Localized orbitals from basis sets augmented with diffuse functions

Ida-Marie Høyvik and Poul Jørgensen

Localized orbitals from basis sets augmented with diffuse functions

Ida-Marie Høyvik[a)] and Poul Jørgensen
qLEAP Center for Theoretical Chemistry, Aarhus University, Langelandsgade 140, 8000 Aarhus C, Denmark

Orbital localization of occupied and virtual Hartree–Fock orbitals generated from basis sets augmented with diffuse functions is performed using the Pipek–Mezey, Boys, powers of the second central moment, and powers of the fourth central moment localizations. The locality of the obtained orbital sets are presented in terms of second and fourth moment orbital spreads. The results show that both local occupied and virtual orbitals may be obtained when using powers of the second central moment and powers of the fourth central moment localizations, while the Pipek–Mezey and Boys localizations fail to produce sets of local virtual orbitals. The locality of the fourth central moment virtual orbitals exhibits a locality similar to the locality of a Boys localization for non-augmented basis sets. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4803456]

I. INTRODUCTION

Localized orthonormal molecular orbitals (MOs) have become an important part of quantum chemistry and during the last sixty years a vast number of localization schemes have been proposed.[1–14] Localized orbitals are determined by carrying out rotations among the occupied and among the virtual Hartree-Fock (HF) orbitals by optimizing a localization function, exploiting that the HF energy is invariant with respect to such rotations. The standard approach has been to determine a set of local occupied MOs using the Boys[1,2] or Pipek–Mezey[4] localization functions and using a Jacobi sweep of iterations.[3] Localizing virtual MOs have proven practically impossible using Jacobi sweeps of iterations. In the absence of local, orthonormal virtual HF orbitals, the redundant non-orthogonal set of projected atomic orbitals (PAOs) have been used in local correlation methods.[15–17] Jansik *et al.*[11] and Høyvik *et al.*[14] have shown that using localization functions containing powers of the second and fourth central moment, respectively, a good locality of both occupied and virtual orbitals may be obtained. In order to generate orthogonal rotations which yield local occupied and local virtual orbitals, the trust region implementation of Høyvik *et al.*[18] was used to optimize the localization functions. Using the trust region method rather than Jacobi sweep of iterations, local virtual HF orbitals may be obtained even for large molecular systems.

In previous orbital localization applications of the trust region implementation of Ref. 18 only standard basis sets such as Dunning's correlation consistent basis sets cc-pVXZ[19] with an increasing cardinal number X were considered, since these are the ones that are used for the description of electron correlation for most molecular properties (e.g., equilibrium geometries and reaction enthalpies[20]). However, for some molecular systems and properties adding diffuse orbitals to the basis set is needed (e.g., for anionic systems and excited states) to add flexibility in the description of the outer valence regions. Also, diffuse functions are needed for an ade-

quate description of interactions with external electric fields[20] and for the description of dispersion energies.

Localization of orbital sets containing diffuse functions have received little attention in quantum chemistry. For the occupied orbitals there is a minimal difference in locality using basis sets with and without diffuse functions. However, adding diffuse functions will complicate the problem of obtaining a set of local virtual orbitals. Basis sets augmented with diffuse functions are per construction more extended in space than non-augmented basis sets and to successfully localize the virtual orbitals for augmented basis sets therefore requires localization functions which impose strict constraints on the spatial extents of the orbitals. Localization functions based on population analyses, e.g., the Pipek–Mezey localization function, do not put explicit restrictions on the spatial extent and these functions may therefore not be very efficient for obtaining localized virtual orbitals.

For standard basis sets and standard localization functions one may obtain sets of orbitals in which a single orbital has a large spatial extent – the so-called outlier. Jansik *et al.*[11] reported outliers when the Boys localization function was used. Jansik *et al.*[11] showed that the problem with outliers could be removed by using powers of the Boys localization function, i.e., powers of the second central moment (PSM). The second central moment of an orbital is often referred to as the variance of the orbital. The power acts as a penalty for the non-local orbitals since the function value will be more affected by their second central moment to some given power than by the second central moment itself. The same reasoning as used by Jansik *et al.*[11] was also used by Høyvik *et al.*[14] who introduced a localization function based on powers of the fourth central moments (PFM). The PFM localization function is aimed at reducing orbital tails, thereby achieving a better spatial locality. For both PSM and PFM, it was shown that a power two was necessary—but also sufficient—to remove outliers, obtaining sets of orbitals where all orbitals exhibited basically the same locality. In Refs. 11 and 14 only non-augmented basis sets were considered. For augmented basis sets it may also be essential to impose a penalty for non-local orbitals to obtain virtual orbitals where all orbitals are local.

---

a)Author to whom correspondence should be addressed. Electronic mail: idamh@chem.au.dk.

