Distributed Gaussian orbitals for molecular calculations: application to simple systems

Stefano Battaglia, David Bouet, Alexis Lecoq, Stefano Evangelisti, Noelia
Faginas Lago, Thierry Leininger, Andrea Lombardi

Distributed Gaussian Orbitals for Molecular Calculations:

Application to Simple Systems

Stefano Battaglia,[a,b] David Bouet,[a] Alexis Lecoq,[a] Stefano Evangelisti,[a,*]

Noelia Faginas-Lago,[b,c,*] Thierry Leininger,[a] and Andrea Lombardi[b,c]

[a] Laboratoire de Chimie et Physique Quantiques - IRSAMC
Université de Toulouse et CNRS
118, Route de Narbonne, F-31062 Toulouse Cedex - France

[b] Dipartimento di Chimica, Biologia e Biotecnologie, Università degli Studi di Perugia
Via Elce di Sotto 8
I-06123, Perugia - Italy

[c] Consortium for Computational Molecular and Materials Sciences (CMS)^2
Via Elce di Sotto 8
I-06123, Perugia - Italy

* E-mail:
noelia.faginaslago@unipg.it
stefano.evangelisti@irsamc.ups-tlse.fr

ABSTRACT

In this article, the possible use of sets of basis functions alternative with respect to the usual atom-centred orbitals sets is considered. The orbitals describing the inner part of the wavefunction (i.e. the region close to each nucleus) are still atomic Gaussian functions: tight Gaussian orbitals having different angular momenta and large exponential coefficients, centred on each nucleus. On the other hand, the outer part of the wavefunction is described through a set of s-type distributed Gaussian orbitals: s-type Gaussians having a *unique fixed exponent*, and whose *fixed* centres are placed on a *uniform mesh of points* evenly distributed in the region surrounding all the atoms of the molecule. The resulting basis sets are applied to various one-electron systems in order to assess the capability to describe different types of one-electron wavefunctions. Moreover, the hydrogen atom and the dihydrogen cation, for which accurate solutions exist, are also considered for comparison, to assess the effectiveness of the proposed approach. Preliminary results concerning the treatment of electron correlation, necessary for a quantitatively correct description of many-electron atoms and molecules, are also presented.

2

1. Introduction

A common practice in quantum chemistry is the expansion of the molecular orbitals (MO) of the system on a set of fixed basis functions. These can be either orbitals having an atomic character, such as Gaussian-type orbitals (GTOs) [1, 2] and Slater-type orbitals (STOs), or functions distributed on the region surrounding the molecule, such as plane waves, wavelets, etc. (see e.g. Ref. [3]). In particular, GTOs, introduced first by Boys in quantum chemistry [4] are nowadays by far the most common type of orbitals used to perform numerical calculations on atoms and molecules. In principle, a single-centre (infinite) set of Gaussian orbitals is enough in order to have a complete basis set, provided the exponents are chosen in a suitable way [5, 6]. From a practical point of view, however, such a single-centre expansion has two major problems:

- The basis set becomes quickly quasi-linear dependent.

- The convergence of the expansion to describe the electrons of atoms whose nuclei are not placed on the centre of the Gaussians is extremely slow.

In practice, therefore, a single-centre Gaussian expansion does not appear to be the most suitable choice to perform actual calculations on multi-atomic systems. In particular, the nuclear cusp of an s-type orbital of a given atom can hardly be described unless the expansion contains Gaussians with very large exponents that are centred on the nucleus of the atom itself. For this reason, a very common computational strategy is to expand the molecular orbitals (MOs) of the system on a set of multi-centred atomic orbitals (AOs) of Gaussian type, centred on each one of the atoms that belong to the molecule. This is the very well known Linear Combination of Atomic Orbitals (LCAO) strategy, first used by Linus Pauling to describe the H_2^+ system, and in a more systematic way by Lennard-Jones, to describe the bonds of atoms belonging to the first main row of the periodic table. The reason why this choice is so effective is that the intra-molecular interactions are relatively weak, and therefore the different atoms maintain their individuality in a molecular system.

In order to avoid extremely large expansions and to limit quasi-redundancy problems connected to the use of very small exponents on neighbouring atoms, a very common practice is to work with linear combinations of the GTOs having *fixed* expansion coefficients. In this way, the contracted Gaussian-type orbitals (CGTOs) are obtained. The set of fixed linear coefficients is usually optimized on atomic calculations involving different states of the atom, and possibly some of its ions. A common practice is to use the atomic natural orbitals (ANOs) of the system in order to define these contraction coefficients [7, 8]. However, this well defined and elegant procedure is by no means the only possibility to obtain a set of CGTOs suitable for molecular calculations.

The LCAO strategy can be generalized by the introduction of distributed basis sets, *i.e.*, basis sets composed of orbitals not centered on any nucleus. To the best of our knowledge, the first attempt to use a set of distributed Gaussian orbitals in order to describe a molecular wavefunction goes back to John Murrell and co-workers in 1974, who proposed such an approach under the name of the Gaussian Cell Model [9]. In their original proposal, a set of single-exponent $1s$ Gaussian orbitals is located at the centers of the cells of a simple-cubic lattice surrounding the molecular system. For a given number of Gaussians, both the lattice constant and the Gaussian exponent are variationally optimized. The conclusion of these authors was that the method was "insufficiently flexible to give good results (...) unless economical techniques can be

developed for handling much larger bases" [9]. At that time, "much larger" was meant to be of the order of 10^3. Notice that calculations with this number of *orbitals* are perfectly feasible with present-day software packages, although such a huge number of *centers* is somehow uncommon for molecular calculations, and can pose some practical problems. In subsequent works, Ralston and Wilson [10] and Wilson [11, 12] reconsidered Murrell's approach by using a substantially larger number of Gaussians (up to 11 Gaussians in each direction), and by considering the use of atom-centered Gaussians basis sets in order to improve the description of the nuclear cusp. Several possibilities were considered by Wilson, involving not only the presence of additional Gaussians on the nuclei, but also of several Gaussians, and of different angular momentum. In this work, we will develop one particular among these possibilities, in which a unique-exponent series of s distributed Gaussians is used, combined with a set of nucleus-centered tight orbitals that describe the inner part of the wave-function in the region of the nuclei.

The essential idea of this approach is to use a set of Gaussian-type orbitals of s type that are evenly distributed in the space surrounding the atoms of the molecule, as in the floating orbitals scheme. We already noticed, however, that the use of atom-centred tight GTOs can hardly be avoided in order to describe the nuclear cusps. Therefore, we construct a mixed basis set according to the following two guidelines:

(1) The orbitals describing the inner part of the wavefunction will be atomic-centred GTOs having different angular momenta and large exponential coefficients.

(2) The outer part of the wavefunction, will be described through a set of s-type distributed GTOs: s-type Gaussians having a *unique fixed exponent* and whose *fixed* centres are placed on a *uniform mesh of points* evenly distributed in the region surrounding all the atoms of the molecule.

Apparently, however, this approach did not receive much attention in subsequent works on the field. Glushkov and Wilson, for instance, used instead a small number of variationally optimized Gaussian functions placed on the inter-nuclear molecular axis in the case of diatomic molecules [13].

A particularly sensitive point is the issue of "practical" linear dependence, *i.e.*, very small eigenvalues of the overlap matrix. In Wilson and co-worker's approach [11], this is treated by orthogonalizing the basis set to the eigenvectors corresponding to eigenvalues smaller than a given threshold (10^{-10}). This procedure can produce artifacts, particularly if derivatives of the energy with respect to the nuclear coordinates are numerically computed. For this reason, in our approach, we decided to avoid linear dependencies by performing a suitable choice of the basis set. With respect to previous investigations, we made the choice to avoid any type of orthogonalization of the basis set in order to avoid quasi linear dependencies. Our choice was done since the procedure of discarding the eigenvectors associated to very small eigenvalues is often numerically unstable. Thus our basis set avoids quasi linear dependencies by using a single exponent for the Distributed set, and an exponent separation between the Atomic and Distributed sets.

Notice that a number of previous attempts to construct mixed basis sets are spread among the literature of last decades and proposed, among others, mixed atom-centred STOs and floating Gaussians [14], the possible addition of bond functions to an atom-

4

centred Gaussian basis [15] or, more recently, the mixed use of local Gaussians and plane waves for applications to correlated periodic systems [16]. Notice also that in recent works by our group, the possibility of using distributed Gaussian orbitals to describe smoothly varying wavefunctions (*i.e.*, in absence of nuclei) has been considered [17, 18]. This strategy is closely related to the floating spherical Gaussian orbital (FSGO) method, developed long time ago by Frost and co-workers and successfully applied to a series of small molecules [19–23], see also [24]. In the FSGO formalism, a series of additional orbitals of s type is added in region of the bond of a molecule in order to improve the wavefunction flexibility in this region. This approach has never become very popular, and a limited number of applications is found in the literature. It should be noticed however, that recently this formalism was applied in the field of molecular dynamics by Perlt *et al.* [25, 26].

