Theor Chim Acta (1994) 87: 465–476

Theoretica Chimica Acta

An analytical expression for interatomic surfaces in the theory of atoms in molecules

P. L. A. Popelier
University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, UK

Received May 19, 1993/Accepted August 8, 1993

Summary. According to the theory of "Atoms in Molecules" as developed by Bader and coworkers a molecule is partitioned into atoms separated by surfaces of zero flux in the gradient of the charge density. For the first time an accurate and explicit analytical expression is given for these interatomic surfaces. They are generated by a system of differential equations which can in principle be solved by using a series expansion. Unfortunately, this expansion has a small radius of convergence and can therefore not be applied in practice. However, by a combined Chebyshev–Fourier fit to a numerically obtained surface, the interatomic surface is globally described to any given accuracy. Finally, the algorithm is tested on a set of simple molecules and on the amide interatomic surfaces of the glycyl residue $|HNCH_2CO|$.

Key words: Electron density – Atoms in molecules – Interatomic surface – Analytical expression

1 Introduction

Molecular electron distributions, obtained either from experiment or from theoretical computation are frequently analyzed by using the theory of "Atoms in Molecules" [1]. This theory partitions the electron density into atoms in a natural way using the gradient vector field of ρ. The topological properties of this field are revealed by tracing *gradient paths*, curves in real space such that at every point of the path the gradient vector $\nabla\rho$ is tangent to the curve. These atoms are separated by surfaces of sometimes very complicated nature, called interatomic surfaces (IAS). These surfaces have to be known in order to perform numerical integrations over the atomic basin to obtain the average atomic properties [2]. For example, a bond energy can be expressed in terms of a surface integral [3], but also atomic multipole moments, volume, force, kinetic and potential energies, torque, power [4] can be evaluated via surface and volume integration. Only an implicit knowledge of the surface boundary is needed, explaining the present absence of any explicit expression to describe the surface globally over some desired range.

It is the purpose of this work to make a first step in the study of the IAS as an object in itself, or in a broader scope, to allow the quantitative study of the gradient vector field with differential geometrical tools. Therefore an analytical expression

must be constructed up to a sufficiently high degree of accuracy in order to be able to differentiate it quickly many times without serious loss in accuracy.

The paper is organized as follows: after a brief review of some crucial concepts it is shown why analytical techniques, although in principle feasible, are unwieldy and useless in practice. Accordingly, a flexible and non-arbitrary fitting algorithm is introduced that leads to a canonical analytical expression irrespective of the molecular coordinate system used. Finally some computational details of the developed algorithm are discussed and illustrated for a set of molecules.

2 The mathematical problem

The general properties of the topology of the gradient vector field of the charge density have been extensively reviewed before [1]. Here we will only focus on the definitions and concepts we need in order to understand the problem at hand. As shown in Fig. 1a, the molecule LiH is partitioned into two atomic basins by the IAS. This surface is a bundle of gradient paths which originate at infinity and terminate at the *bond critical point.* This point is a critical point (because $\nabla\rho = \mathbf{0}$)

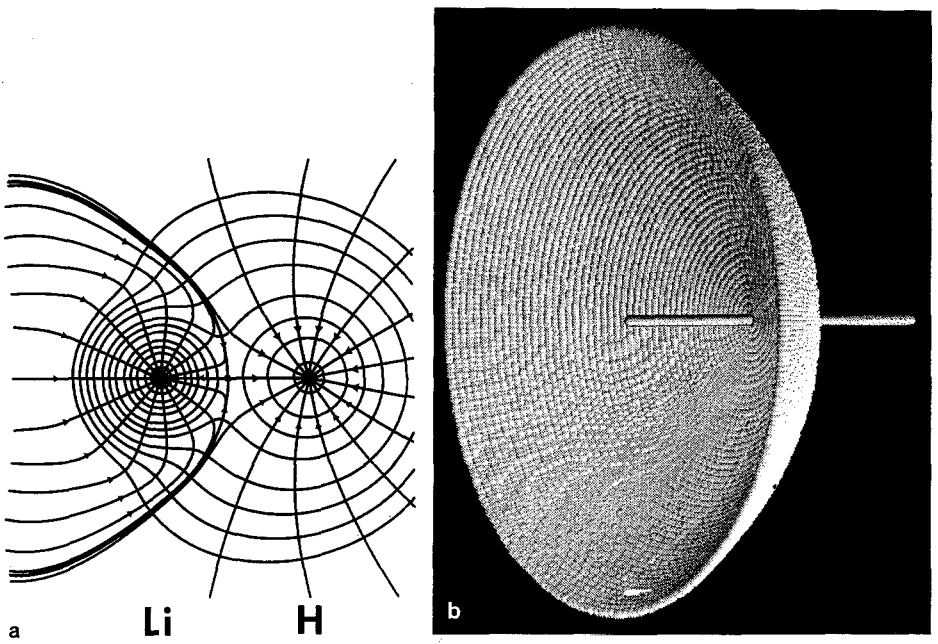

Fig. 1a. Gradient vector field of LiH superimposed on the family of iso-density contour lines. The outer contour value corresponds to 0.001 a.u. and the values increase in order 2×10^n, 4×10^n, 8×10^n with n increasing in steps of unity from $n = -3$. The sense of the gradient paths is denoted by arrows and the bond critical point is marked by a dot. They constitute the intersection between the interatomic surface and the plane of the paper. This picture illustrates the general topological features involved in the construction and appearance of an interatomic surface. **Fig. 1b.** A three-dimensional display of the same interatomic surface in LiH and its associated bond path intersecting the surface at the (3, −1) bond critical point. The lithium atom is situated at the endpoint of the bondpath on the concave side of the surface. The reader is asked to compare this view with the intersection shown in **Fig. 1a**

