Communications on Quantum Similarity (2): A Geometric Discussion on Holographic Electron Density Theorem and Confined Quantum Similarity Measures

R. CARBÓ-DORCA, E. BESALÚ
Institut de Química Computacional, Universitat de Girona, Girona 17071, Catalonia, Spain

Received 1 December 2009; Accepted 8 February 2010
DOI 10.1002/jcc.21537
Published online 21 April 2010 in Wiley InterScience (www.interscience.wiley.com).

Abstract: The so-called holographic electron density theorem (HEDT) is analyzed from an algebraic perspective, and a brief analytical point of view is also given. The connection of the HEDT with quantum similarity measures (QSM) over electronic density functions (DF) is studied using GTO functions, atomic ASA DF, and promolecular ASA DF. Restricted integration of QSM over a box of finite side length is discussed for all this DF. This work emphasizes the geometric aspects of HEDT, but for the sake of completeness, some analytical insight based on a general Taylor series expansion is also given at the end.


Key words: holographic electron density theorem (HEDT); electronic density functions (DF); quantum similarity measures (QSM); atomic shell approximation (ASA); confined QSM; general Taylor series expansion (TSE) of DF

Introduction

Literature Antecedents on HEDT

At the turn of past century, Mezey1 described from a topological point of view the holographic theorem involving electronic density (HEDT)* functions, based in previous work of various authors, starting from an early DFT paper by of Riess and Münch2 and an old crystallographic paper by Karle and Hauptmann.3 One important aspect of the difference between the works of Munch and Reiss and Mezey is that the previous authors result could not apply for complete molecular densities but only for a finite, bounded domain of space, so it was valid only for artificial molecular models. However, Mezey proposed a general framework applicable to complete molecules, by circumventing the limitations of analytic continuation theorems for compact sets. Mezey, using the fact that beyond some distance of the origin, electron density converges uniformly to zero, connected this density function (DF) behavior with the Alexandrov one-point compactification theorem. Doing the analysis in an abstract four-dimensional space, Mezey could prove the HEDT for real molecules, without artificial boundaries. Moreover, Mezey in this mentioned work also connected the HEDT set up with quantum similarity measures and indices.

Scope of This Article

This study, the second of a series of communications on quantum similarity (QS) theoretical and computational background,4 pretends first to analyze the HEDT, reformulating the formal proof from an analytical and geometric point of view and extending it to any DF and state. As far as the authors know, few attempts have been put forward to apply the HEDT, see for example refs. 5–11. For this reason, practical use of the theorem implications to assess its possible impact in Quantum Similarity Measures (QSM) will be discussed in this work, and thus some computational examples will be given with all the programs and needed information provided via specific websites.

According to the considerations of these previous lines, a brief and schematic mathematical introduction to the problem will be given first. Then, two simple introductory examples will be presented and analyzed: they are based on a spherical and a Cartesian 1s type GTO function, in a similar way as it has been done in the first paper of this series4 but constraining the integrations to a finite volume. This introduction will be followed by computation algorithms of constrained similarity measures between two 1s GTO within a finite volume, opening in this way the study of the atomic ASA density functions behavior. Constrained molecular QSM will be studied next, providing quite a deep insight about their properties. A chapter about constrained triple QSM will follow to enlarge the computational point of view of the constrained algorithms. Following this line

*Although this is the customary literature naming of the subject of this article, the present authors consider that it will be perhaps more adequate to use the term: Holographic Theorem of the Electron Density (HTED).

Correspondence to: R. Carbó-Dorca; e-mail: quantumqsar@hotmail.com
Contract/grant sponsor: Spanish Ministerio de Educación y Ciencia; contract/grant number: CTQ2009-09370/BQU

© 2010 Wiley Periodicals, Inc.
of though, a schematic outline for general GTO overlap QSM constrained integrals will be given. Finally, for the sake of completeness, an analytical argument based on a general Taylor series expansion (TSE) will be used to show a forgotten aspect of HEDT.

**Euclidean Spaces and DF**

One can start the algebraic and geometrical analysis of the HEDT with the well known structure of a real Euclidian vector space \( E_n(\mathbb{R}) \) of dimension \( N = 3n \). This is appropriate to describe the position coordinates of a whole set of \( n \) electrons, for instance, just taking every point in \( E_n(\mathbb{R}) \) as a column vector: \( \mathbf{r} \).

Moreover, any \( n \)-th order DF, \( \rho(\mathbf{r}) \), with \( \mathbf{r} \in E_n(\mathbb{R}) \), can be considered as a mapping of such an Euclidian space over the positive real numbers, that is: \( \rho : E_n \rightarrow \mathbb{R}^+ \).

**Bounded DF Parts and the HEDT**

Necessarily, HEDT formulation deals with parts of the DF which are bounded, that is: contained within a closed surface, which can be defined as an isodensity contour; among other choices which will be later described. For example, one can write such bounded DF part, \( \rho^{(B)}(\mathbf{r}) \) in the following way.

First, defining an iso-density surface by means of the rule:

\[
\forall \mathbf{r}_s \in E_n : \rho(\mathbf{r}_s) = \alpha \in \mathbb{R}^+ \rightarrow S(\mathbf{r}_s) = \alpha
\]

this defines a volume \( V^{(B)} \) bounded by \( S(\mathbf{r}_s) \):

\[
V^{(B)} = \left\{ \mathbf{r}^{(B)} | \rho(\mathbf{r}^{(B)}) \geq \alpha \right\}
\]

which in turn describes a bounded DF part:

\[
\forall \mathbf{r} \in V^{(B)} : \rho^{(B)}(\mathbf{r}) = \rho(\mathbf{r}) \wedge \forall \mathbf{r} \notin V^{(B)} : \rho^{(B)}(\mathbf{r}) = 0.
\]