The formalism resulting from the approach described above is here presented and applied to the description of some one-electron systems, to validate is as a method to handle different types of one-electron wavefunctions. In particular, the proposed approach will be applied to simple analytically soluble systems, like the harmonic oscillator and the hydrogen atom. The dihydrogen cation, for which very precise solutions exist, will also be considered, as an example of multi-centre expansion. The introduction of electron correlation, necessary to describe many-electron atoms and molecules, is given in Section 7 along with some preliminary results from applications to simple systems. Notice that, to our knowledge, the mixed LCDAO approach has never been applied to correlated functions containing more than two electrons. More complete work is postponed to a forthcoming paper.

This article is organized as follow: in Section 2 we illustrate the basic ideas leading to a linear combination of distributed and atomic orbitals; in Section 3 the Gaussian basis sets required by this approach are formally presented; in Section 4, the problem of the quasi-completeness and quasi-linear dependence of the basis set is addressed; Section 5 describes the details of the numerical calculations, while in Section 6 results concerning one-electron systems are presented and discussed; In Section 7, applications to many-electron systems are considered. Finally, in Section 8, some conclusions are drawn, and the perspective use of such a type of basis sets for realistic molecular calculations is considered.

2. Linear Combination of Distributed and Atomic Orbitals

Following one of the approaches examined by Ralston and Wilson in their article [10], we will consider in this work two types of orbitals that are combined for the construction of basis sets for the calculation of the wavefunctions of atomic and molecular systems:

(1) Atomic-centred orbitals: these are orbitals whose centres are placed on an atomic nucleus. In this work, we will consider GTOs whose exponents form a geometric series (even-tempered Gaussians).

(2) Distributed orbitals: evenly distributed identical Gaussian orbitals that occupy the whole molecular region. For the sake of simplicity, only uniformly spaced s-type orbitals having a unique exponent will be considered.

The orbitals at point (1) are the type of orbitals of the usual LCAO approach. Although these orbitals do not form a complete basis set, by choosing the ratio of the series sufficiently close to one, it is possible to construct a suitable linear combination to approximate an atomic orbital to a very high accuracy, provided that the centers of the Gaussians are located on the nuclei. They are very effective for describing the core and inner valence orbitals of an atom. Since the inner atomic shells remain essentially unchanged when a given atom combines with other atoms and gives rise to a molecular structure, the LCAO approach gives a good description of the inner region of the orbitals even in a molecular system. Things are much more complicated for the external regions of the atoms, where, because of the bonding mechanism, the molecular wavefunction looses its atomic character. For smoothly varying orbitals, one can adopt the strategy illustrated at point (2) and describe the orbitals through Gaussian functions evenly distributed in the molecular region. As an example, these orbitals can be used to describe a set of electrons placed in a smooth external potential (for instance, the electrons in a quantum dot). However, this approach fails to describe an essential element of any molecule: the cusp of the s orbitals on each nucleus, where the wavefunction derivative has a singularity. For this reason, for atoms, molecules or real solids, this mixed approach seems to be preferable: the simultaneous use of tight atomic orbitals in order to describe the inner region of the wavefunction, with distributed orbitals for the description of the outer region. This technique for the calculation of atomic and molecular orbitals will be indicated as linear combination of distributed and atomic orbitals (LCDAO).

As it will be discussed in detail in the next two sections, a crucial step is represented by the choice of the specific Gaussian-type functions. For nuclear-centred atomic orbitals, there is a large amount of literature concerning even-tempered GTOs and their ability to describe the orbitals of an atom in a satisfactory way. For distributed Gaussians, on the other hand, the situation is much less clear, so we will here focus on this aspect of the problem. It turns out that the key parameter in this case is given by the dimensionless product ξ (formally introduced in the next section) of the Gaussian exponent (whose dimension is a length to the power of minus two) and the square of the inter-centre distance. The situation is illustrated in Figures 1, 2 and 3.
If the product is small (Figure 1), the sum of the two Gaussians is a single-peak function, and these functions are in principle suitable to describe a smooth MO. However, if this product is too small, the two Gaussians have a very large overlap, and they quickly become linearly dependent. If the product is too large, on the other hand, the sum of the two Gaussians has a double-peak shape (Figure 2). In this case, a linear

combination of such GTOs will necessarily have a large oscillatory behaviour while describing any molecular function. This fact introduces a large spurious contribution to the kinetic energy of the system, and the interpolation would be impossible. The border between these two regimes is found when the product equals two: the sum of two Gaussians with identical exponent still has a single peak, but the second derivative in the mid-point between the two functions is zero. This situation is illustrated in Figure 3. In order to describe a smooth function, Gaussians having a product smaller than 2 have to be used. At the same time, the product should not be much smaller than 1, otherwise we face problems of quasi-linear dependence of the basis set [27].

3. Gaussian Basis Sets

In this section, the properties of the Gaussian orbitals used in this article will be described and detailed. We consider a non-normalized three-dimensional (3D) Gaussian function of the type

$$g_i(\mathbf{r}) = g_{\alpha_i, \mathbf{r}_i}(\mathbf{r}) = \exp\left(-\alpha_i \|\mathbf{r} - \mathbf{r}_i\|^2\right) \tag{1}$$

where α_i and $\mathbf{r}_i$ are the exponent and the centre of the Gaussian, respectively. The overlap integral between two Gaussians g_i and g_j is given by

$$S_{ij} = S(\alpha_i, \alpha_j, d_{ij}) = \left(\frac{\pi}{\alpha_i + \alpha_j}\right)^{3/2} \exp\left(-d_{ij}^2 \frac{\alpha_i \alpha_j}{\alpha_i + \alpha_j}\right) \tag{2}$$

where d_{ij} is the distance between the two centres $\mathbf{r}_i$ and $\mathbf{r}_j$, $d_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$. For a single Gaussian, the previous equation gives the square norm of the orbital as

$$S_{ii} = S(\alpha_i, \alpha_i, 0) = \left(\frac{\pi}{2\alpha_i}\right)^{3/2} \tag{3}$$

Two cases are particularly relevant for the present work, and will be treated in detail:

(1) The two Gaussians have the same centre (and in general different exponents). In this case, the overlap becomes

$$S(\alpha_i, \alpha_j, 0) = \left(\frac{\pi}{(\alpha_i + \alpha_j)}\right)^{3/2} \tag{4}$$

(2) The two Gaussians have identical exponents (and in general different centres). In this second case, one gets

$$S(\alpha, \alpha, d_{ij}) = \left(\frac{\pi}{2\alpha}\right)^{3/2} \exp\left(-d_{ij}^2 \alpha/2\right) \tag{5}$$

In the following, it is convenient to work with normalized Gaussian orbitals, by defining the new functions

$$\bar{g}_i(\mathbf{r}) = \bar{g}_{\alpha_i, \mathbf{r}_i}(\mathbf{r}) = \left(\frac{2\alpha_i}{\pi}\right)^{3/4} \exp\left(-\alpha_i(\|\mathbf{r} - \mathbf{r}_i\|^2)\right) \tag{6}$$

The overlap between two normalized Gaussians, denoted as $\bar{S}(\alpha_i, \alpha_j, d_{ij})$, becomes then

$$\bar{S}(\alpha_i, \alpha_j, d_{ij}) = \left(\frac{2\alpha_i}{\pi}\right)^{3/4} \left(\frac{2\alpha_j}{\pi}\right)^{3/4} \left(\frac{\pi}{\alpha_i + \alpha_j}\right)^{3/2} \exp\left(-d_{ij}^2 \frac{\alpha_i \alpha_j}{\alpha_i + \alpha_j}\right) \tag{7}$$

The two particular cases considered earlier therefore become

(1) Two Gaussians with the same centre:

$$\bar{S}(\alpha_i, \alpha_j, 0) = \left(\frac{2\alpha_i}{\pi}\right)^{3/4} \left(\frac{2\alpha_j}{\pi}\right)^{3/4} \left(\frac{\pi}{\alpha_i + \alpha_j}\right)^{3/2}$$

$$= \left(\frac{4\alpha_i\alpha_j}{(\alpha_i + \alpha_j)^2}\right)^{3/4} \tag{8}$$

$$= \left(\frac{2}{1 + \alpha_i/\alpha_j} \frac{2}{1 + \alpha_j/\alpha_i}\right)^{3/4}$$

