Atomic Shell Approximation: Electron Density Fitting Algorithm Restricting Coefficients to Positive Values

Pere Constans and Ramon Carbó*

Institute of Computational Chemistry, University of Girona, Albereda 3-6, 17071 Girona, Spain

Expressing molecular electron densities as linear combinations of Guassian $1S$ spherical functions extends accurate computations of *ab initio* quantum molecular similarity measures to large systems. These approximate least squares fitted functions lose the density of probability statistical meaning since some coefficients might be negative. A new algorithm using positive coefficients only and thus providing an electron density picture as a superposition of atomic shells is presented.

1. INTRODUCTION

Quantum molecular similarity measures (QMSM) based on first-order density functions are defined as[1]

$$z_{AB} \equiv \int\int \rho_A(1)\hat{\Theta}(1,2)\rho_B(2)\, dV_1\, dV_2 \qquad (1)$$

where ρ_A and ρ_B are the corresponding first-order density functions or electron probability densities for two arbitrary molecules A and B, and $\hat{\Theta}$ is a definite positive operator. The measure z_{AB} depends on the relative positioning of the molecules, being assumed in this definition that the mutual arrangement maximizes the integral (eq 1). This optimization, which requires multiple evaluations of many center integrals, is highly time-consuming and limits the use of *ab initio* QMSM to small molecular systems. This difficulty has been overcome fitting the electron densities to linear combinations of spherically symmetric functions, centered at the nuclear coordinates.[2] This approximation reduces the computation times several orders of magnitude without a significant loss of accuracy. However the resulting fitted functions, since some of the coefficients might be negative, cannot have, strictly speaking, the probabilistic meaning of the original electron density functions. In this paper, a new method is presented, the atomic shell approximation, which gives accurate fitted functions restricting the expansion coefficients to positive values only. In this way, a picture of the molecular electron density composed by a superposition of atomic shells, whose occupations are the variational coefficients, is obtained.

2. ATOMIC SHELL APPROXIMATION

The spatial electron density functions $\rho(\mathbf{r})$ arise from averaging the molecular wave function modules over the space-spin coordinates $\mathbf{x}$ of all the N electrons but one[3]

$$\rho(\mathbf{x}_1) = N\int \Psi^*(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)\Psi(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)\, d\mathbf{x}_2 ... d\mathbf{x}_N \qquad (2)$$

integrating over the remaining spin coordinate s_1 and removing superfluous subindices

$$\rho(\mathbf{r}) = \int \rho(\mathbf{x}_1)\, ds_1 \qquad (3)$$

Within the LCAO approach, electron densities are given as double sums over pairs of basis functions, in the form

$$\rho(\mathbf{r}) = \sum_{ij} D_{ij}\varphi_i^*(\mathbf{r})\varphi_j(\mathbf{r}) \qquad (4)$$

where D_{ij} are the density matrix coefficients and $\varphi_i(\mathbf{r})$ and $\varphi_j(\mathbf{r})$ are the atomic orbitals. For free atomic systems, as Unsöld has shown,[4] the wave function sum of squares corresponding to a subshell is spherically symmetric. Consequently, the electron density function for closed-shell atoms could be expressed through an integral Gaussian transform over the radial coordinate in the following manner:

$$\rho(\mathbf{r}) = \int_0^\infty f(\zeta)e^{-\zeta r^2}\, d\zeta \qquad (5)$$

which can be approximated by the finite sum, in the form

$$\rho(\mathbf{r}) \approx \sum_i w_i f_i e^{-\zeta_i r^2} \qquad (6)$$

Expression 6 describes the electronic distribution as a combination of exponential Gaussian functions. Assuming that the coefficients of these functions are all positive-valued, no negative occupancies, the electronic density appears as a superposition of spherical atomic shells. Extending this approach to molecular systems by placing atomic shells at nuclear coordinates describes the theoretical background of the atomic shell approximation (ASA). The electron densities for molecules are expressed in the ASA as

$$\rho_{ASA}(\mathbf{r}) = \sum_a \sum_{i \in a} n_i S_i(\mathbf{R}_a - \mathbf{r}) \qquad (7)$$

where $S_i(\mathbf{R}_a - \mathbf{r})$ is a Gaussian $1S$ function centered at the nuclear coordinate $\mathbf{R}_a$ of the atom a in the molecule. Defining the basis functions as

$$S_i(\mathbf{R}_a - \mathbf{r}) \equiv \left(\frac{\zeta_i}{\pi}\right)^{3/2} e^{-\zeta_i(\mathbf{R}_a - \mathbf{r})^2} \qquad (8)$$

and taking into account the normalization condition

$$\sum_a \sum_{i \in a} n_i \int S_i(\mathbf{R}_a - \mathbf{r})\, d\mathbf{r} = \sum_a \sum_{i \in a} n_i = N \qquad (9)$$

it comes into view that the expansion coefficients n_i are the occupation numbers for the corresponding shells. This spherical representation of electron densities appears in the literature in fittings of experimental densities from X-ray diffractions[5] and is in the heart of the isotropic atom–atom potentials.[6] Other authors have also used spherical n S-like orbitals to construct MO wavefunctions with interesting results,[7] so the present approach seems sufficiently well documented to merit some study.

Two aspects of the problem are discussed in the following sections. The first one consists on how to obtain appropriate shells and occupation numbers, discussed in section 2.1 and illustrated with the examples of argon atom and water molecule. The second aspect deals with the systematic construction of the functions given by eq 7. Section 2.2. describes how to construct and optimize atomic ASA functions, and then, in section 2.3, by means of these optimized atomic functions, the extension to molecular systems is developed.