and is denoted by (3, − 1) where the first integer refers to the number of non-zero eigenvalues of the Hessian of ρ. The second integer is the "sums of signs" of the eigenvalues after according + 1 to a positive eigenvalue and − 1 to a negative.

The bond critical point is the origin of the IAS as made clear by Fig. 1b. Classically, the following equation describes the IAS:

$$\nabla\rho \cdot \mathbf{n}(\mathbf{r}) = 0 \quad \forall \mathbf{r} \in \text{surface}. \tag{1}$$

In words this equation states that the gradient of the electron density is perpendicular to the normal **n** of the surface. Thus the gradient of ρ is parallel to the surface, or finally, the surface *is* a bundle of gradient paths. The only condition that makes the above equation describe the IAS in particular is that it holds for *all* points, a fact which is not enough stressed.

The whole topology of a charge distribution is contained in a system of three ordinary differential equations (ODEs), Eq. (2):

$$\frac{d\mathbf{r}}{dl} = \frac{\nabla\rho(\mathbf{r})}{\| \nabla\rho(\mathbf{r}) \|}, \tag{2}$$

where **r** is a position vector and l is the pathlength. This equation expresses in differential form that the tangent to a gradient path is a normalized gradient vector. Finding an expression for the gradient paths and ultimately the IAS (which is a bundle of selected paths) reduces to solving the initial value problem in Eq. (2). It is a property of this initial value problem that one needs to specify just one (non-critical) point in real space to completely determine the solution for the unique gradient path through that point.

The complexity of this equation is entirely determined by ρ, which is in general an extensive linear combination of primitive Gaussian functions. The mathematical literature on these systems of ODEs is mostly concerned either with qualitative descriptions of the solution (i.e., their topology) or with entirely numerical recipes to trace the gradient paths. Closed analytical expressions for this equation are out of the question except for the local behaviour of a trajectory in the neighbourhood of a critical point as explained below.

An alternative initial value problem can be formulated by introducing the more abstract independent variable s. This new path parameter is related to l by the infinitesimal relation $ds = dl/\| \nabla\rho \|$, so that Eq. (2) becomes

$$\frac{d\mathbf{r}}{ds} = \nabla\rho(\mathbf{r}). \tag{3}$$

In order to describe a complete path in space, s has to be varied from $-\infty$ to $+\infty$ (see Ref. [1, p. 103]). This fact prompts the idea of natural coordinates [2] s, θ and ϕ, where the latter two angles are the usual angular spherical coordinates. By varying these coordinates over their complete range one does describe the whole atomic subspace.

The initial value problem as given in Eq. (3) is suitable for analyzing the local behaviour of a trajectory in the neighbourhood of a bond critical point with position vector $\mathbf{r}_c$. The standard mathematical technique employed here is linearisation [5], where the charge density $\rho(\mathbf{r})$ is expanded around $\mathbf{r}_c$ in a Taylor series, truncated after the second-order terms:

$$\rho(\mathbf{r}) \approx \rho(\mathbf{r}_c) + \sum_i (\rho_i)_{r=r_c}(r_i - r_{c,i}) + \tfrac{1}{2}\sum_{i,j} (\rho_{ij})_{r=r_c}(r_i - r_{c,i})(r_j - r_{c,j}), \tag{4}$$

where $\mathbf{r}(r_1, r_2, r_3)$, $\mathbf{r}_c(r_{c,1}, r_{c,2}, r_{c,3})$, $\rho_i \equiv \partial\rho/\partial r_i$ and $\rho_{ij} \equiv \partial^2\rho/\partial r_i \partial r_j$. Note that the second group of terms vanishes because $\nabla\rho = \mathbf{0}$ at the bond critical point. Substituting this expansion into the system of ODEs and diagonalizing the Hessian yields three uncoupled first-order differential equations, whose solution is

$$r'_i = r'_{i,c} + (r'_{i,o} - r'_{i,c})\exp(\lambda_i s) \quad (i = 1, 2, 3). \tag{5}$$

The subscript o refers to the chosen initial point for the trajectory, λ denotes the eigenvalue of the Hessian and the primes emphasize that the coordinates of the position vector are expressed with respect to the eigenvector axes system. If $\mathscr{U}$ represents the unitary matrix transforming the original molecular frame to the eigenvector frame, then the coordinates with respect to the original molecular frame are given by $\mathbf{r} = \mathscr{U}^{-1}\mathbf{r}'$ [6].