(2) Two Gaussians with identical exponents:

$$\bar{S}(\alpha, \alpha, d_{ij}) = \left(\frac{2\alpha}{\pi}\right)^{3/4} \left(\frac{2\alpha}{\pi}\right)^{3/4} \left(\frac{\pi}{2\alpha}\right)^{3/2} \exp\left(-d_{ij}^2\alpha/2\right)$$

$$= \exp\left(-d^2\alpha/2\right) \tag{9}$$

Notice that Equation (8) is manifestly symmetric under the exchange $\alpha_i \leftrightarrow \alpha_j$. Moreover, this equation implies that $\bar{S}(\alpha_i, \alpha_j, 0)$ is a *function of the ratio α_i/α_j only*. The overlap $\bar{S}(\alpha, \alpha, d_{ij})$, on the other hand, depends only on the product $d_{ij}^2\alpha$. These expressions can be further simplified if we consider some particular series of orbitals. For case (1), if a geometric series for the exponents α_i's is used, $\alpha_j = \alpha_i\gamma^{j-i}$, and we set $\alpha_i = \alpha$ and $j - i = k$, we get

$$\bar{S}(\alpha, \alpha\gamma^k, 0) = \left(\frac{4\gamma^k}{(1 + \gamma^k)^2}\right)^{3/4} \tag{10}$$

where γ is the ratio of the series.
Noticeably, $\bar{S}(\alpha, \alpha\gamma^k, 0)$ does not depend on α. In case (2), by assuming $d_{ij} = (j-i)l = kl$ (with $k \in \mathcal{N}$), and $\xi = \alpha l^2$, one has

$$\bar{S}(\alpha, \alpha, kl) = \exp(-k^2\xi/2) \tag{11}$$

In this case, the overlap does not depend separately on the exponent and the inter-centre distance, but only on k and the product ξ.

Equations (10) and (11) have deep consequences on the quasi-degeneracy properties of this Gaussian basis set. In particular, Equation (10) shows that the overlap between two normalized Gaussians tends to zero if the ratio between their exponents is very different from one. This means that the two Gaussians are almost mutually orthogonal, and they approach exact orthogonality if the ratio tends to zero (or infinity). If the ratio approaches unity, on the other hand, the overlap goes to one, and the two Gaussians become the same function.
Equation (11) implies that for any non-zero fixed exponent α and first inter-centre distance l, the overlap will go to zero for sufficiently large values of k: well separated Gaussians tend to be orthogonal in this case. The first-neighbour overlap, on the other hand, will become arbitrarily close to one for small values of the product ξ.

Unfortunately, as it is often the case for non-orthogonal basis sets, completeness is associated to quasi-degeneracy: in order to be able to describe accurately a function, one gets into troubles of quasi-linear dependence of the basis sets. The connection between these two properties of the basis set will be made evident by considering the numerical investigations of the two next sections.

4. Completeness and Linear Dependence

From a computational point of view, two issues have a crucial importance for the practical applicability of the present scheme, in order to perform actual accurate calculations. These are the quasi-completeness and quasi-linear dependence of the basis set. For non-orthogonal basis sets, these two aspects are often closely related, since, in trying to improve the description of an orbital, the basis set is pushed toward linear dependency

Quasi-completeness means that, given a molecular orbital, the basis set must be able to reproduce such orbital to an (in principle) arbitrary precision. We will not discuss here about the mathematical aspects of such a property, assuming a pragmatic approach. In practice, a basis set will be considered to be accurate enough, or "quasi-complete", with respect to a given system if the energy of the system is reproduced up to a fraction of a $\mu Hartree$.

For very large Gaussian expansions ($n \to \infty$), and either $\gamma \to 1$ or $\xi \to 0$ for the two cases, respectively, the numerical evidence seems to indicate that one gets quasi-complete basis sets. These two conditions, however, imply that the overlap between two first-neighbour normalized Gaussians tends to one, and the basis set becomes quasi-linear dependent

$$\lim_{\gamma \to 1} \bar{S}(\alpha, \alpha\gamma, 0) = \lim_{\gamma \to 1} \left(\frac{4\gamma}{(1+\gamma)^2} \right)^{3/4} = 1 \tag{12}$$

and

$$\lim_{\xi \to 0} \bar{S}(\alpha, \alpha, l) = \lim_{\xi \to 0} \exp(-\xi/2) = 1 \tag{13}$$

From a numerical point of view, therefore, γ and ξ cannot be chosen arbitrarily close to one and zero, respectively, since in these cases the overlap matrix becomes singular. This would mean that the smallest eigenvalue of the $\bar{S}_{i,j}$ matrix, $\bar{S}_{\min}$, becomes arbitrarily close to zero, and numerical calculations would be impossible. In a next section, it will be shown that there exists a *finite range of exponent values* that allow a good approximation of an atomic or molecular orbital without being plagued by quasi-linear dependence.
We address here the problem of the quasi-singularity of the metric. The analysis becomes much simpler if we assume an infinite number of Gaussians, by working in the limit $n \to \infty$. The smaller and largest eigenvalues of the overlap matrix are given in this case by the following expressions

$$\bar{S}_{\min} = \bar{S}(0) + 2 \sum_{k=1}^{\infty} (-1)^k \bar{S}(k) \tag{14}$$

and

$$\bar{S}_{\max} = \bar{S}(0) + 2 \sum_{k=1}^{\infty} \bar{S}(k) \tag{15}$$

By using Equations (10) and (11), Equation (14) becomes

$$\bar{S}_{\min}(\gamma) = 1 + 2\sum_{k=1}^{\infty}(-1)^k\left(\frac{4\gamma^k}{(1+\gamma^k)^2}\right)^{3/4} \qquad (16)$$

for the one-centre expansion, and

$$\bar{S}_{\min}(\xi) = 1 + 2\sum_{k=1}^{\infty}(-1)^k\exp(-k^2\xi/2) \qquad (17)$$

for the distributed-centre one.

The minimum eigenvalue is particularly relevant in the present investigation, since it is precisely $\bar{S}_{\min}$ that, for numerical reasons, must not be too close to zero. For instance, the general-purpose quantum chemistry code MOLPRO [28] does not allow metric eigenvalues smaller than $1.0 \cdot 10^{-6}$. In Figure 4, the lowest eigenvalue $\bar{S}_{\min}(\gamma)$ given by equation (16) is reported as a function of γ. The eigenvalue is a growing function of γ that slowly saturates toward the asymptotic limit of one. The threshold value of 10^{-6} is reached for $\gamma = 1.5$. This leaves a relatively large interval for the choice of the ratio of even-tempered sequences. In Figure 5, a similar behaviour is illustrated for $\bar{S}_{\min}(\xi)$. In this case, it appears that the region to which ξ must belong is much narrower than in the previous case. In fact, one must consider that, for a simple cubic arrangement of the Gaussian mesh, the 3D lowest eigenvalue is the *third power* of $\bar{S}_{\min}(\gamma)$ given in Equation (17), since this equation refers to the one-dimensional case. Therefore, in practice the "safe region" is $0.7 < \xi < 1.4$. Nevertheless, as it will be shown in the numerical applications presented in this article, this narrow region is enough to choose a set of distributed Gaussians that describe to a very high degree of accuracy the orbitals of a molecular system.

5. Numerical Investigations: Computational Details

In this work, the proposed scheme is tested through the application to two one-electron systems: the hydrogen atom and the hydrogen dimer cation. The hydrogen atom is an analytically soluble system. Its spherical symmetry implies that only orbitals of s type are required for the atomic expansion. On the other hand, for the H_2^+ molecular cation, that is the simplest molecular system, an exact solution is not known. Because of the lack of isotropy around each nucleus, an accurate wavefunction description requires in this case the introduction of non spherically symmetric orbitals. In both cases, the presence of the nuclear cusp requires the presence in the basis set of atomic s functions having very large exponents. Unless otherwise stated, all calculations involving one-electron systems were carried out using a self-developed software, whereas those involving many-electron systems were performed using the MOLPRO program package [28].

5.1. Hydrogen Atom

The hydrogen atom wavefunction requires the presence of Gaussians with a high exponent, in order to approximate the nuclear cusp. For this reason it can be described by using either atomic or a combination of distributed and atomic GTOs.

A set of s-type even-tempered Gaussian orbitals is suitable to describe the hydrogen atom wavefunction. The basis set exponents are chosen as

$$\alpha_k = \gamma^k \alpha_0 \tag{18}$$

where k is an *integer* number, either positive, negative or zero. The sequence of exponents defining the Gaussian functions starts at $k = 0$ and is expanded in both positive and negative directions in order to describe both the nuclear cusp and the diffuse region of the electron density. To limit the number of independent parameters to be considered, very large Gaussian expansions were taken until stable values of the energy were reached. In this way, the results coincide, up to an error of about 10^{-10} *Hartree*, with the infinite-expansion ones. We did not investigate the dependence of the approximation quality on the origin α_0 of the even-tempered sequence, that has been arbitrarily chosen to be equal to the best single Gaussian exponent, $\alpha_0 = \frac{8}{9\pi}$ (see Appendix). Notice that, in the limit of $\gamma \to 1$, the choice of α_0 becomes irrelevant.

