fields,³⁰ machine learning force fields,³¹ or embeddings.³² In these applications it is instrumental that orbital centers change continuously along a reaction trajectory. Smooth dependence of localized orbitals on nuclear coordinates is not the least important in the virtual subspace, with the added difficulty of converging standard procedures optimizing a localization function.³³

We consider occupied LMOs here due to their role as proxies for simplified reduced models. When propagating a system in time, occupied orthogonal LMOs are prone to exhibit unphysical jumps if the system is subject to oscillations or perturbations around a structure with multiple localization minima. This leads to jumps in calculated properties, such as the interaction energies predicted by machine learning force fields utilizing LMO centers. Discontinuity of LMO trajectories has raised concerns before and got

mitigated by geometric symmetrization,³⁰ by introducing an extra condition of orbital similarity,^{34,35} by modifying the localization constraint in derivative calculation³⁶ or by abandoning the concept of localized orbitals.³⁷ The former treatments may be beset by arbitrary parametrization or additional rotations, while the latter relies on density discretizations. We here demonstrate that relaxing the requirement of linear independence facilitates a smooth transition of the LMOs and LMO centers. Turning to an overcomplete set of LMOs, symmetry related equivalence requires no special attention, relieving the need for any additional formal or algorithmic considerations.

In this work we rely on projected atomic orbitals (PAO), advocated first by Pulay for constructing virtual LMOs.^{38,39} The advantage of projected AOs is the iteration free algorithm producing LMOs. This feature is shared by just a handful of alternative methods, e.g.^{40–42} As emphasized in a recent paper,⁴² it is not only the columns of matrix $\mathbf{PS}$ that allow to formulate LMOs, but also those of the density matrix $\mathbf{P}$. In current computational chemistry, PAOs have become an inherent component of local correlation methods, designed for linear scaling.^{43–50} Use of PAOs in the occupied space has been also documented^{51–54} though less common. The number of native PAOs agreeing with the dimension of the AO basis, application of PAOs mostly involves redundancy filtering.

The possibility of using an overcomplete set of functions in Hartree-Fock and many-body theories, has been considered already in the 1990’s.⁵⁵ First implementations are due to Olsen and coworkers who formulated AO based Coupled-cluster²² and Møller-Plesset²³ techniques by working with the full set of PAOs both in the occupied and the virtual space. Use of frame theory is implicit in these studies, without reference to the terminology of the field. The present work parallels those of Olsen and coworkers in intentionally keeping all PAOs generated. Our aim of working with PAOs is however not computational efficiency but mitigation of a conceptual shortcoming of orthogonal LMOs.

Below, essential notions of frame theory are introduced in Section 2. Following this general, math style presentation, Section 3 discusses various possibilities for redundant

localized molecular orbitals, still from a formal point of view. A practical advantage of overcomplete LMO sets is presented in Section 4.

2 General theory

The material presented below is an excerpt of frame theory. We aim for a concise and precise presentation, focusing on our application. More stress is put on practice and less on mathematical rigor. For strict mathematical formulation we refer to Refs.^{12–14}

2.1 Frames

Allowing set $\{\varphi_\mu\}_{\mu=1}^M$ to be linearly dependent, we consider $M > N$, with N standing for the dimension of the space. Let us define the so-called ***frame-operator*** of this set as

$$\hat{F} = \sum_{\mu=1}^M |\varphi_\mu\rangle\langle\varphi_\mu|. \quad (8)$$

Operator $\hat{F}$ is obviously hermitian, and non-idempotent for overlap effects.

In order for the redundant set $\{\varphi_i\}_{i=1}^M$ be a ***frame***, the associated frame-operator $\hat{F}$ has to satisfy¹² the so-called ***frame condition***

$$A \hat{I}_N \leq \hat{F} \leq B \hat{I}_N, \quad (9)$$

with $A \leq B$ being finite, positive numbers.

Matrix $\mathbf{X}$, introduced in (1) and (2), is now an $N \times M$ rectangular, possessing the singular-value decomposition

$$\mathbf{X} = \mathbf{U} \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{0} \end{pmatrix} \mathbf{V}^\dagger, \quad (10)$$

with $\mathbf{U}$ and $\mathbf{V}$ being of different dimensions: $\mathbf{U}$ is $N \times N$ and $\mathbf{V}$ is $M \times M$. Matrix $\boldsymbol{\Sigma}$ is diagonal $N \times N$, with positive diagonal entries ordered as $0 < \sigma_N \leq \dots \sigma_2 \leq \sigma_1 < \infty$.

Matrix representation of the frame operator (8) on $\{e_i\}_{i=1}^N$ takes the form $\mathbf{F} = \mathbf{X} \mathbf{X}^\dagger$.

The SVD of (10) allows to write

$$\mathbf{F} = \mathbf{U} \mathbf{\Sigma}^2 \mathbf{U}^\dagger, \quad (11)$$

showing that the bounds in (9) are given by $A = \sigma_N^2$ and $B = \sigma_1^2$.

Linear dependence manifests in the eigenvalue decomposition of the overlap reading as

$$\mathbf{S} = \mathbf{X}^\dagger \mathbf{X} = \mathbf{V} \begin{pmatrix} \mathbf{\Sigma}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{V}^\dagger. \quad (12)$$

The $M \times M$ matrix $\mathbf{S}$ is obviously singular, its rank being only N .

Singular nature of the overlap prohibits the usual construction of the Löwdin-orthogonalized (4) set which requires to construct a negative power of $\mathbf{S}$. Eliminating redundancy of $\{\varphi_\mu\}$ is possible e.g. by Gram-Schmidt orthogonalization, Cholesky-decomposition^{2,3} or Löwdin's canonical orthogonalization.⁷ We here pursue the alternative strategy of keeping the linearly dependent set and aiming for well defined expansion of vectors in terms of the M -dimensional set of $\{\varphi_\mu\}$ or a linear transformation of it.

2.1.1 Löwdin-set via the frame-operator

Contrary to the singular overlap matrix, the frame operator $\hat{F}$ is invertible in the redundant case also. This is apparent from its matrix representation (11), showing positive definite property. It therefore serves to construct the Löwdin-set as

$$|\varphi_\mu^{\overline{L}}\rangle = \hat{F}^{-1/2} |\varphi_\mu\rangle, \quad (13)$$

in analogy with (4). Overlined L in superscript reminds that vectors of (13) are not orthonormal, *vide infra*. Expansion of any vector in terms of $\{\varphi_\mu^{\overline{L}}\}$, $\mu = 1, \dots, M$ is

facilitated by writing the N -dimensional identity as

$$\sum_{\mu=1}^M |\varphi_{\mu}^{\bar{L}}\rangle \langle \varphi_{\mu}^{\bar{L}}| = \hat{F}^{-1/2} \hat{F} \hat{F}^{-1/2} = \hat{I}_N. \quad (14)$$

Expression (14) is the analogue of (7). According to the terminology of frame theory, a frame with a frame-operator reproducing the identity is called a **Parseval-frame**.

This work pays special attention to Parseval-frames for their following property. Parseval-frames facilitate expansion of an arbitrary vector ψ as

$$|\psi\rangle \equiv \hat{I}_N |\psi\rangle = \sum_{\mu=1}^M |\varphi_{\mu}^{\bar{L}}\rangle \langle \varphi_{\mu}^{\bar{L}}|\psi\rangle = \sum_{\mu=1}^M |\varphi_{\mu}^{\bar{L}}\rangle y_i^{\bar{L}}, \quad (15)$$

in close analogy with the expansion in an orthonormal basis, in spite of the fact the $\varphi_{\mu}^{\bar{L}}$ are nonorthogonal. Note however, that expansion coefficients $y_i^{\bar{L}} = \langle \varphi_{\mu}^{\bar{L}}|\psi\rangle$ are not uniquely associated with ψ and $\{\varphi_{\mu}^{\bar{L}}\}$, as a consequence of linear dependence. See Appendix A.1 for more detail on this aspect.

Equation (13) gives a recipe for generating a Parseval-frame from an arbitrary frame, we are therefore concerned with this transformation in the following. In QC practice the Löwdin-vectors of (13) are obtained with the help of the overlap matrix, whenever $\mathbf{S}$ is invertible. For singular $\mathbf{S}$ a generalized overlap inverse is considered, as described in the following.

2.1.2 Löwdin-set via the pseudo overlap-inverse

Let us first take the Moore-Penrose (MP) *pseudo inverse*⁵⁶ of the $N \times M$ matrix $\mathbf{X}$ of (10), defined in SVD form as

$$\bar{\mathbf{X}} = \mathbf{V} \begin{pmatrix} \Sigma^{-1} \\ \mathbf{0} \end{pmatrix} \mathbf{U}^{\dagger}. \quad (16)$$

While $\bar{\mathbf{X}}\mathbf{X} \neq \mathbf{I}$ for the MP pseudo inverse, the following four fundamental properties

$$\bar{\mathbf{X}}\mathbf{X}\bar{\mathbf{X}} = \bar{\mathbf{X}} \quad (17a)$$

$$\mathbf{X}\bar{\mathbf{X}}\mathbf{X} = \mathbf{X} \quad (17b)$$

$$\mathbf{X}\bar{\mathbf{X}} = \bar{\mathbf{X}}^\dagger \mathbf{X}^\dagger \quad (17c)$$

$$\bar{\mathbf{X}}\mathbf{X} = \mathbf{X}^\dagger \bar{\mathbf{X}}^\dagger \quad (17d)$$

are straightforward to verify, based on (10) and (16). The MP inverse of the overlap, $\bar{\mathbf{S}}$ is obtained next in the form

$$\bar{\mathbf{S}} = \bar{\mathbf{X}}\bar{\mathbf{X}}^\dagger = \mathbf{V} \begin{pmatrix} \boldsymbol{\Sigma}^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{V}^\dagger \quad (18)$$

fulfilling relations (17), with obvious substitutions of $\bar{\mathbf{S}}$ and $\mathbf{S}$.

