

# ADAPTIVE ALGORITHM FOR ELECTRONIC STRUCTURE CALCULATIONS USING REDUCTION OF GAUSSIAN MIXTURES

GREGORY BEYLKIN, LUCAS MONZÓN AND XINSHUO YANG

ABSTRACT. We present a new adaptive method for electronic structure calculations based on novel fast algorithms for reduction of multivariate mixtures. In our calculations, spatial orbitals are maintained as Gaussian mixtures whose terms are selected in the process of solving equations.

Using a fixed basis leads to the so-called “basis error” since orbitals may not lie entirely within the linear span of the basis. To avoid such an error, multiresolution bases are used in adaptive algorithms so that basis functions are selected from a fixed collection of functions, large enough to approximate solutions within any user-selected accuracy.

Our new method achieves adaptivity without using a multiresolution basis. Instead, as part of an iteration to solve nonlinear equations, our algorithm selects the “best” subset of linearly independent terms of a Gaussian mixture from a collection that is much larger than any possible basis since the locations and shapes of the Gaussian terms are not fixed in advance. Approximating an orbital within a given accuracy, our algorithm yields significantly fewer terms than methods using multiresolution bases.

We demonstrate our approach by solving the Hartree-Fock equations for two diatomic molecules,  $\text{HeH}^+$  and  $\text{LiH}$ , matching the accuracy previously obtained using multiwavelet bases.

## 1. INTRODUCTION

We present a new adaptive method for electronic structure calculations based on algorithms for reduction of multiresolution multivariate mixtures [12]. While we represent solutions using a linear combination of Gaussians (which has a long history in quantum chemistry), these Gaussians are not selected in advance to be used as an approximate basis but are generated in the process of solving equations. Thus, we avoid the so-called “basis error” usually associated with methods using Gaussians. Our approach can also be characterized as a “gridless” or a “meshfree” method.

Using Gaussians to find solutions of quantum chemistry problems has its origins in seminal papers [13, 24, 29] and is motivated by the fact that integrals involving these functions can be evaluated efficiently. In these early quantum chemistry papers the authors proposed to use linear combinations of Gaussians whose exponents and coefficients were found (or optimized) via Newton’s method in order to capture the correct behavior near the nuclear cusps and the correct rate of decay. However, this approach proved unsustainable as problems became larger. Instead, the construction of a basis for spatial orbitals has been performed off-line and the resulting sets of functions were then used as a fixed basis, leading to the so-called “basis error” if the actual solution is not well approximated within the linear span of such a basis.

The use of these Gaussian bases have revolutionized computational quantum chemistry in spite of the absence of a systematic way for controlling accuracy or providing guaranteed error bounds. In fact selecting an appropriate basis set became an art form requiring insight into the underlying solution. However, once a basis set is selected, the accuracy of the solution obtained using such a basis is ultimately limited. Since the equations being solved, e.g. the Hartree-Fock equations or the Kohn–Sham equations of density functional theory (DFT) (see e.g. [26]), only provide approximate solutions, the limitation in accuracy makes it difficult to separate the impact of using approximate equations from the approximate methods for solving them. For this reason, the use of adaptive methods to avoid the loss of accuracy caused by basis sets is highly desirable in this field.

The advent of multiresolution analysis (see e.g. [14, 3]) laid a conceptual foundation for adaptive methods but it took some time before practical adaptive algorithms were developed using multiwavelets [1]; see [19, 31, 32]. A central element for the success of the multiresolution algorithms can be traced to the fact that physically significant (integral) operators arising in problems of quantum chemistry are naturally represented by radial kernels, which, in turn, can be accurately approximated by a linear combination of Gaussians. The key advantage of using such approximations is that they yield a separated representation in which operators are efficiently applied one direction at a time. Without a separated representation, multiresolution operators would be too expensive to be practical in dimensions three and higher.

Multiresolution methods systematically refine basis functions (i.e., numerical grids) in the vicinity of the cusp-type singularities while using relatively few basis functions elsewhere. These methods have proven successful in efficiently computing highly accurate solutions, achieving guaranteed error bounds and eliminating the basis error. This approach has been implemented at ORNL in the software package MADNESS (Multiresolution ADaptive Numerical Environment for Scientific Simulation, see [18]), which is now considered the most accurate approach in this field [20]. Wavelets are also used in BigDFT [16] (cf. a mixed-basis method using plane waves and atom-centered radial polynomials [25] and interlocking multi-center grids [27], as examples).

However, since wavelets (or multiwavelets) do not resemble the spatial orbitals, a large number of basis functions is needed to represent solutions, e.g. an individual orbital may require  $\approx 2 \cdot 10^5$  basis functions in three dimensions. Moreover, such local refinement schemes do not take advantage of the essential simplicity of the spatial orbitals far from the nuclei and require boundary conditions to limit the computational domain. While adaptive multiresolution methods are sufficiently fast to be used within one-particle theories of quantum chemistry, an advancement towards two particle theories or solving Schrödinger’s equation had a limited success due to computational costs (see e.g. [7, 8]).

Hence it is of interest to develop new adaptive schemes and compare them with MADNESS. As it was demonstrated in [4], it is possible to iteratively solve equations of quantum chemistry using new algorithms based on nonlinear approximation of functions to reduce the number of terms in intermediate representations (without resorting to Newton’s method). The results of this approach are presented in [4] and are based on using Slater-type orbitals.

In this paper we present a new adaptive method that uses linear combinations of Gaussians to represent the solutions. We demonstrate the new approach by solving the Hartree-Fock equations for  $\text{HeH}^+$  and  $\text{LiH}$  so that we can compare the resulting representations with those in [19, 4]. As in [19, 4], we formulate the problem using integral equations and use a convergent iteration to solve them. In contrast to the above approaches, we use linear combinations of Gaussians to accurately approximate not only operators and potentials but also the functions on which they operate. As a result, all integrals can be computed explicitly and exactly by simply updating the parameters of the Gaussians involved. The computational effort thus moves from that of approximating and computing integrals to that of maintaining a reasonable number of terms in intermediate representations of solutions within the iterative scheme. Using our approach, computing a single integral (e.g. involving a Green’s function, potential and a wave function) can easily yield a Gaussian mixture with  $\approx 10^6$  terms, most of which are linearly dependent (within a user-selected accuracy). We use the reduction algorithm from [12, Algorithm 1] to reduce the number of terms after each operation by finding the “best” linearly independent terms, thus maintaining a reasonably small number of them. This algorithm has computational complexity  $\mathcal{O}(r^2N + p(d)rN)$ , where  $N$  is the original number of terms and  $r$  is the number of so-called skeleton terms selected by the algorithm and  $p(d)$  is the cost of computing the inner product between Gaussians as a function of dimension (in this paper  $d = 3$  and  $p(d)$  is a small constant). There is no underlying grid to maintain (thus, “gridless” or “meshfree” method) and there is no need to impose boundary conditions to reduce the size of the computational domain as in [19].

Since we only present results using examples of two small molecules, it is natural to ask if our approach generalizes to large molecules. The obvious obstacle is the quadratic dependence of the reduction algorithm on the number of skeleton terms (clearly, the final number of terms to represent solutions for large molecules is expected to be large). The key to addressing this problem is to avoid the unnecessary application of the reduction algorithm to sets of functions that are obviously linearly independent. Algorithmically, it means splitting terms of a Gaussian mixture into groups that are likely to have linear dependence (and never perform a global reduction). This approach allows us to maintain a reasonable number of skeleton terms within each group. Moreover, the resulting reduction algorithm is trivially parallel since the reduction within each group is independent of other groups. We demonstrate subdivision into groups using both location and scale in our examples and show that the accuracy of the solution improves gradually due to a convergent iteration.