How should the initial points be selected to construct the bundle of gradient paths constituting the IAS? A practical choice is a small circle around the bond critical point in a plane with the proper orientation. This plane is determined by the two eigenvectors of the Hessian of ρ corresponding to the two negative eigenvalues of this matrix, which we call the *eigenplane.*

It is worthwhile to mention that the following very simple system of ODEs also contains an "interatomic surface" or *separatrix* in mathematical language; $\dot{x} = -x$, $\dot{y} = 1 - x^2 - y^2$ [5]. This system has been useful as a "pilot" system in preliminary research to investigate the performance of various analytical procedures.

3 The analytical approach

A very general method to solve differential equations (and systems of them) is the Picard method [7]. Although it appears mostly in a purely mathematical context (e.g. existence theorems) it has some practical use. When applied to the "pilot" system it fails by serious convergence problems in the series expansion that is obtained.

There is a rich literature on the solution of *second*-order differential equations, also known as boundary value problems. Beside the variationally based Rayleigh-Ritz [8] method, there is the projection (or Galerkin) method, [8] but for general cases the finite element technique [8] is among the most widespread in functional analysis. An initial value problem can be treated as a boundary problem where one integration limit is moved parametrically to infinity. This route is not very feasible because of extremely complicated integrals.

Rather, two more direct methods can be applied: the Taylor series substitution method [9] and Adomian's decomposition method [10]. The latter approach has been implemented and the results will be discussed below but it is useful to explain briefly the formalism in the former method to understand the decomposition method better.

The key idea is to expand the three unknown solutions $r_i(t)$ $(i = 1, 2, 3)$ by a Taylor series in the independent parameter t which can be s or l (see above):

$$r_i(t) = \sum_{n=0} \frac{1}{n!} \frac{d^n r_i}{dt^n} t^n. \tag{6}$$

The solution is known if all derivatives can be computed and evaluated at the initial point. The zero-th derivative is the initial point itself and the first derivative is trivial because it is given by the ODE's themselves $dr_i/dt = \rho_i f (i = 1, 2, 3)$ where f is 1 or 1 $\|\nabla\rho\|$, depending on whether one solves Eq. (2) or Eq. (3). Higher

derivatives can then be obtained by differentiating both sides. For example, the second derivative is

$$\left(\frac{d^2 r_i}{dt^2}\right) = \frac{d\rho_i}{dt} = \sum_j \frac{\partial(\rho_i f)}{\partial r_j}\frac{r_j}{dt}. \tag{7}$$

If one opts for $f = 1$ (as in ODE Eq. (2)), the partial derivatives of $\rho_i f$ up to an arbitrary order are easy to evaluate (see Appendix). If however $f = 1/\|\nabla\rho\| = (\rho_x^2 + \rho_y^2 + \rho_z^2)^{-1/2}$ the rule for differentiation of compound functions leads to a proliferation of terms and combinatorically complicated summations. The discovery of a recursion formula in this new context considerably reduced the code for the implementation but in view of the discussion below it is questionable of this route could ever compete with the fitting algorithm presented in the next section.

Consider the general equation of the form $L(u) + N(u) = x(t)$ where L is a linear operator and N represents a nonlinear operator. In the decomposition method the solution $u(t)$ is expressed into components with respect to the parameter λ or $u(t) = \sum_{n=0}^{\infty} \lambda^n u_n(t)$. The non-linear term appearing in the equation to be solved is similarly expressed as $N(u) = \sum_{n=0}^{\infty} \lambda^n A_n$. The A_n polynomials are defined such that each A_n depends only on u_i for i from 0 to n, so $A_0 = A_0(u_0)$, $A_1 = A_1(u_0, u_1)$, $A_2 = A_2(u_0, u_1, u_2)$, etc. These polynomials are defined as

$$\frac{1}{n!}\frac{d^n}{d\lambda^n} N(u(\lambda))|_{\lambda=0}$$

and are determined by the form of N. In our case L is total differentiation with respect to the independent variable t (or $L = d/dt$) and since the ODE is autonomous it follows that $x(t) = 0$. The approximating solutions are then given by $u_n = L^{-1}A_{n-1} = \int_0^1 dt' A_{n-1}$ $(n \geqslant 1)$ where u_0 is a known initial point.