The bounded DF \( \rho^{(B)}(\mathbf{r}) \) could also be redefined throughout a logical Kronecker delta (see for example refs. 13–16), by means of:

\[
\forall \mathbf{r} \in E_n : \rho^{(B)}(\mathbf{r}) = \delta \left( \mathbf{r} \in V^{(B)} \right) \rho(\mathbf{r})
\]

(2)

Along with the bounded volume \( V^{(B)} \), it also coexists an unbounded one: \( V^{(U)} \), which has the following properties:

\[
E_n = V^{(B)} \cup V^{(U)} \wedge V^{(B)} \cap V^{(U)} = \emptyset
\]

Inducing a unbounded DF residual part definition:

\[
\forall \mathbf{r} \in E_n : \rho^{(U)}(\mathbf{r}) = \delta \left( \mathbf{r} \notin V^{(B)} \right) \rho(\mathbf{r})
\]

such that the sum of both bounded and unbounded DF parts produces the following tautology:

\[
\rho(\mathbf{r}) = \rho^{(B)}(\mathbf{r}) + \rho^{(U)}(\mathbf{r})
\]

(3)

\[
\rho^{(B)}(\mathbf{r}) = \rho(\mathbf{r}) - \rho^{(U)}(\mathbf{r}) = \left[ 1 - \delta \left( \mathbf{r} \notin V^{(B)} \right) \right] \rho(\mathbf{r}).
\]

(4)

An alternative definition of the bounded DF part can be described using the residual unbounded DF part in the following manner:

This and the previous tautology (3) indicates that the DF can be decomposed as the sum of a bounded and a residual unbounded parts, which possess disjoint domains and ranges. Such a function construction looks like the topological version of the algebraic direct sum decomposition of a vector space. Within the vector semispace where the DF belong see refs.,15 and 18, for example, eqs. (2) and (4) correspond to the decomposition of the DF, represented as a vector, into the sum of two orthogonal vectors, the ones associated to the DF bounded and unbounded parts:

\[
\rho^{(B)}(\mathbf{r}) = \rho(\mathbf{r}) - \rho^{(U)}(\mathbf{r}) = \left[ 1 - \delta \left( \mathbf{r} \notin V^{(B)} \right) \right] \rho(\mathbf{r}).
\]

(4)

Equations (2) and (4) possess a formal structure such that both bear the explicit presence of the total DF and cannot be separated from it, a fact which permits to consider these equations as an algebraic proof of the HEDT; within any arbitrary order bounded DF part there is present information about the whole DF. Both eqs. (2) and (4) are explicit definitions which can be also interpreted as a projection of the DF over a given volume in Euclidean space, the logical Kronecker deltas acting as a projection device.

Therefore, as in the previous arguments and reasoning, there is not taken any state nor order constraint on the nature of the involved DF, the HEDT can be applied to ground or excited state DFs’ of any order. Because of this fact, the HEDT may be reformulated in the following alternative way: every bounded part of any DF can be considered as a projection of the global DF values into an appropriate bounded volume of a suitable Euclidian space.

---

1Given \( L \), a symbol corresponding to any logical or mathematical expression, then a logical Kronecker delta, which can be written as: \( \delta(L) \), can be translated in any high level language as an if \( (L) \) instruction. If \( L = \text{false} \), then \( \delta(L) \) = 0 and if \( L = \text{true} \), then \( \delta(L) \) = 1. Thus, \( \delta(L) \) is a generalization of the well-known Kronecker’s delta algebraic symbol: \( \delta_{ij} \), which in the formalism of a logical Kronecker’s delta could be written alternatively as: \( \delta_{ij} \equiv \delta(i = j) \).
Moreover, the bounded volume $V^{(B)}$ does not need to be in general wrapped into an isodensity envelope, as it can be constructed by any closed surface, an appropriate $N$-dimensional sphere or spheroid, a cylinder or a box, for instance. Then, the only additional problem which poses these alternatives is that definition (1) shall be accordingly modified.

**A Simple Example**

The discussion on the practical use of the previously expressed ideas starts with the simplest example one can imagine to setting up the posterior computational development, adequate to describe QSM within bounded volumes, that is: confined or bounded QSM.

**Definitions**

One can use the freedom of choice of the bounded regions to illustrate these previous remarks on HEDT with a schematic example. Suppose a one particle DF composed with a 1s GTO:

$$\gamma(r) = N(\lambda) \exp\left(-\lambda |r|^2\right),$$

therefore, the adequate Euclidean vector space will be $E_3$. One can consider such a GTO Minkowski normalized:

$$\langle \gamma(r) \rangle = 1 \rightarrow N(\lambda) = \left(\frac{\lambda}{\pi}\right)^{\frac{3}{2}}.$$

Defining a sphere of radius $R$ as a bounded volume around the GTO origin, then the DF bounded part will be defined as:

$$\gamma^{(B)}(r) = \delta\left(|r|^2 < R^2\right) \gamma(r)$$

and the boundary surface can be easily constructed:

$$\forall r_s \in E_3 : |r_s|^2 = R^2 \rightarrow \gamma(r_s) = N(\lambda) \exp\left(-\lambda R^2\right) = S(r_s)$$

in this case as in isolated single atoms, corresponding to an isodensity envelope.