2.1. Atomic Shell Approximation (ASA) Method. Once truncation (eq 6) is done, defining a finite function basis set, the best approximation to the density $\rho(\mathbf{r})$ in the least squares sense will be the one obtained minimizing the quadratic error integral function:

$$\epsilon^2(\mathbf{n}) \equiv \int (\rho(\mathbf{r}) - \rho_{ASA}(\mathbf{r}))^2 \, d\mathbf{r} \qquad (10)$$

Computing the set of coefficients $\mathbf{n}$ that minimizes $\epsilon^2(\mathbf{n})$ implies a simple resolution of a linear equation system. However, the solution obtained might have some negative coefficients and therefore may violate the ASA assumption. In this section, in order to overcome this drawback, even-tempered basis functions will be used in the least squares fitting (LSF) solution as a way to span nearly complete atomic function spaces. Examining the LSF solution, one can realize that acceptable fitted densities will lie in subspaces from the nearly complete function space. The ASA algorithm will be finally presented as a way to localize such subspaces and calculate acceptable and accurate fitted electron densities.

2.1.1. Nearly Complete Atomic Basis Sets. Nearly complete spaces of Gaussian functions have been easily generated using an even-tempered series of exponents

$$\zeta_i = \alpha\beta^i \qquad (11)$$

Selecting the exponents in a geometric sequence was first recommended by Reeves[8] in 1963 and later developed by Ruedenberg *et al.* in the context of orbital optimization. Performing a simple two-dimensional search over the generators α and β gives no significant difference with respect to the fully variational solution found when optimizing all the exponent series. Even-tempered sequences are a simple and elegant way to construct truncated basis sets, avoiding cumbersome nonlinear optimizations, and at the same time taking control over linear dependence.[9]

2.1.2. Least Squares Fitting. The best fitted density function in a least squares sense is computed minimizing the quadratic error integral function defined in eq 10, which can be expanded as

$$\epsilon^2(\mathbf{n}) = c - 2\sum_i n_i t_i + \sum_{ij} n_i n_j s_{ij} \qquad (12)$$

where it has introduced the notation:

$$c = \int \rho^2(\mathbf{r}) \, d\mathbf{r} \qquad (13)$$

$$t_i = \int \rho(\mathbf{r})S_i(\mathbf{r}) \, d\mathbf{r} \qquad (14)$$

and

$$s_{ij} = \int S_i(\mathbf{r})S_j(\mathbf{r}) \, d\mathbf{r} \qquad (15)$$

Since $\epsilon^2(\mathbf{n})$ is a positive definite quadratic form, its unrestricted minimum $\mathbf{n}_0$ is given by solving the linear equations set

$$\mathbf{S}\mathbf{n}_0 = \mathbf{t} \qquad (16)$$

The normalization constraint (eq 9) can be imposed to the fitted probability function solving an analogous equation system, using the Lagrange multipliers formalism, given by

$$\mathbf{S}\mathbf{n}'_0 = \mathbf{t}' \qquad (17)$$

being $\mathbf{n}'_0$ the restricted minimum solution and

$$\mathbf{t}' = \mathbf{t} + \lambda\mathbf{m} \qquad (18)$$

with the vector $\mathbf{m}$ containing the restriction conditions

$$m_i = \int S_i(\mathbf{r}) \, d\mathbf{r} = 1 \qquad (19)$$

and the Lagrange multiplier λ is expressed by

$$\lambda = (N - \mathbf{m}^T\mathbf{S}^{-1}\mathbf{t})(\mathbf{m}^T\mathbf{S}^{-1}\mathbf{m})^{-1} \qquad (20)$$

2.1.3. Compatible Subspaces. The best fitting function is given solving the linear system (eq 17). The coefficients, in terms of the Cramer's rule, are given by

$$n'_{0,j} = (S_{1j}t'_1 + S_{2j}t'_2 + ... + S_{nj}t'_n) \det|\mathbf{S}|^{-1} \qquad (21)$$

where S_{ij} is the cofactor of the element s_{ij} in the metric matrix $\mathbf{S}$. Since $\mathbf{S}$ is positive definite, and consequently $\det|\mathbf{S}|$ is always positive valued, the ASA function will fulfill the following set of inequalities:

$$S_{1j}t'_1 + S_{2j}t'_2 + ... + S_{nj}t'_n \geq 0 \qquad (22)$$

For a given molecular geometry and the attached density function $\rho(\mathbf{r})$, the terms implied in eq 21 will only depend on the exponents of the shell basis set. The set of inequalities (eq 22) establishes intricate relationships connecting all the exponents, in order to guarantee the coefficients $n'_{0,j}$ to be positive, as the statistical meaning of $\rho(\mathbf{r})$ requires. In the last instance, this means that the ASA solution functions (eq 8) belong to a subset of all the functions generated by the even-tempered series (eq 11). In other words, generating a complete set of exponents is an easy way to determine accurate ASA solutions, determining which subspace gives the best fitted function, fulfilling inequalities (eq 22) simultaneously.

2.1.4. Positive Restricted Fitting Algorithm. Supposing the density $\rho(\mathbf{r})$ for an atomic or molecular system is known and a nearly complete set of n basis functions (eq 8) is available, computation of the ASA density minimizing $\epsilon^2(\mathbf{n})$ requires the determination of the vector of coefficients $\mathbf{n}'_0$ by solving eq 17:

$$\mathbf{n'}_0 = \mathbf{S}^{-1}\mathbf{t'} \tag{23}$$

The positive restricted fitting algorithm developed in our laboratory is an iterative method to solve the above linear system together with the n inequality restrictions. Since $\epsilon^2(\mathbf{n})$ is a quadratic form, its minimum $\mathbf{n'}_0$ can be expressed in terms of an arbitrary vector $\mathbf{n}$, by

$$\mathbf{n'}_0 = \mathbf{n} - \mathbf{S}^{-1}\nabla\epsilon^2(\mathbf{n}) \tag{24}$$

where the gradient at $\mathbf{n}$ is given by

$$\nabla\epsilon^2(\mathbf{n}) = 2(\mathbf{n}\mathbf{S} - \mathbf{t'}) \tag{25}$$