We also observe that the Hartree–Fock setup does not take into account the electron–electron cusps whereas it is well known that incorporating them significantly improves the accuracy of energy calculations. In our approach, it would be natural to introduce a linear combination of Gaussian geminals (originally proposed in the seminal papers [13, 29]). Using Gaussian geminals to improve accuracy has a long history in quantum chemistry (see [23, 2] on this topic and references therein). We note that incorporating geminal functional forms into the adaptive approach of this paper appears possible (but by no means trivial) since the reduction algorithm for multivariate mixtures discussed in [12] (and on which the approach of this paper is based) is applicable to much more general functions than those used in our two examples. We also note that a different approach to incorporate cusps and to address basis set error was recently proposed in [17] by using a range-separated DFT formulation. We plan to address adaptivity involving electron–electron interactions and to demonstrate our approach on large molecules elsewhere.

We start by briefly stating the Hartree–Fock equations in Section 2 and demonstrating that the functional form of our approximate solution, a Gaussian mixture, is maintained when solving the integral form of these equations via iteration (we note that the same reasoning applies to the Kohn–Sham equations). Then we briefly review approximations of operators and potentials that we use in Section 4. We then turn to the reduction algorithm in Section 3 and describe a subdivision scheme which is critical to the practical use of our approach. We present examples and comparisons in Section 4 and discuss the results in Section 5

## 2. SOLVING THE HARTREE-FOCK EQUATIONS

We use an integral version of the Hartree–Fock equations (see e.g. [30]) and present these equations for the reader’s convenience in order to emphasize a combination of properties that make our approach possible. We note that our approach is equally applicable to the Kohn–Sham equations considered in [19]. Our method is based on computing orbitals as Gaussian mixtures and using the reduction algorithm in [12, Algorithm 1] to keep the number of terms under control.

Briefly, the occupied orbitals are the lowest  $N$  eigenfunctions of the Hartree–Fock operator

$$(2.1) \quad \mathcal{F} \phi_j(\mathbf{r}) = E_j \phi_j(\mathbf{r}), \quad j = 1, \dots, N,$$

where  $\mathcal{F} = -\frac{1}{2}\Delta + \mathcal{V}_{tot}$ ,  $\mathcal{V}_{tot} = \mathcal{V}_{ext} + 2\mathcal{J} - \mathcal{K}$ . The eigenvalues  $E_j < 0$  are referred to as orbital energies and are negative for the occupied orbitals. The external potential  $\mathcal{V}_{ext}$  accounts for attraction of electrons to the nuclei at locations  $\mathbf{R}_l$ ,  $l = 1, \dots, L$ ,

$$(2.2) \quad (\mathcal{V}_{ext}\phi)(\mathbf{r}) = \left( \sum_{l=1}^L \frac{Z_l}{\|\mathbf{r} - \mathbf{R}_l\|} \right) \phi(\mathbf{r}), \quad Z_m < 0,$$

where  $L$  is the number of nuclei,  $Z_m$  is the charge of the  $m$ th nucleus and  $\|\cdot\|$  is the standard Euclidean vector norm. The Coulomb operator  $\mathcal{J}$  describes the potential created by orbitals  $\phi_i(\mathbf{r})$ ,

$$(\mathcal{J}\phi)(\mathbf{r}) = \phi(\mathbf{r}) \left( -4\pi\Delta^{-1} \left( \sum_{i=1}^N |\phi_i(\mathbf{r})|^2 \right) \right),$$

and the exchange operator  $\mathcal{K}$  is defined as

$$(\mathcal{K}\phi)(\mathbf{r}) = \sum_{i=1}^N \phi_i(\mathbf{r}) (-4\pi\Delta^{-1} (\phi_i^* \phi)),$$

where  $\phi^*$  indicates the complex conjugate of the function  $\phi$ . The orbitals are obtained solving the  $N$  coupled equations (2.1) and the total energy,  $E_{tot}$ , is calculated by adding the nucleus–nucleus repulsion energies to the total electron energy

$$E_{tot} = \sum_{j=1}^N \left( E_j + \left\langle \left( -\frac{1}{2}\Delta + \mathcal{V}_{ext} \right) \phi_j, \phi_j \right\rangle \right) + \sum_{k=1}^L \sum_{l>k}^L \frac{Z_l Z_k}{\|R_l - R_k\|}.$$

In order to obtain integral equations for (2.1), we follow [21, 22, 19, 4] and use that the operator  $-\Delta + \mu^2$  has Green’s function

$$\mathcal{G}_\mu(\mathbf{r} - \mathbf{r}') = \frac{1}{4\pi} \frac{e^{-\mu\|\mathbf{r} - \mathbf{r}'\|}}{\|\mathbf{r} - \mathbf{r}'\|}.$$

That is,

$$(-\Delta + \mu^2) \mathcal{G}_\mu(\mathbf{r} - \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}').$$

Note that the kernel of  $\Delta^{-1}$  in the equations above is

$$\mathcal{G}_0(\mathbf{r} - \mathbf{r}') = \frac{1}{4\pi} \frac{1}{\|\mathbf{r} - \mathbf{r}'\|}$$

and that (2.1) is equivalent to

$$(2.3) \quad \phi_j(\mathbf{r}) = -2\mathcal{G}_{\mu_j} * (\mathcal{V}_{tot} \phi_j)(\mathbf{r}), j = 1, \dots, N,$$

where  $\mu_j^2 = -2E_j$ . We solve (2.3) via an iteration in which  $\mu_j$  are changing and approach  $\sqrt{-2E_j}$ . There are four observations (properties) that make our approach possible as follows.

- (1) The iteration to solve (2.3) is convergent (see additional comments below).
- (2) All potentials and Green’s functions can be accurately and efficiently approximated by linear combinations of Gaussians. This property of potentials and Green’s functions is the foundation of the approach in MADNESS, see e.g. [19, 18].
- (3) Since in our approach we represent orbitals  $\phi_j$  via linear combinations of Gaussians and potentials and Green’s functions are approximated by linear combinations of Gaussians, all integrals are evaluated explicitly.
- (4) In order to keep the number of terms in the representations of the orbitals  $\phi_j$  under control, the key tool in our approach is a reduction algorithm that selects a subset of the “best” linearly independent terms from a large number of terms which result from applying operators.

The integral form of the Kohn-Sham equations has the same properties (cf. [19], where properties 1 and 2 are essential) so that our approach can be used for these equations as well.

**2.1. An iteration to solve integral equations.** We solve (2.3) via the following iteration. We initialize using a collection of Gaussians centered inside the convex polyhedron defined by nuclear centers (see Remark 4).

Then, at step  $m$ , we first apply the operator  $\mathcal{V}_{tot}^{(m)}$  to the current approximation of the orbitals and compute

$$(2.4) \quad \begin{pmatrix} \mathcal{V}_{tot}^{(m)} \phi_1^{(m)} \\ \mathcal{V}_{tot}^{(m)} \phi_2^{(m)} \\ \vdots \\ \mathcal{V}_{tot}^{(m)} \phi_N^{(m)} \end{pmatrix} \leftarrow \begin{pmatrix} (\mathcal{V}_{ext} + 2\mathcal{J}^{(m)} - \mathcal{K}^{(m)}) \phi_1^{(m)} \\ (\mathcal{V}_{ext} + 2\mathcal{J}^{(m)} - \mathcal{K}^{(m)}) \phi_2^{(m)} \\ \vdots \\ (\mathcal{V}_{ext} + 2\mathcal{J}^{(m)} - \mathcal{K}^{(m)}) \phi_N^{(m)} \end{pmatrix}$$

where

$$\begin{aligned} (\mathcal{V}_{ext} \phi_j^{(m)}) (\mathbf{r}) &= \left( \sum_{l=1}^L \frac{Z_l}{\|\mathbf{r} - \mathbf{R}_l\|} \right) \phi_j^{(m)} (\mathbf{r}), \\ (\mathcal{J}^{(m)} \phi_j^{(m)}) (\mathbf{r}) &= \phi_j^{(m)} (\mathbf{r}) \left( -4\pi\mathcal{G}_0 * \left( \sum_{i=1}^N |\phi_i^{(m)} (\mathbf{r})|^2 \right) \right) \end{aligned}$$

and

$$(\mathcal{K}^{(m)} \phi_j) (\mathbf{r}) = \sum_{i=1}^N \phi_i^{(m)} (\mathbf{r}) \left( -4\pi\mathcal{G}_0 * (\phi_i^{(m)} \phi_j^{(m)}) \right)$$