**Fraction of Total Density over a Sphere**

The fraction of total density over a sphere of radius $R$ is obviously better computed in spherical coordinates in this oversimplified case. Then, the fraction of total density over a normalized 1s GTO, once integrating over the angular part, becomes:

$$\theta(R|\lambda) = 4\pi \left(\frac{\lambda}{\pi}\right)^{\frac{3}{2}} \int_0^R r^2 \exp\left(-\lambda r^2\right) dr$$

$$= 4\pi R^2 \left(\frac{\lambda}{\pi}\right)^{\frac{3}{2}} \int_0^1 \tilde{r}^2 \exp\left(-\lambda \tilde{r}^2\right) d\tilde{r}$$

$$= 4\pi R^2 \left(\frac{\lambda}{\pi}\right)^{\frac{3}{2}} F_1(\lambda R^2)$$

where the integral family:

$$F_m(w) = \int_0^1 s^m \exp\left(-ws^2\right) ds$$

corresponds to the well-known incomplete gamma function, see for example refs. 19–22. In fact, the incomplete gamma function of zero-th order can be expressed in terms of the well-known Fehler integral or error function$^{23}$ because it is easy to show that:

$$F_0(w) = \frac{1}{2} \left(\frac{\pi}{w}\right)^{\frac{3}{2}} \text{erf}\left(w^\frac{1}{2}\right)$$

and as one can write the recurrence relationship:

$$F_1(w) = \frac{1}{2w} (F_0(w) - \exp(-w))$$

$$= \frac{1}{2w} \left(\frac{1}{2} \left(\frac{\pi}{w}\right)^{\frac{3}{2}} \text{erf}\left(w^\frac{1}{2}\right) - \exp(-w)\right)$$

one can describe the first order incomplete gamma function in terms of the error function. Although a warning shall be given now as it is well known that proceeding in this way, numerical errors can appear in computational practice.$^{19}$ But despite of this, one can take the recurrence above just to show the role of the error function in constrained problems involving GTO.

**Fraction of Total Density over a Box**

Within the Cartesian GTOs usual computational framework, the easiest way to obtain the needed information about constrained integrals appears to be the possibility to use a cubic (or elongated) box wrapping the density in some volume of 3D space.

For example, the fraction of total density within a cubic box of length sides equal to $L = 2R$ can be computed with the third power of the integral:

$$v(R|\lambda) = 2 \left(\frac{\lambda}{\pi}\right)^{\frac{3}{2}} \int_0^R \exp\left(-\lambda x^2\right) dx$$

$$= 2 \left(\frac{1}{\pi}\right)^{\frac{3}{2}} \int_0^R \exp\left(-y^2\right) dy \rightarrow \left(\frac{1}{\pi}\right)^{\frac{1}{2}} dy = dx \wedge R = R^{\lambda \frac{3}{2}} = \text{erf}(R) \equiv v(R)$$

which being the 1s GTO normalized, becomes a function dependent of a parameter involving the radius of the sphere inscribed into the cubic box and the exponent. It will be called here the reduced radius: $R = R^{\lambda \frac{3}{2}}$. One can see without effort that such a function of the radius is simply the error function.

This result is different from the one shown in Equation (5), but one shall obviously expect some different outcome when comparing spherical or cubic integration volumes. In fact, the difference between both volumes can be easily expressed as:

$$\Delta = (8 - \frac{4}{3} \pi) R^3 \approx 3.8112R^3,$$

and the ratio between both volumes becomes: $P = \frac{\Delta}{\Delta} \approx 1.9099$. That is: when using a cubic box of side $2R$ as a constraining integration volume, such a
volume is roughly twice the one of an inscribed sphere of radius $R$. The constrained computations which follow are computed within a cubic box of side $2R$ and the results are given in this way. However, dividing them by 2 one could approximately guess how the result will be when a spherical constraint of radius $R$ is used.

Of course, one can also use the incomplete gamma function reversing equality (6) or for instance a polynomial approach to compute it, but the error function is an intrinsic function in fortran 95 and thus, the error function is preferable from the computational point of view for programming simplicity than the fortran 95 and thus, the error function is preferable from the computational point of view for programming simplicity than the

Accordingly, the probability fraction contained within the cubic box is only dependent of the reduced radius $R_j$, which can be taken as a unique parameter representing the spherical volume wrapped within the box. It is easy to see that with progressive values of the reduced radius one obtains for instance: $v(1) = 0.84270; v(2) = 0.999532; v(3) = 0.99998; \ldots$ Now, defining the unbounded residual probability content as: $u(R_j) = 1 - v(R_j)$, this function varies with the reduced radius like: $u(1) = 0.15730; u(2) = 0.00468; u(3) = 0.00002; \ldots$ The final GTO density fraction shall be computed with the third powers of the integral $v(R_j)$ values given above. Consequently, the residual probability content will tend to quite irrelevant values at relatively small radius, a behavior which can be attributed to the well-known error function trends and thus to the use of GTO.

Such a property precludes the results on atoms and molecules described under such kind of atomic basis sets, which will be shown in the next sections below.

**QSM between Constrained DF Parts**

The above result, concerning simplified DF consisting of a unique 1s GTO, can be put forward adding another function with different exponent to obtain the QSM (see for example refs. 26–72 for the theoretical development) within a bounded volume instead of the whole space, as Mezey has suggested when defining HEDT.

The product of two normalized 1s Gaussian functions located at the same origin now will be written by the composite Gaussian function:

$$\gamma(r)\eta(r) = \frac{1}{\pi^3} (\lambda \mu)^2 \exp\left(-\frac{\lambda + \mu}{\mu^2}\right)$$

so the similarity measure between both GTO within a cubic box of sides equal to $L = 2R$ can be written as the cube of the integral:

$$\zeta(R; \mu) = \langle \gamma(x)\eta(x) | R \rangle = \frac{2}{\pi} (\lambda \mu)^2 \int_0^R \exp\left(-\frac{\lambda + \mu}{\mu^2}\right) dx$$

which can be simplified, as it was in the above section when single 1s GTO case was analyzed. Now using a reduced radius having the scale factor being computed with the sum of the exponents of both functions:

$$\zeta(R; \mu) = \frac{2}{\pi} \left(\frac{\lambda \mu}{\pi (\lambda + \mu)}\right) \int_0^{R_{\mu}} \exp\left(-\frac{\lambda^2 - \mu^2}{\mu^2}\right) dy$$

$$= \left(\frac{\lambda \mu}{\pi (\lambda + \mu)}\right) \frac{\exp\left(-\frac{\lambda^2 - \mu^2}{\mu^2}\right)}{\mu^2}$$