Choosing the arbitrary point $\mathbf{n}$ with all the components positive and taking the direction

$$\mathbf{p} = \mathbf{S}^{-1}\nabla\epsilon^2(\mathbf{n}) \tag{26}$$

corresponding to the shortest approaching path from the point $\mathbf{n}$ to the minimum $\mathbf{n'}_0$, it is possible to define a new point $\mathbf{n'}_1$ in $\mathbf{p}$ given by

$$\mathbf{n'}_1 = \mathbf{n} - \xi\mathbf{p} \tag{27}$$

The parameter $\xi \in [0, 1]$ is the largest step through the descending path that keeps positive the coefficients. Analyzing every component at the intersection:

$$0 = n_i - \xi_i p_i, \forall i \tag{28}$$

a subset of ξ_i values can be defined

$$\xi_k^{(+)} = n_k p_k^{-1} \wedge p_k > 0 \ \forall k \tag{29}$$

for the positive components of the approaching path $\mathbf{p}$ only, giving the maximum step for the considered component. Obviously, no restriction exists if a component p_j is negative because the corresponding coefficient $n'_{j,1}$ always will be positive. Then, taking ξ as

$$\xi \equiv \min_k(1, \xi_k^{(+)}) \tag{30}$$

forces the new point $\mathbf{n'}_1$ to have positive or zero components. Since any path conducting to the minimum is always descending in a positive defined quadratic form, the new set of coefficients will decrease the function $\epsilon^2(\mathbf{n})$.

At this step of the iterative process, the functions with null coefficients and with a positive slope at $\mathbf{n'}_1$ are discarded. This is so because they would have negative coefficients in a differential steepest descent displacement from $\mathbf{n'}_1$. Afterwards, a new approaching path is computed:

$$\mathbf{p}_r = \mathbf{S}_r^{-1}\nabla_r\epsilon^2(\mathbf{n'}_{1,r}) \tag{31}$$

The dimension of the problem has been reduced as indicated in expression 31 by the subindices r. In the way previously shown, a new step ξ and a new point $\mathbf{n'}_{2,r}$ are computed. Then, after expanding $\mathbf{n'}_{2,r}$ to a whole dimensioned vector $\mathbf{n'}_2$, maintaining the original zero values for the discarded functions, a computation of the gradient at this improved $\mathbf{n'}_2$ is performed, closing the second iteration. The process stops when ξ is one, a minimum is reached in one possible subspace, and all the slopes of shells with zero occupancies are positive, the conditions of a restricted minimum. In this

Table 1. Schematic Description of the ASA Algorithm

```
COMPUTE INTEGRALS c, t, and S
INITIALIZE n and ∇ε²(n)

DO
      FOR i (IF nᵢ = 0 and ∇ᵢε²(n) > 0  DISCARD SHELL i)
      ESTABLISH REDUCED DIMENSION tᵣ, Sᵣ, and Sᵣ⁻¹
      COMPUTE nᵣ¹
      EXPAND nᵣ¹ to n¹
      IF λ < 1 DO CONTINUE
      IF (FOR nᵢ¹ = 0  ∇ᵢε²(n¹) > 0) DO EXIT
END DO

n = n¹
COMPUTE q

introducing now normalization condition

DO
      FOR i (IF nᵢ = 0 and ∇ᵢε²(n) > 0  DISCARD SHELL i)
      ESTABLISH REDUCED DIMENSION tᵣ, Sᵣ, and Sᵣ⁻¹
      COMPUTE q'
      COMPUTE n'ᵣ¹
      EXPAND n'ᵣ¹ to n'¹
      IF λ < 1 DO CONTINUE
      IF (FOR n'₀¹ = 0  ∇ᵢε²(n'¹) > 0) DO EXIT
END DO
```

manner, not only a minimum is found in a problem subspace, accomplishing

$$\mathbf{n'}_{0,r} = \mathbf{S}_r^{-1}\mathbf{t'}_r \tag{32}$$

but also the best subspace, i.e., the best fitting function from all possible combinations of basis set functions, is obtained.

Referring to the computational efficiency of this algorithm, two considerations must be taken into account. First, it is worthwhile to realize that an important computational simplification can be introduced removing the normalization constraint introduced in eq 16 during the localization of compatible subspaces. Since the original density function strictly obeys the electron normalization, any flexible enough fitting expansion will freely reproduce this constraint, and therefore, this imposition practically does not influence the final selection of functions. The Lagrange multiplier can be introduced once this first selection is done, allowing further iterations if necessary. Table 1 schematizes the complete algorithm. The second consideration refers to eq 24, which might give numerical inaccuracies, reflected in abnormally large values for the gradient components. In such a case, the solution could be refined since the compatible subspace is already determined, solving directly the linear system (eq 32). Even though the number of matrix inversions to be performed during the iterative procedure is important, the computational cost for this restricted fitting is not greater than the simple least squares fitting. This is because symmetric matrix inversion is a fast process compared to integral evaluation.

2.1.5. Argon Atom and Water Molecule as Examples. In order to illustrate the ASA method presented above, two special examples computed using even-tempered functions taken from ref 9b will be discussed. The argon, a third-row closed-shell atom, having a perfectly spherical electron distribution expressible in terms of eq 5, was chosen first. The fitting to the HF/6-311G* argon density was performed using 50 initial Gaussian functions. Restricting coefficients to positive values does not prevent it from reaching an acceptable minimum, as indicated by the small quadratic

ATOMIC SHELL APPROXIMATION

J. Chem. Inf. Comput. Sci., Vol. 35, No. 6, 1995 **1049**

Table 2. Comparison of a Nonrestricted Fitted Density and ASA Density for the Ar Atom