With expression (18) of the generalized overlap-inverse it is easy to prove that $\{\varphi_\mu^{\bar{L}}\}$ of (13) arise as

$$|\varphi_\mu^{\bar{L}}\rangle = \sum_{\nu=1}^M |\varphi_\nu\rangle \bar{S}_{\nu\mu}^{1/2} . \quad (19)$$

The important difference between $\{\varphi_\mu^L\}$ of the linearly independent and $\{\varphi_\mu^{\bar{L}}\}$ of the overcomplete case is to be emphasized here. While (4) gives an orthonormal set, the overlap of vectors in (19) is non-unit. The overlap matrix of $\{\varphi_\mu^{\bar{L}}\}$ is an M -dimensional representation of the N -dimensional unit, according to

$$\langle \varphi_\mu^{\bar{L}} | \varphi_\nu^{\bar{L}} \rangle = \left(\mathbf{V} \begin{pmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{V}^\dagger \right)_{\mu\nu} \quad (20a)$$

$$= \sum_{\lambda=1}^N V_{\mu\lambda} V_{\nu\lambda}^* \neq \delta_{\mu\nu} . \quad (20b)$$

In the orthonormal case, unitary nature of $\mathbf{V}$ allows to write $\delta_{\mu\nu}$ on the rhs of the above, since $M = N$ holds then (note that $\mathbf{V}$ is $M \times M$ unitary).

2.1.3 Löwdin-sets via the reflexive overlap-inverse

If substituting the zero block in the middle factor in (16) by an arbitrary $(M - N) \times N$ matrix Ξ , according to

$$\bar{\bar{\mathbf{X}}} = \mathbf{V} \begin{pmatrix} \Sigma^{-1} \\ \Xi \end{pmatrix} \mathbf{U}^\dagger, \quad (21)$$

the arising $\bar{\bar{\mathbf{X}}}$ is a weaker inverse than the MP pseudo inverse, fulfilling criteria (17a)-(17c) listed in Section 2.1.2, but failing in (17d).

In parallel with (18) a generalized inverse can be composed for $\mathbf{S}$ as

$$\bar{\bar{\mathbf{S}}} = \bar{\bar{\mathbf{X}}} \bar{\bar{\mathbf{X}}}^\dagger = \mathbf{V} \begin{pmatrix} \Sigma^{-2} & \Sigma^{-1} \Xi^\dagger \\ \Xi \Sigma^{-1} & \Xi \Xi^\dagger \end{pmatrix} \mathbf{V}^\dagger. \quad (22)$$

It is easy to check that the $\mathbf{S}$ and $\bar{\bar{\mathbf{S}}}$ pair fulfills criteria (17a)-(17b), but violates (17c)-(17d) for a general Ξ . This is an even weaker generalized inverse, for which the *reflexive inverse* terminology⁵⁶ is adopted here.

Let us examine the Löwdin-set obtained via the reflexive overlap-inverse (22) as

$$|\varphi_\mu^{\bar{\bar{L}}}\rangle = \sum_{\nu=1}^M |\varphi_\nu\rangle \bar{\bar{S}}_{\nu\mu}^{1/2}. \quad (23)$$

The frame-operator of $\{\varphi_\mu^{\bar{\bar{L}}}\}$ in (23) reads

$$\sum_{\mu=1}^M |\varphi_\mu^{\bar{\bar{L}}}\rangle \langle \varphi_\mu^{\bar{\bar{L}}}| = \hat{I}_N. \quad (24)$$

To see that the operator in (24) is indeed unit, we can rely on $\bar{\bar{\mathbf{X}}} \bar{\bar{\mathbf{S}}} \bar{\bar{\mathbf{X}}}^\dagger = \mathbf{I}_N$, an obvious consequence of (10) and (22). The set $\{\varphi_\mu^{\bar{\bar{L}}}\}$ of (23) therefore represents a Parseval-frame. Flexibility in the reflexive inverse $\bar{\bar{\mathbf{S}}}$, originating in the arbitrariness of Ξ , leads to distinct, different Parseval-frames, obtained via (23). In spite of being nonorthonormal, each of these sets facilitate an expansion of the unit in the form of (24). The Löwdin-set obtained with the MP pseudo inverse corresponds to $\Xi = \mathbf{0}$. The MP inverse is the

only set related to the original frame by the inverse square root of the frame-operator according to (13). This set also fulfills a parallel of Löwdin's resemblance theorem,^{57–59} in that the Parseval-frame closest to the original frame in sum of difference squared norm sense is the Löwdin-frame as generated by the MP pseudo inverse.⁶⁰

2.2 Trace formulae with frames

We now examine the trace of a Hermitian operator $\hat{T}$ in the N -dimensional space expressed with the orthonormal basis as

$$\text{Tr } \hat{T} = \sum_{i=1}^N \langle e_i | \hat{T} | e_i \rangle . \quad (25)$$

Let us insert the unit in the form

$$\hat{I}_N = \sum_{\mu\nu=1}^M |\varphi_\mu\rangle \bar{S}_{\mu\nu} \langle \varphi_\nu| \quad (26)$$

before and after $\hat{T}$ in (25) to get

$$\text{Tr } \hat{T} = \sum_{\mu\nu\lambda\sigma=1}^M \bar{S}_{\mu\nu} \langle \varphi_\nu | \hat{T} | \varphi_\lambda \rangle \bar{S}_{\lambda\sigma} \underbrace{\sum_{i=1}^N \langle \varphi_\sigma | e_i \rangle \langle e_i | \varphi_\mu \rangle}_{S_{\sigma\mu}} \quad (27)$$

Using $\bar{S}S\bar{S} = \bar{S}$, c.f. (17a) the above simplifies to

$$\text{Tr } \hat{T} = \sum_{\nu\lambda=1}^M \langle \varphi_\nu | \hat{T} | \varphi_\lambda \rangle \bar{S}_{\lambda\nu} .$$

Factorizing $\bar{S}$ as $\bar{S}^{1/2} \bar{S}^{1/2}$ we can write the above with the Löwdin-vectors as

$$\text{Tr } \hat{T} = \sum_{\mu=1}^M \langle \varphi_\mu^L | \hat{T} | \varphi_\mu^L \rangle . \quad (28)$$

Formula (28) is valid for the vectors obtained with the MP pseudo inverse $\bar{S}$ as well as those generated by the weaker, reflexive inverse $\bar{\bar{S}}$. Equivalence of (25) and (28) is an

important result of frame theory, often exploited in the forthcoming Sections.

The above derivation highlights key elements of the theory to be used later: expression (26) of the unit allowing to introduce $\varphi_{\mu}^{\bar{L}}$, as well as properties (17a)-(17b) of the generalized overlap-inverse. With this background we now proceed to examine some particular sets of LMOs.

3 Localized molecular orbitals as frames

Localized molecular orbitals in the occupied space represent a solution of the generalized self-consistent field (SCF) equations. Even if expanded on the basis of M atomic orbitals (AO), they span a subspace of the dimension of the occupied space, denoted as N_o , with $N_o \leq M$ generally. Localization algorithms, e.g. Foster-Boys,⁶¹ Edmiston-Ruedenberg^{62,63} or Pipek-Mezey (PM)⁶⁴ often supply exactly N_o localized orbitals, by seeking unitary transformations within the occupied space. There exist schemes providing more orbitals than the dimension of the space, e.g. projected AOs advocated by Pulay as localized virtuals.³⁸ The recently surveyed LMOs defined by the columns of the density matrix (P-LMOS) or its square root,⁴² also belong to this category. Below we examine some M -dimensional Parseval-frames derived from PAOs. Occupied LMOs are considered in the discussion, but the statements apply to the set of virtual LMOs as well, *mutatis mutandis*.

We consider $\{\chi_{\mu}\}$ as AOs and $\{\psi_i\}$ the N_o -dimensional set of orthonormal canonical MOs, expanded on AOs as

$$|\psi_i\rangle = \sum_{\mu=1}^M |\chi_{\mu}\rangle C_{\mu i} . \quad (29)$$

Equation

$$\mathbf{C}^{\dagger} \mathbf{S} \mathbf{C} = \mathbf{I}_{N_o} \quad (30)$$

is a consequence of the orthonormal property of $\{\psi_i\}$, with $\mathbf{S}$ denoting the AO overlap

matrix. We are working in the N_o -dimensional space, with the identity reading as

$$\hat{I}_{N_o} = \sum_{i=1}^{N_o} |\psi_i\rangle\langle\psi_i| = \sum_{\mu\nu=1}^M |\chi_\mu\rangle P_{\mu\nu} \langle\chi_\nu| , \quad (31)$$

and

$$P_{\mu\nu} = \sum_{i=1}^{N_o} C_{\mu i} C_{\nu i}^*$$

standing for the elements of the SCF density matrix, $\mathbf{P}$. Matrix $\mathbf{C}$ collecting $C_{\mu i}$ is considered to be of dimension $M \times N_o$, i.e. columns belonging to virtual canonical MOs are omitted. This allows to write

$$\mathbf{P} = \mathbf{C}\mathbf{C}^\dagger . \quad (32)$$

In the following we examine some LMO sets from the frame point of view. We focus on the concepts necessary for our applications and defer to Appendices some essential notions, with less practical relevance. In particular, it is in Appendix A.2 that we give quantities $\mathbf{X}$, $\mathbf{F}$ and $\overline{\overline{\mathbf{X}}}$ introduced in Section 2 for the LMOs considered below.

3.1 Projected AOs: PS-LMOs

Traditional projected AOs emerge by letting identity (31) of the occupied space act on AOs, yielding

$$|\phi_\mu^{PS}\rangle = \sum_{\nu\lambda=1}^M |\chi_\nu\rangle P_{\nu\lambda} \langle\chi_\lambda|\chi_\mu\rangle = \sum_{\nu=1}^M |\chi_\nu\rangle (\mathbf{PS})_{\nu\mu} , \quad (33)$$

justifying the PS-LMO nomenclature. The set of $\{\phi_\mu^{PS}\}$ represents a frame, as shown in Appendix A.2.

Elements of the overlap of PS-LMOs are obtained by

$$\langle\phi_\mu^{PS}|\phi_\nu^{PS}\rangle = \sum_{\lambda\sigma} (\mathbf{SP})_{\mu\lambda} \langle\chi_\lambda|\chi_\sigma\rangle (\mathbf{PS})_{\sigma\nu} = (\mathbf{SPSPS})_{\mu\nu} = (\mathbf{SPS})_{\mu\nu}$$

where

$$\mathbf{PSP} = \mathbf{P} \quad (34)$$

has been used as a consequence of (30) and (32), reflecting the idempotency of matrix $\mathbf{PS}$.