(complex conjugation is removed noting that in our calculations orbitals are real). Next, we compute entries of the matrix

$$(2.5) \quad H_{ij}^{(m)} = -\frac{1}{2} \langle \Delta \phi_i^{(m)}, \phi_j^{(m)} \rangle + \langle \mathcal{V}_{tot}^{(m)} \phi_i^{(m)}, \phi_j^{(m)} \rangle, \quad i, j = 1, \dots, N,$$

and evaluate its eigenvalues, which we set to be approximations of the orbital energies  $E_i^{(m)}$ ,  $i = 1, \dots, N$ . We then compute

$$(2.6) \quad \mu_i^{(m)} = \sqrt{-2E_i^{(m)}}$$

and a new set of functions

$$(2.7) \quad \begin{pmatrix} \tilde{\phi}_1^{(m+1)} \\ \tilde{\phi}_2^{(m+1)} \\ \vdots \\ \tilde{\phi}_N^{(m+1)} \end{pmatrix} \leftarrow -2 \begin{pmatrix} \mathcal{G}_{\mu_1^{(m)}} * (\mathcal{V}_{tot}^{(m)} \phi_1^{(m)}) \\ \mathcal{G}_{\mu_2^{(m)}} * (\mathcal{V}_{tot}^{(m)} \phi_2^{(m)}) \\ \vdots \\ \mathcal{G}_{\mu_N^{(m)}} * (\mathcal{V}_{tot}^{(m)} \phi_N^{(m)}) \end{pmatrix}.$$

Finally, we orthogonalize (and normalize) the resulting functions  $\tilde{\phi}_j^{(m+1)}$ ,  $j = 1, \dots, N$ , yielding the updated orbitals  $\phi_j^{(m+1)}$ ,  $j = 1, \dots, N$ .

We observe that as long as the Green's functions, the Coulomb potentials and the orbitals are represented as Gaussian mixtures, then each operation described above maintains this form, i.e. the result is again a Gaussian mixture. However the number of terms grows rapidly and we show how to control the number of terms in Section 3.

*Remark 1.* This iteration is known to be convergent for computing bound states (see discussion in [19]) although we are not aware of a rigorous mathematical proof of this fact. An argument can be made that if the potential  $\mathcal{V}_{tot}$  is in the so-called Rollnik class (see [28]) then, for any fixed  $-\mu_i^2 < 2E_i$ ,  $i = 1, \dots, N$ , the norms of relevant operators are less than 1. This makes this iteration a fixed point iteration. The nuclear Coulomb potentials in  $\mathcal{V}_{ext}$  just miss being in the Rollnik class because of their slow decay away

from the nuclei. However, any truncation of these potentials that is sufficient for computing the bound states (at an arbitrarily large distance from nuclei) puts them back into the Rollnik class. In this paper we simply accept the fact that the iteration is convergent. We do truncate the nuclear Coulomb potentials at infinity in our approximations (see below) as is done (explicitly or implicitly) in all numerical methods for computing bound states.

**2.2. Accurate approximation of Green's functions and potentials.** We briefly recall the key approximation results in [9, 10] for the power functions  $r^{-\alpha}$ ,  $\alpha > 0$ , and the Green's function for the non-oscillatory Helmholtz equations,

$$\mathcal{G}_\mu(r) = \frac{1}{(2\pi)^{3/2}} \left(\frac{\mu}{r}\right)^{1/2} K_{\frac{1}{2}}(\mu r) = \frac{1}{4\pi} \frac{e^{-\mu r}}{r},$$

where  $K_{\frac{1}{2}}$  is the modified Bessel function of the second kind.

**Theorem 2.** [10, Theorem 5] *For any  $\alpha > 0$ ,  $\delta > 0$  and  $1/e \geq \varepsilon > 0$ , there exist a step size  $h$  and integers  $M$  and  $N$  such that*

$$|r^{-\alpha} - \mathcal{G}_0(r; M, N, h, \tau)| \leq r^{-\alpha} \varepsilon, \quad \text{for all } \delta \leq r \leq R,$$

where

$$(2.8) \quad \mathcal{G}_0(r; M, N, h, \tau) = \frac{h}{\Gamma(\alpha/2)} \sum_{n=M+1}^N e^{\alpha(hn-\tau)/2} e^{-e^{hn-\tau} r^2},$$

and  $\tau$  is any number  $0 \leq \tau < h$ .

The error estimates are based on discretizing the integral

$$\frac{1}{r^\alpha} = \frac{1}{\Gamma(\frac{\alpha}{2})} \int_{-\infty}^{\infty} e^{-r^2 e^t + \frac{\alpha}{2} t} dt$$

using (an infinite) trapezoidal rule

$$(2.9) \quad \frac{1}{\Gamma(\frac{\alpha}{2})} \int_{-\infty}^{\infty} e^{-r^2 e^t + \frac{\alpha}{2} t} dt \approx \frac{h}{\Gamma(\frac{\alpha}{2})} \sum_{n \in \mathbb{Z}} e^{\alpha(hn-\tau)/2} e^{-e^{hn-\tau} r^2},$$

where the step size  $h$  satisfies

$$h \leq \frac{2\pi}{\log 3 + \frac{\alpha}{2} \log(\cos 1)^{-1} + \log \varepsilon^{-1}}$$

and  $\varepsilon$  is any user-selected accuracy. The proof in [10, Theorem 5] allows one to shift the grid by any  $\tau$ ,  $0 \leq \tau < h$ .

For a given accuracy  $\varepsilon$ , power  $\alpha$  and a range of values  $r$ , the infinite sum (2.9) is then truncated due to the exponential or super-exponential decay of the integrand at  $\pm\infty$ , to yield a finite sum approximation in that range. The number of terms is estimated as

$$N - M \leq \frac{1}{10} (2 \log \varepsilon^{-1} + \log \alpha + 2) (\log \delta^{-1} + \frac{1}{\alpha} \log \varepsilon^{-1} + \log \log \varepsilon^{-1} + \frac{3}{2}).$$

This approximation provides an analytic construction of a multiresolution, separated representation for the Poisson kernel and the Coulomb potential in any dimensions. In practice many terms with small exponents in (2.8) can be combined using [9, Section 6] to reduce the total number of terms further. In our examples we use the following approximation of the Poisson kernel and the Coulomb potential via a linear combination of 146 Gaussians:

$$(2.10) \quad \left| r^{-1} - \sum_{n=1}^8 w_n e^{-\eta_n r^2} - \frac{h}{\Gamma(\frac{1}{2})} \sum_{n=-51}^{87} e^{(hn-\tau)/2} e^{-e^{hn-\tau} r^2} \right| \leq r^{-1} \varepsilon, \quad \text{for all } 10^{-7} \leq r \leq 10^5,$$

where  $h = 0.40994422603935795$ ,  $\tau = 0.192967891816239$ ,  $\varepsilon = 10^{-10}$  and exponents  $\eta_n$  and coefficients  $w_n$  are given in Table 1.

$n$	$\eta_n$	$w_n$
1	$2.1073876854180 \cdot 10^{-12}$	$3.2630674210379 \cdot 10^{-6}$
2	$1.8365780986634 \cdot 10^{-11}$	$3.1058837221013 \cdot 10^{-6}$
3	$4.7777245228151 \cdot 10^{-11}$	$2.8014247111005 \cdot 10^{-6}$
4	$8.5624630300630 \cdot 10^{-11}$	$2.5227064974618 \cdot 10^{-6}$
5	$1.3289239111902 \cdot 10^{-10}$	$2.7039982943831 \cdot 10^{-6}$
6	$2.0054640049463 \cdot 10^{-10}$	$3.2761422288967 \cdot 10^{-6}$
7	$3.0217586807074 \cdot 10^{-10}$	$4.0205002817225 \cdot 10^{-6}$
8	$4.5529860118663 \cdot 10^{-10}$	$4.9351231646262 \cdot 10^{-6}$

TABLE 1. Exponents and weights for the first eight terms in (2.10).