The distributed basis functions were chosen as equally spaced s-type Gaussians, having a unique exponent. Again, in order to limit the number of parameters to be considered, the exponent was chosen equal to $\alpha_0 = \frac{8}{9\pi}$. The centres of the Gaussians are located at the corners of a simple cubic network, whose edges contains $2N+1$ functions. Integer values of N from zero to six were considered, giving rise to cubic arrangements of Gaussians going from 1 to 2197 atomic orbitals. The distance d between two first-neighbour centres is related to α_0 via the equation

$$\alpha_0 d^2 = \xi \tag{19}$$

where ξ is a parameter close to one.

In the case of LCDAO basis sets, we used 30 s-type atom-centred even-tempered Gaussians for the atomic basis set, having exponents of the type $\alpha_k = \gamma^k \alpha_0$, with

k being a *positive* integer starting at $k = 1$ and the ratio γ equal 2. These Gaussians therefore characterize the inner part of the electron density, being a sequence of increasingly tight functions.

We also investigated the dependence of the results on the position of the atomic centre within the cube of distributed Gaussians. To this purpose, the nucleus was moved from the central position towards the cube surface along the three independent directions $\mathbf{e}_x$, $\mathbf{e}_x + \mathbf{e}_y$ and $\mathbf{e}_x + \mathbf{e}_y + \mathbf{e}_z$. For each independent direction, the energy was calculated at 11 equally spaced points between the position of two Gaussians.

5.2. The H_2^+ Dimer

The hydrogen dimer cation was studied within the LCDAO approach only. A large box containing 9·9·21 distributed orbitals with exponent $\alpha_0 = \frac{8}{9\pi}$ was chosen, in combination with atomic orbitals given by an even-tempered series of Gaussian functions, starting from $\alpha_1 = 2\alpha_0$ and with ratio γ equal to 2. In view of the large number of atomic and distributed functions involved in the expansion, this basis set can be considered converged as far as the s exponents are concerned.

For the atomic part, higher angular momentum functions were also included in the expansion in order to correctly describe the wavefunction in the bonding region. In particular, the dissociation was performed using a basis comprising 30 s-, 10 p- and 5 d-type orbitals, for a total of 90 atom-centred Gaussians (Cartesian Gaussian functions were used for the d-type orbitals). The dependence on the atomic orbitals was investigated at three different internuclear distances, namely at $R = 1, 2, 4$ *bohr*.

The reference values were obtained with a standard $v6z$ basis set expansion [28–30], having $6s5p4d3f2g1h$ contracted orbitals and using the MOLPRO program package.

6. One-Electron Systems

6.1. Hydrogen Atom

The system that will be used to investigate the quality of the one-centre atomic basis set is the hydrogen atom. Being a one-electron system that admits an analytical solution, the error in the energy is exactly known, and due to the basis set incompleteness only. If we assume an infinite value for the mass of the nucleus, the ground state wavefunction depends only on the radial coordinate. For the sake of simplicity, we will limit our investigation to s-type orbitals only. In this case, the energies are given (in atomic units) by

$$\epsilon_n = -\frac{1}{2n^2} \tag{20}$$

while the corresponding wavefunctions are

$$\psi_n(r,\theta,\phi) = \psi_n(r) = \exp(-r) \tag{21}$$

By using a one-centre even-tempered Gaussian expansion, it is possible to approximate the hydrogen atom wavefunction to an arbitrary accuracy. The results are shown in Table 1 as a function of the even-tempered ratio γ. Notice how the single-centre expansion of ratio $\gamma = 2$ provides an energy of -0.499999998129 *Hartree*. If, on the other hand, a ratio $\gamma = 1.8$ is used, the total energy becomes -0.499999999949 *Hartree*.
The number of basis functions required to converge to the prescribed accuracy is inversely proportional to γ because for large values of the ratio, extremely tight and very diffuse Gaussians are generated, which do not significantly contribute to the wavefunction anymore.

In Table 2, the energy of the hydrogen atom obtained within the LCDAO approximation is reported for different values of the product ξ. In order to keep the error associated to the one-centre expansion sufficiently small, a ratio $\gamma = 2$ was chosen for the even-tempered series. The nearest-neighbour distance in the mesh is set to

$$d = \sqrt{\frac{\xi}{\alpha_0}} = \frac{\sqrt{8\xi}}{3\sqrt{\pi}} \tag{22}$$

and is listed in the last row of Table 2 for each value of ξ.
Note how the value of ξ limits the achievable accuracy of a distributed basis set for a fixed value of the exponent α_0. An accuracy in the $\mu Hartree$ range is only achieved in the case $\xi = 0.7$. For $\xi = 1.4$, the energy starts to saturate already with a cubic mesh with 5 Gaussians on each edge, whereas for $\xi = 0.7$, the energy gains significant digits up to $2N + 1 = 9$. This can be understood from the directly proportional relationship between ξ and the distance d between the Gaussians: the smaller ξ is, the more compact is the cubic mesh, ultimately providing a better achievable accuracy. On the other hand, too smaller values of ξ cannot be chosen without incurring into linear dependency issues (cf. Section 4).

The dependence of the energy on the position of the atom within the box of distributed Gaussians was investigated too, using a cubic mesh of $11 \cdot 11 \cdot 11$ basis functions

and setting $\xi = 0.7$. This box is sufficiently large such that surface effects do not appear. Moving the atom from the centre of the cube in three independent directions, a difference δ in the order of 10^{-5} *Hartree* was observed, as reported in Figure 6. In principle, only one parameter plays a role here, which is the distance d between the distributed Gaussians. The smaller the value of d (set according to ξ), the less sensitive is the energy to the actual position of the atom in the cubic mesh. In practice nevertheless, for small boxes, surface effects might appear, in particular for systems described by a diffuse wavefunction. Small d values imply compact boxes, thus one has to find the right trade-off between ξ and the size of the box of distributed Gaussian in order to balance the errors coming from the different sources.

6.2. The Hydrogen-Dimer Cation

In Figure 7, the dissociation curve of the H_2^+ cation within the LCDAO approximation is shown and compared to the all-*s* and *spdfgh* curves. The minimum is very close to the internuclear distance $R = 2.0$ *bohr* with all the three basis sets. The dissociation energy, on the other hand, is affected by the basis set in a more noticeable way. Clearly, the all-*s* expansion does not have much flexibility, such that a large fraction of the dissociation curve lies above the other two cases. On the other hand, the LCDAO approximation is able to practically reproduce the reference curve very accurately at all internuclear distances. In fact, by using the distributed Gaussian expansion one gets for the total energy, at a distance of $R = 2.0$ *bohr*, the value $E = -0.602631$ *Hartree*, while the *spdfgh* value at the same distance is $E = -0.602632$ *Hartree*. By using an all-*s* basis set, one gets the substantially higher value $E = -0.590900$ *Hartree*.

The dependence of the energy on the GTOs centred on the two nuclei is reported in Table 3. One can see how the presence of *p*-type orbitals is necessary in order to obtain energies coinciding with the reference at *mHartree* accuracy at least. Actually, addition of *d*-type orbitals improves the results compared to the *spdfgh* basis set, as highlighted by the bold digits in Table 3. Even though the effect of the d functions decreases with increasing interatomic distance, it remains significant at the equilibrium. These d functions probably explain the remaining difference of our converged value with that obtained previously by Ralston and Wilson[10].

A second parameter which plays a role regarding the accuracy of the expansion is the square section size of the distributed Gaussians mesh. Accordingly, to study this effect, the energy was computed at the same three distances R as before, but with boxes of different sizes starting from $3 \cdot 3 \cdot 21$ up to $9 \cdot 9 \cdot 21$ and with the largest atom-centred basis. The energies obtained in this case are reported in Table 4. As observed for the Gaussian cell model [10], the energy saturates quite fast with respect to the mesh size, and already for the $5 \cdot 5 \cdot 21$ box the energy is better than the reference one at $R = 1$ and $R = 4$, whereas virtually the same at $R = 2$. These results imply that one has to find an ideal trade-off between the number of distributed and atom-centred Gaussian functions in order to keep the corresponding errors at the same order of magnitude. Note that in all calculations presented for the H_2 cation the length of the box was kept fixed to account for all internuclear distances, thus resulting in a large number of Gaussians which effectively have no contribution. This was made to reduce the number of parameters to consider in this study. A more effective approach would only require to have a box of distributed Gaussians around each atom, substantially

decreasing the size of the basis.

6.3. Computational Complexity

Clearly, the success of any computational approach strongly relies on the efficiency with which calculations can be carried out. The last column of Tables 1 to 4 reports for each case the number of Gaussian-type orbitals in the basis. Evidently, in terms of raw number of basis functions, the LCDAO approach appears rather unfavourable. Nevertheless, a more scrupulous analysis reveals a different picture.