The Carbo similarity index, $c(R; \mu)$, between both GTO could be readily calculated, as the associated self-similarities are $\zeta(R; \lambda)$ and $\zeta(R; \mu)$, thus:

$$c(R; \lambda; \mu) = \zeta(R; \lambda)(\zeta(R; \lambda)\zeta(R; \mu))^\frac{1}{2}$$

one of these:

$$r(\lambda; \mu) = \left(\frac{2(\lambda \mu)^2}{\lambda + \mu}\right)^{\frac{1}{2}}$$

corresponds to the square root of the ratio between geometric and arithmetic means of the involved GTO exponents: equals the Carbo index of two univariate GTO, when the similarity measures are computed over the whole range of the electron position coordinates (see the part (1) of these communications on QS$^3$). The other factor:

$$s(R; \mu) = \frac{\exp\left(R_{\mu}\right)}{\exp\left(R_{\mu}\right)}$$

and

$c(R; \lambda; \mu)$ can be interpreted as a Carbo index of the error functions, evaluated with the reduced radius of the sphere, included in turn into the cubic box and scaled by the square root of the corresponding sums of the involved exponents. When both exponents become the same, then the confined similarity index part will be unity and the same can be said as well at infinite confinement radius, where the three error functions become unity.

The final Carbo similarity index shall be computed as a third power: $C = c(R; \lambda; \mu)$ due that the integrals deduced above are computed over one of the three dimensional coordinates.

However, this last remark permits to obtain more diverse results, employing parallelepipedal boxes inscribing spheroids, instead of a cubic box inscribing a sphere; in general, whenever the integration radius is different for each space direction, one can write the overall Carbo similarity index as a product of three terms:

$$C = c(R_1; \lambda; \mu)c(R_2; \lambda; \mu)c(R_3; \lambda; \mu).$$
An Atomic Shell Approximation (ASA) General Situation

Another related question which can be associated to HEDT, consists in the computation of the probability content associated to any ASA DF\textsuperscript{63–65} which can be written as a convex linear combination of 1s GTO functions, like the ones discussed and used in the previous section. The ASA DF for atoms can be supposedly normalized in the form of a shape function\textsuperscript{18}:

\[ \rho_A(r) = \sum_{\mu} w_\mu N(z_\mu) \exp\left(-z_\mu |r|^2\right) \wedge \sum_{\mu} w_\mu = 1 \wedge \langle \rho_A \rangle = 1. \]

The bounded DF probability contained within a cubic box of side length equal to \( L = 2R \), inscribing a sphere of radius \( R \) can be easily written as:

\[ \langle \rho_A | R \rangle = \sum_{\mu} w_\mu \nu \langle R | z_\mu \rangle^3 = \sum_{\mu} w_\mu \text{erf}^3 \left(R_{\|,z_\mu}\right). \]

The probability content of an ASA atom inside a cubic box corresponds to a convex linear combination of error functions of the reduced radius associated to each GTO, forming the ASA DF of the considered atom. Such a result can be seen as providing interesting information about atomic behavior because that ASA DF are quite accurate approximations of \textit{ab initio} atomic DF.

Therefore, from this simple result for any atom, one can approximately know which radius will provide the adequate fraction of the nuclear charge contained within the corresponding cubic box.

Bounded Similarity Measures Between Two ASA Atomic DF

An ASA similarity measure between two bounded atoms within the same cubic box of side length equal to \( L = 2R \) can be easily computed with the result already obtained in eq. (7), one can write:

\[ \langle \rho_A | R \rangle = \sum_{\mu} \sum_{\nu} w_\mu w_\nu \frac{\text{erf}^3 \left(R_{\|,z_\mu}\right)}{\text{erf}^3 \left(R_{\|,z_\nu}\right)} \]

Actually, within each term of this bilinear sum, two factors appear: one corresponding to the QSM between two GTO and being independent of the integration reduced radius; the other factor weights each integral by the corresponding error function, correcting the former integral term about the fact the integration is performed within a cubic box of finite side length. When the radius of the inscribed sphere becomes infinite, the error functions become unity and the restricted formula reverts to the usual similarity measure form.

### Table 1. Assorted Constrained ASA Atomic QSS at the 3-21G Basis Set Level for Several Radius (\( R \)) of a Sphere, Inscribed into a Constraining Cubic Box of Side 2\( R \).

<table>
<thead>
<tr>
<th>Radius</th>
<th>C</th>
<th>S</th>
<th>Mn</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>30.998875</td>
<td>803.520042</td>
<td>3484.333178</td>
</tr>
<tr>
<td>1.0</td>
<td>31.230626</td>
<td>805.415980</td>
<td>3498.072017</td>
</tr>
<tr>
<td>1.5</td>
<td>31.354548</td>
<td>805.708030</td>
<td>3499.469040</td>
</tr>
<tr>
<td>2.0</td>
<td>31.383955</td>
<td>805.805192</td>
<td>3499.491104</td>
</tr>
<tr>
<td>2.5</td>
<td>31.387620</td>
<td>805.823483</td>
<td>3499.491723</td>
</tr>
<tr>
<td>3.0</td>
<td>31.387883</td>
<td>805.825639</td>
<td>3499.491723</td>
</tr>
<tr>
<td>3.5</td>
<td>31.387894</td>
<td>805.825803</td>
<td>3499.491723</td>
</tr>
<tr>
<td>4.0</td>
<td>31.387894</td>
<td>805.825811</td>
<td>3499.491723</td>
</tr>
<tr>
<td>( \infty )</td>
<td>31.387894</td>
<td>805.825811</td>
<td>3499.491723</td>
</tr>
</tbody>
</table>

The fortran 90 program for computing atomic QSS can be downloaded from the website \[73\]. The coefficients for the ASA functions fitted to a 6-311G basis set can be downloaded from the general ASA website \[74\]. The web exponents are multiplied by two, when used within the programs furnished here.\textsuperscript{3}

\textsuperscript{3}The coefficients for the ASA functions fitted to a 6-311G basis set can be downloaded from the ASA atomic program website \[73\] or other choices can be found in the ASA website \[74\]. As in the 3-21G case the exponents are multiplied by two in the programs furnished.
Radius Inscribed into a Constraining Cubic Box of Sides \(2R\) for Atoms O, Ar, and Se. QSS Values Are Scaled by \(Z^{-2}\).