A similar result holds for the Green’s function for the non-oscillatory Helmholtz equation,

$$(2.11) \quad \mathcal{G}_\mu(r) = \frac{1}{4\pi} \frac{e^{-\mu r}}{r} = (4\pi)^{-\frac{3}{2}} \int_{-\infty}^{\infty} e^{-\frac{r^2 t^2}{4} - \mu^2 e^{-t} + \frac{t}{2}} dt,$$

where the integrand has a super-exponential decay at  $\pm\infty$ . An accurate and efficient approximation is obtained by discretizing this integral representation. We use

$$\mathcal{G}_\mu(r; M, N, h) = (4\pi)^{-3/2} h \sum_{l=M}^N e^{-\mu^2 e^{-hl} + \frac{hl}{2}} e^{-\frac{\|r\|^2 e^{hl}}{4}}$$

where the step size  $h$  is selected to achieve the desired accuracy  $\varepsilon$ ,

$$|\mathcal{G}_\mu(r) - \mathcal{G}_\mu(r; M, N, h)| \leq r^{-1} \varepsilon, \quad \text{for all } \delta \leq r \leq R.$$

Note that this integral is discretized for a range of the parameter  $\mu$  so that only the weights (coefficients) change as we change  $\mu$ . In our examples we approximate the Green’s function  $\mathcal{G}_\mu(r)$  (2.11) with

$$(2.12) \quad \mathcal{G}_\mu(r; -20, 120, 0.38190954773869346734)$$

resulting in  $\varepsilon = 10^{-10}$  for  $r$  in the range  $10^{-7} \leq r \leq 10^5$ .

### 3. REDUCTION ALGORITHM WITH SUBDIVISION OF TERMS INTO GROUPS

We approximate orbitals using Gaussian mixtures (cf. the multiwavelets in [19, 31, 32] and the Slater-type orbitals in [4]). While using a fixed basis set of Gaussians selected in advance has a long history in quantum chemistry, in contrast with previous methods we do not choose a set of Gaussians in advance (nor do we use a particular multiresolution basis of Gaussians although such an approach is possible using results in [11]). We allow the iterative algorithm for solving the equations in Section 2.1 to “select” the necessary basis functions. After each operation that yields a significant increase in the number of Gaussians, we prune the resulting Gaussian mixture by using the reduction algorithm in [12, Algorithm 1] that chooses the “best” linearly independent subset of terms, the so-called skeleton terms. By estimating the error of the solution, we terminate the iteration when the desired accuracy is achieved.

Conceptually our method is simpler and, also, less technical than either [19, 31, 32] or [4] and, as far as we know, is novel. Clearly, the key to enabling our approach is the reduction algorithm described and analyzed in [12]. This reduction algorithm has complexity  $\mathcal{O}(r^2 N + p(d) r N)$ , where  $N$  is the initial number of terms,  $r$  is the number of skeleton terms and  $p(d)$  is the cost of computing the inner product between the terms of a Gaussian mixture ( $p(d)$  is a small constant in dimension  $d = 3$ ). If the number of skeleton terms,  $r$ , is small, the computational cost of the reduction algorithm is completely satisfactory.

However, we cannot assume that the number of skeleton terms (which are used to represent orbitals) is small for a large molecule.

In this paper we introduce an important modification of the reduction algorithm to avoid an excessive computational cost due to the quadratic dependence on the number of skeleton terms. The modification stems from a simple observation that, if two sets of functions are known to be linearly independent in advance, an attempt to find skeleton terms is wasteful. In many applied problems a multivariate mixture with a large number of terms can be split into subgroups of terms so that each subgroup reflects the local behavior of the function. For example, it is well understood that orbitals have cusps at nuclear centers but otherwise are smooth functions (in fact it is this structure of solutions that underpins the success of multiresolution methods such as [19, 31, 32]). Well-localized terms associated with cusps decay rapidly away from a nuclear center so that these terms will be linearly independent (or even nearly orthogonal) from similar terms at other nuclear centers. This suggests an acceleration technique for the reduction algorithm based on a hierarchical subdivision of the terms of the mixture by scale and location. Given such a subdivision, reduction is performed only within a local group. In such an approach the reduction of terms within each group is independent from other groups and, thus, the computation is trivially parallel.

In [12] we consider Gaussian terms in the form of  $L^2$ -normalized multivariate normal distribution  $N(\mathbf{x}, \mathbf{s}_k, \Sigma_k)$ ,

$$(3.1) \quad \begin{aligned} g_k(\mathbf{x}, \mathbf{s}_k, \Sigma_k) &= (\det(4\pi\Sigma_k))^{1/4} N(\mathbf{x}, \mathbf{s}_k, \Sigma_k) \\ &= \frac{1}{(\det(\pi\Sigma_k))^{1/4}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{s}_k)^T \Sigma_k^{-1} (\mathbf{x} - \mathbf{s}_k)\right), \end{aligned}$$

where the vector  $\mathbf{s}_k$  defines its location and the symmetric positive definite matrix  $\Sigma_k$  controls its shape. In the setup of this paper, all matrices  $\Sigma_k$  are diagonal and, therefore, functions represented as Gaussian mixtures are in separated form (see [5, 6]). In fact, in our examples it is sufficient to use rotationally symmetric Gaussians so that their shape is controlled by a single scalar parameter.

We subdivide terms by grouping them by their scale (shape) and location. We consider a single (global) group of flat Gaussians. i.e. Gaussians that have a significant support around more than one nuclear center and control behavior of orbitals far away from nuclear centers. The rest of the terms we split into groups according to their proximity to nuclear centers using nuclear centers as ‘‘seeds’’ in a Voronoi-type decomposition: all terms are split into groups by their proximity to the nuclei, i.e. a term located at  $\mathbf{s}_k$  belongs to a group associated with a given nucleus at location  $\mathbf{R}_l$  if it is the closest to it among all other nuclei, i.e.  $\|\mathbf{R}_l - \mathbf{s}_k\| \leq \min_{l' \neq l} \|\mathbf{R}_{l'} - \mathbf{s}_k\|$ . Such subdivision ensures that the cost of reduction is proportional to the number of nuclei since the number of skeleton terms associated with each group is expected to be roughly the same given the structure of functions they represent. In our experiments we observed that the number of skeleton terms in these groups is still fairly large (e.g. several thousands terms) so further subdivision by scale and location is appropriate. Towards this end, we associate with each nucleus center and its group the distance  $s_l^{\max} = \max_k \|\mathbf{R}_l - \mathbf{s}_k\|$  and subdivide each group further by considering two parameters: (i) the radial intervals  $[\frac{m}{2^j} s_l^{\max}, \frac{m+1}{2^j} s_l^{\max}]$ ,  $m = 0, 1, \dots, 2^j - 1$  that restrict the locations of the terms; (ii) the ranges of exponents  $[4^{-j-1} \sigma^{\text{far}}, 4^{-j} \sigma^{\text{far}}]$ , where  $\sigma^{\text{far}}$  is a parameter (see discussion below), identifying the shape of the terms. Within each radial interval  $[\frac{m}{2^j} s_l^{\max}, \frac{m+1}{2^j} s_l^{\max}]$ , terms may have significantly different shapes determined by the shape matrices  $\Sigma_k$  and, thus, cannot be linearly dependent. Effectively, it is a multiresolution subdivision strategy.