As the LCDAO approximation is a Gaussian basis set expansion only, the actual computational complexity of any calculation is still dictated by the quantum chemical method used, e.g. $\mathcal{O}(\mathcal{N}^5)$ for MP2, $\mathcal{O}(\mathcal{N}^6)$ for CCSD, and so on. On the other hand, as these methods scales with the number of basis functions, it is important to understand how this number changes with respect to the system size.

Let us assume that each atom must be embedded in a mesh of M^3 Gaussians to be correctly described. For the sake of simplicity, we consider a "cubic" system of linear dimension N. More specifically, if D is the dimension of the system, then for $D = 1$ we have a linear chain composed by N atoms, for $D = 2$ we have a square lattice made by N^2 atoms and for $D = 3$ we have a cube made by N^3 atoms. Let us call d the distance between the centres of two adjacent Gaussians as before and l that between two atoms. The point is that usually $d \leq l$ such that the total number of Gaussians is much smaller than the total number of atoms N^D times M^3, since many functions are "shared" by several atoms. Let us start by considering a one-dimensional system, where we note that the total number of Gaussians needed along the chain within the two atoms at the extremities is $(N - 1)\frac{l}{d} + 1$. To this number we should add extra $(M - 1)/2$ Gaussians on each end of the chain, summing to $(N - 1)\frac{l}{d} + M$. Note that this number represents the "Gaussian centres" along one dimension. The total number of Gaussians for a one-dimensional system is then simply given by multiplying the number of centres with the number of Gaussians required in a two-dimensional section, i.e. M^2, yielding

$$N_g \propto \left[(N - 1)\frac{l}{d} + M \right] \cdot M^2$$

For a two-dimensional system, the Gaussian centres are simply the squared value of the linear centres, i.e. $[(N - 1)\frac{l}{d} + M]^2$. This number should be now multiplied with M Gaussians required for each of the centre, for a total number of functions given by

$$N_g \propto \left[(N - 1)\frac{l}{d} + M \right]^2 \cdot M$$

In the three-dimensional case, the total number of Gaussians is simply the cubed value of the linear centres, i.e. $[(N - 1)\frac{l}{d} + M]^3$. However, this number should not be multiplied by anything else, since we already accounted for all Gaussians at the border. Thus, the general scaling of Gaussians required within the LCDAO approach, representing an upper bound to the actual number required, is given by

$$N_g \propto \left[(N - 1)\frac{l}{d} + M \right]^D \cdot M^{3-D}$$

where D is the dimension of the system as introduced before. Note how the latter formula reduces to M^3 in the trivial case of a single atom.

Substituting diffuse basis functions of high angular momenta with a set of distributed s-type Gaussian to fill the space brings some computational advantages. Calculation of integrals of high angular momenta relies on the use of recursive relations [31, 32] which are at least an order of magnitude slower than computing integrals over s-type orbitals, for which simple analytical formulas exist [33]. Therefore, one can expect the LCDAO to be much more efficient than the LCAO approach in that respect. Secondly, the fact that the distributed Gaussians are taken relatively compact and far away from each other, implies that a large portion of integrals will contribute insignificantly to the calculation and hence can be discarded by employing prescreening techniques. Lastly, the regular pattern of the 3-dimensional grid of Gaussians is likely to generate a large number of linear dependencies in the bi-electronic integral matrix, which can be therefore efficiently discarded by techniques such as the Cholesky decomposition [34–36].

7. Many-Electron Systems

The present, preliminary investigation is mainly focused on one-electron systems. However, our final aim is the description of atoms, (large) molecules, and possibly solids. Therefore, although the systematic treatment of many-electron systems is postponed to future works, it is important to give a preliminary overview of the feasibility of the proposed approach. For this reason, we report in this section the results of calculations on a few many-electron systems.

The first case that will be discussed is given by two isoelectronic systems, *i.e.*, the helium atom and the H^- negative ion. The Helium atom is a very compact system, due to the closed-shell character and the occupation of $1s$ orbital only. It can be described by a relatively small number of distributed function. On the other hand, because of the excess of negative charge, the H^- system has a very diffuse wavefunction, and thus requires a large number of distributed Gaussians for an accurate description. Notice that this large number could be substantially reduced at the price of using more diffuse distributed Gaussians, i.e. by taking a smaller value of α_0. However, since our philosophy is to use a *unique* type of Gaussians for the distributed functions, we did not explore this possibility.

In Table 5, the Hartree-Fock (HF) and full configuration interaction (FCI) total energies for the He atom, obtained by using different boxes, are compared with one-centre calculations with a large size basis set containing $28s15p13d11f9g7h5i3j$ uncontracted Gaussians. Both the HF and FCI energies converge quickly as a function of the size of the box, and the difference between the $2N + 1 = 7$ and $2N + 1 = 9$ FCI energies is less than $2 \cdot 10^{-7}$ *Hartree*. With respect to the one-centre expansion, and by using the largest basis set, the HF error is less than 10^{-6} *Hartree*, while the FCI error is of the order of $3 \cdot 10^{-6}$ *Hartree*. The cpu time to obtain these energies is at the moment substantial, since the programs we used are not adapted to such distributed basis sets, where a large number of four-centre two-electron integrals need to be computed, but are not efficiently prescreened or approximated. In the largest case ($2N + 1 = 9$), the calculation required about ten hours on a medium-size computer.

In Table 6, the FCI energy for the hydrogen anion is reported for a series of boxes of different size and compared with high-accuracy calculations. The situation is similar to the previous one, except for the fact that the wavefunction is more diffuse due to the excess of negative charge. It can be seen from the table that a cubic box with edge size $2N + 1 = 9$ is able to reproduce the reference energy of this system up to one *mHartree*.

As a more challenging example, we considered the beryllium atom and the poly-atomic system Be_4, whose results, obtained using a completely decontracted cc-pVDZ basis set for Be[37], are reported in Table 7 and Table 8, respectively.

The calculations with atom-centred orbitals were performed by using the complete set of orbitals for each beryllium atom. The LCDAO calculations, on the other hand, were done by taking those atomic functions having a value of the exponent larger than 0.5. Different box sizes were explored for the distributed Gaussians, all having the exponent $\alpha_0 = 8/9\pi$: for both systems the box sizes from 1 to 9 were considered. For Be_4, we used a T_d symmetry geometry, where the HF equilibrium distance was chosen for all calculations. We performed Hartree-Fock and coupled cluster singles and doubles (CCSD) calculations on both systems as well as complete active space self-consistent field (CASSCF) calculations, with a (2,4) valence active space for the atom,

19

and a comparatively smaller (8,10) active space for the tetramer. In the atomic case, a good $\mu Hartree$ accuracy is achieved for the uncorrelated method and the largest box, whereas the approaches introducing electron correlation deviate significantly from the reference energy.

For the beryllium tetramer, a similar accuracy to the single atom is observed for boxes up to a size of $2N+1 = 7$. However, the CASSCF energy is particularly underestimated for this system. On the other hand, remarkably, the distributed-orbital expansion is capable to outperform the LCAO approach by more than 2 $mHartree$ in the CCSD case.

8. Future Perspectives and Conclusions

In this contribution we investigated the combination of distributed and atomic Gaussian orbitals for the description of a molecular system. In a previous article, it had been shown that distributed Gaussian functions are well suited to describe orbitals having a smooth dependence on the spatial coordinates, as it is the case for a harmonic oscillator [18]. The presence of the nuclear cusp in molecular systems, on the other hand, requires the use of tight atom-centred Gaussians having large exponents. For this reason, a mixed approach, combining space-fixed distributed Gaussians for the valence region and atom-centred Gaussians to describe the inner shells can be seen as an interesting strategy.

We notice that, in the present approach, diffuse orbitals are never introduced in the distributed basis set expansion, and this fact has several advantages. In fact, the results obtained with this strategy should be very little affected by the basis set superposition error (BSSE). The BSSE is mainly due to the presence in the basis set of diffuse orbitals centred on one atom, that artificially improve the description of neighbouring atoms only in the case of short inter-atomic distances. Also the quasi-linear dependence problems at short inter-atomic distances, associated to the use of very large basis sets, should be substantially reduced. Despite the large number of orbitals in the basis set expansion, their compactness and organized position in a regular mesh can be exploited by prescreening and decomposition techniques applied to very large basis sets, which substantially reduce the number of significant two-electron integrals to compute. Moreover, regarding efficiency, the calculation of integrals over s-type GTOs is substantially faster than over functions of higher angular momenta.

At the moment, the main bottleneck for a systematic exploration of the proposed approach is represented by the number of centres of the distributed Gaussians, that are seen in standard *ab initio* codes as dummy atoms not bearing any charge. In fact, while the total number of orbitals in many general-purpose codes can easily be of several thousands, the total number of atoms, and therefore of Gaussian centres, is often limited to a few hundreds. For this reason, in the many-electron applications we were not able to go beyond a box of $9 \cdot 9 \cdot 9$ Gaussians, that is not enough to converge the total energy to more than four digits in general. Work is in progress in order to overcome this limitation, and be able to obtain results using substantially larger distributed basis sets.