In this paper we consider basis sets which have no linear dependence when including diffuse basis functions. The treatment of linear dependencies in the basis may be performed using standard techniques as implemented in most electronic structure programs and will be discussed elsewhere. We have determined a set of HF orbitals for a basis containing diffuse functions and we discuss how to localize the orbitals using orthogonal transformations determined by minimizing appropriate localization functions using the trust region implementation of Høyvik et al.[18] We consider Dunning's augmented correlation consistent basis sets[19] and exclude diffuse functions for the hydrogen atoms. We use four different localization functions: Pipek–Mezey,[4] Boys,[1,2] power two of the second central moment,[11] and power two of the fourth central moment.[14]

The paper is organized as follows. In Sec. II we list the localization functions and localization measures used, while results are presented in Sec. III. A summary and some concluding remarks are presented in Sec. IV

II. THEORY

In this section we will present the Pipek–Mezey,[4] Boys,[1] PSM with power two,[11] and PFM with power two[14] localization functions. The Pipek–Mezey localization function for a set of N orthonormal occupied or virtual HF orbitals is defined by

$$\xi^{PM} = -\sum_{p=1}^{N}\sum_{A}\left(Q_{pp}^{A}\right)^2, \tag{1}$$

where the sum over A runs over atomic centers in the molecule and Q_{pp}^A is the population of orbital $|p\rangle$ on center A. The population matrix $\boldsymbol{Q}^A$ may be implemented using the Mulliken population analysis[21] or the Löwdin population analysis.[22] In this paper we use an implementation in which $\boldsymbol{Q}^A$ is computed in a Löwdin framework,

$$Q_{ij}^{A} = \sum_{\mu \in A}[\boldsymbol{S}^{\frac{1}{2}}\boldsymbol{C}]_{\mu i}[\boldsymbol{S}^{\frac{1}{2}}\boldsymbol{C}]_{\mu j}, \tag{2}$$

where $\mu \in A$ denotes a summation over all atomic basis functions on center A. Using $\boldsymbol{Q}^A$ computed in a Löwdin framework has previously been shown to yield more localized orbitals for the virtual space[23] than computing $\boldsymbol{Q}^A$ in the Mulliken framework.

The PSM localization function is defined by

$$\xi_{m}^{SM} = \sum_{p}\langle p|(\hat{\boldsymbol{r}} - \bar{\boldsymbol{r}}_p)^2|p\rangle^{m}, \tag{3}$$

where m is an integer, $\hat{\boldsymbol{r}}$ contains the $\hat{x}$, $\hat{y}$, and $\hat{z}$ Cartesian coordinates, and $\bar{\boldsymbol{r}}_p = \langle p|\hat{\boldsymbol{r}}|p\rangle$. For m = 1 in Eq. (3) we recover the Boys localization function. The PFM with power two localization function is defined as

$$\xi^{FM} = \sum_{p}\langle p|(\hat{\boldsymbol{r}} - \bar{\boldsymbol{r}}_p)^4|p\rangle^{2}. \tag{4}$$

The minimization of the localization functions is carried out using the trust region algorithm as described by Ref. 18.

To determine the locality of an orbital locality measures which directly reflect the spatial extent of the orbital $|p\rangle$ have to be used. Both the size of the bulk and the tail of the orbital have to be tested. The second central moment spread, σ_2, and the fourth central moment spread, σ_4, complement each other well as localization measures. σ_2 is the conventional orbital spread being the square root of the second central moment (variance),

$$\sigma_2^p = \sqrt{\langle p|(\hat{\boldsymbol{r}} - \bar{\boldsymbol{r}}_p)^2|p\rangle}, \tag{5}$$

and reflects the locality of the bulk of the orbital. σ_4 is the fourth root of the fourth central moment,

$$\sigma_4^p = \sqrt[4]{\langle p|(\hat{\boldsymbol{r}} - \bar{\boldsymbol{r}}_p)^4|p\rangle}, \tag{6}$$

and puts more emphasis on orbital tails.[14]

To characterize the locality of an entire set of orbitals $\{|p\rangle\}$ we report the locality of the least local orbital in the set, i.e., "the weakest link." Average locality is not a good locality measure since a set of orbitals where the locality of the orbitals on average is good may contain a single or few orbitals which are only semi-local or completely delocalized across the molecular system. By the term semi-local we refer to orbitals which are more local than canonical MOs, but which have a large spatial extent compared to the extent obtainable by using state of the art localization functions. When reporting the locality of a set of orbitals we therefore always report the locality of the orbital with the largest σ_2 and σ_4 values in the set. The largest σ_2 for an orbital in a set of orbitals $\{|p\rangle\}$ is denoted σ_2^{max},

$$\sigma_2^{max} = \max_{p}\left[\sigma_2^p\right], \tag{7}$$

and, similarly, for the σ_4 values,

$$\sigma_4^{max} = \max_{p}\left[\sigma_4^p\right]. \tag{8}$$

Small orbital spreads for all orbitals in a set of occupied and a set of virtual orbitals are important in order to use the orbitals in local correlation calculations. Two orbitals, $\phi_p(\boldsymbol{r})$ and $\phi_q(\boldsymbol{r})$, both having small second and fourth moment orbital spreads, will have a charge distribution $\phi_p^*(\boldsymbol{r})\phi_q(\boldsymbol{r})$, which decays rapidly with the distance between the center of charge for the orbitals. Two-electron integrals contains two charge distributions and two-electron integrals will therefore be banded in a local basis where the width of the bands will depend on the size of the orbital spreads. If a single or more orbitals in a set of local HF orbitals have large second and fourth moment orbital spreads the width of the bands will be large, preventing the orbitals from being efficient for describing local correlation effects.