In the spirit of Section 2.1.2, we could proceed by taking the generalized MP inverse of $\mathbf{SPS}$. A reflexive inverse is however sufficient for our purpose and in the case of $\mathbf{SPS}$, a suitable choice is provided by matrix $\mathbf{P}$. By ample use of (34), we find that

$$\mathbf{P SPS P} = \mathbf{P} \quad (35a)$$

$$\mathbf{SPS P SPS} = \mathbf{SPS} \quad (35b)$$

both hold, allowing to sidestep numerical pseudo inversion. Reflexive inverse relations (35) represent the basis of the biorthogonal approach worked out in Ref.²³ With the notation of Section 2.1.3, we can summarize (35) as

$$\overline{\overline{\mathbf{SPS}}} = \mathbf{P} .$$

Presently we profit of this relation for writing the unit of the occupied subspace as

$$\hat{I}_{N_o} = \sum_{\mu\nu=1}^M |\phi_\mu^{PS}\rangle P_{\mu\nu} \langle \phi_\nu^{PS}| . \quad (36)$$

The double sum on the rhs of (36) can be reduced to a single sum by introducing the Löwdin-set of $\{\phi_\mu^{PS}\}$, written as

$$|\phi_\mu^{PS-L}\rangle = \sum_{\nu=1}^M |\chi_\nu^{PS}\rangle P_{\nu\mu}^{1/2} = \sum_{\nu=1}^M |\chi_\nu\rangle (\mathbf{PSP}^{1/2})_{\nu\mu} , \quad (37)$$

where the square root of the reflexive overlap inverse and (33) have been used. This leads to

$$\hat{I}_{N_o} = \sum_{\mu=1}^M |\phi_\mu^{PS-L}\rangle \langle \phi_\mu^{PS-L}| , \quad (38)$$

which is a frame-operator of the Parseval-frame $\{\phi_\mu^{PS-L}\}$. Since $\mathbf{P}$ is positive semidefinite, its square root, needed for the Löwdin-set in (37), surely exists. In fact, with the SVD of $\mathbf{C}$ at hand, $\mathbf{P}^{1/2}$ can be readily constructed based on $\mathbf{P} = \mathbf{C}\mathbf{C}^\dagger$. This point is elaborated further in Appendix A.3.

3.2 P-LMOs

Let us examine P-LMOs defined as

$$|\phi_\mu^P\rangle = \sum_{\nu=1}^M |\chi_\nu\rangle P_{\nu\mu} . \quad (39)$$

As seen in Appendix A.2, $\mathbf{X}$ matrices of $\{\phi_\mu^P\}$ and $\{\phi_\mu^{PS}\}$ fulfill the generalized inverse relation between (10) and (21). It immediately follows that the overlap matrix of $\{\phi_\mu^P\}$ is given by $\mathbf{P}$, as a reflexive inverse of the overlap of $\{\phi_\mu^{PS}\}$. Alternatively, by substitution of (39) and using (34) we see that

$$\langle\phi_\mu^P|\phi_\nu^P\rangle = \sum_{\lambda\sigma} P_{\mu\lambda}\langle\chi_\lambda|\chi_\sigma\rangle P_{\sigma\nu} = (\mathbf{PSP})_{\mu\nu} = P_{\mu\nu} .$$

With the help of $\overline{\overline{\mathbf{P}}} = \mathbf{SPS}$, the unit of the occupied space is expressed with $\{\phi_\mu^P\}$ as

$$\hat{I}_{N_o} = \sum_{\mu\nu=1}^M |\phi_\mu^P\rangle (\mathbf{SPS})_{\mu\nu} \langle\phi_\nu^P| . \quad (40)$$

We now proceed by transforming $\{\phi_\mu^P\}$ to obtain a Löwdin-set in the form

$$|\phi_\mu^{P-L}\rangle = \sum_{\nu=1}^M |\phi_\nu^P\rangle (\mathbf{SPS})_{\nu\mu}^{1/2} , \quad (41)$$

facilitating expression

$$\hat{I}_{N_o} = \sum_{\mu=1}^M |\phi_\mu^{P-L}\rangle \langle\phi_\mu^{P-L}| \quad (42)$$

of the unit of the occupied subspace. Square root of $\mathbf{SPS}$ is expected to exist based on the positive semidefinite property of this matrix.

Obviously, $\{\phi_\mu^{P-L}\}$ and $\{\phi_\mu^{PS-L}\}$ are two different Parseval-frames spanning the occupied subspace. We shall be investigating these two frames in our applications. It is instructive to compare the two frames by relating expressions (38) and (42) to the rhs of (31) leading to

$$\sum_{\mu=1}^M |\phi_\mu^{PS-L}\rangle \langle \phi_\mu^{PS-L}| = \sum_{\mu\nu\lambda=1}^M |\chi_\mu\rangle (\mathbf{PSP}^{1/2})_{\mu\nu} (\mathbf{P}^{1/2}\mathbf{SP})_{\nu\lambda} \langle \chi_\lambda| \quad (43)$$

and

$$\sum_{\mu=1}^M |\phi_\mu^{P-L}\rangle \langle \phi_\mu^{P-L}| = \sum_{\mu\nu\lambda=1}^M |\chi_\mu\rangle (\mathbf{P}(\mathbf{SPS})^{1/2})_{\mu\nu} ((\mathbf{SPS})^{1/2}\mathbf{P})_{\nu\lambda} \langle \chi_\lambda| , \quad (44)$$

respectively. We see that the difference between (43) and (44) is essentially the way the middle factor $\mathbf{P}$ in (31) is augmented using (34) and decomposed in a symmetric manner.

3.3 Further symmetric decompositions of $\mathbf{P}$

A simple symmetric decomposition of the middle factor in (31) is given by $\mathbf{P} = \mathbf{P}^{1/2}\mathbf{P}^{1/2}$ giving rise to

$$|\phi_\mu^{P^{1/2}}\rangle = \sum_{\nu=1}^M |\chi_\nu\rangle P_{\nu\mu}^{1/2} \quad (45)$$

and allowing to write

$$\hat{I}_{No} = \sum_{\mu\nu=1}^M |\phi_\mu^{P^{1/2}}\rangle \langle \phi_\nu^{P^{1/2}}| ,$$

indicating that the $P^{1/2}$ -LMOs defined in (45) also represent a Parseval-frame of the occupied subspace.

Though not immediately obvious, the set of (45) is not new, $P^{1/2}$ -LMOs in fact agree with PS-L LMOs. To show that, we observe

$$\mathbf{PSP}^{1/2} = \mathbf{P}^{1/2} . \quad (46)$$

One argument for the above is that the effect of $\mathbf{PS}$ is unit when it transforms in the

occupied subspace, which is ensured by the razor property⁴² of $\mathbf{P}^{1/2}$, meaning that transformation by $\mathbf{P}^{1/2}$ eliminates virtual components. A rigorous proof, based on the SVD of $\mathbf{C}$ is given in Appendix A.3. Using (46) on the rhs of (37) we get

$$|\phi_\mu^{PS-L}\rangle = \sum_{\nu=1}^M |\chi_\nu\rangle (\mathbf{PSP}^{1/2})_{\nu\mu} = \sum_{\nu=1}^M |\chi_\nu\rangle P_{\nu\mu}^{1/2} = |\phi_\mu^{P^{1/2}}\rangle. \quad (47)$$

An alternative, symmetric treatment of $\mathbf{P}$ in (31) is offered by the Cholesky-decomposition $\mathbf{P} = \mathbf{LL}^\dagger$, generating a Cholesky-frame with vectors

$$|\phi_\mu^{Ch}\rangle = \sum_{\nu=1}^M |\chi_\nu\rangle L_{\nu\mu}. \quad (48)$$

The M -dimensional Cholesky-frame is however never considered in practice, since zero norm ϕ_μ^{Ch} occur for $\mu > N$ by construction,³ and they are not even computed. For this reason, Cholesky-decomposition of the density matrix is better regarded a redundancy filtering technique than a genuine frame approach.

3.4 Trace formulae with frame LMOs

Dimension of the occupied space, i.e. rank of operator $\hat{I}_{N_o}$ is the simplest trace formula, written with the orthonormal canonical orbitals as

$$\sum_{i=1}^{N_o} \langle \psi_i | \psi_i \rangle = N_o.$$

According to Section 2.2 the above can be rewritten as a trace of the overlap of Parseval-frame LMOs. Taking e.g. Löwdin-transformed PS and P-LMOs, introduced in (37) and (41), respectively, we have

$$\sum_{\mu=1}^M \langle \phi_\mu^{PS-L} | \phi_\mu^{PS-L} \rangle = \sum_{\mu=1}^M \langle \phi_\mu^{P-L} | \phi_\mu^{P-L} \rangle = N_o.$$

A remarkable feature of the above expressions is that N_o is obtained as a sum of M non-negative numbers. Noting that $N_o \leq M$, it suggests to interpret the squared norm

of Parseval-frame LMOs

$$\begin{aligned} n_{\mu}^{PS-L} &= \langle \phi_{\mu}^{PS-L} | \phi_{\mu}^{PS-L} \rangle, \mu = 1, \dots, M \\ n_{\mu}^{P-L} &= \langle \phi_{\mu}^{P-L} | \phi_{\mu}^{P-L} \rangle, \mu = 1, \dots, M \end{aligned} \quad (49)$$

as partial occupation numbers. This is an obvious generalization from the canonical, diagonal case for the nondiagonal overlap matrix of Parseval-frame LMOs.