no.	exponent	unrestricted coefficients	restricted coefficients	gradient[a]
1	0.88334E−01	−0.0134380	0	0.49E−04
2	0.14506E+00	0.0474144	0.0365563	0
3	0.23823E+00	−0.0448422	0.0129552	0
4	0.39124E+00	0.2131020	0	0.15E−05
5	0.64251E+00	0.1151697	0.4471984	0
6	0.10551E+01	0.2331811	0	0.17E−03
7	0.17328E+01	−0.0032735	0	0.51E−03
8	0.28457E+01	−0.0945963	0	0.57E−03
9	0.46734E+01	0.0003625	0	0.24E−03
10	0.76750E+01	0.2164541	0.0993295	0
11	0.12604E+02	0.1619073	0.3041425	0
12	0.20699E+02	0.1023362	0.0033951	0
13	0.33993E+02	−0.0408238	0.0069699	0
14	0.55825E+02	0.0169620	0	0.22E−04
15	0.91679E+02	−0.0048141	0	0.13E−04
16	0.15056E+03	0.0415560	0.0400088	0
17	0.24725E+03	0.0205718	0.0213351	0
18	0.40605E+03	0.0182108	0.0178979	0
19	0.66684E+03	0.0080615	0.0079777	0
20	0.10951E+04	0.0029715	0.0033165	0
21	0.17984E+04	0.0025877	0.0021454	0
22	0.29535E+04	−0.0003329	0.0000779	0
23	0.48504E+04	0.0008494	0.0005417	0
24	0.79656E+04	−0.0001891	0	0.21E−05
25	0.13081E+05	0.0001194	0.0000237	0
26	0.21483E+05	−0.0000089	0.0000335	0
27	0.35280E+05	0.0000199	0	0.73E−05
28	0.57939E+05	−0.0000113	0	0.81E−05
29	0.95151E+05	0.0000073	0.0000010	0
30	0.15626E+06	−0.0000023	0.0000008	0
31	0.25662E+06	0.0000015	0	0.15E−04
32	0.42143E+06	−0.0000008	0	0.35E−04
33	0.69210E+06	0.0000004	0	0.47E−04
34	0.11366E+07	−0.0000002	0	0.47E−04
35	0.18665E+07	0.0000001	0	0.41E−04
36	0.30653E+07	−0.0000001	0	0.33E−04
37	0.50341E+07	0.0000000	0	0.25E−04
38	0.82673E+07	0.0000000	0	0.18E−04
39	0.13576E+08	0.0000000	0	0.12E−04
40	0.22296E+08	0.0000000	0	0.91E−05
41	0.36616E+08	0.0000000	0	0.63E−05
42	0.60133E+08	0.0000000	0	0.44E−05
43	0.98754E+08	0.0000000	0	0.30E−05
44	0.16218E+09	0.0000000	0	0.21E−05
45	0.26633E+09	0.0000000	0	0.14E−05
46	0.43739E+09	0.0000000	0	0.10E−05
47	0.71831E+09	0.0000000	0	0.69E−06
48	0.11796E+10	0.0000000	0	0.47E−06
49	0.19372E+10	0.0000000	0	0.33E−06
50	0.31814E+10	0.0000000	0	0.22E−06

[a] The gradients correspond to the ASA fitting.

Table 3. ASA Function for Water Molecule

oxygen		hydrogen	
exponents	occupations	exponents	occupations
0.40992E+00	0.1310652	0.27870E+00	0.0189095
0.82656E+00	0.2908240	0.20101E+01	0.0410369
0.11737E+01	0.1131416	0.27941E+01	0.0004820
0.16666E+01	0.0850993	0.10430E+02	0.0000066
0.23666E+01	0.0742525	0.14498E+02	0.0012018
0.19403E+02	0.0265906	0.75231E+02	0.0000368
0.27552E+02	0.0605241		
0.39124E+02	0.0068292		
0.55556E+02	0.0527216		
0.11202E+03	0.0229198		
0.15907E+03	0.0058868		
0.32075E+03	0.0047711		
0.45547E+03	0.0014198		
0.13041E+04	0.0004945		
0.18519E+04	0.0000900		
0.75296E+04	0.0000145		
0.10692E+05	0.0000079		

molecular environments, the value of the quadratic error has been found to be 1.8×10^{-4}, greater than the isoelectronic neon system, which yielded a value of 1.3×10^{-6}. The relative error in the self-similarity is 0.027%, sufficient for similarity measure purposes. The error of the function at nuclei is 0.9% for hydrogen and 0.03% for oxygen. This greater error of the function at the hydrogen coordinates has been usual in all the molecules tested. They may be due to the small values of the molecular function over hydrogens with respect to heavier atoms. Nevertheless, increasing the number of initial functions improves these values considerably. Shell exponents and normalized occupations for the ASA fitted density are listed in Table 3.

When comparing the number of contributing functions or shells per atom, there are 19 functions for atomic argon, and in the water molecule, there are 17 for oxygen and six for hydrogen atoms. A reduction on the number of initial functions is highly desirable, since the resulting density may be used in similarity maximizations, but it is remarkable that no arbitrary assumption has been done in the number of functions needed to properly describe different atoms except, obviously, the size of the initial functional space.

2.2. Generation of Complete Basis Sets in Atomic Systems. The even-tempered series of exponents appearing in the literature are optimized to describe atomic orbitals of a given symmetry. These basis sets are therefore specially designed to span wave functions for atoms. Although the use of nearly complete sets should in principle span any function belonging to the same class of symmetry, in this contribution, generators were optimized in order to have as good quality as possible for the fitted densities. The optimization is performed mainly in the way proposed by Rudenberg *et al.*, in ref 9b. In the next section, peculiarities concerning this optimization are introduced, and some results are presented.

2.2.1. Atomic Basis Sets Spanning Electron Densities. A successful approach to a complete basis set implies a dependence of the values α and β in eq 11 with respect to the basis size n. Ruedenberg *et al.* postulate this functional dependence as

$$\ln \ln \beta = b \ln n + b' \qquad (33)$$

and

error integral value, $\epsilon^2(n)$, which is 9.4×10^{-6}. Also, the error when computing the self-similarity for argon atom using the ASA fitted function appears to be $2.5 \times 10^{-4}\%$, and the error of the function at nucleus, the most different local value, is 0.18%. The list of argon exponents defining the basis functions and its resulting occupation numbers are both presented in Table 2. They can be compared with the standard least squares fitting coefficients. The density function has been scaled in such a way that the summation over the occupation numbers equals one.