<table>
<thead>
<tr>
<th>(R)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>(\infty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>1.269257</td>
<td>1.272816</td>
<td>1.272837</td>
<td>1.272837</td>
<td>1.272837</td>
</tr>
<tr>
<td>Ar</td>
<td>3.718922</td>
<td>3.721324</td>
<td>3.721360</td>
<td>3.721360</td>
<td>3.721360</td>
</tr>
<tr>
<td>Se</td>
<td>8.339201</td>
<td>8.339670</td>
<td>8.339699</td>
<td>8.339699</td>
<td>8.339699</td>
</tr>
</tbody>
</table>

appropriate size it is accurately contained information on atomic self-similarity measures over the whole DF.

**Molecular QSM**

Such atomic behavior even if solely found between QSM under ASA approach may be expected to approximately hold within molecular QSM in some extent. In fact, when considering a DF promolecular description assembling ASA atomic parts, the approximate molecular DF is expressed as (see for example ref. 50):

\[
\rho_M(r) = \sum_{A \in M} Z_AR_A(r - r_A)
\]

where the coefficient set \(\{Z_A\}\) corresponds to some atomic populations or just plain nuclear charges and the set \(\{r_A\}\) contains the atomic position coordinates in Euclidian three dimensional space. An overlap QSM between two molecules \(\{P, Q\}\) and constrained within a cubic box of side length equal to \(L = 2R\) can be written as:

\[
\langle \rho_P \rho_Q \rangle = \sum_{A \in P} \sum_{B \in Q} Z^P_A Z^Q_B \left\langle \rho^P_A \rho^Q_B \right\rangle
\]

also, the integrals between two ASA atomic centers \(\{A, B\}\), each one located at the corresponding molecular structure, can be expressed furthermore as:

\[
\left\langle \rho^P_A \rho^Q_B \right\rangle = \sum_{\mu \in A} \sum_{\nu \in B} \omega^A_\mu \omega^B_\nu \left\langle \rho^P_A (r - r^P_A) \rho^Q_B (r - r^Q_B) \right\rangle
\] (8)

Moreover, the integral between the GTO basis functions in Cartesian coordinates becomes:

\[
\left\langle \rho^P_A (r - r^P_A) \rho^Q_B (r - r^Q_B) \right\rangle = \frac{8}{\pi^3} \left(x^A_\mu x^B_\nu \right)^2 \int_0^R \int_0^R \int_0^R \exp(-x^A_\mu |r - r^P_A|^2 - x^B_\nu |r - r^Q_B|^2) dx d\nu
dr = \frac{8}{\pi^3} \left(x^A_\mu x^B_\nu \right)^2 \left\langle \rho^P_A \rho^Q_B \right\rangle
\]

\[
\rho_C = \frac{x^A_\mu r_A + x^B_\nu r_B}{x^A_\mu + x^B_\nu},
\] (9)

in the way that one can write the basic integral needed as:

\[
\left\langle \rho^A_\mu \rho^B_\nu \right\rangle = \exp \left(-\frac{x^A_\mu x^B_\nu}{x^A_\mu + x^B_\nu} |r_A - r_B|^2 \right)
\]

\[
\int_0^R \int_0^R \int_0^R \exp(-\left(x^A_\mu + x^B_\nu \right) |r - r_C|^2) dx
dr = \exp \left(-\frac{x^A_\mu x^B_\nu}{x^A_\mu + x^B_\nu} |r_A - r_B|^2 \right) \left(\frac{\pi}{4 x^A_\mu + x^B_\nu} \right)^{\frac{3}{2}} \text{erf} \left(R_{x^A_\mu + x^B_\nu}\right)
\]

The integral above can be now expressed as the product of three parts like:

\[
\left\langle \rho^A_\mu \rho^B_\nu \right\rangle = \left\langle \rho^A_\mu \right\rangle \left\langle \rho^B_\nu \right\rangle \left\langle \rho^A_\mu \rho^B_\nu \right\rangle
\]

\[
= \frac{1}{8} \exp \left(-\frac{x^A_\mu x^B_\nu}{x^A_\mu + x^B_\nu} |r_A - r_B|^2 \right) \left(\frac{\pi}{4 x^A_\mu + x^B_\nu} \right)^{\frac{3}{2}} \text{erf} \left(R_{x^A_\mu + x^B_\nu}\right)
\]

thus, the complete integral is computed via the product:

\[
\left\langle \rho^A_\mu \rho^B_\nu \right\rangle = \left\langle \rho^A_\mu \right\rangle \left\langle \rho^B_\nu \right\rangle \left\langle \rho^A_\mu \rho^B_\nu \right\rangle
\]

\[
= \frac{1}{8} \exp \left(-\frac{x^A_\mu x^B_\nu}{x^A_\mu + x^B_\nu} |r_A - r_B|^2 \right) \left(\frac{\pi}{4 x^A_\mu + x^B_\nu} \right)^{\frac{3}{2}} \text{erf} \left(R_{x^A_\mu + x^B_\nu}\right)
\]

therefore, the final integral can be written as:

\[
\left\langle \rho^A_\mu (r - r^P_A) \rho^B_\nu (r - r^Q_B) \right\rangle = \exp \left(-\frac{x^A_\mu x^B_\nu}{x^A_\mu + x^B_\nu} |r_A - r_B|^2 \right) \left(\frac{x^A_\mu x^B_\nu}{x^A_\mu + x^B_\nu} \right)^{\frac{3}{2}} \text{erf} \left(R_{x^A_\mu + x^B_\nu}\right)
\] (10)

Hence, every term in the QSM between two molecules, as in the atomic case, can be written within the promolecular ASA framework as the third power of the error function with the reduced radius of the cubic box inscribed sphere integration as a variable, multiplied by a factor, which coincides with the overlap QSM over the whole space, see for example part 1 of this series. The error functions become in this way extra weighting factors in the bilinear form (8), expressing the QSM between two atoms of the different structures involved; they become unity when the integration inscribed sphere radius is infinite and the constrained integral transforms into the unconstrained one.