A visualization of frame LMOs is given in Fig. 1, taking the π -MO space of the allyl anion as a minimalist example. A two-dimensional, orthonormal basis spanning the occupied π -electron space is provided by the highest energy, occupied canonical MOs (CMO) or e.g. Boys LMOs. In both cases the 4 electrons are distributed equally among the two MOs, as indicated by MO populations n in Fig. 1. In case of the Löwdin-set of P- and PS-LMOs, the occupied π -MO space is spanned by three frame LMOs, derived from the projection of the p_z AOs of the three carbon atoms. Frame LMOs can be related to an orthonormal basis according to the general formula of (1). Taking e.g. the CMO basis, expansion of P-L and PS-L LMOs is obtained as

$$|\phi_{\mu}\rangle = \sum_i |\psi_i\rangle X_{i\mu} \quad (50)$$

with the interrelation matrix reading as

$$\mathbf{X} \approx \begin{pmatrix} 0.7 & 0.5 & 0.5 \\ 0.0 & 0.7 & -0.7 \end{pmatrix}. \quad (51)$$

Expansion coefficients of P-L and PS-L LMOs agree for the digits shown in (51), their difference amounts to ± 1 in the second decimal. As (51) and Fig. 1 reflect, one frame LMO, ϕ_1 is simply a downscaled version of the bonding π -CMO, ψ_1 . The other two are linear combinations of the bonding and anti-bonding π -CMOs, showing great resemblance to Boys LMOs. Similarly to ϕ_1 , ϕ_2 and ϕ_3 are not normalized either, as obvious from (51). Their partial occupation number, as obtained from (49) by multiplication by 2 for spin-summation, are approximately 1 for ϕ_1 and 1.5 for ϕ_2

and ϕ_3 , giving a total sum of 4.

This minimalist illustration gives an impression on the nature of P-L and PS-L frame LMOs, but does not exemplify their potential utility. The latter aspect is deferred to Section 4

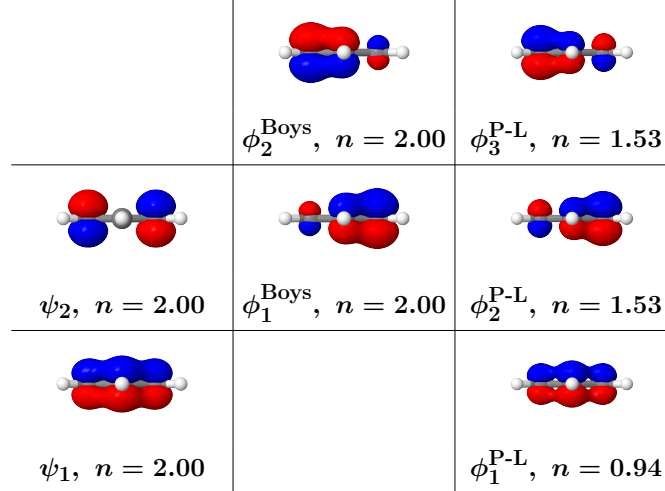


Figure 1: Molecular orbitals of the allyl anion at equilibrium geometry, spanning the occupied π -electron space. Canonical MOs ψ and localized MOs ϕ are drawn with cutoff parameter 0.05. Spin-summed, integer occupation number for orthonormal CMOs and Boys LMOs as well as partial occupation number for the P-L LMOs is denoted by n . Orbital populations and contours of the PS-L set exhibit a qualitatively identical picture to the P-L set, shown.

3.4.1 Dipole moment

We take the example of the electronic dipole moment, evaluated with the Fermi-vacuum as $\langle \text{HF} | \hat{r} | \text{HF} \rangle$, with $\hat{r}$ standing for the coordinate operator, to analyse determinantal expectation value of one-electron operators expressed with frame LMOs.

Again starting from the canonical orbitals' trace formula

$$\langle \text{HF} | \hat{r} | \text{HF} \rangle = \sum_{i=1}^{N_o} \langle \psi_i | \hat{r} | \psi_i \rangle$$

the same can be expressed with Parseval-frame LMOs, e.g. ϕ_μ^{PS-L} as

$$\langle \text{HF} | \hat{r} | \text{HF} \rangle = \sum_{\mu=1}^M \langle \phi_\mu^{PS-L} | \hat{r} | \phi_\mu^{PS-L} \rangle , \quad (52)$$

due to the equivalence of (25) and (28) shown in Section 2.2. The above and further expressions of this Section relating to Parseval-frame LMOs are written with the PS-L set. We suffice to mention that the same apply to P-L LMOs as well as to any other Parseval-frame LMO set. The advantage of Parseval-frames in the terminology of Ref.⁵⁵ is that the nontensorial treatment exhibits the simplicity of the tensorial one, manifesting in the lack of overlap matrices in (52). The reason behind is that the metric of Parseval-frames is nonunit but idempotent, c.f. (20).

Individual terms of the sums in (52) do not have a direct chemical meaning, since the LMOs are not normalized. Introducing the LMO *centroid* as

$$R_\mu^{PS-L} = \frac{\langle \phi_\mu^{PS-L} | \hat{r} | \phi_\mu^{PS-L} \rangle}{\langle \phi_\mu^{PS-L} | \phi_\mu^{PS-L} \rangle} , \quad (53)$$

we can expand each term in (52) by the squared norm of the corresponding LMO, yielding

$$\langle \text{HF} | \hat{r} | \text{HF} \rangle = \sum_{\mu=1}^M \frac{\langle \phi_\mu^{PS-L} | \hat{r} | \phi_\mu^{PS-L} \rangle}{\langle \phi_\mu^{PS-L} | \phi_\mu^{PS-L} \rangle} n_\mu^{PS-L} = \sum_{\mu=1}^M R_\mu^{PS-L} n_\mu^{PS-L} , \quad (54)$$

by taking into account (49). In this last equation the classical expression is recovered, with Parseval-frame LMO centroids and partial occupation numbers.

4 A benefit of redundant LMOs

We employ the Parseval frames of PS-L and P-L LMOs, c.f. (37) and (41) in our application. The PS-L and P-L LMOs are produced based on atomic centers, which is a key point in avoiding equivalent sets of LMOs. While the the need of any ad hoc symmetrization algorithms is relieved this way, one has to handle an overcomplete set

in return. This raises the question whether redundancy is a necessary or avoidable burden in this situation. Let us address this question in two shots.

First, seeking a closest orthonormal (ON) set to a given Parseval frame in a representative example, multiple equivalent closest ON sets can be found at the high-symmetry arrangement. Upon distortion these sets become nonequivalent. The pathological discontinuity therefore manifests along the reaction trajectory with the closest ON orbitals. Based on this finding, relaxing orthogonality appears necessary for obtaining a smooth transition. The more difficult question of relaxing linear independence is addressed in the discussion of the numerical example of Section 4.2.

The distortion of benzene between D_{3h} structures through D_{6h} symmetry is a well known example where standard procedures yield equivalent sets of orthogonal LMOs at the D_{6h} point. Localization extrema can be assigned to the Kekulé-structures of the molecule. We examine a different bond rearrangement in Section 4.2 below, with the similar feature of two equivalent localized minima at the C_{2v} high-symmetry point of the trajectory. Of the three doubly occupied, orthogonal LMOs, two are of the same type and the third is different in the example of Section 4.2. The analogy between the D_{6h} benzene and the C_{2v} semibullvalene of Section 4.2 is that a single set of LMOs corresponding to just one Kekulé-structure does not represent the molecular point group. In such a case, standard localization algorithms are expected to produce multiple equivalent extrema, resulting in LMO discontinuity during trajectory tracking.

4.1 Computation details

While no iteration procedure is associated with PS-L and P-L frames, numerical evaluation of one matrix square-root is required when generating the Löwdin-set from the native LMOs. The former calls for $(\mathbf{SPS})^{1/2}$ while the latter needs $\mathbf{P}^{1/2}$, according to the alternative formula of (47). Both $\mathbf{SPS}$ and $\mathbf{P}$ are of rank N_o and allow for a decomposition as a product of an $M \times N$ matrix and its adjoint, c.f. (A.5) and (A.6). This facilitates a computation cost of the square root scaling as $\mathcal{O}(M^2N)$ via the thin SVD of the factors.⁶⁵ This is to be compared with the $\mathcal{O}(M^3)$ procedure characterizing the

square root evaluation of full rank $M \times M$ matrices, with the help of iterative matrix multiplication algorithms.⁶⁶

In order to keep the cost of matrix manipulations manageable, we project only the tightest core and valence set of the AOs in this work to generate native P- and PS-LMOs. This ensures that the resulting Parseval frame represents the occupied space while keeping its dimension at the minimal basis level. This procedure admittedly introduces an ambiguity in the calculation scheme, which is accepted at the current stage with plans for waiving this feature in follow up works.

4.2 Semibullvalene bond rearrangement

Semibullvalene tautomerisation provides an example for a geometrical distortion involving bond cleavage and formation. Molecular structure at the initial point of the reaction is shown in Fig. 2. Double bonds C1-C2 and C6-C7 and single bond C3-C5 are cleaved along the reaction path, whereas double bonds between C2-C3 and C5-C6 and single bond between C1-C7 are formed, resulting in the tautomer pair of the structure in Fig. 2. Tautomers exhibit C_s symmetry and the system passes through a C_{2v} saddle point at halfway of the reaction trajectory. Hartree-Fock computations were performed in cc-pVDZ basis. Orientation of the molecule is fixed by the principal axis system. Optimized geometry points, provided in Supplementary Information, were taken from Nudged Elastic Band⁶⁷ trajectory calculations. The reaction path is discretized for 15 points, Fig. 2 displays geometry point 1.

Contours of Boys and PM LMOs are shown in Fig. 3 for the middle, C_{2v} symmetry point of the reaction path. The two sets of localized orbitals, shown in separate columns are equivalent only at this point of the trajectory. In the reactant/product regime, one or the other set is more favorable. As the orbitals are localized at distinctly different regions in the two sets, a sudden change in the character of LMOs occurs around the mid point of the trajectory.

This effect is depicted in Fig. 4, showing the displacement of LMO centroids along the trajectory. Step size of Boys and PM LMO centroids clearly stand out between

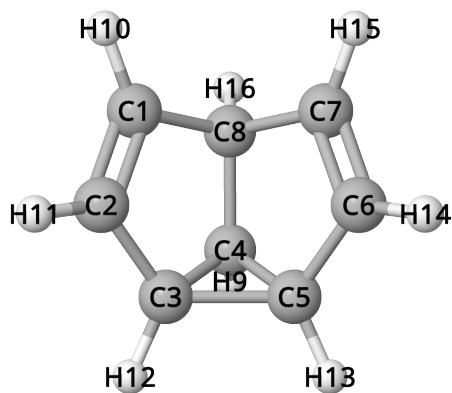


Figure 2: Atom numbering of semibullvalene. Molecular structure is drawn at the initial point of the reaction trajectory. During the reaction double bonds $C1=C2$, $C6=C7$ and single bond $C3-C5$ are rearranged to $C2=C3$, $C5=C6$ and $C1-C7$, respectively.