Explicit expressions for A_n are given by Adomian expanded up to $n = 10$ [10]. Since we have a nonlinear vector function $\mathbf{N}$ (components N_i) of more than one variable the A_n's had to be computed by successive applications of the chain and product rule. As this becomes a formidable task for high n the derivation was only accomplished up to $n = 4$. If we denote the desired solutions by r_i $(i = 1, 2, 3)$ then $r_i = \sum_{n=0}^{\infty} r_{i,n}(t) = \sum_{n=0}^{\infty} (1/n!) c_{i,n} t^n$. The coefficients c are listed up to fifth order below:

$$\begin{aligned}
c_{i,0} &= r_{i,0},\\
c_{i,1} &= N_i,\\
c_{i,2} &= \sum_{j=1}^{3} (\partial_j N_i)_0 c_{j,1},\\
c_{i,3} &= \sum_{j,k=1}^{3} (\partial_{jk}^2 N_i)_0 c_{j,1} c_{k,1} + (\partial_j N_i)_0 c_{j,2},\\
c_{i,4} &= \sum_{j,k,l=1}^{3} (\partial_{jkl}^3 N_i)_0 c_{j,1} c_{k,1} c_{l,1} + 3(\partial_{jk}^2 N_i)_0 c_{j,2} c_{k,1} + (\partial_j N_i)_0 c_{j,3},\\
c_{i,5} &= \sum_{j,k,l,m=1}^{4} (\partial_{jklm}^4 N_i)_0 c_{j,1} c_{k,1} c_{l,1} c_{m,1} + 6(\partial_{jkl}^3 N_i)_0 c_{j,1} c_{k,1} c_{l,2}\\
&\quad + 3(\partial_{jk}^2 N_i)_0 c_{j,2} c_{k,2} + 4(\partial_{jk}^2 N_i)_0 c_{j,1} c_{k,3} + (\partial_j N_i)_0 c_{j,4},
\end{aligned} \tag{8}$$

where $\partial_{ij\ldots l}^n$ is a shorthand notation for $\partial_{ij\ldots l}^n / \partial r_i \partial r_j \ldots \partial r_l$.

As a test for the performance of these analytical expressions the gradient paths up to fourth order were plotted for the IAS of LiH. The radius ε of the initial point circle proved to be a crucial parameter for the quality of the gradient path at finite order as opposed to the fitting algorithm discussed below. If ε is set to 0.001 a.u. the analytical path matched the true (numerical) path with a maximum deviation of 0.03 a.u. over a range of 0.9 a.u. ($s \in [-300,0]$). It will be demonstrated below that with a fitted polynomial of the same order an interval roughly four times larger is approximated more than an order of magnitude more accurately. Techniques such as the Euler transformation [10] or Padé approximants [11, 12] to accelerate convergence were not helpful in this test case.

In summary, these results are disappointing and show that the power of an analytical approach, namely that it can in principle determine the track of a gradient path from knowledge of all the derivatives at only one point, leads to its ultimate defeat for these heavily non-linear ODE's. This applies to the elegant Carleman linearization as well [13]. Adding to this the bad performance of the "pilot" system in other unsuccessful attempts, one is inclined to abandon this approach and move towards a fitting procedure. The only practical use of the formulae in Eq. (8) may lie in a combined numerical-analytical approach in which the steplength of some numerical integrating algorithm can now be increased.

4 Theory of the fitting procedure

It is desirable to have a fitting method that can approximate the true surface as closely as needed without introducing arbitrary parameters. As shown below a most natural option is the combined Chebyshev–Fourier basis set.

The norm with respect to which the Chebyshev fit performed is called the *uniform* norm L_∞ and is defined as $L_\infty = \|f - p\|_\infty = \max_{a \leqslant x \leqslant b} |f(x) - p(x)|$ where $f(x)$ is the function that has to be approximated and $p(x)$ is the approximating polynomial for the interval $[a, b]$ and a given weight function $w(x)$.

As a result of Weierstrass's theorem we can improve a fit to an IAS to any desired degree of accuracy. The approximating polynomial $p(x)$ is constructed as a linear combination of Chebyshev polynomials $T_n(x)$, defined by $T_n = \cos(n \arccos(x))$. The first few are given in Eq. (9):

$$T_0(x) = 1,\ T_1(x) = x,\ T_2(x) = 2x^2 - 1,\ T_3(x) = 4x^3 - 3x,\ T_4(x) = 8x^4 - 8x^2 + 1. \tag{9}$$

A surface is a two-dimensional manifold that can be represented by a vector function $\mathbf{f}(u, v)$. Each position vector of a point on the surface is given by $\mathbf{r} = \mathbf{f}(u, v) = f_1(u, v)\mathbf{e}_1 + f_2(u, v)\mathbf{e}_2 + f_3(u, v)\mathbf{e}_3$, where u and v are the two independent parameters required to describe the entire surface and $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is a set of orthonormal basis vectors. One coordinate should express a periodic dependence, i.e. it is taken to be an angle, called θ, describing the initial points on the circle in the eigenplane. The second coordinate is chosen to be the pathlength (called l) describing the radial behaviour of the gradient paths. This option benefits from the reference to the local axis system centered at the bond critical point in that the expression obtained is independent of the molecular axis system.

The u-parameter curve is given by $\mathbf{f}(u, \theta_0)$ where u is l (θ_0 is a constant) and is called a *ray*. It describes the trajectory of a gradient path in the IAS from a given position on the initial point circle. In similar vein the θ-parameter curve is given by

$\mathbf{f}(l_0, \theta)$ (l_0 is a constant) and is called a *ring*. In general surfaces are often plotted as a twofold family of curves, each dependent on one parameter only. An IAS can be represented by a network of rings and rays.