III. RESULTS

In this section we present orbitals obtained by optimizing the localization functions described in Sec. II and for comparison we also present the PAOs. Results are presented for arachidic acid ($C_{20}O_2H_{40}$) and alanine(20), an alpha helix consisting of 20 alanine residues ($C_{60}N_{20}O_{21}H_{102}$). Dunning's augmented correlation consistent basis sets[19] aug-cc-pVXZ

TABLE I. $\sigma_2^{\max}$ and $\sigma_4^{\max}$ values (in a.u.) for the occupied orbitals for arachidic acid (aa) and alanine(20) (ala(20)), obtained using the ξ^{PM}, ξ_1^{SM}, ξ_2^{SM}, and ξ^{FM} localization functions.

	ξ^{PM}		ξ_1^{SM}		ξ_2^{SM}		ξ^{FM}	
Molecule/basis	$\sigma_2^{\max}$	$\sigma_4^{\max}$	$\sigma_2^{\max}$	$\sigma_4^{\max}$	$\sigma_2^{\max}$	$\sigma_4^{\max}$	$\sigma_2^{\max}$	$\sigma_4^{\max}$
aa/aug-cc-pVDZ	1.79	2.23	1.58	2.21	1.58	2.21	1.58	2.19
aa/aug-cc-pVTZ	1.78	2.28	1.58	2.20	1.58	2.21	1.58	2.18
aa/aug-cc-pVQZ	1.79	2.27	1.58	2.20	1.58	2.20	1.57	2.18
ala(20)/aug-cc-pVDZ	1.91	2.46	1.74	2.32	1.62	2.23	1.62	2.18

with X = D, T, and Q are used for arachidic acid and aug-cc-pVDZ is used for alanine(20). For the hydrogen atoms we use the cc-pVXZ bases.

In Secs. III A and III B the locality for the occupied and virtual orbitals, respectively, is presented in terms of the locality of the least local orbitals (the orbitals which have $\sigma_2 = \sigma_2^{\max}$ and $\sigma_4 = \sigma_4^{\max}$). In Sec. III C, the σ_2 and σ_4 values for each individual orbital for arachidic acid in an aug-cc-pVQZ is plotted for the atomic orbitals (AOs), PAOs, ξ^{PM}, ξ_1^{SM}, ξ_2^{SM}, and ξ^{FM}. In Sec. III D the least local virtual orbitals for the augmented bases (aug-cc-pVXZ) are compared to the least local virtual orbitals obtained using the non-augmented bases (cc-pVXZ). Finally, in Sec. III E convergence results and timings are shown for arachidic acid using the aug-cc-pVDZ basis set.

A. Comparison of the localization functions for the least local occupied orbitals

The values for the localization measures $\sigma_2^{\max}$ and $\sigma_4^{\max}$ for the occupied orbitals for arachidic acid and alanine(20) are given in Table I. As seen from Table I the locality does not change with increasing cardinal number and with the presence of diffuse functions in the basis set. Local occupied orbitals are obtained for all localization functions. The occupied ξ^{PM} orbitals have slightly higher $\sigma_2^{\max}$ and $\sigma_4^{\max}$ values than do the orbitals obtained using ξ_1^{SM}, ξ_2^{SM}, and ξ^{FM}, but not significantly so. The biggest differences in $\sigma_2^{\max}$ and $\sigma_4^{\max}$ values are found for ala(20)/aug-cc-pVDZ for ξ^{PM} and ξ^{FM} and are 0.29 a.u. and 0.28 a.u., respectively. Both arachidic acid and alanine(20) have simple covalent bond structures and the orbitals are easy to localize.

B. Comparison of the localization functions for the least local virtual orbitals

The values for the localization measures $\sigma_2^{\max}$ and $\sigma_4^{\max}$ for the virtual orbitals obtained for arachidic acid and ala-nine(20) are given in Table II. In contrast to the results for the occupied orbitals, the locality of the least local virtual orbitals differ dramatically depending on which localization function is used. Looking at the results for ξ^{PM} we see that the ranges of $\sigma_2^{\max}$ and $\sigma_4^{\max}$ values are 8.43–10.36 a.u. and 10.23–13.17 a.u., respectively, for the molecules/basis sets in Table II. For ξ_1^{SM} the $\sigma_2^{\max}$ and $\sigma_4^{\max}$ values range from 6.94–8.91 a.u. and 8.51–10.64 a.u., respectively. Thus, the Boys localization function (ξ_1^{SM}) yields slightly more local virtual orbitals than the Pipek–Mezey function (ξ^{PM}), but both generate virtual orbitals which are far less local than the occupied orbitals (see Table I). For the PAOs the $\sigma_2^{\max}$ and $\sigma_4^{\max}$ ranges are 6.28–7.28 a.u. and 6.75–7.69 a.u. and the PAOs are thus more local than the ξ^{PM} and ξ_1^{SM} orbitals. For the ξ_2^{SM} localization the $\sigma_2^{\max}$ and $\sigma_4^{\max}$ values range from 3.91–4.67 a.u. and 5.78–6.67 a.u., respectively. That is, the values are approximately half the size compared to the results for ξ^{PM} and ξ_1^{SM}. Further improvement in locality is obtained when using the PFM localization function ξ^{FM}. For ξ^{FM} the $\sigma_2^{\max}$ and $\sigma_4^{\max}$ values range from 3.75–4.22 a.u. and 5.07–5.65 a.u., respectively. The $\sigma_2^{\max}$ values for the ξ_2^{SM} and ξ^{FM} virtual orbitals show that the bulk of the orbitals have a much smaller spatial extent than do the ξ^{PM} and ξ_1^{SM} virtual orbitals.