Another illustrative example is a simple molecular system: the water molecule. It has been computed, as before, at the HF/6-311G** level of theory. The fitting basis functions consisted in a set of 50 initial shells per atom in order to allow some comparison with the previous example. As can be expected from the loss of sphericity of atoms in

1050 *J. Chem. Inf. Comput. Sci., Vol. 35, No. 6, 1995*

CONSTANS AND CARBÓ

$$\ln \alpha = a \ln(\beta - 1) + a' \qquad (34)$$

The parameters α and β are optimized for different sizes of the basis and the constants in eqs 33 and 34 are obtained by least squares fitting. The values given by these equations, called regularized even-tempered parameters, differ very little from the optimized ones, having the interesting advantage of allowing a different basis sizes and quality fittings exploration in the implementation of the ASA.

The infinite basis set size limit of even-tempered parameters

$$\alpha \rightarrow 0 \qquad (35)$$

and

$$\beta \rightarrow 1 \qquad (36)$$

can be used as the starting point in the optimization. Considering a set of 50 functions near to the complete basis limit, taking the values 0.01 and 1.2 for the initial α_0 and β_0 parameters respectively, then a steepest descent scheme has been employed to decrease the quadratic error integral (eq 10) with respect to $\ln \alpha$ and $\ln \beta$. Due to the discontinuous nature of the ASA fitting, no gradient has been computed. Instead, a discrete search with a small step, $\delta = 0.0005$, has proved to be sufficient. Four values: $\epsilon^2(\ln \alpha_0, \ln \beta_0)$, $\epsilon^2(\ln \alpha_0 + \delta, \ln \beta_0)$, $\epsilon^2(\ln \alpha_0, \ln \beta_0 + \delta)$ and $\epsilon^2(\ln \alpha_0 + \delta, \ln \beta_0 + \delta)$ are computed, taking the best one as the starting point $(\ln \alpha_1, \ln \beta_1)$ in a next search. The minimization stops when the quadratic error integral no longer decreases. Once the optimal values are determined, the size of the basis is reduced by one function and, using the previous values, the optimal α and β for the $n - 1$ basis set are determined. At the end of this process, when only two functions remain in the basis, a linear regression is performed, and the constants b, b', a and a' together with the respective uncertainties are determined. Table 4 recompiles the constant values b, b', a, and a' and the corresponding uncertainties, in italics, for atoms up to argon. The electron densities have been computed using the 6-311G* basis set, at the Hartree–Fock level of theory, in order that the even-tempered parameters describe a reasonably good density. The values presented are quite independent of a particular atom in a given period, and at the same time, no significant differences are obtained in the quality of the ASA density if parameters from ref 9b are used and extended basis sets employed. Nevertheless, the regularized parameters in Table 4 give good quality fittings even when few starting functions are used, a desirable and necessary election when dealing with large molecules.

2.3. Computational Results.
Three computational examples are given in this section to illustrate the robustness and precision of the presented fitting algorithm. In the first example, several standard basis sets are used in order to remark the independence of the fitting algorithm with respect to the wave function accuracy level. In the second one, a large molecular system, with respect to the number of involved atoms and orbital functions, is presented with the purpose to indicate that ASA is a general procedure, extensible to huge molecules. Finally, in the last one, HF and ASA densities are compared when computing MQSM in order to complete this development and to validate the fitted densities for an accurate and inexpensive computation of *ab initio* quantum similarities.

Table 4. Atomic Constants for Generating Even-Tempered Sets of Exponents To Span Electron Densities (Uncertainties Are in Italics)

	b		b'		a		a'	
H	−0.6870	0.016	1.2599	0.050	0.5131	0.014	−3.9829	0.013
He	−0.7144	0.045	1.4289	0.045	0.5322	0.013	−3.9818	0.012
Li	−0.6771	0.013	1.4779	0.042	0.5789	0.014	−3.9969	0.011
Be	−0.6857	0.010	1.5127	0.032	0.5840	0.014	−3.9958	0.012
B	−0.6802	0.010	1.5360	0.033	0.6027	0.014	−3.9989	0.012
C	−0.6869	0.011	1.5909	0.034	0.6164	0.014	−4.0010	0.012
N	−0.6906	0.010	1.6164	0.032	0.6247	0.015	−4.0018	0.013
O	−0.7091	0.008	1.6745	0.027	0.6209	0.016	−3.9995	0.014
F	−0.7277	0.008	1.7323	0.028	0.6224	0.016	−3.9958	0.014
Ne	−0.7568	0.009	1.8131	0.030	0.6181	0.017	−3.9892	0.016
Na	−0.7375	0.006	1.7863	0.020	0.6331	0.016	−3.9970	0.015
Mg	−0.7499	0.006	1.8250	0.020	0.6356	0.017	−3.9943	0.016
Al	−0.7569	0.006	1.8448	0.020	0.6368	0.017	−3.9925	0.016
Si	−0.7673	0.007	1.8774	0.023	0.6387	0.017	−3.9900	0.017
P	−0.7662	0.006	1.8732	0.020	0.6395	0.018	−3.9902	0.017
S	−0.7578	0.006	1.8579	0.019	0.6472	0.017	−3.9931	0.017
Cl	−0.7536	0.006	1.8539	0.020	0.6537	0.017	−3.9949	0.016
Ar	−0.7693	0.006	1.8926	0.020	0.6525	0.017	−3.9897	0.017

Molecular densities are, in all cases, computed at the Hartree–Fock level of theory using the *Gaussian 92* ensemble of programs.[10] The ASA algorithm has been implemented in a Fortran program developed in our laboratory, called ASAC, which fits electron density functions taken from the *Gaussian 92* output files using even-tempered basis functions generated from parameters given in Table 4.