The fitting process occurs in two stages as described by the general formulae below, given for the general case of three components f_i fitted with respect to (l, θ). First the rays are fitted, one by one, with respect to a basis of Chebyshev polynomials:

$$f_i(l, \theta) = \sum_n^N c_{in}(\theta) T_n(l). \tag{10}$$

The coefficients c_{in} are still parametrically dependent on the angular variable θ. Since the c_{in} are periodic functions of θ, they can be adequately expanded in a Fourier expansion, so that the total expression becomes

$$f_i(l, \theta) = \sum_{m,n}^{M,N} (d_{inm} \cos(m\theta) + d'_{inm} \sin(m\theta)) T_n(l), \tag{11}$$

where d_{inm}, d'_{inm} are the constants to be obtained [14].

The approximation formula for an arbitrary function $F(l)$ in $[-1, 1]$ is [15]

$$F(l) \approx \sum_{k=1} c_k T_{k-1}(l) - \tfrac{1}{2} c_1, \tag{12}$$

where the coefficients c_k are computed by Eq. (13)

$$c_k = \frac{2}{N} \sum_{m=1}^{N} F(l_m) T_{k-1}(l_m). \tag{13}$$

The value of l_m is one of the N zeros of $T_N(l)$ and is given by $l_m = \cos(\pi(m - \frac{1}{2})/N)$. The approximation can readily be extended to a general interval $[a, b]$ by the following change of variable: $s = 2(l - a - b)/(b - a)$. The latter two expressions yield the points within $[a, b]$ where the function $F(l)$ should be sampled radially. Gradient paths are traced from the bond critical point with some numerical ODE-integrator and at given intervals (specific for Chebyshev fitting) points on the IAS are calculated and stored. For the angular variable θ, equidistant sampling is performed.

A nice property of the Chebyshev fit is that if a fitted polynomial (typically of fairly high order N) is truncated to a polynomial of lower degree n, the latter is the most accurate approximation of degree n [15]. A theorem yielding L_∞ error bounds for the truncated polynomial informs us that the loss in using this polynomial as a best approximation is small for an arbitrary function [16]. At a qualitative level the above statement assures that even after drastic truncations the error is still reasonable as observed in all test cases.

Another useful phenomenon is that the individual Chebyshev coefficients rapidly converge for an increasing number of sampled points [17]. This fact will be illustrated numerically below. It is important in that it encourages one to use by default a fairly high number of sampled points (typically up to 40) and then truncate the expansion according to required fitting precision.

Finally, the quality of the fit is estimated by computing one discrete error measure and three continuous error measures L_1, L_2 and L_∞. In the following we focus on global error estimate, i.e. single numbers condensing the partial errors per ray and per component of the fitted vector function. The discrete measure is the familiar root mean square (rms) error defined as $(1/N \sum_i^N \sum_j^3 (g_{ij} - f_j(l_i, \theta_i))^2)^{1/2}$ where g_{ij} is the j-th component of the i-th grid point on the IAS and f_j is the j-th

component of the fitted function at that point. For an untruncated expansion the fit is exact for the grid points and therefore the rms error vanishes. It effectively measures how much the analytical rays deviate from the given data points, but it does not disclose what happens between those points.

With the continuous measures we are able to check the interpolated points and can for example probe for any spurious wiggling. The integrals appearing in the definition of L_1, L_2, and L_∞ are approximated by a sum of typically a few hundred terms. Therefore equally spaced points on the rays are accurately computed and stored.

5 Computation

The IAS are fitted using the program MORPHY [18]. The wavefunctions of the optimized test molecules were obtained by the package GAUSSIAN 92 [19] at the HF/6-31G**//HF/6-31G** [20] level, except for the blocked glycyl residue, α-(formylamino)-propanamide, computed at HF/6-311 + + G**//HF/6-31 + G* [21] (203 basisfunctions).

Prior to any fit it is important to check that the data set to be fitted is reliable. For example, it is well known that the smoother a given function, the quicker L_∞ decreases or the better the fit for a given polynomial degree [16]. It has been observed in our case that the error L_∞ can drop by several orders of magnitude if (slight) wiggling of the rays of the IAS is removed.

The quality of the fit is reliably determined by the *global* measures defined in Eq. (14), provided N is large enough and the traced paths are accurate:

$$
\begin{aligned}
L_1 &= \|r - f(l,\theta)\|_1 \approx \frac{1}{N}\sum_i^N |r_{i,j} - f_{i,j}|, \\
L_2 &= \|r - f(l,\theta)\|_2 \approx \frac{1}{N}\left(\sum_i^N |r_{i,j} - f_{i,j}|^2\right)^{1/2}, \\
L_\infty &= \|r - f(l,\theta)\|_\infty \approx \max_{i,j} |r_{i,j} - f_{i,j}|, \quad (j = 1, 2, 3).
\end{aligned}
\tag{14}
$$