In the examples in Section 4, we initialize the spatial orbitals so that every term in the mixture has a diagonal shape matrix  $\Sigma_k = \text{diag}(\sigma_k, \dots, \sigma_k) = \sigma_k I$ , where  $I$  is the identity matrix. This forces all Gaussians involved in further computations to be rotationally symmetric, i.e. to have  $\Sigma_k$  of the same structure. As already stated, we first identify a global group by considering all terms with  $\sigma_k \in [\sigma^{\text{far}}, \infty)$ , where  $\sigma^{\text{far}} > 0$  is a parameter. Terms assigned in this group are significant in the vicinity of more than

one nuclear center and capture the far-field behavior. The rest of the terms we divide into groups by their proximity to the nuclei and scale (shape). Within each group associated with nuclei  $\mathbf{R}_l$ , we identify a Gaussian term  $g_k$  as a member of a subgroup, denoted by  $G_{jm}^l$  if

$$\|\mathbf{s}_k - \mathbf{s}_l^{\max}\| \in \left[ \frac{m}{2^j} s_l^{\max}, \frac{m+1}{2^j} s_l^{\max} \right] \quad \text{and} \quad \sigma_k \in \left[ 4^{-j-1} \sigma^{\text{far}}, 4^{-j} \sigma^{\text{far}} \right],$$

where  $j = 0, 1, \dots, J$ . The index  $m$  controls the location and the index  $j$  controls the scale (shape) of a term. The role of these parameters is similar to those labeling basis functions in Multiresolution Analysis (MRA).

*Remark 3.* In general, subdivision into groups may differ in detail from the one we use for our examples in this paper. This depends on the type of functions we want to represent. In particular, the number of indices controlling location of terms (and, therefore, subgroups) can be larger than what we use in our examples (e.g. there could be three such indices).

We note that, for large  $j$  and  $m$ , the corresponding group  $G_{jm}^l$  consists of terms with a small essential support and located (relatively) far away from the nucleus center  $\mathbf{R}_l$ . Such terms (and, therefore, groups) are not needed to represent orbitals since orbitals have cusps only at the nuclei. To reduce the number of unnecessary groups, we do not maintain such groups.

In order to capture the cusp behavior, we introduce in the vicinity of nuclei an additional set of groups  $G_{j,0}^l$ ,  $j = J+1, \dots, \tilde{J}$ , so that  $g_k \in G_{j,0}^l$  if

$$\|\mathbf{s}_k - \mathbf{s}_l^{\max}\| \in \left[ 0, \frac{s_l^{\max}}{2^j} \right] \quad \text{and} \quad \sigma_k \in \left[ 4^{-j-1} \sigma^{\text{far}}, 4^{-j} \sigma^{\text{far}} \right].$$

In our computation we ignore any terms that are not included in the groups  $\left\{ G_{jm}^l \right\}_{j=0, \dots, J}^{m=0, \dots, 2^j-1}$  or  $\left\{ G_{j,0}^l \right\}_{j=J+1, \dots, \tilde{J}}$  for all nucleus locations  $\mathbf{R}_l$ . Selecting  $J = 4$  and  $\tilde{J} = 26$ , the total number of groups (including the global group) is 107. It is worth noting that, when reducing a Gaussian mixture according to the group subdivision, some of the groups might be empty.

To estimate how subdivision into groups speeds up computations, let us consider a multivariate mixture with  $N$  terms where the number of linearly independent terms is  $r$ . Let us assume that we can subdivide these terms into  $m$  groups with an equal number of terms (for simplicity of the estimate). We assume that the number of linearly independent terms in each group is then at most  $r/k$ , where  $k > 1$ . The cost of performing the reduction algorithm for one group is then  $\mathcal{O} \left( \left( \frac{r}{k} \right)^2 \frac{N}{m} + p(d) \frac{r}{k} \frac{N}{m} \right)$  and, thus, the total cost to reduce all groups is  $\mathcal{O} \left( \left( \frac{r}{k} \right)^2 N + p(d) \frac{r}{k} N \right)$ . Depending on  $k$ , we get a significant speed-up factor and note that the reduction of independent groups of terms is trivially parallel (a property that we did not use it in our computations). With the subdivision into groups described above, the number of groups is naturally proportional to the number of nuclear centers. Since the solution has a cusp at the nuclear center, we can consider  $r/k$  to be a constant (i.e., each cusp locally requires roughly the same number of terms). Hence, the reduction algorithm with subdivision into groups is linear in the number of nuclear centers.

*Remark 4.* By examining operators in (2.4)-(2.7), we observe that if we start an iteration with a Gaussian mixture with diagonal shape matrices  $\Sigma_k$ , then all resulting Gaussian mixtures involved in the computation will have diagonal shape matrices as well. Moreover, location of the terms is restricted to a box defined by positions of nuclei. We observe that, in (2.4)-(2.7), convolutions and multiplications produce new Gaussians (with different shifts and/or shape matrices). While convolution does not change the center of the resulting Gaussian terms, multiplication of two Gaussians generates a Gaussian centered at a different location. Specifically, we have for the product of two normal distributions

$$(3.2) \quad N(\mathbf{x}, \mu_1, \Sigma_1) N(\mathbf{x}, \mu_2, \Sigma_2) = N(\mu_1, \mu_2, \Sigma_1 + \Sigma_2) \cdot N\left(\mathbf{x}, \mu_c, (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}\right)$$

where

$$\mu_c = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1} (\Sigma_1^{-1} \mu_1 + \Sigma_2^{-1} \mu_2).$$

If  $\Sigma_1 = \text{diag}(\alpha_1, \alpha_2, \alpha_3)$ ,  $\Sigma_2 = \text{diag}(\beta_1, \beta_2, \beta_3)$ ,  $\mu_1 = (x_1, x_2, x_3)$  and  $\mu_2 = (y_1, y_2, y_3)$ , then we have

$$\mu_c = \left( \frac{\beta_1}{\alpha_1 + \beta_1} x_1 + \frac{\alpha_1}{\alpha_1 + \beta_1} y_1, \frac{\beta_2}{\alpha_2 + \beta_2} x_2 + \frac{\alpha_2}{\alpha_2 + \beta_2} y_2, \frac{\beta_3}{\alpha_3 + \beta_3} x_3 + \frac{\alpha_3}{\alpha_3 + \beta_3} y_3 \right).$$

Hence, all three coordinates of the new center  $\mu_c$  are in between the corresponding coordinates of  $\mu_1$  and  $\mu_2$ . This implies that all Gaussian terms resulting from iteration (2.4)-(2.7) will have their centers inside a box defined by the minimum and the maximum of individual coordinates of nuclear centers  $\mathbf{R}_i$  in (2.2) (in fact, this result holds in any dimension). Moreover, if the Gaussians are rotationally invariant, then all terms will have their centers inside a convex polyhedron defined by the nuclear centers  $\mathbf{R}_i$ . We always initialize the iteration using Gaussians centered inside the convex polyhedron defined by the nuclear centers.

We note that the restricted location of terms of a Gaussian mixture in our approach cannot be achieved when using wavelet or multiwavelet bases, which force a computational box that is much larger than the box defined by nuclear centers.

#### 4. EXAMPLES OF COMPUTATIONS FOR TWO SMALL MOLECULES

4.1. **Helium Hydride, HeH<sup>+</sup>**. First we solve the Hartree–Fock equation for HeH<sup>+</sup>

$$(4.1) \quad \left( -\frac{1}{2}\Delta + \mathcal{V}_{ext} - 4\pi\Delta^{-1}(|\phi|^2) \right) \phi = E\phi,$$

with the potential

$$\mathcal{V}_{ext}(\mathbf{r}) = \frac{Z_1}{\|\mathbf{r} - \mathbf{R}_1\|} + \frac{Z_2}{\|\mathbf{r} - \mathbf{R}_2\|}.$$

where  $Z_1 = -1$ ,  $Z_2 = -2$ ,  $\mathbf{R}_1 = (0, 0, -0.7)$  and  $\mathbf{R}_2 = (0, 0, 0.7)$ .