2.3.1. Analyzing Different Wave Function Basis.
The fitting procedure presented here on one hand overcomes the need of having different sets of functions depending on the particular densities to be fitted and on the other hand avoids arbitrary assignments with respect to the number of functions needed for a proper description of the molecule. Two aromatic molecules have been computed with different basis sets, from the simple minimal basis to nearly Hartree–Fock limit quality, with a full geometry optimization. The first molecule corresponds to benzene and the second one to 4H,6H-1,2,3,5-diazodiborene. Tables 5 and 6 include the number of basis functions and primitives in the SCF computation, the starting number of fitting functions and shells, i.e., the final contributing functions. In both molecules, the number of initial functions is 270, corresponding to assigning 15 functions in hydrogen atoms and 30 functions in boron, carbon, and nitrogen. Shells on hydrogen and carbon are also specified for benzene. The initial assignation of atomic functions reduces the basis set size for computational efficiency and has no significative effect on the proportion (no. of shells on C)/(no. of shells on H). Increasing the number of initial functions increases the number of final selected shells, but for a nearly saturated basis set, a great growth in the number of functions is required to add an extra shell. However, when the basis size is augmented, the shell exponents can more flexibly be improved, minimizing the quadratic error even if the number of shells is not enlarged. As can be seen in Tables 5 and 6, the number of shells increases accordingly to the quality of the wave function basis set. Nevertheless, some anomalous behavior of the Dunning–Huzinaga basis set D95[11] is remarkable especially for 4H,6H-1,2,3,5-diazodiborene molecule: this basis set doubles, differing from the other ones used, the number of core functions. Tables 5 and 6 also present the quadratic error integral values and errors in self-similarity, giving a measure of the quality of the fitting.

ATOMIC SHELL APPROXIMATION

J. Chem. Inf. Comput. Sci., Vol. 35, No. 6, 1995 **1051**

Table 5. Details of Fitting Computations Analyzing Different Wave Function Basis Sets for Benzene Molecule

	basis sets					
no. of functions	STO-3G	3-21G	6-21G	D95	6-31G*	6-311G**
basis functions	36	66	66	72	102	150
primitives	108	108	126	168	192	240
fitting functions	270	270	270	270	270	270
shells	84	96	114	126	114	132
shells on H	6	5	5	5	5	6
shells on C	8	11	14	16	14	16
ϵ^2	1.233E−05	2.239E−05	2.223E−05	1.844E−05	3.742E−05	3.905E−05
error in S(A,A)	0.01157%	0.02743%	0.02668%	0.02321%	0.03777%	0.03914%

Table 6. Details of Fitting Computations Analyzing Different Wave Function Basis Sets for 4H,6H-1,2,3,5-Diazodiborene Molecule

	basis sets					
no. of functions	STO-3G	3-21G	6-21G	D95	6-31G*	6-311G**
basis functions	36	66	66	72	102	150
primitives	108	108	126	168	192	240
fitting functions	270	270	270	270	270	270
shells	79	86	105	125	111	116
ϵ^2	3.867E−05	3.262E−05	3.235E−05	3.046E−05	4.130E−05	4.382E−05
error in S(A,A)	0.03828%	0.03171%	0.03153%	0.02837%	0.03810%	0.04085%

Table 7. Atomic Constants for Generating Even-Tempered Sets of Exponents to Span Electron Densities for Cobalt and Rhenium (Uncertainties Are in Italics)

	b		b′		a		a′	
Co	−0.7908	*0.014*	1.9194	*0.043*	0.6690	*0.017*	−3.9721	*0.018*
Re	−0.7663	*0.013*	2.0389	*0.042*	0.7298	*0.016*	−4.0254	*0.018*

In all basis sets tested, this quality is uniform, validating the ASA method as a robust and systematic procedure for fitting densities.

2.3.2. Organometallic Compound as a Large Molecular Example.

This example corresponds to the fitting of the density for a cobalt—rhenium cluster compound and indicates that spherical atomic shells constitute a natural way to represent molecular electron densities, independent of the angular momentum of the orbitals composing the wave function. The electron density of the hexacarbonylbis(μ-dicyclohexylphosphanido)cobaltrhenium[12] molecule has been computed at the HF/STO-3G level of theory for its experimental geometry. This molecule presents a great variety of chemical bond types. The transition elements are linked through a metal—metal bond and to the phosphorus atoms. Two carbonyls are also linked to the penta-coordinate cobalt, while the remaining four are bonded to the hepta-coordinate rhenium. The organic part presents triple bonds in carbonyl groups, and C-heteroatom contacts link the four cyclohexanes to phosphorus atoms. For cobalt and rhenium, the optimization of the even-tempered parameters has been performed using the HF/STO-3G density, obtaining the values shown in Table 7. The number of initial fitting functions on every atom appears in Table 8. It is interesting for this huge molecule to reduce the number of initial functions, sacrificing a little bit of precision, but getting a more portable fitted function. The total number of shells in the fitted density is 466, while the number of terms for the exact density is 921 × 921, considering symmetric terms, and 921 being the number of primitives. The quadratic integral error for the normalized density appears too in Table 8 together with the values of the self-similarity. The quadratic error is greater than in previous examples on aromatic molecules, probably due to a loss of sphericity in the densities around the atoms

Table 8. Details of Fitting Computation for Cobalt—Rhenium Cluster

[CoRe(C₁₂H₂₂P)₂(CO)₆]	
experimental geometry	84 atoms
HF/STO-3G density	404 electrons
	307 basis functions
	921 primitives
958 fitting functions	7 on H
	15 on C and O
	20 on P
	35 on Co and Re
466 shells fitted density	10 on P and Co
	14 on Re
ϵ^2	2.240E−4
exact self-similarity	0.79545
fitted self-similarity	0.79522
error in self-similarity	0.02847%

produced by rhenium, although the relative error in the self-similarity is comparable to the other molecules presented.

2.3.3. Applying ASA to QMSM.

This subsection ends the series of examples showing that ASA fitted densities reproduced the *ab initio* similarity measures with a small error. The results of computing exact QMSM are compared with the measures using fitted densities in a set of small molecules, concretely a series of fluoro- and chloro-substituted methanes. The electronic densities for these molecules has been computed at the HF/6-31G* level of theory at its optimized geometry. The fitting has been performed using 15 functions on hydrogen atoms, 25 functions on carbon and fluorine, and 35 functions on chlorine, using the even-tempered parameters from Table 4. Special care has been taken in maximizing the similarity integral in order to avoid local maxima. Similarity measures given by eq 1 are six variable functions, three of them corresponding to relative translations of the molecules and the other three to the relative orientation. First of all, an exploration over translations is performed, placing every atom in a molecule to the atomic coordinates of the other molecule, the maximum of the electron density. For these molecules, this gives 25 exploring points representing all reasonable arrangements involving translations. Once a pair of atoms is superimposed, the function (eq 1) becomes a two-variable function, i.e., two angles are enough to describe the mutual arrangement. A grid of 72 points is taken to