These reference paths are numerically computed by the modified midpoint method [15], which is an improved version of the second-order Runge–Kutta method. The complete interval to be traced is subdivided according to the step length, typically between 0.01 and 0.02 a.u. This steplength is deliberately kept small in order to obtain a sufficient number of test points (N in the equation above) in between the sampled points to probe the quality of the fitted paths. The accuracy of the integration is determined by doubling the number of intermediate steps per subinterval until its endpoint changes by less than a preset accuracy value, 10^{-5} a.u. in our case. The radius of the initial point circle is set to 0.001 a.u. throughout. Once the Chebyshev coefficients are obtained the analytical expression is *not* converted into a Taylor polynomial because evaluation of the latter gives rise to gigantic errors for orders higher than ten. Rather Clenshaw's recurrence [15] is used to evaluate the Chebyshev polynomial directly. Furthermore, in all reported results each ray has been traced up to the 0.001 a.u. contour surface of the electron density. Tests have established that the global error is drastically reduced if the interval lengths for the Chebyshev fit are made to depend on the angular variable θ.

Linear molecules provide an interesting test case to focus on the purely radial fitting since for a cylindrically symmetrical charge density the IAS is a *surface of revolution* where the *profile* can be represented by any ray [22]. The equation of the surface then reduces to

$$x = f_1(s)\cos(\theta), \qquad y = f_2(s)\sin(\theta), \qquad z = f_3(s). \tag{15}$$

As a concrete example of an analytical expression we display in Eq. (16) the function f_3 for LiH where the molecular axis coincides with the z-axis. The linear combination of Chebyshev polynomials is obtained by truncating a fit of thirty terms to ten terms, for an integrator accuracy of 10^{-7} a.u.:

$$\begin{aligned} f_3(s) = {} & 0.60 + 7.96\times 10^{-1}T_1(s) + 1.80\times 10^{-1}T_2(s) - 1.58\times 10^{-2}T_3(s) \\ & - 4.69\times 10^{-3}T_4(s) + 1.63\times 10^{-4}T_5(s) + 3.91\times 10^{-4}T_6(s) \\ & - 1.18\times 10^{-5}T_7(s) - 4.59\times 10^{-5}T_8(s) + 6.55\times 10^{-6}T_9(s). \end{aligned} \tag{16}$$

Note that the Chebyshev coefficients are rapidly decreasing with increasing order. For this particular fit $L_1 = 1\times 10^{-5}$ a.u., $L_2 = 8\times 10^{-6}$ a.u., $L_\infty = 1\times 10^{-5}$ a.u. and the rms error is 8×10^{-6} a.u. To illustrate the aforementioned phenomenon of the stability of lower order Chebyshev coefficients, an untruncated expression of

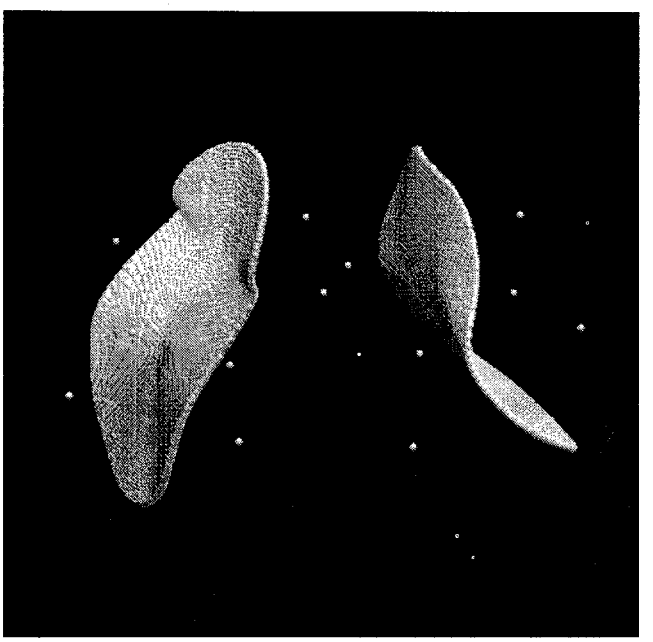

Fig. 2. The two interatomic surfaces delineating the glycyl residue in α-(formylamino)-propanamide. The two surfaces shown are constructed from the corresponding analytical expression, obtained by a fit to 20 traced rays and 20 rings. Each white dot represents a nucleus, the one in the center being the α carbon of the glycyl residue $|HNCH_2CO|$. The three dots at the right denote the NH_2 group and the ones at the left the HCO group. No discrepancy between the numerical and analytical interatomic surface can be observed at maximum magnification on a high resolution screen (using AVS). The global continuous errors for this picture are given in Table 1

Table 1. The norms[a] L_1, L_2 (both with unit weight) and L_∞ for a fit with 20 Chebyshev coefficients and 20 Fourier coefficients[b]. The integrator accuracy is 10^{-5} a.u. and the rays are traced up to the 0.001 a.u. electron density contour