As described in Section 2.1, our basic approach involves recasting (4.1) as an integral equation which we solve iteratively. The iteration (see below) is convergent and, while the rate of convergence is slower than quadratic, it is sufficiently fast so that only a few dozen iterations are required (see discussion in [19] and references therein). We represent the spatial orbital  $\phi(\mathbf{r})$  as a Gaussian mixture

$$\phi(\mathbf{r}) = \sum_{k=1}^K c_k g_k(\mathbf{r}, \mathbf{s}_k, \Sigma_k), \quad \mathbf{r} \in \mathbb{R}^3,$$

that is gradually constructed via iterations. In each step of the iterative solution, the number of terms in representing  $\phi(\mathbf{r})$  grows and we use [12, Algorithm 1] to control their number.

We approximate potentials  $Z_1/\|\mathbf{r} - \mathbf{R}_1\|$  and  $Z_2/\|\mathbf{r} - \mathbf{R}_2\|$  and the Poisson kernel using (2.10) and the Green's function  $\mathcal{G}_\mu(r)$  in (2.11) using (2.12).

	$N_{global}$	$N_{groups}$	$N_{max}$	$N_{min}$	$N_{ave}$
$\phi$	137	28	69	2	50.9
$\mathcal{V}_{tot}\phi$	137	70	70	4	32.8

TABLE 2. The number of terms in the global group  $N_{global}$ , the number of non-empty groups  $N_{groups}$  and the maximum  $N_{max}$ , minimum  $N_{min}$  and average  $N_{ave}$  number of terms in these non-empty groups in Gaussian mixtures representing  $\phi$  and  $\mathcal{V}_{tot}\phi$  at the final step.

For the reader's convenience, we explicitly describe the iteration for solving (4.1) ,

$$\begin{aligned}
(4.2) \quad \mathcal{V}_{tot}^{(m)}\phi^{(m)} &\leftarrow \left( \mathcal{V}_{ext} - 4\pi\mathcal{G}_0 * \left( |\phi^{(m)}|^2 \right) \right) \phi^{(m)} \\
E^{(m)} &\leftarrow \left\langle -\frac{1}{2}\Delta\phi^{(m)} + \mathcal{V}_{tot}^{(m)}\phi^{(m)}, \phi^{(m)} \right\rangle, \\
\mu^{(m)} &\leftarrow \sqrt{-2E^{(m)}}, \\
\phi^{(m+1)} &\leftarrow -2\mathcal{G}_{\mu^{(m)}} * \left( \mathcal{V}_{tot}^{(m)}\phi^{(m)} \right), \\
\phi^{(m+1)} &\leftarrow \frac{\phi^{(m+1)}}{\|\phi^{(m+1)}\|_2}, \quad m = 0, 1, \dots,
\end{aligned}$$

where  $*$  denotes convolution.

To start the iteration, we use the single Gaussian

$$\phi^{(0)}(\mathbf{r}) = \pi^{-\frac{3}{4}} e^{-\frac{\|\mathbf{r}\|^2}{2}}.$$

We choose  $\varepsilon = 10^{-6}$  as the error tolerance for the reduction of terms in each group and terminate the iteration if  $|E^{(m)} - E^{(m-1)}| < 4 \times 10^{-6}$ . After 20 iterations we obtain the orbital energy  $E = -1.6605545$  and the total energy  $E_{tot} = -2.9325704$ . Using the MADNESS software [15] yields orbital energy  $E = -1.66053903$  and total energy  $-2.93256741$  so that, using these values as reference, our computations have absolute error  $1.5 \cdot 10^{-5}$  for the orbital energy and  $3.0 \cdot 10^{-6}$  for the total energy. As a result of the iteration, the total number of terms representing the orbital  $\phi(\mathbf{r})$  is 1563. This number of terms is significantly smaller than the  $\approx 2 \cdot 10^5$  terms required when using multiwavelets [19, 31, 32] and a factor  $\approx 2$  larger than that in [4], where the number of Slater-type orbitals is 637. However, our approach is less technical than that in [4] and, for this reason, it is easier to extend it to deal with large molecules. In Table 2, we show the number of terms in the global group  $N_{global}$ , the number of non-empty groups  $N_{groups}$  and the maximum  $N_{max}$ , minimum  $N_{min}$  and average  $N_{ave}$  number of terms in these non-empty groups in  $\phi$  and  $\mathcal{V}_{tot}\phi$  at the final step.

In Figure 4.1, we display the spatial orbital  $\phi(\mathbf{r})$  on the line  $\mathbf{r} = (0, 0, x)$  connecting the two nuclei locations  $\mathbf{R}_1$  and  $\mathbf{R}_2$ .

**4.2. Lithium Hydride, LiH.** For our second example, we consider the Hartree-Fock equations for lithium hydride, LiH. The Hartree-Fock equations in this case are

$$(4.3) \quad \mathcal{F}\phi_j(\mathbf{r}) = E_j\phi_j(\mathbf{r}), \quad j = 1, 2,$$

where  $\mathcal{F} = -\frac{1}{2}\Delta + \mathcal{V}_{ext} + 2\mathcal{J} - \mathcal{K}$ ,

$$\begin{aligned}
\mathcal{J}\phi_j &= \phi_j \left( -4\pi\Delta^{-1} \left( |\phi_1|^2 + |\phi_2|^2 \right) \right), \\
\mathcal{K}\phi_j &= \phi_1 \left( -4\pi\Delta^{-1} (\phi_1^* \phi_j) \right) + \phi_2 \left( -4\pi\Delta^{-1} (\phi_2^* \phi_j) \right),
\end{aligned}$$

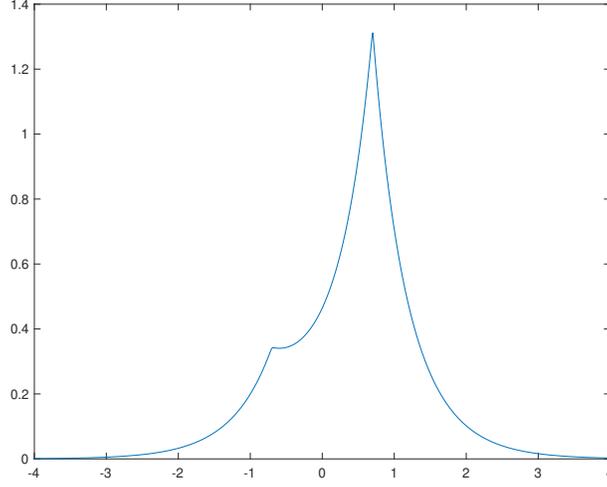


FIGURE 4.1. Plot of the spatial orbital  $\phi(\mathbf{r})$  for helium hydride  $\text{HeH}^+$  on the line  $\mathbf{r} = (0, 0, x)$  connecting the two nuclear centers  $\mathbf{R}_1 = (0, 0, -0.7)$  and  $\mathbf{R}_2 = (0, 0, 0.7)$ .

and

$$\mathcal{V}_{ext}(\mathbf{r}) = \frac{Z_1}{\|\mathbf{r} - \mathbf{R}_1\|} + \frac{Z_2}{\|\mathbf{r} - \mathbf{R}_2\|}.$$

For lithium hydride we have  $Z_1 = -3$ ,  $Z_2 = -1$ ,  $\mathbf{R}_1 = (-3.15/2, 0, 0)$  and  $\mathbf{R}_2 = (3.15/2, 0, 0)$ . We approximate potentials  $Z_1/\|\mathbf{r} - \mathbf{R}_1\|$  and  $Z_2/\|\mathbf{r} - \mathbf{R}_2\|$  and the Poisson kernel using (2.10) and the Green's function  $\mathcal{G}_\mu(r)$  in (2.11) using (2.12).