Boys				PM	

Figure 3: Boys and Pipek-Mezey (PM) LMO contours at the C_{2v} symmetry point of the reaction trajectory of semibullvalene. Orbitals undergoing the most distortion along the reaction path are shown. One column of the table corresponds to one extremum of the localization function. Two columns under the same method show the LMOs of two equivalent extrema, c.f. Kekulé-structures.

geometry points 8 and 9. While LMO centers displace less than $0.1 \text{ \AA}$ between other points of the trajectory, a step of cca. $1 \text{ \AA}$ for the rearranging double bond and cca. $2 \text{ \AA}$ for the single bond is unacceptably large, causing artifacts in semiempirical approaches exploiting these LMOs.

Orbital contours of P-L LMOs, shown in Fig. 5 for four points along the trajectory

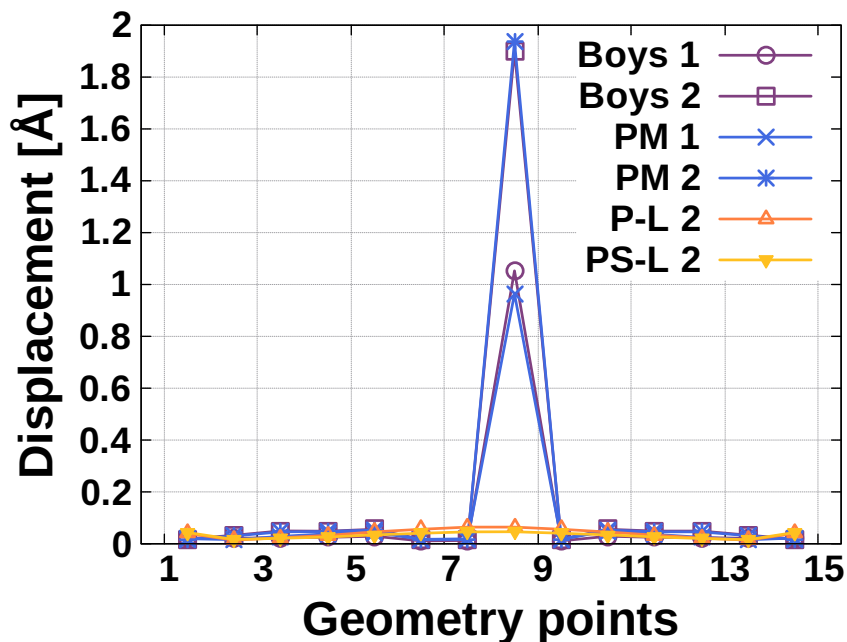


Figure 4: Displacement of LMO centroids between neighboring points of the reaction trajectory of semibullvalene computed in cc-pVDZ basis set. Orbitals were tracked based on maximum overlap from point to point for Boys, PM, P-L and PS-L schemes. Index 1 refers to the orbital in the top row, index 2 stands for the orbital in the bottom row of Figure 3 for Boys and PM LMOs. For PAO related orbitals, index 2 stands for the 1st row of Figure 5.

clearly transition smoothly along the reaction path, at difference with Boys and PM. Comparing the change of plots in the same row of Fig. 5, they do not seem to be larger between points 6 and 10 than between 1 and 6 or between 10 and 15. A parallel of Fig. 5 for PS-L LMOs, included in Supplementary Information, allows to draw the same conclusion. Smooth change of P-L and PS-L frame LMOs is also illustrated in Fig. 4, showing displacements of LMO centroids for one orbital, undergoing a relatively large distortion in Fig. 5. Step size of these, as well as other P-L and PS-L LMO centroids remain on the same level and relatively small all along the reaction.

The P-L LMOs drawn in Fig. 5 are obtained from the tightest valence p AO set of the two types of carbon atoms involved in bond rearrangement. From the two atoms shown in Fig. 5, all others' plots can be generated by symmetry relations, e.g. those of C3 are obtained from C1 by reflection along the C_s plane of the molecule and swapping trajectory points $1 \leftrightarrow 15$ and $6 \leftrightarrow 10$. The character of these frame LMOs is clearly neither

atomic nor diatomic, they rather tend to localize around three neighboring nuclei. It is important to note that individual projected AOs are in general not invariant to rotation, it is a *full* $p, d, \dots$ shell that allows to draw orientation independent conclusions. It is for this reason, that a full valence p set is included in Fig. 5, even if some of the orbitals hardly show any variation.

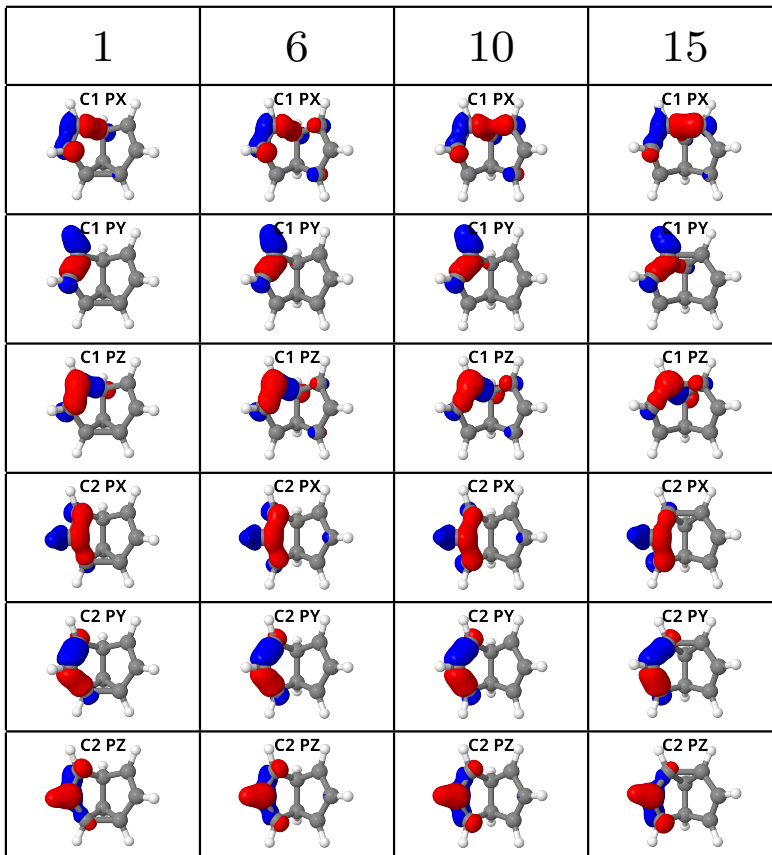


Figure 5: Contours of P-L LMOs at selected points (step No. indicated in header) of the tautomerization trajectory of semibullvalene. Frame LMOs obtained by projecting the tightest valence p AOs of atoms C1 and C2 are shown. See atom numbering in Fig. 2. Atom and AO labels refer to the AO used for projection. The molecule is oriented along its principal axis system. See text for more explanation.

Let us now address the general question what makes frame LMOs transition smoothly as opposed to Boys or PM. In the simplest picture there are 6 electrons participating in the tautomerization, allowing for 3 doubly occupied orthogonal LMOs.

A single set of three LMOs, shown in one column under PM in Fig. 3 is not invariant under symmetry operations of the C_{2v} point group. Thinking in a dot diagram representation and examining atom C1, it is a must to decide whether its electron participating in the process contributes to the double bond with C2 or to the single bond with C7. Midway solution is not possible from the part of orthogonal LMOs, even for complete equivalence of the two LMO sets, i.e. two Kekulé-structures. Banana bonds of the Boys procedure, shown in Fig. 3 increase the number of electrons and LMOs involved, but the essence remains the same.

The frame LMO concept allows for more than N_o LMOs and releases the requirement on individual occupations, fixing only their sum at the number of electrons. In the present scheme, PAOs adhering to the dimension of the minimal basis, the number of valence electrons agrees with the number of valence frame occupied LMOs. Accordingly, in a uniform distribution, all valence frame LMO occupations would be 1. As a consequence, the number of frame LMOs expected to describe the process is increased at least to 6. With 6 localized orbitals it is possible to represent the molecular point group, at difference with the 3 orthogonal LMOs. Thinking again in a dot diagram representation, 6 frame LMOs bring the freedom of describing the contribution of a single electron to the rearranging bond system. Atom C1 is now given the freedom of gradually changing the commitment of its affected electron from the double bond with C2 to the single bond with C7, with possibilities for varying midway solutions. As Fig. 5 indicates, the real situation is more complicated than the simplified picture given above. In fact it is hard to assign one single projected frame LMO of atom C1 to the bond rearrangement. It is rather the altogether 18 p AO related frame LMOs of the affected carbon atoms, which describe the process. The basic finding, however, remains.

The above discussion hints the necessity of linear dependence in achieving a smooth transition of LMOs, as three LMOs localized around 6 carbon atoms is not enough to form an invariant subspace under all symmetry operations of the C_{2v} point group. Smooth LMO transition is facilitated by the increased number of frame orbitals with a

simultaneous decrease in their occupations.