From Table II we see that the $\sigma_2^{\max}$ and $\sigma_4^{\max}$ values for ξ_2^{SM} and ξ^{FM} localization in general decrease with an increasing cardinal number, while for the PAOs the $\sigma_2^{\max}$ and $\sigma_4^{\max}$ values increase with an increasing cardinal number. The improved locality for an increasing cardinal number is due to the higher flexibility introduced when including more functions. We also note that both the $\sigma_2^{\max}$ and $\sigma_4^{\max}$ values are lower for ξ^{FM} than for ξ_2^{SM}. The fact that the $\sigma_2^{\max}$ values are smaller is non-intuitive as σ_2 values are minimized directly for a ξ_2^{SM} localization. The σ_4 values are minimized directly in a ξ^{FM} minimization and hence, when minimizing σ_4 values, small σ_2 values are a favorable side effect and this is discussed in Ref. 14.

TABLE II. $\sigma_2^{\max}$ and $\sigma_4^{\max}$ values (in a.u.) for the virtual orbitals for arachidic acid (aa) and alanine(20) (ala(20)), obtained using the ξ^{PM}, ξ_1^{SM}, ξ_2^{SM}, and ξ^{FM} localization functions and the PAOs.

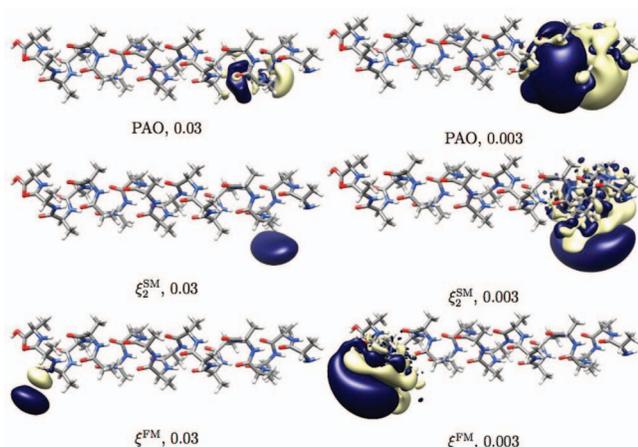|	ξ^{PM}		ξ_1^{SM}		ξ_2^{SM}		ξ^{FM}		PAO	
Molecule/basis	$\sigma_2^{\max}$	$\sigma_4^{\max}$	$\sigma_2^{\max}$	$\sigma_4^{\max}$	$\sigma_2^{\max}$	$\sigma_4^{\max}$	$\sigma_2^{\max}$	$\sigma_4^{\max}$	$\sigma_2^{\max}$	$\sigma_4^{\max}$
aa/aug-cc-pVDZ	8.43	13.17	7.66	10.53	4.42	6.60	3.96	5.65	6.33	6.75
aa/aug-cc-pVTZ	9.54	13.13	7.31	10.36	4.67	6.43	3.75	5.17	6.85	7.26
aa/aug-cc-pVQZ	10.36	12.51	8.91	10.64	3.91	5.78	3.77	5.07	7.28	7.69
ala(20)/aug-cc-pVDZ	8.46	10.23	6.94	8.51	4.48	6.67	4.22	5.26	6.28	6.75

FIG. 1. The least local PAO (top), the least local virtual ξ_2^{SM} (middle), and ξ^{FM} (bottom) orbitals for alanine(20) in an aug-cc-pVDZ basis plotted using contour values 0.03 (left) and 0.003 (right). The visualization is made using the UCSF Chimera program package.[25]