Molecule	Surface	L_1	L_2	L_∞	Molecule	Surface	L_1	L_2	L_∞
LiH	Li|H	9(−6)	4(−5)	6(−4)	CO_2	C|O	2(−5)	2(−5)	4(−5)
BH	B|H	3(−5)	3(−5)	1(−4)	BeH_2	Be|H	1(−4)	1(−4)	3(−4)
LiF	Li|F	1(−7)	6(−7)	8(−6)	H_2O	O|H	2(−6)	2(−6)	2(−5)
HF	H|F	2(−6)	1(−6)	2(−6)	H_2CO	C|O	4(−5)	4(−5)	1(−4)
NaCl	Na|Cl	1(−8)	2(−8)	4(−7)	H_2CO	C|H	3(−7)	5(−7)	5(−6)
CO	C|O	2(−4)	2(−4)	4(−4)	CH_3OH	C|H	2(−6)	4(−6)	4(−5)
OH^-	O|H	1(−6)	1(−6)	2(−6)	CH_3OH	C|O	6(−6)	7(−6)	4(−5)
C_2H_2	C|H	2(−8)	2(−8)	4(−8)	CH_3OH	O|H	2(−6)	2(−6)	2(−5)
HCN	C|N	2(−5)	2(−5)	3(−5)	Glycyl	H(C=O)|N	1(−4)	2(−4)	9(−4)
HCN	H|C	3(−8)	3(−8)	5(−8)	Glycyl	C(=O)|NH_2	5(−6)	5(−6)	2(−5)

[a] Expressed in the form $x(y)$ for $x \times 10^y$

[b] For linear molecules there is just 1 Fourier coefficient

ten terms yields the same terms up to three digits except for $c_7 = 1.23 \times 10^{-5}$, $c_8 = -4.43 \times 10^{-5}$, $c_9 = -8.58 \times 10^{-6}$.

Rather than listing Chebyshev coefficients for the other test molecules only the global errors are shown in Table 1. All results were obtained in a few minutes per molecule on a STELLAR GS1025 computer except the glycyl residue which required about twenty minutes.

Finally, the possibility of an analytical surface integral [23] has been investigated. Again, in principle it is possible but it is unlikely to be faster than a numerical method. The bottleneck in evaluating the integrals is the appearance of non-negligible high order terms as an argument of the exponential functions appearing in the integrand. To our knowledge this general type of integral cannot be solved exactly [24]. Perhaps in combination with a quadrature technique for volume integrals the analytical expressions for the IAS may prove useful, since it can provide *everywhere* accurate intersection points between the interatomic surface(s) and the integration rays.

6 Conclusion

An explicit analytical expression has been obtained for the boundaries of an atom which is the first necessary step in a quantitative study of interatomic surfaces. This work shows that, in principle, the actual analytical integration of the central system of ordinary differential equations is possible. Individual gradient paths lying in the interatomic surface can be described by a series expansion using the derivatives of the charge density at *one single initial point*. This procedure is therefore classifiable as an extrapolation method. Although this method can solve the system in theory, in practice it is cumbersome and exhibits slow convergence. Therefore an interpolation method has been applied on data obtained by a numerical integrator. Attention has been paid to the straightforward extendibility of the set of basis functions, the possibility of arbitrarily accurate fits, and potential transferability of Chebyshev coefficients by introducing the local Hessian frame. Clearly, this

method is superior to the purely analytical approach since much better convergence is obtained. It can handle a molecule of any size and opens the route for a differential geometrical approach to the toplogy of the electron density.

Acknowledgements. The author wishes to thank Dr. RFW Bader for his critical reading of the manuscript and Dr. A Stone for his valuable comments.

Appendix

Computation of a general derivative of ρ of any order n

Usage of Gauss's product theorem avoids the application of the chain rule which is needed when the derivation passes the intermediate stage of molecular orbitals. Indeed, the total charge density can be broken down as follows [25]:

$$\rho = \sum_{i,j=1}^{N} \mathscr{P}_{ij} G_{ij}, \tag{A1}$$

$$G_{ij} = \exp\left(\frac{-\alpha_i \alpha_j}{\alpha_i + \alpha}(\mathbf{A}_i - \mathbf{A}_j)^2\right) \prod_{k=1}^{3} G_{P_{ij},k}, \tag{A2}$$

$$G_{P_{ij},k} = \sum_{m_k=0}^{l_i+l_j} f_{m_k,ij} r_{k,P_{ij}}^{m_k} \exp(-\gamma_{ij} r_{k,P_{ij}}^2), \tag{A3}$$

where

$$\gamma_{ij} = \alpha_i + \alpha_j, \qquad \mathbf{r}_{P_{ij}} = \mathbf{r} - \mathbf{P}_{ij}, \qquad \mathbf{P}_{ij} = \frac{\alpha_i \mathbf{A}_i + \alpha_j \mathbf{A}_j}{\mathbf{A}_i + \mathbf{A}_j}$$

and

$$f_{m_k,ij} = \sum_{r=0}^{l_i} \sum_{s=0}^{l_j} (A_{i,k} - P_{ij,k})^{(l_i-r)} (A_{j,k} - P_{ij,k})^{(l_j-s)} \frac{l_i! l_j!}{(l_i-r)!(l_j-s)! r! s!}.$$