Table 9. QMSM for Fluoro- and Chloro-Substituted Methanes Using HF/6-31G* Densities

	exact similarity measures								
	CH$_4$	CH$_3$F	CH$_3$Cl	CH$_2$F$_2$	CH$_2$Cl$_2$	CHF$_3$	CHCl$_3$	CF$_4$	CCl$_4$
CH$_4$	0.318451	0.326555	0.554690	0.226270	0.343399	0.173171	0.248676	0.140276	0.194908
CH$_3$F		0.466397	0.676687	0.321310	0.419139	0.242239	0.303666	0.193919	0.238099
CH$_3$Cl			1.520926	0.471593	0.941128	0.361133	0.681356	0.292503	0.534325
CH$_2$F$_2$				0.400045	0.292202	0.292223	0.211903	0.228921	0.166286
CH$_2$Cl$_2$					1.147685	0.223730	0.713691	0.181281	0.450922
CHF$_3$						0.337171	0.163000	0.270797	0.127983
CHCl$_3$							0.898022	0.132078	0.627679
CF$_4$								0.288587	0.104292
CCl$_4$									0.733634

Table 10. QMSM for Fluoro- and Chloro-Substituted Methanes Using ASA Fitted Densities

	ASA similarity measures								
	CH$_4$	CH$_3$F	CH$_3$Cl	CH$_2$F$_2$	CH$_2$Cl$_2$	CHF$_3$	CHCl$_3$	CF$_4$	CCl$_4$
CH$_4$	0.318310	0.326918	0.554706	0.226500	0.343411	0.173325	0.248683	0.140383	0.194912
CH$_3$F		0.466250	0.677045	0.321529	0.419302	0.242620	0.303755	0.194383	0.238154
CH$_3$Cl			1.520880	0.471990	0.941192	0.361449	0.681470	0.292761	0.534466
CH$_2$F$_2$				0.399950	0.292403	0.292371	0.212041	0.229202	0.166393
CH$_2$Cl$_2$					1.147660	0.223865	0.713772	0.181446	0.451105
CHF$_3$						0.337100	0.163152	0.270862	0.128113
CHCl$_3$							0.898010	0.132228	0.627749
CF$_4$								0.288530	0.104438
CCl$_4$									0.733620

Table 11. Carbó Indices for Fluoro- and Chloro-Substituted Methanes Using HF/6-31G* Densities

	exact Carbó indices								
	CH$_4$	CH$_3$F	CH$_3$Cl	CH$_2$F$_2$	CH$_2$Cl$_2$	CHF$_3$	CHCl$_3$	CF$_4$	CCl$_4$
CH$_4$	1	0.84733	0.79703	0.63394	0.56802	0.52848	0.46501	0.46272	0.40324
CH$_3$F		1	0.80344	0.74386	0.57288	0.61086	0.46921	0.52857	0.40704
CH$_3$Cl			1	0.60458	0.71233	0.50429	0.58301	0.44150	0.50583
CH$_2$F$_2$				1	0.43123	0.79567	0.35354	0.67374	0.30694
CH$_2$Cl$_2$					1	0.35965	0.70300	0.31499	0.49141
CHF$_3$						1	0.29622	0.86812	0.25732
CHCl$_3$							1	0.25944	0.77331
CF$_4$								1	0.22665
CCl$_4$									1

Table 12. Carbó Indices for Fluoro- and Chloro-Substituted Methanes Using ASA Fitted Densities

	ASA Carbó indices								
	CH$_4$	CH$_3$F	CH$_3$Cl	CH$_2$F$_2$	CH$_2$Cl$_2$	CHF$_3$	CHCl$_3$	CF$_4$	CCl$_4$
CH$_4$	1	0.84859	0.79724	0.63480	0.56817	0.52912	0.46513	0.46322	0.40334
CH$_3$F		1	0.80400	0.74457	0.57320	0.61198	0.46943	0.52997	0.40720
CH$_3$Cl			1	0.60517	0.71239	0.50481	0.58311	0.44195	0.50598
CH$_2$F$_2$				1	0.43159	0.79626	0.35381	0.67472	0.30718
CH$_2$Cl$_2$					1	0.35991	0.70309	0.31531	0.49162
CHF$_3$						1	0.29653	0.86851	0.25762
CHCl$_3$							1	0.25977	0.77340
CF$_4$								1	0.22700
CCl$_4$									1

explore this rotational space. Once this rotational arrangement is done for all pairs of atoms, the configuration with maximum value is taken as the leading point to the global maximum. Then the variable metric algorithm of Broyden—Fletcher—Goldfarb—Shano,[13] a variant of the Davison—Fletcher—Powell method, is used to completely reach the maximum in its whole six-dimensional space. Analytical gradients have been used. This divides approximately by an amount of 12 the computational timing with respective central differences and, as it is well known, reducing the required number of iterations with respect to numerical derivatives.

Results presented in Tables 9 and 10 reveal that the mean of percentage errors comparing exact-fitted QMSM is 0.056%, with a standard deviation of 0.05. The measures

between the lightest pairs of molecules, such as CH$_4$—CH$_4$, have a greater error, probably attributed to the absence of core electrons in hydrogen atoms and in the significant distance, about 0.1 Å, from nuclear positions to the attractor or the maximum of density around hydrogen. Therefore, for molecules containing hydrogens, the error could even be smaller, displacing the Gaussian function at the attractor coordinates. Nevertheless, an error of 0.2% in the worst case is a completely assumable difference, and further refinements in the fitting function are not required for QMSM purposes. The corresponding Carbó's similarity indices are listed in Tables 11 and 12. As in the measures, ASA indices present small percentage errors, with a mean of 0.066% and a standard deviation of 0.06.

3. CONCLUSIONS

The main task of this paper has been to construct a computer program ASAC for the fitting of any molecular electron density in an automatic and routinary manner, being strictly respectful with the statistical interpretation of electronic density as a probability density function. The immediate use of these fitted functions, the ASA densities, is to extend *ab initio* quantum molecular similarity measures to large molecular systems. The accuracy of these fast measures has been demonstrated all thorough the present work.