9. Appendix : Best Gaussian Orbital for the Hydrogen Atom

Let us consider a Gaussian orbital ϕ_α, centred on the origin of the coordinates. The square of its norm is

$$\|\phi_\alpha\|^2 = \langle\phi_\alpha|\phi_\alpha\rangle = \left(\frac{\pi}{2\alpha}\right)^{3/2} \tag{23}$$

The kinetic and potential mean values on the Gaussian orbital are

$$\langle T\rangle_\alpha = \langle\phi_\alpha|\hat{T}|\phi_\alpha\rangle = \frac{3}{2}\alpha\left(\frac{\pi}{2\alpha}\right)^{3/2} \tag{24}$$

and

$$\langle V\rangle_\alpha = \langle\phi_\alpha|\hat{V}|\phi_\alpha\rangle = -\left(\frac{\pi}{\alpha}\right) \tag{25}$$

respectively (see, for instance, the book by Szabo and Ostlund[33]). The mean value of the energy becomes then

$$E_\alpha = \frac{3}{2}\alpha - \left(\frac{\alpha}{\pi}\right)^{1/2} 2^{3/2} \tag{26}$$

The best variational value for α is found by imposing the first derivative of E_α with respect to α equal to zero, yielding

$$\frac{d}{d\alpha}E_\alpha = \frac{3}{2} - \left(\frac{2}{\pi}\right)^{1/2}\alpha^{-1/2} \tag{27}$$

By setting $\frac{d}{d\alpha}E_\alpha = 0$ we obtain the optimum value for α, namely $\frac{8}{9\pi}$.

References

[1] F. Jensen, "Atomic orbital basis sets," *WIREs: Comput. Mol. Sci.*, vol. 3, pp. 275–295, 2013.

[2] J. Grant Hill, "Gaussian basis sets for molecular applications," *Int. J. Quantum Chem.*, vol. 113, pp. 21–34, 2013.

[3] D. Moncrieff and S. Wilson, "Distributed basis sets of s-type gaussian functions in molecular electronic structure calculations," *Mol. Phys.*, vol. 82, pp. 523–530, 1994.

[4] S. F. Boys, "Electronic wave functions. i. a general method of calculation for stationary states of any molecular system," *Proc. R. Soc. London Ser. A*, vol. 200, p. 542, 1950.

[5] B. Klahn and W. A. Bingel, "The convergence of the rayleigh-ritz method in quantum chemistry," *Theor. Chim. Acta*, vol. 44, pp. 9–26, 1977.

[6] B. Klahn and W. A. Bingel, "Completeness and linear independence of basis sets used in quantum chemistry," *Int. J. Quantum Chem.*, vol. 11, no. 6, pp. 943–957.

[7] P.-O. Widmark, P.-A. Malmqvist, and B. O. Roos, "Density matrix averaged atomic natural orbital (ano) basis sets for correlated molecular wave functions," *Theor. Chim. Acta*, vol. 77, pp. 291–306, 1990.

[8] K. Pierloot, B. Dumez, P.-O. Widmark, and B. O. Roos, "Density matrix averaged atomic natural orbital (ano) basis sets for correlated molecular wave functions," *Theor. Chim. Acta*, vol. 90, pp. 87–114, 1995.

[9] L. M. Haines, J. N. Murrel, B. Ralston, and D. Woodnutt, "Gaussian Cell Model for Molecular Orbitals," *J. Chem. Soc., Faraday Trans. II*, vol. 24, p. 1794, 1974.

[10] B. Ralston and S. Wilson, "Distributed basis sets of s-type Gaussian Function in molecular electronic structure calculations. Gaussian Cell Model revisited," *J. Mol. Struct.: THEOCHEM*, vol. 341, pp. 115–121, 1995.

[11] B. Ralston and S. Wilson, "Distributed basis sets of s-type Gaussian Function in molecular electronic structure calculations. Part 2. Gaussian Cell Model," *J. Mol. Struct.: THEOCHEM*, vol. 357, no. 1-2, pp. 37–48, 1995.

[12] S. Wilson, "Distributed basis sets of s-type Gaussian functions for molecular electronic structure calculations: Applications of the Gaussian cell model to one-electron polycentric linear molecular systems," *Int. J. Quantum Chem.*, vol. 60, no. 1, pp. 47–57, 1996.

[13] V. Glushkov and S. Wilson, "Distributed Gaussian Basis Sets Variationally Optimized s-Type Sets for H_2, LiH, and BH," *Int. J. Quantum Chem.*, vol. 89, no. 4, pp. 237–247, 2002.

[14] G. B. Bacskay and J. W. Linnett, "Mixed basis functions in molecular quantum mechanics," *Theor. Chim. Acta*, vol. 26, pp. 1–13, 1972.

[15] D. Neisius and G. Verhaegen, "Bond functions for ab initio calculations on polyatomic

molecules. molecules containing c, n, o and h," *Chem. Phys. Lett.*, vol. 78, pp. 147–152, 1981.

[16] G. H. Booth, T. Tsatsoulis, K.-L. C. Garnet, and A. Grünes, "From plane waves to local gaussians for the simulation of correlated periodic systems," *J. Chem. Phys.*, vol. 145, p. 08411, 2013.

[17] A. Diaz-Marquez, S. Battaglia, G. L. Bendazzoli, S. Evangelisti, T. Leininger, and A. Berger, "Signatures of Wigner localization in one-dimensional systems," *J. Chem. Phys.*, vol. 148, no. 12, p. 124103, 2018.

[18] L. Brooke, S. Evangelisti, T. Leininger, P.-F. Loos, N. Suaud, and A. Berger, "Localized Electrons without an Explicit Confining Potential," *J. Mol. Model.*, vol. 24, no. 8, p. 216, 2018.

[19] A. A. Frost, B. H. Prentice III, and R. A. Rouse, "A simple floating localized orbital model of molecular structure," *J. Am. Chem. Soc.*, vol. 89, p. 3064, 1967.

[20] A. A. Frost, "Floating spherical gaussian orbital model of molecular structure. i. computational procedure. lih as an example," *J. Phys. Chem.*, vol. 47, p. 3707, 1967.

[21] A. A. Frost, "Floating spherical gaussian orbital model of molecular structure. ii. one- and two-electron-pair systems," *J. Phys. Chem.*, vol. 47, p. 3714, 1967.

[22] A. A. Frost, "A floating spherical gaussian orbital model of molecular structure. iii. first-row atom hydrides," *J. Phys. Chem.*, vol. 72, p. 1289, 1968.

[23] A. A. Frost and R. A. Rouse, "A floating spherical gaussian orbital model of molecular structure. iv. hydrocarbons," *J. Am. Chem. Soc.*, vol. 90, p. 1965, 1968.

[24] A. A. Frost, R. A. Rouse, and L. Vescelius, "A floating spherical gaussian orbital model of molecular structure v. computer programs," *Int. J. Quantum Chem.*, vol. 2, no. S2, pp. 43–66.

[25] E. Perlt, M. Brüssel, and B. Kirchner, "Floating orbital molecular dynamics simulations," *Phys. Chem. Chem. Phys.*, vol. 16, pp. 6997–7005, 2014.

[26] E. Perlt, C. Apostolidou, M. Eggers, and B. Kirchner, "Unrestricted floating orbitals for the investigation of open shell systems," *Int. J. Chem.*, vol. 8, pp. 194–202, 2016.

[27] B. Klahn and W. A. Bingel, "Completeness and linear independence of basis sets used in quantum chemistry," *Int. J. Quantum Chem.*, vol. 11, p. 943, 1977.

[28] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Győrffy, D. Kats, T. Korona, R. Lindh, A. O. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and M. Wang, "MOLPRO, version 2015.1, a package of ab initio programs," 2015.

[29] A. K. Wilson, T. van Mourik, and T. H. Dunning Jr., "Gaussian basis sets for use in correlated molecular calculations. vi. sextuple zeta correlation consistent basis sets for boron through neon," *J. Mol. Struct.: THEOCHEM*, vol. 388, pp. 339–349, 1996.

[30] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, "Molpro: a general-purpose quantum chemistry program package," *WIREs Comput. Mol. Sci.*, vol. 2, pp. 241–253, 2012.

[31] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory*. Wiley, 2012.

[32] S. Reine, T. Helgaker, and R. Lindh, "Multi-electron integrals," *WIREs: Comput. Mol. Sci.*, vol. 2, no. 2, pp. 290–303, 2012.

[33] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Courier Corporation, 2012.

[34] S. Wilson, "Universal basis sets and cholesky decomposition of the two-electron integral matrix," *Comp. Phys. Comm.*, vol. 58, pp. 71–81, 1990.

[35] D. W. O'Neal and J. Simons, "Application of cholesky-like matrix decomposition methods to the evaluation of atomic orbital integrals and integral derivatives," *Int. J. Quantum.*

Chem., vol. 36, pp. 673–688, 1989.