To visualize the difference in bulk locality, we have in Figure 1 for alanine(20) in the aug-cc-pVDZ basis plotted the least local PAO ($\sigma_2 = 6.28$ a.u., $\sigma_4 = 6.75$ a.u.) and the least local virtual orbitals for ξ_2^{SM} ($\sigma_2 = 4.48$ a.u., $\sigma_4 = 6.67$ a.u.), and ξ^{FM} ($\sigma_2 = 4.22$ a.u., $\sigma_4 = 5.26$ a.u.) for contour values 0.03 (left) and 0.003 (right). From Figure 1 it is clear that the bulk of the least local PAO has a larger spatial extent (Figure 1, top and left) than the bulk of the least local ξ_2^{SM} and ξ_2^{FM} orbitals (Figure 1, middle and bottom, left), as reflected also by the σ_2 values. Further, the spatial extent of the ξ_2^{SM} (Figure 1, middle) and ξ^{FM} (Figure 1, bottom) orbitals is directed outside the molecular system. This happens because there is no flexibility outside the molecular system to rotate the orbitals in a beneficial fashion. From a locality point of view, it is advantageous that the orbital extent is directed outwards – rather than inwards – if the orbitals are to be used in, for example, local correlation methods. In contrast, the PAO extends inwards. From Figure 1 we also see a faster tail decay for ξ^{FM} than for ξ_2^{SM} as reflected by the σ_4^{max} values of 5.26 a.u. and 6.67 a.u., respectively.

C. Locality measures for each orbital in the orbital sets

In this section we consider the locality of the individual virtual orbitals for arachidic acid in the aug-cc-pVQZ basis. In Figure 2 we have plotted the σ_2 and σ_4 values for all 2745 virtual orbitals, obtained using the ξ^{PM}, ξ_1^{SM}, ξ_2^{SM}, and ξ^{FM} localization functions. In Figure 3 the σ_2 and σ_4 values of the 2833 AOs and PAOs are plotted. Figure 2 illustrates that the orbital sets obtained using ξ^{PM} and ξ_1^{SM} contain many outliers, i.e., many orbitals which have large σ_2 and σ_4 values. The Pipek–Mezey and Boys virtual orbitals are local compared to canonical HF MOs ($\sigma_2^{max} = 19.77$ a.u., $\sigma_4^{max} = 21.65$ a.u.), but compared to ξ_2^{SM} and ξ^{FM} the locality is rather poor. The set of PAOs in Figure 3 is seen to exhibit a very similar locality to the AOs. As seen from Figure 3 the set of PAOs contain many orbitals with small σ_2 and σ_4 values. However, there are also many orbitals with quite large σ_2 and σ_4 values (as re-
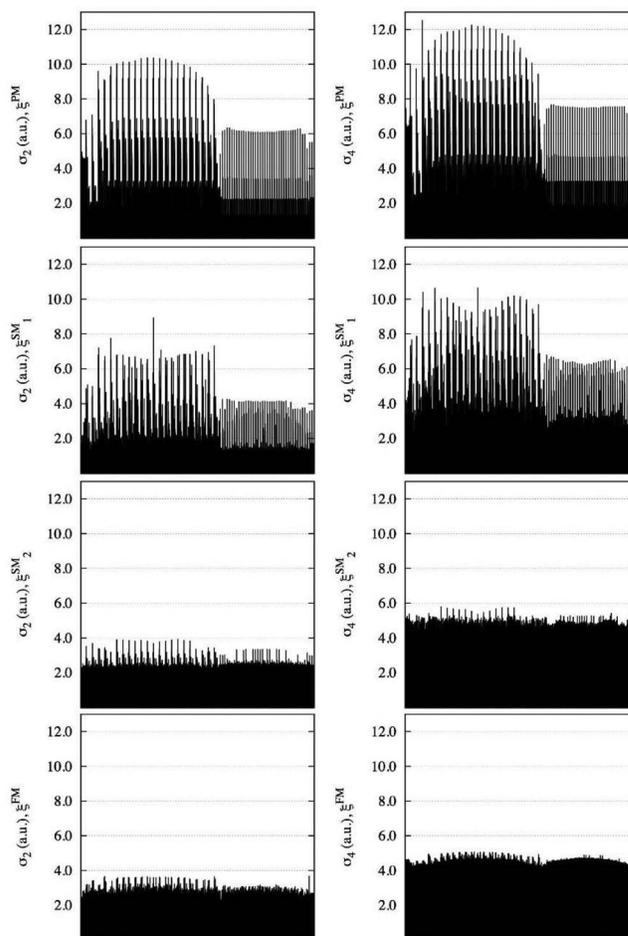

FIG. 2. σ_2 (left) and σ_4 (right) values plotted for all virtual orbitals of arachidic acid in an aug-cc-pVQZ basis obtained using ξ^{PM}, ξ_1^{SM}, ξ_2^{SM}, and ξ^{FM}.

flected in Table II, where the σ_2^{max} and σ_4^{max} are presented). In contrast, Figure 2 shows that for ξ_2^{SM} there are some orbitals with slightly larger σ_2 values than most other orbitals in the set. These are outliers, however not to the same extent as seen for, e.g., ξ^{PM} where the σ_2 values of the most and least local orbitals differ by approximately a factor 5. For ξ^{FM} the same features are seen as for ξ_2^{SM} but with even less pronounced outliers. Thus, all the ξ^{FM} orbitals exhibit almost identical locality and it is therefore reasonable to conclude that it is not possible to localize the set further using higher central moments.