A remark about the center of the cubic box, which one can choose to perform the bounded integration, is due now. When two molecules are compared by means of a QSM as the one discussed here, one of the two molecular structures (or both) can be translated and rotated in order that some of their atomic coordinate features are coincident. In the context of

---

**Table 2.** Variation of 6-311G Atomic ASA QSS with \(R\), a Sphere Inscribed into a Constraining Cubic Box of Sides \(2R\) for Atoms O, Ar, and Se. QSS Values Are Scaled by \(Z^{-2}\).
constrained DF integrals, one can move one molecule with respect
to the other if a pair of atoms, each belonging to
every molecule, is coincident, then nothing opposes to use this
joint atomic coordinates as the Cartesian common origin of both
molecules. Afterward, the integration box can be constructed
centered at the same time at this common origin.

However, in the molecular QSM when the translation
providing a change of origin is the same for both molecules, the final
integral result in terms of the radius of the inscribed sphere
appears to be independent of the chosen origin of coordinates.

It is easy to prove such invariance upon any change of the
common origin of the involved molecules: the integral (10) has
a unique term depending of both atomic coordinates, one
belonging to molecule $P$ and the other to molecule $Q$; such term
is a squared distance, subtracting a third coordinate vector: $t$, say, from both atomic center coordinates one has:

$$\forall t : |r^P_\lambda - r^Q_\mu|^2 = |(r^P_\lambda - t) - (r^Q_\mu - t)|^2.$$ 

If every two atom integral is invariant, the whole molecular
QSM remains invariant.

A program has been built up to test constrained QS measures
between two molecular structures, which for simplicity and to
avoid the superposition problem, are chosen to be the same. Thus,
they are completely superposed in any case within any program
run. The computational procedure of this program consists into the
calculation of molecular QSS measures to avoid the additional
problem of molecular superposition, as commented before.

Hence, in this way, there will be no possibility that the
superposition problem get entangled with the constrained evaluation
of QSS measures. As QSM are origin independent under
the previously commented circumstances, the computational orig-
inal has been taken as the one provided by the native atomic
dimensions of both molecules. However, program coherence has
been testing changing the origin at the atomic sites and checking
the results provided by the program, which yield the same fig-
ures irrespective of the origin.

Table 3 below presents some computational tests carried out
on assorted molecules of various sizes as sample results. Here, a
third ASA basis set is also used; it has been fitted to Huzi-
naga, ab initio atomic HF basis set computations. Molecular
coordinates are obtained from HF computations under 3-21G
basis set within Spartan program, except for water and formal-
dehyde; in this case, the needed data information has been
obtained from Snyder and Basch calculations. Promolecular
ASA DFs have been calculated using the Mulliken populations
as weighting factors for the ASA atomic functions, constructed
within the Huzinaga’s basis set in all cases.

Constrained Triple Similarity Measures

Triple QSM have been defined several years ago and had been
found valuable to use as molecular descriptors, while in recent
times they have become part of the general theory connecting QS
and quantum QSPR. So, it seems interesting to have some pre-
liminary insight about how constrained triple QSM behave.

First of all, it will be studied the simple situation where three
1s GTO are centered at the origin of coordinates. Then, as in
the former two GTO case one can write:

$$\varphi(r)\gamma(r)\eta(r) = \frac{1}{\pi^3} \left(\frac{\kappa \lambda \mu}{\pi}\right)^\frac{3}{2} \exp\left(-\left(\kappa + \lambda + \mu\right)^2/r^2\right)$$

then, the monovariate integral to be evaluated is:

$$\tau(R|\kappa; \lambda; \mu) = \langle \varphi(x)\gamma(x)\eta(x)|R| \rangle$$

$$= \frac{2}{\pi} \left(\frac{\kappa \lambda \mu}{\pi}\right)^\frac{3}{2} \int_0^R \exp\left(-\left(\kappa + \lambda + \mu\right)^2\right) dx$$

which easily leads to:

$$\tau(R|\kappa; \lambda; \mu) = \frac{2}{\pi^2} \left(\frac{\kappa \lambda \mu}{\kappa + \lambda + \mu}\right)^\frac{3}{2} \int_0^{R_{K+\mu+\mu}} \exp\left(-y^2\right) dy$$

$$= \frac{1}{\pi} \left(\frac{\kappa \lambda \mu}{\kappa + \lambda + \mu}\right)^\frac{3}{2} \text{erf}\left(R_{K+\lambda+\mu}\right)$$

Such integral can be readily used to assess the influence of the
box constraint when atomic triple QSM measures are com-
puted. A program is available to compute within ASA atomic
triple density QSS measures. Several computations on assorted
atoms using the 3-21G and 6-311G ASA basis sets are presented
in the Table 4 below.