This kind of fitting procedure, the ASA method, gives us a picture of molecular electron densities as a superposition of spherical atomic shells. This suggests that these simple densities could improve the computations of molecular volumes and surfaces based on the electronic densities. Moreover, this shows that atoms in molecular environments do not significantly lose the sphericity of the electronic distribution or that molecular densities can be spanned, in a first approach, through spherical functions. Further improvements in this approximation can be introduced using spheroidal functions or allowing exponents to vary with respect to the three spatial Cartesian coordinates rather than keeping them equal. A worthwhile future work will analyze post-Hartree—Fock densities in order to corroborate these ideas with nearly real densities.

At the time of writing this paper, some preliminary results obtained in our laboratory on the application of quantum similarity to the study of conformational analysis seems to indicate that a positive restricted fitting is superior to a simple least squares fitting when modeling a promolecule electron distribution.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) (a) Carbó, R.; Leyda, L.; Arnau, M. How Similar is a Molecule to Another? An Electron Density Measure of Similarity between two Molecular Structures. *Int. J. Quantum Chem.* **1980**, *17*, 1185−1189. (b) Carbó, R.; Calabuig, B. Molecular Quantum Similarity Measures and N-Dimensional Representation of Quantum Objects. I. Theoretical Foundations. *Int. J. Quantum Chem.* **1992**, *42*, 1681−1693. (c) Carbó, R.; Calabuig, B. Molecular Quantum Similarity Measures and N-Dimensional Representation of Quantum Objects. II. Practical Applications. *Int. J. Quantum Chem.* **1992**, *42*, 1695−1709. (d) Carbó, R.; Calabuig, B.; Vera, L.; Besalú, E. Molecular Quantum Similarity: Theoretical Framework, Ordering Principles, and Visualization Techniques. *Adv. Quantum Chem.* **1994**, *25*, 253−313. (e) Besalú, E.; Carbó, R.; Mestres, J.; Solà, M. Foundations and Recent Developments on Molecular Quantum Similarity. *Top. Curr. Chem.* **1995**, *173*, 31−62.

(2) Mestres, J.; Solà, M.; Duran, M.; Carbó, R. On the Calculation of *Ab Initio* Quantum Molecular Similarities for Large Systems: Fitting the Electron Density. *J. Comput. Chem.* **1994**, *15*, 1113−1120.

(3) (a) Löwdin, P. O. Quantum Theory of Many-Particle Systems. I. Physical Interpretation by Means of Density Matrices, Natural Spin-Orbitals, and Convergence Problems in the Method of Configurational Interaction. *Phys. Rev.* **1955**, *97*, 1474−1489. (b) McWeeny, R. The Density Matrix in Many-Electron Quantum Mechanics. I. Generalized Product Functions. Factorization and Physical Interpretation of the Density Matrices. *Proc. R. Soc. London*, **1959**, *A253*, 242−259.

(4) Unsöld, A. Beiträge zur Quantenmechanik de Atome. *Ann. Physik.* **1927**, *82*, 355−393.

(5) (a) Coppens, P.; Pautler, D.; Griffin, J. F. Electron Population Analysis of Accurate Diffraction Data. II. Application of One-Center Formalisms to Some Organic and Inorganic Molecules. *J. Am. Chem. Soc.* **1971**, *93*, 1051−1058. (b) Schwarz, W. H. E.; Lagenbach, A.; Birlenbach, L. Density Matrices from position and momentum densities. *Theor. Chim. Acta* **1994**, *88*, 437−445.

(6) Hirschfelder, J. O.; Curtiss, Ch. F.; Bird, R. B. *Molecular Theory of Gases and Molecules*; John Wiley & Sons: New York, 1964.

(7) (a) Paoloni, L.; Giambiagi, M. S.; Giambiagi, M. On the symmetry properties of the AO basis for the semiempirical MO-LCAO description of aromatic systems. *Atti Soc. Nat. Mat. Modena* **1969**, *C*, 89−105. (b) Frost, A. A. Floating Spherical Gaussian Model of Molecular Structure. I. Computational Procedure. LiH as an Example. *J. Chem. Phys.* **1967**, *47*, 3707. (c) Moncrieff, D.; Wilson, S. Distributed basis sets of s-type Gaussian functions in molecular electronic structure calculations. *Mol. Phys.* **1994**, *82*, 523−530.

(8) Reeves, C. M.; Harrison, M. C. Use of Guassian Functions in the Calculation of Wavefunctions for Small Molecules. II. The Ammonia Molecule. *J. Chem. Phys.* **1963**, *39*, 11−17.

(9) (a) Ruedenberg, K.; Raffenetti, R. C.; Bardon, D. *Energy, Structure and Reactivity*; Proceedings of the 1972 Boulder Conference on Theoretical Chemistry; Wiley: New York, 1973; 164 pp. (b) Schmidt, M. W.; Ruedenberg, K. Effective convergence to complete orbital bases and to the atomic Hartree—Fock limit through systematic sequences of Gaussian primitives. *J. Chem. Phys.* **1979**, *71*, 3951−3962. (c) Feller, D. F.; Ruedenberg, K. Systematic Approach to Extended Even-Tempered Orbital Bases for Atomic and Molecular Calculations. *Theor. Chim. Acta* **1979**, *52*, 231−251.

(10) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. *Gaussian 92*, Revision B; Gaussian, Inc.: Pittsburgh, PA, 1992.

(11) Dunning, T. H.; Hay, P. J. *Mod. Theor. Chem.* **1976**, *1*, 1−28.

(12) Flörke, U.; Haupt, H. J. The First Dinuclear Cobalt-Rhenium Cluster Compound: Hexacarbonylbis(μ-dicyclohexylphosphanido)cobaltrhenium. *Acta Crystallogr.* **1993**, *C49*, 1906−1908.

(13) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes. The Art of Scientific Computing*; Cambridge University Press: New York, 1988; Chapter 10.