D. Comparison to non-augmented basis results for the virtual orbitals

Localization results for arachidic acid using the cc-pVXZ basis sets, X = D, T, and Q, and alanine(20) in a cc-pVDZ basis are shown in Table III. The ξ_1^{SM}, ξ_2^{SM}, ξ^{FM}, and PAO results for arachidic acid are taken from Ref. 14. Comparing the results of the augmented bases (Table II) with the results for the non-augmented bases (Table III) we see that the inclusion of diffuse basis functions leads to orbitals which are less local for all localization functions considered. However, it is seen that whereas ξ^{PM} and ξ_1^{SM} are sensitive towards the inclusion of diffuse functions, ξ_2^{SM}, ξ^{FM}, and PAOs are less
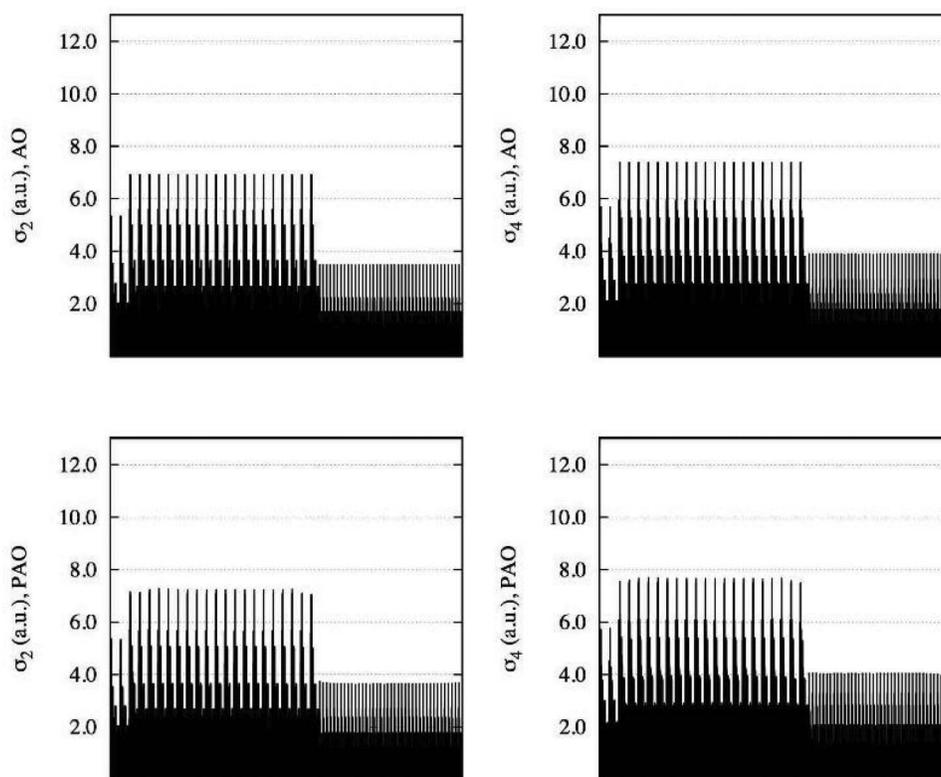
FIG. 3. σ_2 (left) and σ_4 (right) values plotted for all AOs and PAOs of arachidic acid in an aug-cc-pVQZ basis.

sensitive. This can be seen from the smaller increase in σ_2^{max} and σ_4^{max} values for ξ_2^{SM}, ξ^{FM}, and PAOs when including diffuse functions (see Tables II and III), compared to the results for ξ^{PM} and ξ_1^{SM}.

If we consider the most local set of orbitals for arachidic acid in an aug-cc-pVQZ basis (the ξ^{FM} localized orbitals), we see that relative to the localized orbitals obtained for a cc-pVQZ basis (Table III) the increase in σ_2^{max} is 1.57 a.u. while the increase in the σ_4^{max} is 2.05 a.u. To put this locality decrease in perspective, the locality of the ξ^{FM} virtual orbitals for arachidic acid in aug-cc-pVQZ basis ($\sigma_2^{max} = 3.77$ a.u. and $\sigma_4^{max} = 5.07$ a.u.) is still slightly more local than the Boys virtual orbitals for arachidic acid in the non-augmented cc-pVQZ basis ($\sigma_2^{max} = 3.81$ a.u. and $\sigma_4^{max} = 5.12$ a.u.).