Within both ASA basis sets, triple QSM involving the den-
sities of three equal atoms converge toward the unconstrained
values quickly than simple overlap atomic QSS, as at an
inscribed sphere radius equal 1.5 the N atom has already con-
verged and heavier atoms converge even at the radius of 1 with
sufficient accuracy as the computed values become sufficiently
nearby to the unconstrained similarity integral obtained with

![Table 3. Variation of Several Molecular QSS Measures under ASA
Promolecular Density Functions with $R$, a Sphere Radius Inscribed into a
Constraining Cubic Box of Sides 2R. Mulliken Atomic Populations Have
Been Employed to Construct the Promolecular DF and ASA Huzinaga’s
Basis Set Is Employed for the Atomic Functions.](https://example.com/table3)

<table>
<thead>
<tr>
<th>Molecule</th>
<th>$R$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>H$_2$O</td>
<td>96.696466</td>
<td>97.079686</td>
<td>97.080480</td>
<td>97.080485</td>
<td>97.080485</td>
<td></td>
</tr>
<tr>
<td>Formaldehyde</td>
<td>118.13432</td>
<td>118.70059</td>
<td>118.70669</td>
<td>118.70672</td>
<td>118.70672</td>
<td></td>
</tr>
<tr>
<td>CH$_3$SH</td>
<td>310.03669</td>
<td>311.31008</td>
<td>311.32023</td>
<td>311.32027</td>
<td>311.32027</td>
<td></td>
</tr>
<tr>
<td>Estradiol</td>
<td>798.77413</td>
<td>802.49791</td>
<td>802.57838</td>
<td>802.57868</td>
<td>802.57868</td>
<td></td>
</tr>
<tr>
<td>XeF$_6$</td>
<td>40996.211</td>
<td>41001.645</td>
<td>41001.646</td>
<td>41001.646</td>
<td>41001.646</td>
<td></td>
</tr>
</tbody>
</table>

---

1. The ASA molecular QSS measures computation program can be down-
loaded from the web site [77].
2. Huzinaga ASA basis set can be also downloaded from the molecular
program web site [77] or from the general ASA website [74].

---

Table 4. Variation of 3-21G and 6-311G Atomic ASA Triple QSM with $R$, a Sphere Radius Inscribed into a Constraining Cubic Box of Sides 2R. QSS Values are Scaled by $Z^{-3}$.

<table>
<thead>
<tr>
<th>$R$</th>
<th>0.5</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>$\infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-21G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>8.600573</td>
<td>8.601237</td>
<td>8.601334</td>
<td>8.601338</td>
<td>8.601338</td>
</tr>
<tr>
<td>S</td>
<td>131.4828</td>
<td>131.4836</td>
<td>131.4836</td>
<td>131.4836</td>
<td>131.4836</td>
</tr>
<tr>
<td>Ca</td>
<td>272.2346</td>
<td>272.2349</td>
<td>272.2349</td>
<td>272.2350</td>
<td>272.2350</td>
</tr>
<tr>
<td>As</td>
<td>1362.201</td>
<td>1362.207</td>
<td>1362.207</td>
<td>1362.207</td>
<td>1362.207</td>
</tr>
<tr>
<td>6-311G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>8.949783</td>
<td>8.950491</td>
<td>8.950592</td>
<td>8.950595</td>
<td>8.950595</td>
</tr>
<tr>
<td>S</td>
<td>1365.313</td>
<td>1365.320</td>
<td>1365.320</td>
<td>1365.320</td>
<td>1365.320</td>
</tr>
<tr>
<td>Ca</td>
<td>284.1529</td>
<td>284.1533</td>
<td>284.1533</td>
<td>284.1533</td>
<td>284.1533</td>
</tr>
<tr>
<td>As</td>
<td>1419.420</td>
<td>1419.423</td>
<td>1419.423</td>
<td>1419.423</td>
<td>1419.423</td>
</tr>
</tbody>
</table>

infinite radius value. This result seems plausible and will be more accursed as the number of density functions in the integrand augments or the involved atoms become heavier.

**Constrained GTO of Arbitrary Order QSM**

Molecular overlap QSM in general will depend of an integral over one of the three variables but containing a quadruple GTO product involved in such calculations. See part (1) of these communications on QS for more details. Preliminarily, the integrals to be evaluated can be expressed by means of a four monovariable functions product of the type:

$$\langle \phi_A^R | \phi_C^P | \phi_B^Q | \phi_D^P \rangle =$$

$$2 \exp \left( -\frac{\rho_A^R}{\rho_B^P} |x_A - x_B|^2 \right) \exp \left( -\frac{\rho_C^P}{\rho_D^P} |x_C - x_D|^2 \right)$$

$$\times \exp \left( -\frac{\rho_A^R + \rho_C^P}{\rho_B^P + \rho_D^P} (x_A^2 + x_C^2) \right)$$

$$\times \sum_k \Theta_k \int_0^\infty x^2 \exp \left( -\omega |x - x_H|^2 \right) dx$$

where: $\{A, B, C, D\}$ are the function centers, $\{F, G\}$ are the two condensed centers for the pairs $\{A, B\} \wedge \{C, D\}$, obtained in a similar manner as in eq. (9), and $\{H\}$ is the condensed center of $\{F, G\}$ obtained in a similar manner. The coefficients and the sum are equivalent to the ones needed in the evaluation of four center repulsion integrals, and the exponent in the integral is the sum of the four involved exponents, as appearing in the denominator of the exponential function exponent in the preceding term. Thus, the needed integral is just:

$$G_{2k}(R; \omega) = \int_0^R x^{2k} \exp \left( -\omega |x - x_H|^2 \right) dx$$

$$= R^{2k+1} \int_0^1 y^{2k} \exp(-\omega R^2 y^2) dy = R^{2k+1} F_k(\omega R^2)$$

an expression related to the incomplete gamma function. It has been already commented that incomplete gamma functions within the development of the evaluation of integrals involving GTO are well-known functions and can be obtained in various ways. In principle, except for a slight computational complication, constrained QS overlap integrals over general GTO can be easily computed by means of well-defined and known algorithms already in use for many years in the applications of GTO basis sets for molecular calculations.

The structure of the general integrals involved into the QS evaluation indicates that as in the ASA case QSS measures will be independent of the origin of coordinates whenever the same constraining conditions are met.

**Analytic Description of the HEDT**

So far carried out the geometrical discussion of the computational implementation of the HEDT applied to the framework of QS measures and to present an as complete discussion as possible of the HEDT, this study will finish with some insight of the analytical description.