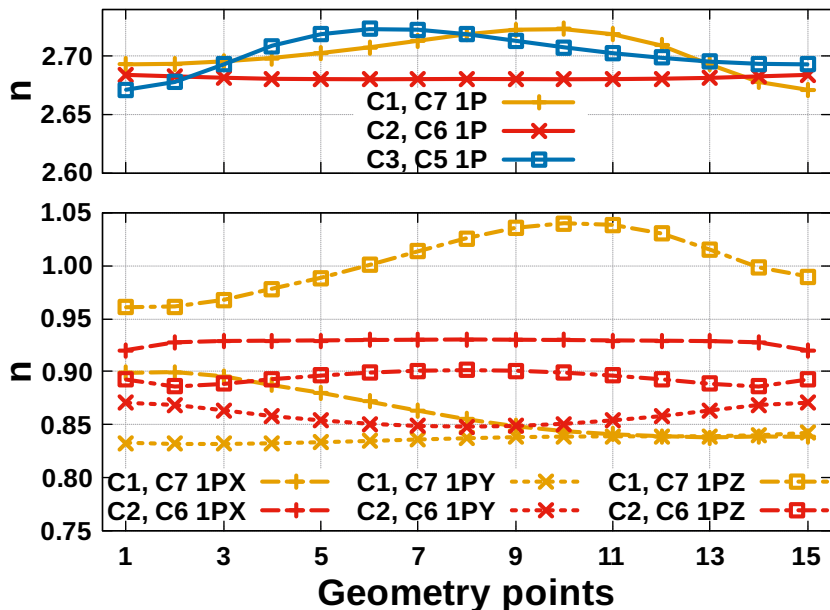


Figure 6: Partial occupation numbers, n of P-L frame LMOs along the tautomerization reaction trajectory of semibullvalene. Atom and AO labels refer to the AO used for projection. Top panel shows occupation numbers summed for the valence p shell of the same center. Bottom panel displays occupation numbers of frame LMOs originating from individual p_x , p_y and p_z AOs.

Partial occupation numbers obtained from (49) and multiplied by 2 for spin-summation, shown in Fig. 6 for P-L and in Fig. 7 for PS-L frame LMOs support the above statements. Individual frame LMO occupations plotted in the bottom panel of Fig. 6 correspond to the LMOs in Fig. 5, with Fig. 7 being analogous. As expected, frame LMO occupations are close to 1. Moreover they vary smoothly along the reaction path, on a scale not larger than 0.1. Interestingly, while frame LMO contours are rather similar for P-L and PS-L, their occupation numbers show a significant difference: those of P-L tend to be smaller than one and those of PS-L are rather larger than one. We attribute this difference to overlap effects and choose not to interpret the occupation numbers or their variations, especially since they change on a fine scale. Note also, that individual LMO occupations are subject to change with molecule orientation. This is the reason for plotting p AO related frame LMO occupations summed for a given center

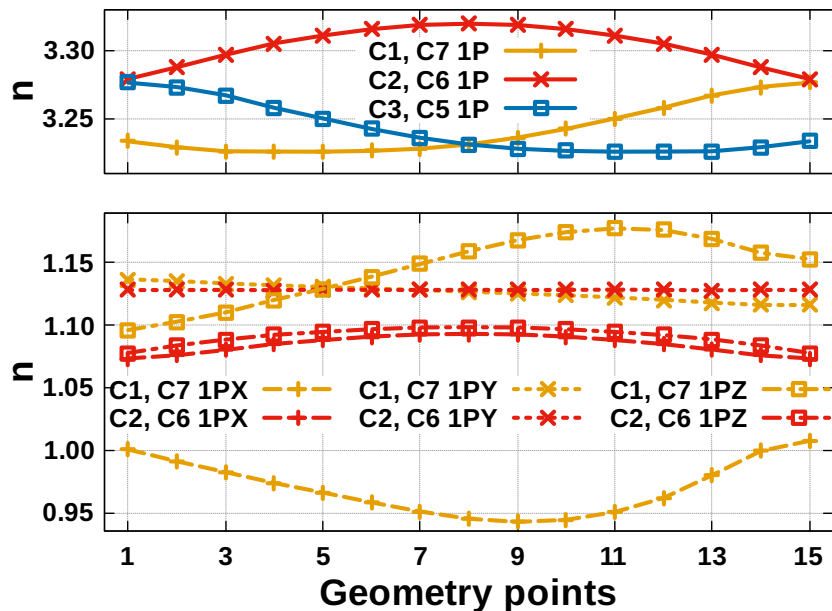


Figure 7: Partial occupation numbers, n of PS-L LMOs along the tautomerization reaction trajectory of semibullvalene. See the legend of Fig. 6 for more.

in the top panels of Figs. 6 and 7. These curves also show a smooth dependence on the reaction coordinate. Let us stress once again, that atom and AO labels in Figs. 6 and 7 refer to the AO used for projection and do not mean that frame LMOs would exhibit atomic character, c.f. Figs 5.

5 Conclusion

We have shown that the notion of frames facilitates the use of projected AO related, overcomplete orbital sets in a way reminiscent of orthogonal bases. In particular, the unit operator of the space spanned by frame orbitals is represented by a single sum of symmetric dyads, when the underlying set is a Parseval-frame. This opens a way of expanding arbitrary orbitals, computing traces and operator matrix elements in terms of overcomplete orbitals in a manner analogous to the orthonormal case.

A finite set of overcomplete orbitals always represents a frame of the space it spans. Generating a Parseval-frame of an initial frame is possible by performing a symmetric

Löwdin-transformation with the help of a generalized inverse of the singular overlap matrix. The Moore-Penrose pseudo-inverse serves for this purpose, as well as weaker inverse choices, satisfying only the first two, so-called reflexive items of the four Penrose conditions.

In this work we have considered previously introduced PS- and P-LMOs as initial frame sets and showed that reflexive inverses are readily available via the idempotency condition of matrix $\mathbf{PS}$, removing the need of numerical inversion. We have also shown that the PS-L Parseval-frame, generated by this reflexive inverse is equivalent to the $P^{1/2}$ LMO set proposed in a former study.⁴²

Dealing with an overcomplete set instead of a linearly independent one brings an inevitable loss of efficiency, which should be compensated by some advantage of the frame. In this work we have shown one example for redundancy helping to overcome a symmetry related ambiguity, which can not be evaded with orthogonal MOs. The problem arises from equivalent localization minima, occurring whenever one set of doubly occupied LMOs, affected in bond rearrangement, fails to form a representation of the molecular point group. We surmise that the problem can not be evaded with neither orthogonal nor overlapping, doubly occupied LMOs. Leaving linear independence behind appears necessary for mitigating this symmetry induced the discontinuity of LMOs along reaction trajectories. The proposed solution by frame LMOs is simple and free from any special adjustment related to the problematic LMO subset. We expect that it may have implications in LMO-based property calculations, e.g., decomposition of electron density, LMO-based electrostatic descriptors, as dipole moments or higher order moments, van der Waals interaction centers, embedding techniques.

While allowing for redundancy, keeping the dimension of the frame set as low as possible is an evident goal. The workaround applied in this study, i.e. projecting from the tightest core and valence AO set, is not regarded an ultimate tool. More robust ways of reducing the number of frame LMOs are under investigation. While rotation invariance of PAOs poses no problem as long as full $p, d, \dots$ shells are considered on a given center, a frame set with rotation invariant individual elements could also have

benefits.

Finally, some remarks on redundancy related numerical instabilities are due. The generalized overlap inverse yielding Parseval-frames is one potential weakness which is presently sidestepped by relying on well defined reflexive inverses. It deserves mention though, that the rank of matrices treated here is exactly the number of occupied MOs. Since this determines the number of nonzero singular values or eigenvalues of $\mathbf{X}$ or $\mathbf{S}$, respectively, even an MP pseudo inverse computation is free from the usual pitfalls associated with a numerical zero threshold.

Existence of alternative coefficient sets representing the same function is another valid concern when dealing with redundant sets. In the context of this study it however does not introduce any conceptual or numerical problem.

The smoothness of the curves shown in Section 4.2 is encouraging on continuous derivatives with respect to nuclear positions of frame LMO related quantities along a reaction trajectory.

Associated Content

Supporting Information available. A parallel of Fig. 5 with PS-L LMOs, projected from the tightest valence p AOs of atoms C1 and C2, at selected points of the reaction trajectory.

Acknowledgements

The work was supported by the National Research, Development and Innovation Office (NKFIH) under the Hungarian National Excellence Programme (NKKP) ADVANCED grant scheme, No. 150283. as well as from the Young Researcher Excellence Programme grant No. FK147031, and the ELTE University Excellence Fund. Support by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (JD) and the EKÖP-25 University Research Scholarship Program of the Ministry for Culture and Innovation

from the source of the National Research, Development and Innovation Fund (AG) is also gratefully acknowledged.

A Appendix

A.1 Nonunique expansion coefficients

Diagonal entries of the overlap (12) are squared norms of $\{\varphi_\mu\}$. Rearranging in the form

$$\mathbf{V}^\dagger \mathbf{S} \mathbf{V} = \begin{pmatrix} \mathbf{\Sigma}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad (\text{A.1})$$

we still have squared norms in the diagonal, now those of vectors

$$\sum_{\mu=1}^M |\varphi_\mu\rangle V_{\mu\lambda} .$$

The zero bottom right block of (A.1) means that

$$\sum_{\mu=1}^M |\varphi_\mu\rangle V_{\mu\lambda} = |0\rangle \quad , \quad \lambda = N+1, \dots, M \quad (\text{A.2})$$

where $|0\rangle$ stands for the vector of zero norm.

To highlight the ambiguity of the linear coefficients, let us expand φ_ν on the set of $\{\varphi_\mu\}$ with the help of (14) to get

$$\begin{aligned} |\varphi_\nu\rangle &\equiv \sum_{\mu=1}^M |\varphi_\mu^{\bar{L}}\rangle \langle \varphi_\mu^{\bar{L}} | \varphi_\nu \rangle = \sum_{\mu\lambda=1}^M |\varphi_\mu\rangle \bar{S}_{\mu\lambda} \langle \varphi_\lambda | \varphi_\nu \rangle = \sum_{\mu=1}^M |\varphi_\mu\rangle (\bar{\mathbf{S}}\mathbf{S})_{\mu\nu} \\ &= \sum_{\mu=1}^M |\varphi_\mu\rangle \sum_{\lambda=1}^N V_{\mu\lambda} V_{\nu\lambda}^* , \end{aligned} \quad (\text{A.3})$$

having used (18) as the definition of $\bar{\mathbf{S}}$. The last step above holds as $\bar{\mathbf{S}}\mathbf{S}$ agrees with the overlap matrix of the Löwdin-vectors in (20). Though (14) ensures that the rhs of the above gives $|\varphi_\nu\rangle$, the coefficient appearing next to $|\varphi_\mu\rangle$ is embarrassingly not $\delta_{\mu\nu}$.