E. Convergence rates and timings

Convergence rates and timings for localizing orbitals are generally similar for augmented basis sets and non-

augmented basis sets. As an example we show convergence rates and timings for arachidic acid in the aug-cc-pVDZ basis. All starting orbitals for the localizations are generated using the three-level HF optimization procedure, which generates a set of orthonormal least-change HF orbitals.[10,24] Starting from canonical HF MOs would require a few more iterations, since the canonical MOs are more delocalized than the least-change HF orbitals. The calculations were run on HP SL6500 nodes with 12 Intel/Westmere cores where the HF calculation took 53 min (wall time). The wall times for localization of the core, valence, and virtual spaces are listed for ξ^{PM}, ξ_1^{SM}, ξ_2^{SM}, and ξ_2^{FM} in Table IV. Convergence rates are presented in Figure 4 for the virtual spaces only, since the core and valence space localizations are very simple. The convergence is reported in terms of gradient norms and function values which are plotted relative to the converged values as a function of the number of iterations in the localization procedure. Due to the vastly different gradient norms in the starting iterations (ranging from $\sim 10^{-2}$ for ξ^{PM} to $\sim 10^7$ ξ_2^{FM}) the gradient norms are

TABLE III. σ_2^{max} and σ_4^{max} values (in a.u.) for the virtual orbitals of arachidic acid (aa) and alanine(20) (ala(20)) obtained using the ξ^{PM}, ξ_1^{SM}, ξ_2^{SM}, and ξ^{FM} localization functions and the PAOs. The ξ_1^{SM}, ξ_2^{SM}, ξ^{FM}, and PAO results for arachidic acid are found in Ref. 14.

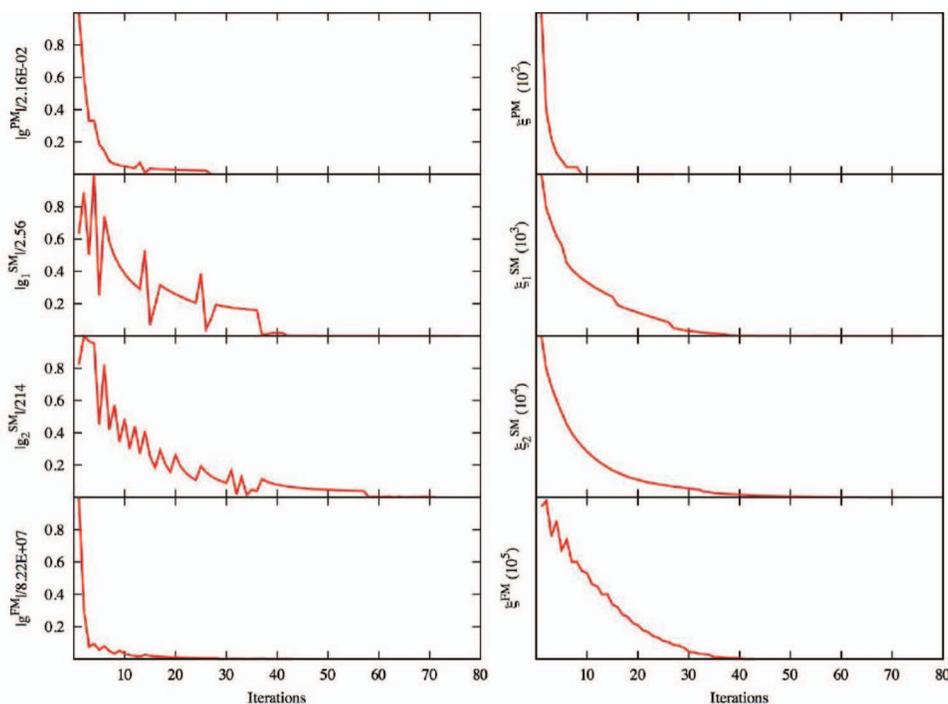| Molecule/basis	ξ^{PM}		ξ_1^{SM}		ξ_2^{SM}		ξ^{FM}		PAO	
	σ_2^{max}	σ_4^{max}	σ_2^{max}	σ_4^{max}	σ_2^{max}	σ_4^{max}	σ_2^{max}	σ_4^{max}	σ_2^{max}	σ_4^{max}
aa/cc-pVDZ	3.33	4.47	3.19	3.99	2.36	3.46	2.26	3.03	3.13	3.79
aa/cc-pVTZ	4.31	5.49	3.57	4.59	2.43	3.66	2.18	2.98	3.49	3.76
aa/cc-pVQZ	5.29	6.62	3.81	5.12	2.53	3.66	2.20	3.02	3.82	4.11
ala(20)/cc-pVDZ	3.55	4.45	3.33	4.27	2.36	3.63	2.39	3.05	3.21	4.36

FIG. 4. Convergence rates for localization of the virtual space for arachidic acid in an aug-cc-pVDZ basis for ξ^{PM} (1 row), ξ_1^{SM} (2 row), ξ_2^{SM} (3 row), and ξ^{FM} (4 row). The convergence rates are illustrated by means of the scaled gradient norms and function values relative to the converged values. The order of magnitudes for the function values and scaling of gradients are indicated in the y-axis labeling.

scaled to be in the range 0.0–1.0. The scaling factors are indicated in the y-axis labeling of Figure 4.

Table IV shows that localization of the core and valence spaces is very fast for all localization functions, while there are some differences in the computational cost for the optimization of the virtual space. ξ^{PM} is the fastest localization function for the virtual space, using 49 s. ξ_1^{SM} and ξ_2^{SM} use a similar amount of time, requiring 2 min 19 s and 2 min 22 s, respectively, and ξ_2^{FM} is the most computationally demanding localization function requiring 7 min 43 s to localize the virtual space.