[36] H. Koch, A. Sánchez De Merás, and T. B. Pedersen, "Reduced scaling in electronic structure calculations using Cholesky decompositions," *J. Chem. Phys.*, vol. 118, no. 21, pp. 9481–9484, 2003.

[37] B. P. Prascher, D. E. Woon, K. A. Peterson, T. H. Dunning Jr, and A. K. Wilson, "Gaussian basis sets for use in correlated molecular calculations. vii. valence, core-valence, and scalar relativistic basis sets for Li, Be, Na, and Mg," *Theor. Chem. Acc.*, vol. 128, pp. 69–82, 2011.

10. Figure Captions

(1) The linear combination of two equal-exponent Gaussian orbitals having $\xi < 2$, arbitrary units.

(2) The linear combination of two equal-exponent Gaussian orbitals having $\xi > 2$, arbitrary units.

(3) The linear combination of two equal-exponent Gaussian orbitals having $\xi = 2$, arbitrary units.

(4) The minimum eigenvalue of the metric for a one-dimensional even-spaced infinite set of identical Gaussians as a function of γ. The inset zooms in to small values of γ. Note that values lower than one are omitted.

(5) The minimum eigenvalue of the metric for a one-dimensional even-spaced infinite set of identical Gaussians as a function of ξ. The inset zooms in to small values of ξ.

(6) Electronic energy error δ with respect to the position of the hydrogen atom in the box of distributed Gaussians. A total of 11 equidistant points were considered for each direction.

(7) The H_2^+ potential energy curve as a function of the internuclear distance.

Table 1. Energy of the hydrogen atom calculated using atom-centred GTOs as a function of the harmonic ratio γ. The exact digits are underlined. The last column shows the total number of GTOs in the basis set. All energies are given in *Hartree*.

γ	Energy	# GTOs
16	$-\underline{0.4}70197942656$	15
8	$-\underline{0.49}1985592311$	19
4	$-\underline{0.499}656436902$	25
2	$-\underline{0.499999998}129$	43
1.8	$-\underline{0.4999999999}49$	51

Table 2. Energy of the hydrogen atom calculated within the LCDAO approximation for different sizes of the box. The exact digits are underlined. The d values represent the distance in *bohr* between the distributed Gaussians along the edges of the cubic mesh. The last column shows the total number of GTOs in the basis set. All energies are given in *Hartree*.

	ξ			
$2N+1$	1.4	1.0	0.7	# GTOs
1	$-\underline{0.48858560}$	$-\underline{0.48858560}$	$-\underline{0.48858560}$	31
3	$-\underline{0.49956512}$	$-\underline{0.49918847}$	$-\underline{0.49864904}$	57
5	$-\underline{0.49990338}$	$-\underline{0.49995258}$	$-\underline{0.49989458}$	155
7	$-\underline{0.49992076}$	$-\underline{0.49999350}$	$-\underline{0.49999371}$	373
9	$-\underline{0.49992125}$	$-\underline{0.49999431}$	$-\underline{0.49999949}$	759
11	$-\underline{0.49992139}$	$-\underline{0.49999449}$	$-\underline{0.49999987}$	1361
13	$-\underline{0.49992142}$	$-\underline{0.49999452}$	$-\underline{0.49999989}$	2227
d	2.2244	1.8800	1.5729	

Table 3. Dependence of the energy with respect to the atom-centred basis at three different internuclear distances of the H$_2^+$ dissociation curve. The exact digits are underlined and boldface digits highlight the lower energy obtained with the LCDAO approach. Energies are given in *Hartree* and distances in *bohr*.

basis	$R = 1.0$	$R = 2.0$	$R = 4.0$	# GTOs
30s	$-$0.44648815	$-$0.60041608	$-$0.54588596	1761
30s10p	$-$0.45174444	$-$0.60248169	$-$0.54607627	1821
30s10p5d	$-$0.45178575	$-$0.60263066	$-$0.54608306	1881
ref	$-$0.45178187	$-$0.60263209	$-$0.54603249	182

Table 4. Dependence of the energy with respect to the section of the cubic mesh at three different internuclear distances of the H_2^+ dissociation curve. The exact digits are underlined and boldface digits highlight the lower energy obtained with the LCDAO approach. Energies are given in *Hartree* and distances in *bohr*.

$2N+1$	$R = 1.0$	$R = 2.0$	$R = 4.0$	# GTOs
3	$-\underline{0.45178}\mathbf{481}$	$-\underline{0.60261706}$	$-\underline{0.54}592175$	369
5	$-\underline{0.45178}\mathbf{574}$	$-\underline{0.60263059}$	$-\underline{0.5460}\mathbf{7888}$	705
7	$-\underline{0.45178}\mathbf{575}$	$-\underline{0.60263063}$	$-\underline{0.5460}\mathbf{8281}$	1209
9	$-\underline{0.45178}\mathbf{575}$	$-\underline{0.60263066}$	$-\underline{0.5460}\mathbf{8306}$	1881
ref	-0.45178187	-0.60263209	-0.54603249	182

Table 5. Energy of the He atom obtained using four different boxes at HF and FCI levels of theory. The exact digits are underlined. Energies are given in *Hartree*.

$2N + 1$	E_{HF}	E_{FCI}
3	$-\underline{2.86165}904$	$-\underline{2.9034}7799$
5	$-\underline{2.861677}82$	$-\underline{2.903501}48$
7	$-\underline{2.861678}42$	$-\underline{2.903502}77$
9	$-\underline{2.861678}44$	$-\underline{2.903502}97$
ref	-2.86167906	-2.90350597

Table 6. FCI energy of the H$^-$ anion obtained using five different box sizes. The exact digits are underlined. Energies are given in *Hartree*.

$2N + 1$	E_{FCI}
1	-0.393695
3	-0.490392
5	$-0.\underline{516930}$
7	$-0.\underline{524322}$
9	$-0.\underline{526522}$
ref	-0.527685

Table 7. Ground state energies of the Be atom obtained within the LCDAO approximation for different box sizes and different methods. Exact digits are underlined, and boldface digits highlight the lower energy obtained with the LCDAO approach. Energies are given in *Hartree*.

$2N+1$	HF	CASSCF(2,4)	CCSD
1	-14.26856824	-14.29070144	-14.28939455
3	$-14.\underline{54}550960$	-14.59399663	-14.59533382
5	-14.56978700	-14.61263992	$-14.\underline{61}465921$
7	$-14.\underline{57233}219$	$-14.\underline{61}484891$	$-14.\underline{61}696444$
9	$-14.\underline{572}\mathbf{48747}$	$-14.\underline{61}496826$	$-14.\underline{61}710109$
ref	-14.57233763	-14.61593821	-14.61782180

Table 8. Ground state energies of the Be$_4$ tetramer obtained within the LCDAO approximation. for different box sizes and different methods. Exact digits are underlined and boldface digits highlight the lower energy obtained with the LCDAO approach. Energies are given in *Hartree*.