To solve (4.3), we initialize orbitals  $\phi_j^{(0)}(\mathbf{r})$ ,  $j = 1, 2$  as

$$\phi_1^{(0)}(\mathbf{r}) = e^{-\frac{\|\mathbf{r} - \mathbf{R}_1\|^2}{20}} \quad \text{and} \quad \phi_2^{(0)}(\mathbf{r}) = e^{-\frac{\|\mathbf{r} - \mathbf{R}_2\|^2}{20}}.$$

There are many possible initializations, e.g., we can use an approximation to  $e^{-\|\mathbf{r} - \mathbf{R}_1\|}$  and  $e^{-\|\mathbf{r} - \mathbf{R}_2\|}$  via Gaussian mixtures. Importantly, the initial approximation should be chosen so that the initial orbital energies are negative. After orthonormalizing the functions  $\phi_j^{(0)}(\mathbf{r})$ , we compute the initial energies  $E_j^{(0)}$  as the eigenvalues of the  $2 \times 2$  matrix defined in (2.5). We then set  $\mu_j^{(0)}$ ,  $j = 1, 2$  using (2.6) and update the orbitals via (2.7).

Finally, when the desired accuracy is achieved, we compute the total energy  $E_{tot}$  as

$$E_{tot} = \sum_{j=1}^2 \left( E_j + \left\langle \left( -\frac{1}{2}\Delta + \mathcal{V}_{ext} \right) \phi_j, \phi_j \right\rangle \right) + \frac{Z_1 Z_2}{\|\mathbf{R}_1 - \mathbf{R}_2\|}.$$

In this example, we choose an error tolerance of  $\varepsilon = 10^{-6}$  for the reduction of terms in each group. We terminate the computation if  $|E_j^{(m)} - E_j^{(m-1)}| < 4 \times 10^{-6}$  for  $j = 1, 2$ . After 28 iterations, we arrived at the orbital energies  $E_1 = -2.451757267$  and  $E_2 = -0.297819313$  computed with absolute errors of  $5.7 \times 10^{-6}$  and  $3.7 \times 10^{-6}$ , respectively, when compared with the reference energies  $E_1 = -2.451763$  and  $E_2 = -0.297823$  obtained using the MADNESS software. The computed total energy  $E_{tot} = -7.9869324$  has an absolute error of  $4.0 \times 10^{-6}$  compared with the value  $E_{tot} = -7.9869364$  evaluated in MADNESS. The total number of terms to represent orbitals  $\phi_1(\mathbf{r})$  and  $\phi_2(\mathbf{r})$  is 2185 and 2569. In Table 3, we show the number of terms in the global group  $N_{global}$ , the number of non-empty groups  $N_{groups}$  and the maximum

	$N_{global}$	$N_{group}$	$N_{max}$	$N_{min}$	$N_{ave}$
$\phi_1$	218	26	108	1	75.7
$\phi_2$	217	31	107	1	75.9
$\mathcal{V}_{tot}\phi_1$	220	68	107	3	43.0
$\mathcal{V}_{tot}\phi_2$	215	69	107	4	46.2

TABLE 3. The number of terms in the global group  $N_{global}$  the number of non-empty groups  $N_{group}$  and the maximum  $N_{max}$ , minimum  $N_{min}$  and average  $N_{ave}$  number of terms in these non-empty groups in  $\phi_1$ ,  $\phi_2$ ,  $\mathcal{V}_{tot}\phi_1$  and  $\mathcal{V}_{tot}\phi_2$  at the final step.

$N_{max}$ , minimum  $N_{min}$  and average  $N_{ave}$  number of terms in the non-empty groups in the Gaussian mixture representations of  $\phi_1$ ,  $\phi_2$ ,  $\mathcal{V}_{tot}\phi_1$  and  $\mathcal{V}_{tot}\phi_2$  at the final step.

The resulting number of terms is again much smaller (about 100 times smaller) than that needed using MADNESS software and by a factor  $\approx 2$  larger than that in [4], where the number of terms to represent the Slater-type orbitals is 1282 and 1327.

In Figure 4.2 we display the spatial orbitals  $\phi_1(\mathbf{r})$  and  $\phi_2(\mathbf{r})$  on the line  $\mathbf{r} = (x, 0, 0)$  connecting the two nuclei locations  $\mathbf{R}_1$  and  $\mathbf{R}_2$ .

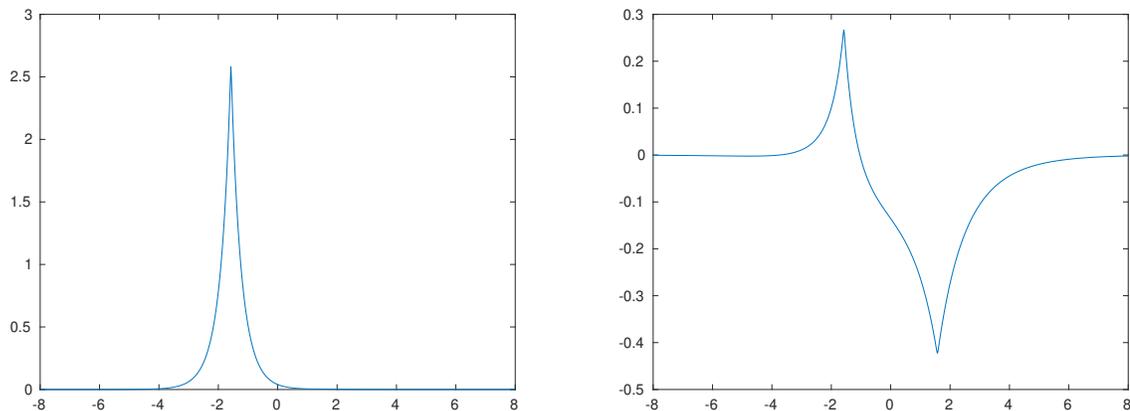


FIGURE 4.2. Plot of the spatial orbitals for lithium hydride, LiH,  $\phi_1(\mathbf{r})$  (left) and  $\phi_2(\mathbf{r})$  (right) on the line  $\mathbf{r} = (x, 0, 0)$  connecting the two nucleus locations  $\mathbf{R}_1 = (-3.15/2, 0, 0)$  and  $\mathbf{R}_2 = (3.15/2, 0, 0)$ .

*Remark 5.* We implemented our code in Fortran90 and compiled it with the Intel Fortran Compiler version 19.0.1.144. All computations were performed on a single core (without any parallelization) of an Intel i7-6700 CPU at 3.4 GHz on a 64-bit Linux workstation with 64 GB of RAM. Currently it takes about 119 seconds to solve the Hartree–Fock equations for HeH<sup>+</sup> and 999 seconds to solve the equations for LiH. We made no attempt to optimize our implementation. While the reduction algorithm that splits Gaussian mixtures into groups (see Section 3) is trivially parallel, we did not take advantage of this property in the current version of our code. Since, for small molecules, the main cost is in reduction, we expect a speedup factor of about 20, given the number of groups in Table 3. Such an acceleration factor should bring the timing close to that of MADNESS. We plan to do a careful speed comparison with existing methods separately.

*Remark 6.* The reduction algorithm we used [12, Algorithm 1] is capable of achieving near full double precision accuracy if certain computations are carried out in quadruple precision. In addition to solving (4.1) and (4.3) using the reduction algorithm implemented in double precision and obtaining orbital energies  $\approx 10^{-6}$ , we continued iterations and switched to the reduction algorithm implemented in quadruple precision to improve the accuracy. We observed that, within several additional iterations, the orbital energies improved by gaining additional accurate digits. However, as pointed out in [12], the quadruple precision implementation is more than 10 times slower than the double precision version (if the same accuracy is required). We do not include these results since, by modifying the reduction algorithm, it should be possible to avoid using quadruple precision. However, we note that, in most practical computations, an accuracy of  $\approx 10^{-6}$  is typically sufficient.

## 5. CONCLUSIONS AND FURTHER WORK

We presented a new adaptive algorithm for computing the electronic structure of molecules. Our approach is novel since we do not select a basis in advance and let the iteration for solving a system of integral equations identify the necessary functions to achieve a desired accuracy. The reduction algorithm developed in [12] selects the “best” subset of linearly independent terms after each operation that generates a large number of terms. These terms are Gaussians (identified by their parameters) and, since potentials and Green’s functions are also represented via Gaussian mixtures, all integrals are evaluated explicitly. Hence, in the algorithm, we only update the parameters of Gaussian mixtures.