In this scheme $\mathscr{P}$ is the density matrix with respect to the primitive Gaussian functions G_i, the angular part of which is expressed as a cubic harmonical function: $G_i = \prod_{k=1}^{3} (r_k - A_k)^{l_k} \exp(-\alpha_i(\mathbf{r} - \mathbf{A})^2)$. This is the general primitive function centered on a point in space with position vector **A**. Note that the parameters $f_{m_k,ij}$ are only dependent on the nuclear position vectors, the exponential coefficients α and the angular quantum numbers l_k, which makes these quantities evaluable prior to any knowledge of P. This equation scheme further shows that any partial derivative of ρ can be written as a simple product of derivatives with respect to *one* coordinate. The n-th derivative of Eq. (A3) yields Eq. (A4) (after omission of indices i and j):

$$\frac{d^n G_k}{d^n r_k} = \sum_{m=0} f_m \frac{d^n}{d^n r_k} (r_k^m \exp(-\gamma r_k^2)) = \sum_{m=0} f_m \frac{d^n}{d^n r_k} E_m(r_k). \tag{A4}$$

Setting E_{-1} to zero the recursive derivative is expressed as (for $m \geqslant 0$)

$$\frac{dE_m}{dr} = mE_{m-1} - 2\gamma E_{m+1}. \tag{A5}$$

Taking advantage of the permutational symmetry in a general partial derivative (the index order is immaterial) differentiation up to 30th order has been implemented in a very compact code.

References

1. Bader RFW (1990) in: Atoms in molecules. A quantum theory. Clarendon, Oxford
2. a. Biegler-König FW, Nguyen-Dang TT, Tal Y, Bader RFW, Duke AJ (1981) J Phys B: At Mol Phys 14:2739, b. Biegler-König FW, Bader RFW, Tang TH (1982) J Comp Chem 13:317
3. Bader RFW, Wiberg KB (1987) in: Erdahl R, Smith VH Jr (eds) Density matrices and density functionals. p 677
4. a. Bader RFW, Nguyen-Dang TT (1981) Adv Quant Chem 14:63. b. Bader RFW, Popelier PLA (1993) Int J Quant Chem 45:189
5. Verhulst F (1990) Nonlinear differential equations and dynamical systems. Springer, Berlin
6. Attempts to pursue this approach beyond second order lead to serious difficulties, one of which is the diagonalization of higher order tensors
7. Reinhard H (1986) Differential equations. Foundations and applications. North Oxford Academic, UK
8. Milne RD (1980) Applied functional analysis. Pitman, MA
9. Nagle RK, Saff EB (1986) Fundamentals of differential equations. Benjamin, USA
10. Adomian GA (1989) Nonlinear stochastic system theory and applictions to physics. Kluwer, Dordrecht
11. Baker GA Jr (1975) Essentials of Padé approximants. Adademic, NY
12. Char BW, Geddes KO, Gonnet GH, Watt SM (1990) Maple reference manual. Watcom Publ, Waterloo, Canada
13. Kowalski K, Steeb WH (1986) Nonlinear dynamical systems and Carleman linearization. World Scientific, Singapore
14. Note that this function is separable, i.e. it can be expressed as a product of two summations, each in one single variable. This facilitates partial differentiation and integration of $f_i(l, \theta)$
15. Press WH, Flannery BP, Tenkolsky SA, Vetterling WT (1986) Numerical recipes. Cambridge Press, UK
16. a. Rivlin TJ (1990) Chebyshev Polynomials. From approximation theory to algebra and number theory, Wiley, US, p. 166ff. b. Dahlquist G Björck (1984) Numerical methods. Prentice-Hall, NJ
17. More precisely, a low order coefficient c_n of the Chebyshev polynomial T_n in an expansion with N sampled points ($N \gg n$) is very close to the same coefficient c_n obtained in another expansion with M sampled points, if M is not too different from N
18. Popelier PLA (1992) Program MORPHY, McMaster Univ.
19. Gaussian 92 (1992) Frisch MJ, Trucks GW, Schlegel HB, Gill PMW, Wong MW, Foresman JB, Johnson BG, Schlegel HB, Robb MA, Replogle ES, Gomperts R, Andres JL, Raghavachari K, Binkley JS, Gonzalez C, Martin RL, Fox DJ, DeFrees DJ, Baker J, Stewart JP, Pople JA, Gaussian Inc, Pittsburgh, PA
20. Hariharan PC, Pople JA (1972) Chem Phys Lett 16:217
21. a. Frisch MJ, Pople JA, Binkley JS (1984) J Chem Phys 80:3265, b. Clark T, Chandrasekhar J, Spitznagle GW, v.R. Schleyer P (1983) J Comp Chem 4:294
22. Struik Dirk J (1961) Differential geometry. Addison-Wesley, London
23. Marsden JE, Tromba JA (1981) Vector calculus. Freeman, San Fran
24. Gradshteyn IS, Ryzhnik IM (1965) Table of integrals, series and products. Academic Press, NJ
25. Daudel R, Leroy G, Peeters D, Sana M (1983) Quantum chemistry. Wiley, UK