It would appear if the unitary nature of $\mathbf{V}$ could be utilized, which is prevented on the rhs of (A.3) by the summation over λ being restricted to $1, \dots, N$. To extend this summation over $N + 1, \dots, M$, one has to remark that this step means adding merely zeroes, due to (A.2). While coefficients $\delta_{\mu\nu}$ appear next to $|\varphi_\mu\rangle$ by this procedure, it is just one particular way of augmenting the rhs of (A.3) without altering $|\varphi_\nu\rangle$.

A.2 On the frame properties of PS-LMOs and P-LMOs

Starting with PS-LMOs, let us rewrite (33) using $\mathbf{P} = \mathbf{C}\mathbf{C}^\dagger$ and (29) to get

$$|\phi_\mu^{PS}\rangle = \sum_{\nu=1}^M \sum_{i=1}^{N_o} |\chi_\nu\rangle C_{\nu i} (\mathbf{C}^\dagger \mathbf{S})_{i\mu} = \sum_{i=1}^{N_o} |\psi_i\rangle (\mathbf{C}^\dagger \mathbf{S})_{i\mu} ,$$

indicating that the $N_o \times M$ matrix of expansion coefficients introduced in Section 2.1 reads for the PS-LMOs as

$$\mathbf{X}^{PS} = \mathbf{C}^\dagger \mathbf{S} .$$

The frame-operator of PS-LMOs

$$\hat{F}^{PS} = \sum_{\mu=1}^M |\phi_\mu^{PS}\rangle \langle \phi_\mu^{PS}|$$

therefore takes the matrix representation

$$\mathbf{F}^{PS} = \mathbf{X}^{PS} \mathbf{X}^{PS\dagger} = \mathbf{C}^\dagger \mathbf{S}^2 \mathbf{C} .$$

As the rank of $\mathbf{C}$ is N_o , the above matrix is positive definite, the set of occupied PS-LMOs accordingly represents a frame.

Turning to P-LMOs, their expansion on the orthonormal basis of canonical MOs reads

$$|\phi_\mu^P\rangle = \sum_{i=1}^{N_o} \sum_{\nu=1}^M |\chi_\nu\rangle C_{\nu i} C_{\mu i}^* = \sum_{i=1}^{N_o} |\psi_i\rangle C_{\mu i}^* ,$$

the $N_o \times M$ matrix of expansion coefficients of P-LMOs accordingly takes the form

$$\mathbf{X}^P = \mathbf{C}^\dagger.$$

The matrix representation of the frame-operator of P-LMOs

$$\hat{F}^P = \sum_{\mu=1}^M |\phi_\mu^P\rangle \langle \phi_\mu^P|$$

is obtained as

$$\mathbf{F}^P = \mathbf{C}^\dagger \mathbf{C},$$

fulfilling the frame condition, since $\mathbf{F}^P$ is positive definite when $\mathbf{C}$ is of rank N_o .

It is easy to show that $\mathbf{X}^{PS\dagger}$ is a generalized inverse of $\mathbf{X}^P$, fulfilling criteria (17a)-(17c). With the notation of Section 2.1.3

$$\mathbf{X}^{PS\dagger} = \overline{\overline{\mathbf{X}^P}}. \quad (\text{A.4})$$

All three criteria implied by (A.4) rest on the orthonormal property of the occupied MOs, expressed as $\mathbf{C}^\dagger \mathbf{S} \mathbf{C} = \mathbf{I}_{N_o}$. In particular, the check of (17a)-(17c) read as

$$\text{a) } \mathbf{X}^{PS\dagger} \mathbf{X}^P \mathbf{X}^{PS\dagger} = \mathbf{S} \mathbf{C} \mathbf{C}^\dagger \mathbf{S} \mathbf{C} = \mathbf{S} \mathbf{C} \mathbf{I}_{N_o} = \mathbf{X}^{PS\dagger}$$

$$\text{b) } \mathbf{X}^P \mathbf{X}^{PS\dagger} \mathbf{X}^P = \mathbf{C}^\dagger \mathbf{S} \mathbf{C} \mathbf{C}^\dagger = \mathbf{I}_{N_o} \mathbf{C}^\dagger = \mathbf{X}^P$$

$$\text{c) } \mathbf{X}^P \mathbf{X}^{PS\dagger} = \mathbf{X}^{PS} \mathbf{X}^{P\dagger} = \mathbf{C}^\dagger \mathbf{S} \mathbf{C} = \mathbf{I}_{N_o}$$

Overlap matrices of PS-LMOs and P-LMOs are accordingly reflexive inverses of each other in the sense of Section 2.1.3. They are given by

$$\mathbf{X}^{PS\dagger} \mathbf{X}^{PS} = \mathbf{S} \mathbf{P} \mathbf{S} \quad (\text{A.5})$$

and

$$\mathbf{X}^{P\dagger} \mathbf{X}^P = \mathbf{P} \quad (\text{A.6})$$

obeying criteria (35).

A.3 $\mathbf{P}^{1/2}$ through the SVD of $\mathbf{C}$

Write the SVD of the $M \times N_o$ coefficient matrix of the occupied orbitals as

$$\mathbf{C} = \mathbf{Z} \begin{pmatrix} \mathbf{\Lambda} \\ \mathbf{0} \end{pmatrix} \mathbf{W}^\dagger ,$$

where the diagonal entries of $\mathbf{\Lambda}$ are assumed to be positive. With the above introduced quantities the spectral decomposition of the density matrix of the occupied space reads

$$\mathbf{P} = \mathbf{C}\mathbf{C}^\dagger = \mathbf{Z} \begin{pmatrix} \mathbf{\Lambda}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Z}^\dagger ,$$

allowing to express the square root of the matrix as

$$\mathbf{P}^{1/2} = \mathbf{Z} \begin{pmatrix} \mathbf{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Z}^\dagger .$$

Let us now introduce matrix $\mathbf{D}$ as

$$\mathbf{D} = \mathbf{Z} \begin{pmatrix} \mathbf{I}_{N_o} \\ \mathbf{0} \end{pmatrix} \mathbf{W}^\dagger .$$

The following equations can be verified by substitution

$$\mathbf{P}^{1/2} = \mathbf{C}\mathbf{D}^\dagger = \mathbf{D}\mathbf{C}^\dagger .$$

Using the orthonormal property of the occupied MOs, expressed as

$$\mathbf{C}^\dagger \mathbf{S} \mathbf{C} = \mathbf{I}_{N_o}$$

we see that

$$\mathbf{P}\mathbf{S}\mathbf{P}^{1/2} = \mathbf{C}\mathbf{C}^\dagger \mathbf{S} \mathbf{C} \mathbf{D}^\dagger = \mathbf{C}\mathbf{D}^\dagger = \mathbf{P}^{1/2} ,$$

as well as

$$\mathbf{P}^{1/2}\mathbf{S}\mathbf{P} = \mathbf{D}\mathbf{C}^\dagger\mathbf{S}\mathbf{C}\mathbf{C}^\dagger = \mathbf{D}\mathbf{C}^\dagger = \mathbf{P}^{1/2}$$

hold. It is also obvious, that $\mathbf{P}^{1/2}\mathbf{S}\mathbf{P}^{1/2}$ is a symmetric projector of the occupied space, since

$$\mathbf{P}^{1/2}\mathbf{S}\mathbf{P}^{1/2} = \mathbf{D}\mathbf{C}^\dagger\mathbf{S}\mathbf{C}\mathbf{D}^\dagger = \mathbf{D}\mathbf{D}^\dagger = \mathbf{Z} \begin{pmatrix} \mathbf{I}_{N_o} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{Z}^\dagger.$$

References

- (1) Löwdin, P.-O. Some aspects on the orthogonalization and configuration interaction problems. *Int. J. Quantum Chem.* **1993**, *48*, 225–232.
- (2) Beebe, N.; Linderberg, J. Simplifications in the generation of multicenter integrals in molecular quantum mechanics. *Int. J. Quantum Chem.* **1977**, *12*, 683.
- (3) Pedersen, T.; Aquilante, F.; Lindh, R. The Cholesky decomposition in electronic structure theory. *Theor. Chim. Acta* **2009**, *124*, 1.
- (4) Dalgaard, E. A quadratically convergent reference state optimization procedure. *Chemical Physics Letters* **1979**, *65*, 559–563.
- (5) Mayer, I. Unitary perturbation theory: a generalization of the two-by-two rotations. *Theor. Chim. Acta* **2000**, *104*, 163–166.
- (6) 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.
- (7) Löwdin, P.-O. Quantum Theory of Cohesive Properties of Solids. *Adv. Phys.* **1966**, *5*, 1.
- (8) Sekino, H.; Maeda, Y.; Yanai, T.; Harrison, R. J. Basis set limit Hartree–Fock and density functional theory response property evaluation by multiresolution multiwavelet basis. *The Journal of Chemical Physics* **2008**, *129*, 034111.

- (9) Lehtola, S. Curing basis set overcompleteness with pivoted Cholesky decompositions. *The Journal of Chemical Physics* **2019**, *151*, 241102.
- (10) Stoychev, G. L.; Auer, A. A.; Gauss, J.; Neese, F. DLPNO-MP2 second derivatives for the computation of polarizabilities and NMR shieldings. *The Journal of Chemical Physics* **2021**, *154*, 164110.
- (11) Daubechies, I.; Grossmann, A.; Meyer, Y. Painless nonorthogonal expansions. *Journal of Mathematical Physics* **1986**, *27*, 1271–1283.
- (12) Christensen, O. *An Introduction to Frames and Riesz Bases*; Applied and Numerical Harmonic Analysis; Birkhäuser, 2016; Vol. 27.
- (13) Casazza, P. G.; Kutyniok, G.; Philipp, F. In *Finite Frames: Theory and Applications*; Casazza, P. G., Kutyniok, G., Eds.; Birkhäuser Basel, 2012; Chapter 1, pp 1–54.
- (14) Morgenshtern, V. I.; Bölcskei, H. In *Mathematical Foundations for Signal Processing, Communications, and Networking*; E. Serpedin, T. C., Rajan, D., Eds.; CRC Press, 2012; Chapter 20, pp 737–789.
- (15) 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. *Journal of Chemical Physics* **1973**, *58*, 1017–1035.
- (16) Feng, H.; Bian, J.; Li, L.; Yang, W. An efficient method for constructing nonorthogonal localized molecular orbitals. *The Journal of chemical physics* **2004**, *120*, 9458–66.
- (17) Usvyat, D.; Schütz, M. Nonorthogonal ultralocalized functions and fitted Wannier functions for local electron correlation methods for solids. *Theoretical Chemistry Accounts* **2005**, *114*, 276–282.