We introduce in this paper a hierarchical subdivision of terms of Gaussian mixtures so that the computational cost of reduction is proportional to the number of nuclear centers. Although we did not implement a parallel version of our approach, it is clear that such an implementation will have a significant impact on the speed since reduction within each group of terms is independent of the other groups and computations are also mostly independent between orbitals.

We note that our method is less technical than that used in MADNESS (see [18]) and in [4] since our only tool is the reduction algorithm. We plan further work on our approach to verify its performance on large molecules.

## ACKNOWLEDGEMENT

We thank the anonymous reviewers for their comments and encouragement.

## REFERENCES

- [1] B. Alpert. A class of bases in  $L^2$  for the sparse representation of integral operators. *SIAM J. Math. Anal.*, 24(1):246–262, 1993.
- [2] G. M. J. Barca and P-F. Loos. Three- and four-electron integrals involving gaussian geminals: fundamental integrals, upper bounds, and recurrence relations. *The Journal of chemical physics*, 147(2):024103, 2017.
- [3] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms, I. *Comm. Pure Appl. Math.*, 44(2):141–183, 1991. Yale Univ. Technical Report YALEU/DCS/RR-696, August 1989.
- [4] G. Beylkin and T. S. Haut. Nonlinear approximations for electronic structure calculations. *Proc. R. Soc. A*, 469(20130231), 2013.
- [5] G. Beylkin and M. J. Mohlenkamp. Numerical operator calculus in higher dimensions. *Proc. Natl. Acad. Sci. USA*, 99(16):10246–10251, August 2002.
- [6] G. Beylkin and M. J. Mohlenkamp. Algorithms for numerical analysis in high dimensions. *SIAM J. Sci. Comput.*, 26(6):2133–2159, July 2005.
- [7] G. Beylkin, M. J. Mohlenkamp, and F. Pérez. Preliminary results on approximating a wavefunction as an unconstrained sum of Slater determinants. *Proceedings in Applied Mathematics and Mechanics*, 7(1):1010301–1010302, 2007.
- [8] G. Beylkin, M. J. Mohlenkamp, and F. Pérez. Approximating a wavefunction as an unconstrained sum of Slater determinants. *Journal of Mathematical Physics*, 49(3):032107, 2008.
- [9] G. Beylkin and L. Monzón. On approximation of functions by exponential sums. *Appl. Comput. Harmon. Anal.*, 19(1):17–48, 2005.

- [10] G. Beylkin and L. Monzón. Approximation of functions by exponential sums revisited. *Appl. Comput. Harmon. Anal.*, 28(2):131–149, 2010.
- [11] G. Beylkin, L. Monzón, and I. Satkauskas. On computing distributions of products of random variables via Gaussian multiresolution analysis. *Appl. Comput. Harmon. Anal.*, 2017. see also arXiv:1611.08580.
- [12] G. Beylkin, L. Monzon, and X. Yang. Reduction of multivariate mixtures and its applications. *J. Comp. Phys.*, 383:94–124, 2019. see also <https://arxiv.org/abs/1806.03226>.
- [13] S. F. Boys. The integral formulae for the variational solution of the molecular many-electron wave equations in terms of gaussian functions with direct electronic correlation. *Proceedings of the Royal Society of London Series A- Mathematical and Physical Sciences*, 258(1294):402, 1960.
- [14] I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Series in Applied Mathematics. SIAM, 1992.
- [15] G. Fann, R. Harrison, G. Beylkin, R. Hartman-Baker, J. Jia, W. A. Shelton, and S. Sugiki. MADNESS applied to density functional theory in chemistry and physics. *J. of Physics, Conference Series*, 78:012018, 2007. doi:10.1088/1742-6596/78/1/012018.
- [16] L. Genovese, A. Neelov, S. Goedecker, T. Deutsch, S.A. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson, and A. Bergman. Daubechies wavelets as a basis set for density functional pseudopotential calculations. *The Journal of chemical physics*, 129(1):014109, 2008.
- [17] E. Giner, B. Pradines, A. Ferté, R. Assaraf, A. Savin, and J. Toulouse. Curing basis-set convergence of wave-function theory using density-functional theory: a systematically improvable approach. *The Journal of chemical physics*, 149(19):194301, 2018.
- [18] R. J. Harrison, G. Beylkin, F. A. Bischoff, J. A. Calvin, G. I. Fann, J. Fosso-Tande, D. Galindo, J.R Hammond, R. Hartman-Baker, J.C. Hill, J. Jia, J.S. S. Kottmann, M-J. Y. Ou, L.E. Ratcliff, M.G. Reuter, A.C. Richie-Halford, N.A. Romero, H. Sekino, W.A. Shelton, B.E. Sundahl, W.S. Thornton, E.F. Valeev, A. Vázquez-Mayagoitia, N. Vence, and Y. Yokoi. MADNESS: a multiresolution, adaptive numerical environment for scientific simulation. *SIAM J. Sci. Comput.*, 38(5):S123–S142, 2016. see also arXiv preprint arXiv:1507.01888.
- [19] R.J. Harrison, G.I. Fann, T. Yanai, Z. Gan, and G. Beylkin. Multiresolution quantum chemistry: basic theory and initial applications. *J. Chem. Phys.*, 121(23):11587–11598, 2004.
- [20] S. R. Jensen, S. Saha, J. A. Flores-Livas, W. Huhn, V. Blum, S. Goedecker, and Luca Frediani. The elephant in the room of density functional theory calculations. *The Journal of Physical Chemistry Letters*, 8(7):1449–1457, 2017.
- [21] M. H. Kalos. Monte Carlo calculations of the ground state of three- and four-body nuclei. *Phys. Rev. (2)*, 128:1791–1795, 1962.
- [22] M. H. Kalos. Monte Carlo integration of the Schrödinger equation. *Trans. New York Acad. Sci. (2)*, 26:497–504, 1963/1964.
- [23] A. Komornicki and H. F. King. A general formulation for the efficient evaluation of n-electron integrals over products of gaussian charge distributions with gaussian geminal functions. *The Journal of chemical physics*, 134(24):244115, 2011.
- [24] J.V.L. Longstaff and K. Singer. The use of gaussian (exponential quadratic) wave functions in molecular problems. ii. wave functions for the ground state of the hydrogen atom and of hydrogen molecule. *Proc. R. Soc. London Ser. A-Math.*, 258(1294):421, 1960.
- [25] F. A. Pahl and N. C. Handy. Plane waves and radial polynomials: a new mixed basis. *Molecular Physics*, 100:3199–3224, 2002.
- [26] R. G. Parr and W. Yang. *Density-Functional Theory of Atoms and Molecules*. Number 16 in The International Series of monographs on Chemistry. Oxford Univ. Press, New york, 1989.
- [27] T. Shiozaki and S. Hirata. Grid-based numerical hartree-fock solutions of polyatomic molecules. *Phys. Rev. A*, 76:040503, Oct 2007.
- [28] B. Simon. *Quantum mechanics for Hamiltonians defined as quadratic forms*. Princeton University Press, 1971.
- [29] K. Singer. The use of gaussian (exponential quadratic) wave functions in molecular problems. i. General formulae for the evaluation of integrals. *Proc. R. Soc. London Ser. A-Math.*, 258(1294):412, 1960.
- [30] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry: Intro to Advanced Electronic Structure Theory*. Dover publications, 1996.
- [31] T. Yanai, G.I. Fann, Z. Gan, R.J. Harrison, and G. Beylkin. Multiresolution quantum chemistry: Analytic derivatives for Hartree-Fock and density functional theory. *J. Chem. Phys.*, 121(7):2866–2876, 2004.
- [32] T. Yanai, G.I. Fann, Z. Gan, R.J. Harrison, and G. Beylkin. Multiresolution quantum chemistry: Hartree-Fock exchange. *J. Chem. Phys.*, 121(14):6680–6688, 2004.