$2N+1$	HF	CASSCF(8,10)	CCSD
1	-53.49773025	-54.01994232	-53.94998949
3	-57.99063568	-58.12724099	-58.19390694
5	-58.32481811	-58.45532242	$-\underline{58.53}694478$
7	$-\underline{58.35}101973$	-58.47700987	$-\underline{58.56}181230$
9	$-\underline{58.35}370683$	-58.47932486	$-\underline{58.56}\mathbf{513432}$
ref	-58.35445711	-58.48017161	-58.56252343
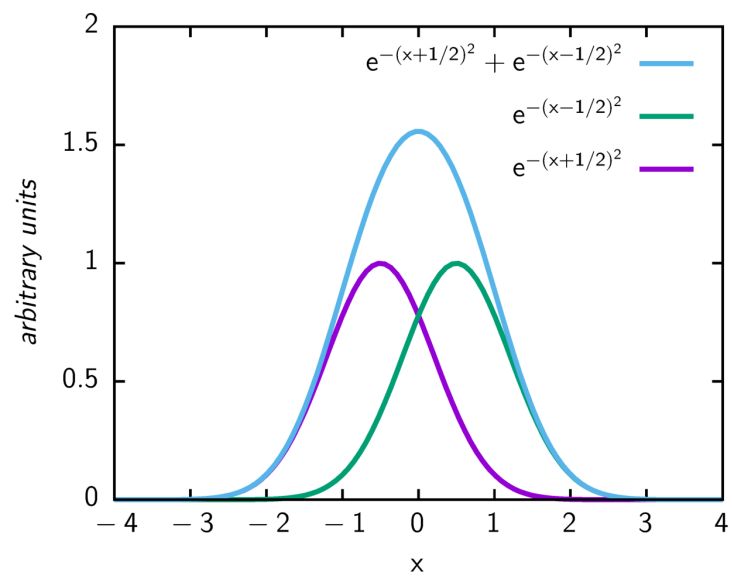
**Figure 1.** The linear combination of two equal-exponent Gaussian orbitals having $\xi < 2$, arbitrary units.
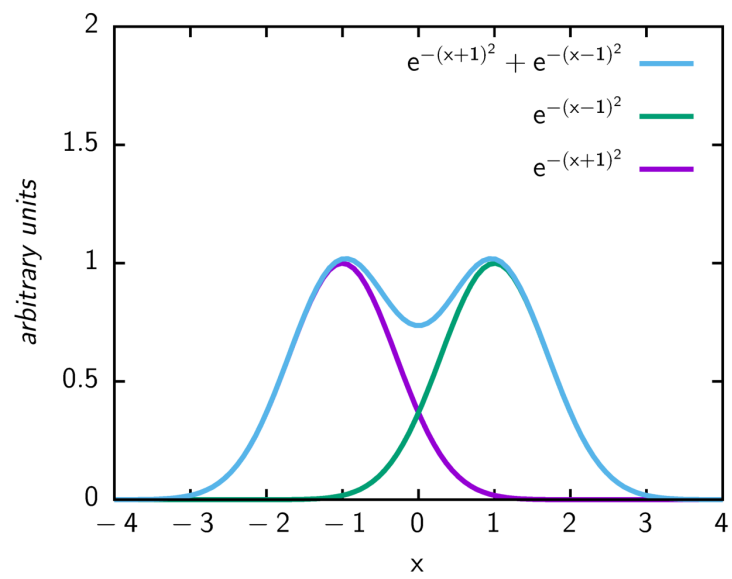
**Figure 2.** The linear combination of two equal-exponent Gaussian orbitals having $\xi > 2$, arbitrary units.
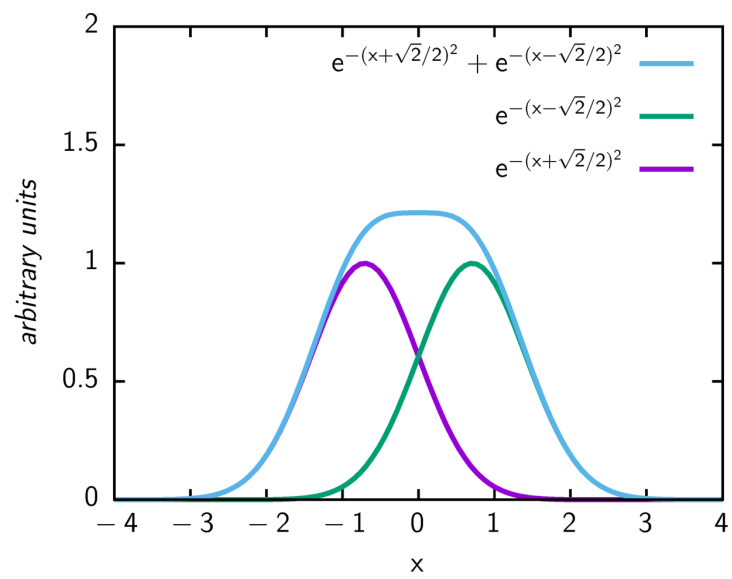
**Figure 3.** The linear combination of two equal-exponent Gaussian orbitals having $\xi = 2$, arbitrary units.
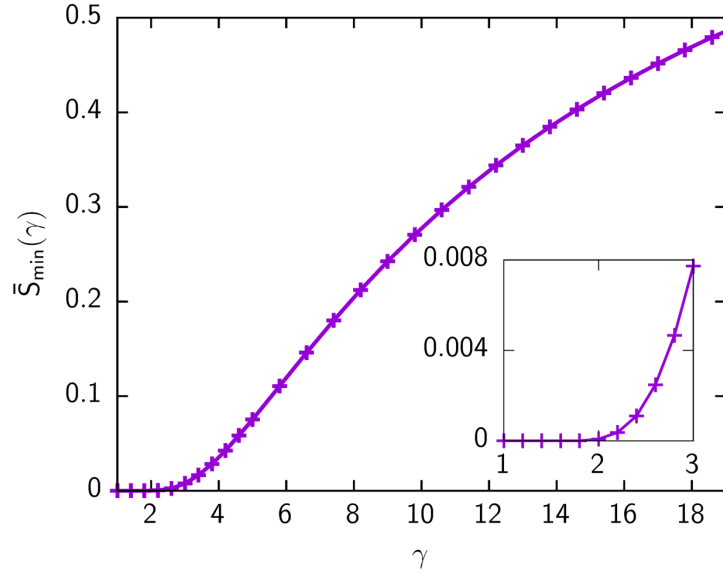
Figure 4. The minimum eigenvalue of the metric for a one-dimensional even-spaced infinite set of identical Gaussians as a function of γ. The inset zooms in to small values of γ. Note that values lower than one are omitted.
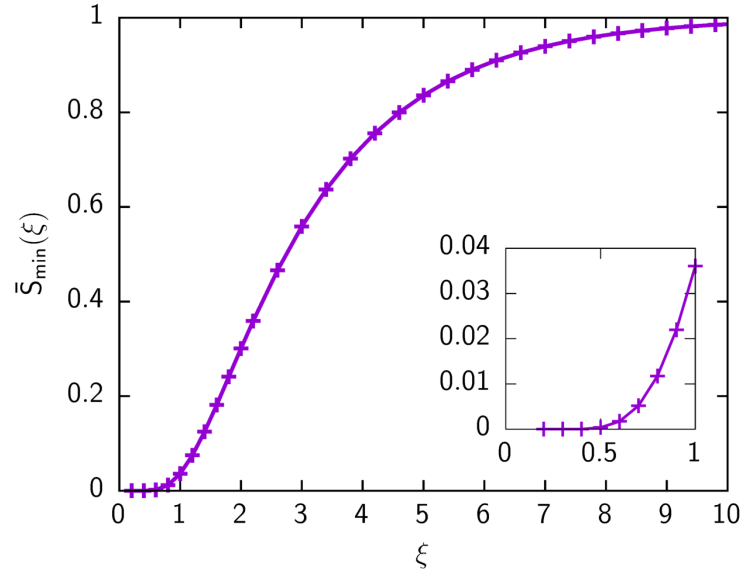
Figure 5. The minimum eigenvalue of the metric for a one-dimensional even-spaced infinite set of identical Gaussians as a function of ξ. The inset zooms in to small values of ξ.
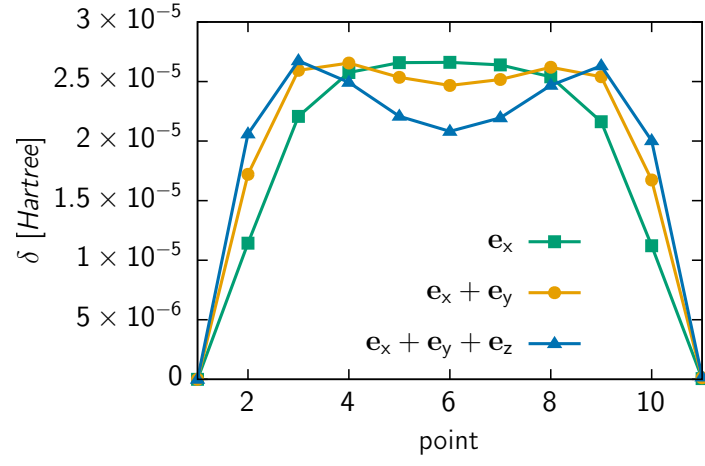
Figure 6. Electronic energy error δ with respect to the position of the hydrogen atom in the box of distributed Gaussians. A total of 11 equidistant points were considered for each direction.
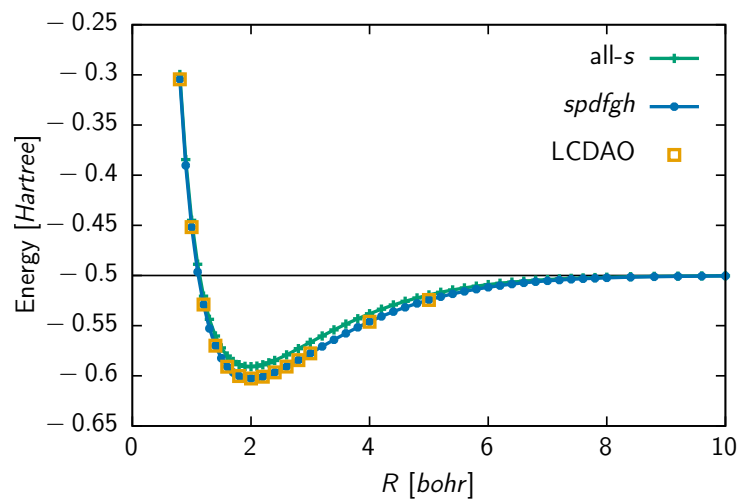
Figure 7. The H_2^+ potential energy curve as a function of the internuclear distance.