- (18) Høyvik, I.-M.; Olsen, J.; Jørgensen, P. Generalising localisation schemes of orthogonal orbitals to the localisation of non-orthogonal orbitals. *Molecular Physics* **2017**, *115*, 16 – 25.
- (19) Høyvik, I.-M. The spectrum of the atomic orbital overlap matrix and the locality of the virtual electronic density matrix. *Molecular Physics* **2020**, *118*, e1765034.
- (20) Luo, Z.; Khaliullin, R. Z. Variable-Metric Localization of Occupied and Virtual Orbitals. *Journal of Chemical Theory and Computation* **2021**, *17*, 5568–5581.
- (21) Small, D. W.; Head-Gordon, M. Tractable spin-pure methods for bond breaking: Local many-electron spin-vector sets and an approximate valence bond model. *The Journal of Chemical Physics* **2009**, *130*, 084103.
- (22) Christiansen, O.; Manninen, P.; Jørgensen, P.; Olsen, J. Coupled-cluster theory in a projected atomic orbital basis. *The Journal of Chemical Physics* **2006**, *124*, 084103.
- (23) Weijs, V.; Manninen, P.; Jørgensen, P.; Christiansen, O.; Olsen, J. General biorthogonal projected bases as applied to second-order Møller-Plesset perturbation theory. *J. Chem. Phys.* **2007**, *127*, 074106.
- (24) Margócsy, Á.; Szabados, Á. Ring coupled cluster doubles at the multireference level. *The Journal of Chemical Physics* **2020**, *152*, 204114.
- (25) Szabados, Á.; Margócsy, Á.; Surján, P. R. Pivot invariance of multiconfiguration perturbation theory via frame vectors. *The Journal of Chemical Physics* **2022**, *157*, 174118.
- (26) Gupta, P. K.; Schienbein, P.; Daru, J.; Marx, D. Terahertz spectra of microsolvated ions: Do they reveal bulk solvation properties? *The Journal of Physical Chemistry Letters* **2019**, *10*, 393–398.
- (27) Rozsa, V. F.; Galli, G. Molecular polarizabilities in aqueous systems from first-principles. *The Journal of Physical Chemistry B* **2021**, *125*, 2183–2192.

- (28) Sit, P. H.-L.; Zipoli, F.; Chen, J.; Car, R.; Cohen, M. H.; Selloni, A. Oxidation state changes and electron flow in enzymatic catalysis and electrocatalysis through Wannier-function analysis. *Chemistry–A European Journal* **2011**, *17*, 12136–12143.
- (29) Silvestrelli, P. L. Van der Waals interactions in DFT made easy by Wannier functions. *Physical Review Letters* **2008**, *100*, 053002.
- (30) Grimme, S.; Bannwarth, C.; Caldeweyher, E.; Pisarek, J.; Hansen, A. A general intermolecular force field based on tight-binding quantum chemical calculations. *The Journal of Chemical Physics* **2017**, *147*.
- (31) Zhang, L.; Wang, H.; Muniz, M. C.; Panagiotopoulos, A. Z.; Car, R.; E, W. A deep potential model with long-range electrostatic interactions. *The Journal of Chemical Physics* **2022**, *156*, 124107.
- (32) Liang, C.; Jeon, J.; Cho, M. Ab initio modeling of the vibrational sum-frequency generation spectrum of interfacial water. *The Journal of Physical Chemistry Letters* **2019**, *10*, 1153–1158.
- (33) Subotnik, J. E.; Dutoi, A. D.; Head-Gordon, M. Fast localized orthonormal virtual orbitals which depend smoothly on nuclear coordinates. *The Journal of Chemical Physics* **2005**, *123*, 114108.
- (34) Toyota, K.; Ehara, M.; Nakatsuji, H. Elimination of singularities in molecular orbital derivatives: minimum orbital-deformation (MOD) method. *Chemical Physics Letters* **2002**, *356*, 1–6.
- (35) Bensberg, M.; Neugebauer, J. Orbital Alignment for Accurate Projection-Based Embedding Calculations along Reaction Paths. *Journal of Chemical Theory and Computation* **2020**, *16*, 3607–3619.
- (36) Pinski, P.; Neese, F. Analytical gradient for the domain-based local pair natural

- orbital second order Møller-Plesset perturbation theory method (DLPNO-MP2). *The Journal of Chemical Physics* **2019**, *150*, 164102.
- (37) Thomas, M.; Brehm, M.; Kirchner, B. Voronoi dipole moments for the simulation of bulk phase vibrational spectra. *Physical Chemistry Chemical Physics* **2015**, *17*, 3207–3213.
 - (38) Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Letters* **1983**, *100*, 151.
 - (39) Boughton, J. W.; Pulay, P. Comparison of the Boys and Pipek–Mezey localizations in the local correlation approach and automatic virtual basis selection. *J. Comput. Chem.* **1993**, *14*, 736–740.
 - (40) Hesselmann, A. Local Molecular Orbitals from a Projection onto Localized Centers. *Journal of Chemical Theory and Computation* **2016**, *12*, 2720–2741.
 - (41) Zhang, C.; Li, S. An efficient localization procedure for large systems using a sequential transformation strategy. *The Journal of Chemical Physics* **2014**, *141*, 244106.
 - (42) Surján, P. R.; Szabados, Á.; Gombás, A. Properties of the simplest localized molecular orbitals: The P-LMOs. *The Journal of Chemical Physics* **2025**, *163*, 234124.
 - (43) Maslen, P. E.; Ochsenfeld, C.; White, C. A.; Lee, M. S.; Head-Gordon, M. Locality and Sparsity of Ab Initio One-Particle Density Matrices and Localized Orbitals. *The Journal of Physical Chemistry A* **1998**, *102*, 2215–2222.
 - (44) Schütz, M.; Hetzer, G.; Werner, H.-J. Low-order scaling local electron correlation methods. I. Linear scaling local MP2. *The Journal of Chemical Physics* **1999**, *111*, 5691–5705.
 - (45) Crawford, T.; King, R. A. Locally correlated equation-of-motion coupled cluster theory for the excited states of large molecules. *Chemical Physics Letters* **2002**, *366*, 611–622.

- (46) Li, S.; Shen, J.; Li, W.; Jiang, Y. The cluster-in-molecule method for electron correlation of large systems. *J. Chem. Phys.* **2006**, *125*, 074109.
- (47) Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *The Journal of Chemical Physics* **2013**, *138*, 034106.
- (48) Kats, D. Speeding up local correlation methods. *The Journal of Chemical Physics* **2014**, *141*, 244101.
- (49) Nagy, P. R.; Samu, G.; Kállay, M. An integral-direct linear-scaling second-order Møller–Plesset approach. *J. Chem. Theory Comput.* **2016**, *12*, 4897.
- (50) Szirmai, A. B.; Hégyel, B.; Tajti, A.; Kállay, M.; Szalay, P. Projected Atomic Orbitals As Optimal Virtual Space for Excited State Projection-Based Embedding Calculations. *Journal of Chemical Theory and Computation* **2024**, *20*, 3420–3425.
- (51) Maslen, P. E.; Head-Gordon, M. Non-iterative local second order Møller–Plesset theory. *Chem. Phys. Letters* **1998**, *283*, 102–108.
- (52) Reinhardt, P. Des interactions locales et leur calcul dans des systèmes quantiques étendus. Habilitation thesis, www.lct.jussieu.fr/pagesperso/reinh/habilitation.pdf, Université Paris VI — Pierre et Marie Curie UFR de Chimie Laboratoire de Chimie Théorique, 2004.
- (53) Damle, A.; Lin, L.; Ying, L. Compressed Representation of Kohn–Sham Orbitals via Selected Columns of the Density Matrix. *Journal of Chemical Theory and Computation* **2015**, *11*, 1463–1469.
- (54) Fuemmeler, E. G.; Damle, A.; DiStasio, R. A. J. Selected Columns of the Density Matrix in an Atomic Orbital Basis I: An Intrinsic and Non-iterative Orbital Localization Scheme for the Occupied Space. *Journal of Chemical Theory and Computation* **2023**, *19*, 8572–8586.

- (55) Head-Gordon, M.; Maslen, P. E.; White, C. A. A tensor formulation of many-electron theory in a nonorthogonal single-particle basis. *The Journal of Chemical Physics* **1998**, *108*, 616–625.
- (56) Israel, A. B.; Greville, T. *Generalized Inverses: Theory and Applications*, (2nd ed.); Springer,: New York, 2003.
- (57) Carlson, B. C.; Keller, J. M. Orthogonalization Procedures and the Localization of Wannier Functions. *Phys. Rev.* **1957**, *105*, 102–103.
- (58) Löwdin, P.-O. Control of the Accuracy of Molecular Wavefunctions by Mixed Variational Principles. *Adv. Quantum Chem.* **1970**, *5*, 185.
- (59) Mayer, I. On Löwdin’s method of symmetric orthogonalization. *Int. J. Quantum Chem.* **2002**, *90*, 63–65.
- (60) Eldar, Y. C.; Forney, G. D. Optimal tight frames and quantum measurement. *IEEE Trans. Inf. Theory* **2001**, *48*, 599–610.
- (61) Foster, J.; Boys, S. The Quantum Theory of Molecular Orbitals. *Rev. Mod. Phys* **1960**, *32*, 300.
- (62) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev.Mod.Phys* **1963**, *35*, 457.
- (63) Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. II. *J. Chem. Phys.* **1965**, *43*, S97.
- (64) 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.* **1989**, *90*, 4916.
- (65) Golub, G. H.; Van Loan, C. F. *Matrix Computations*, 4th ed.; Johns Hopkins University Press: Baltimore, MD, USA, 2013.

- (66) Higham, N. J. Stable iterations for the matrix square root. *Numerical Algorithms* **1997**, *15*, 227–242.
- (67) Henkelman, G.; Uberuaga, B. P.; Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* **2000**, *113*, 9901–9904.

TOC Graphic