The convergence rates illustrated in Figure 4 show that including diffuse functions poses no severe challenges, the optimization algorithm[18] is able to optimize all four localization functions relatively easily. The function values converges steadily as the gradient norm decreases towards zero.

IV. SUMMARY

Sets of occupied and virtual Hartree–Fock orbitals generated from Dunning's augmented correlation consistent basis sets aug-cc-pVXZ for X = D, T, and Q, have been

localized using the Pipek–Mezey, Boys, power two of the second central moment (ξ_2^{SM}) and power two of the fourth central moment localization functions (ξ^{FM}). All four localization functions gave local occupied orbitals for the molecules and bases presented. However, the Pipek–Mezey and Boys localization functions are shown to give virtual orbitals which are considerably less local than the occupied orbitals. The PAOs showed some improvement in locality compared to the Boys and Pipek–Mezey virtual orbitals, but the PAOs are seen to suffer from poor bulk locality. Contrarily, the ξ_2^{SM} and ξ^{FM} localization functions yield both local occupied and virtual orbitals. The ξ_2^{SM} and ξ^{FM} virtual orbitals are more local than the PAOs, with the ξ^{FM} virtual orbitals being the most local. The obtained ξ^{FM} virtual orbitals are less local than ξ^{FM} virtual orbitals generated from non-augmented basis sets. Further, the ξ_2^{SM} and ξ^{FM} localization functions give improved locality upon increasing the cardinal number, while the other schemes give worse locality as the cardinal number is increased.

The results show that it is possible to obtain both local occupied and virtual Hartree–Fock orbitals for augmented basis sets when using a robust localization function. It is also seen that further improvement in locality for the augmented basis sets beyond what is obtained with ξ^{FM} cannot be expected for localizations based on higher central moments.

ACKNOWLEDGMENTS

TABLE IV. Timings (in minutes and seconds) for the different localization schemes.

	Core space (s)	Valence space (s)	Virtual space
ξ^{PM}	0.01	0.47	49 s
ξ_1^{SM}	0.01	0.03	2 min 19 s
ξ_2^{SM}	0.01	0.03	2 min 22 s
ξ^{FM}	0.03	1.37	7 min 43 s

[1] S. F. Boys, Rev. Mod. Phys. **32**, 296 (1960).

[2] S. F. Boys, *Quantum Theory of Atoms, Molecules and Solid State* (Academic, New York, 1966), p. 253.

[3] C. Edmiston and K. Ruedenberg, Rev. Mod. Phys. **35**, 457 (1963).

[4] J. Pipek and P. G. Mezey, J. Chem. Phys. **90**, 4916 (1989).

[5] W. von Niessen, J. Chem. Phys. **56**, 4290 (1972).

[6] J. E. Subotnik, A. D. Dutoi, and M. Head-Gordon, J. Chem. Phys. **123**, 114108 (2005).

[7] S. Dubillard, J. Rota, T. Sause, and K. Fægri, J. Chem. Phys. **124**, 154307 (2006).

[8] F. Aquilante, T. Pedersen, A. S. de Meras, and H. Koch, J. Chem. Phys. **125**, 174101 (2006).

[9] V. Weber and J. Hutter, J. Chem. Phys. **128**, 064107 (2008).

[10] M. Ziółkowski, B. Jansík, P. Jørgensen, and J. Olsen, J. Chem. Phys. **131**, 124112 (2009).

[11] B. Jansík, S. Høst, K. Kristensen, and P. Jørgensen, J. Chem. Phys. **134**, 194104 (2011).

[12] J. Ciupka, M. Hanrath, and M. Dolg, J. Chem. Phys. **135**, 244101 (2011).

[13] P. de Silva, M. Giebultowski, and J. Korchowiec, Phys. Chem. Chem. Phys. **14**, 546 (2012).

[14] I.-M. Høyvik, B. Jansik, and P. Jørgensen, J. Chem. Phys. **137**, 224114 (2012).

[15] S. Saebø and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993).

[16] M. Schütz, G. Hetzer, and H.-J. Werner, J. Chem. Phys. **111**, 5691 (1999).

[17] M. Schütz and H.-J. Werner, J. Chem. Phys. **114**, 661 (2001).

[18] I.-M. Høyvik, B. Jansik, and P. Jørgensen, J. Chem. Theory Comput. **8**, 3137 (2012).

[19] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).

[20] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic Structure Theory*, 1st ed. (Wiley, 2000).

[21] R. Mulliken, J. Chem. Phys. **23**, 1833 (1955).

[22] P. Lowdin, J. Chem. Phys. **18**, 365 (1950).

[23] I.-M. Høyvik and P. Jørgensen, "Pipek-Mezey localization of occupied and virtual orbitals," J. Comput. Chem. (to be published).

[24] I.-M. Høyvik, B. Jansik, K. Kristensen, and P. Jørgensen, J. Comput. Chem. **34**, 1311 (2013).

[25] E. Pettersen, T. Goddard, C. Huang, G. Couch, D. Greenblatt, E. Meng, and T. Ferrin, J. Comput. Chem. **25**, 1605 (2004).