Other than to the topological discussion, the HEDT can be related to the well-known TSE of a function in the neighborhood of a given point $P$ in space, provided that the series converge in the whole space. Actually, the assumptions of Mezey’s when describing HEDT are even stronger than the ones which can be attached to such a TSE description. For instance, Mezey in the mentioned work advocates for the need to know all the density function properties inside a closed and finite domain $d$ containing a fixed point $P$. On the other side, the TSE around a point $P$ only requires the knowledge of the function derivatives at this point. In other words, the TSE only needs the knowledge of analytical properties attached to the neighborhood of point $P$, whereas Mezey’s HEDT requirements need the full knowledge of the density function inside a finite (not differential) domain containing $P$.

The TSE of any DF, being much less restrictive than Mezey’s theorem, will be less affected by the presence of the nonanalytical points excluded by HEDT. If these Coulomb singularities (nonanalyticities) are barred from the domain $d$, as indicated in Mezey’s work, particularly they can be also excluded from a small neighborhood of any point $P$ not presenting one of such singularities, allowing the series expansion to be constructed at this point. This can be seen from another point of view: the appropriate mathematical properties required by the HEDT inside the domain $d$ will ensure the convergence of the series for all positions in space, except at the singularity points of measure zero, which are nondifferentiable points.

The points of measure zero related to nucleus-electron Coulomb singularities can be handled when expressed in terms of multidimensional delta functions centered in turn at every nucleus $I$ of the molecule. For instance, when studying first order density functions one can write:

$$\rho_{\text{nuclear}}(r) = \sum_I Z_I \delta(r - r_I),$$

also for higher order density functions one can use an appropriate number of delta function products.
This density nuclear term is always forgotten when expressing the total first order electron density of a molecule, which for completion sake shall be expressed as:

$$\rho(r) = \rho_{\text{nuclear}}(r) - \rho_{\text{electronic}}(r).$$  (12)

To have a well behaved and smooth nuclear contribution to the DF, one can use a recent description and analysis made by one of us, so there could be no problem at the moment to develop a TSE of any DF.

The proposal in this context consists to accept the fact that a reproduction of the molecular DF: $\rho(r)$, starting from a basic or primitive knowledge, can be achieved without knowing all the function information within a given domain. The DF can be expanded by means of a TSE using the following formal scheme:

$$\rho(r) = \sum_{p=0}^{\infty} \frac{1}{p!} \left[ \left( \frac{\partial^{p} \rho(r)}{\partial r^{p}} \right) * \bigotimes_{i=1}^{p} (r_i - r_{i,p}) \right].$$  (13)

Where in eq. (13), the partial derivative part consists of a tensorial expression, containing all the possible combinations of derivatives of the density function with respect the electron coordinates up to order $p$. The notation $\rho(r_p)$ means that the derivatives are evaluated at point $P$. The operation between brackets affecting both tensors is an inward tensor product involving the derivative tensor and the tensorial product of the coordinate differences, with a final complete sum of the elements of the resultant tensor.

Expression (13) can be also written in terms of Nested Summation Symbols. Employing any of the alternative expressions, even using the one present in eq. (13), it is not so difficult to see how every term in the TSE resembles the formal structure, which can be generally attached to the density functions of $p$-th order, see for example refs. 71, 72, 83. Indeed, observing the form of each of the TSE expansion terms, it can be structured as:

$$\left\langle \frac{\partial^{p} \rho(r_p)}{\partial r^{p}} \right\rangle * \left[ \bigotimes_{i=1}^{p} (r_i - r_{i,p}) \right] \equiv \langle P * Q(r_i = 1, p) \rangle,$$

being $P$ and $Q$, $p$-th order tensors, the first one has constant components, the second one possess components depending of the position coordinates of $p$ electrons. This formally corresponds to the same expressions as they are obtained, when dealing with the density function sequence; see, for example, refs. 71, 72, and 83.

Provided that the previous expression expands the molecular DF, it must be noted that the TSE formulation illustrates a general algorithm to obtain the DF anywhere in space, when enough numerical precision is available. This kind of information is not given by the HEDT, which only states that the full knowledge of the function arises from any finite domain information but do not provides any tool to achieve it. Thus, HEDT may just be considered an existence theorem. Perhaps, the development of this last analytical section in this work could correspond to a general HEDT formulation because the order of the DF entering the TSE is arbitrary and no other condition than the smoothness of the DF at any point in electronic position space is needed. Thus, from this analytical point of view, HEDT can be reformulated as follows: knowing the derivatives of any DF at an arbitrary analytical point in the electronic coordinate space, the whole DF can be computed by means of a TSE. This constitutes a trivial statement, which the authors are reluctant to call it a theorem.

Concluding Remarks

Some annotations are due at this stage of the description and discussion of the HEDT characteristic features seen from this point of view:

1. applies to any DF order.
2. applies to any state DF.
3. will still apply whenever the bounded volume constructed on the DF can be made as the union of bounded disjoint volumes.

The residual unbounded DF can be defined accordingly.

In fact, when molecular bounded isodensity surfaces are computed and displayed, there is some threshold first order DF value: $\zeta_0$, which will display all the atoms in the molecule as disjoint bounded volumes. An alternative procedure of this situation might consist into constructing spheres around the molecular atoms until a radius is reached, containing in general an individual threshold percent of the atomic nuclear charge, in such an additional way that the associated spheres do not overlap. These spheres one by one or in diverse unions will also contain the presence of the total DF.

It seems at the light of this discussion that HEDT could correspond to a general feature of probability density functions, which can be usefully used in the quantum mechanical study of electronic systems and furnishing QS with a new set of computational techniques to evaluate constrained QSM. Also, HEDT seems to be related to the topological formulation of the analytical possibility to represent any DF as a TSE.

Finally, the results which can be obtained from the QSS computation of atoms and molecules preclude the easy computation of constrained ab initio DF, at least within the usual HF and DFT schemes: atoms and molecules in a box, with just a simple GTO integral reprogramming. It will be worth testing if the MOs nature, constructed as linear combinations of GTO, used with high frequency as basis sets in usual molecular computations, generally include the characteristic behavior found in this QSS study, when used in a constrained integration. Another side of the problem to be studied in our opinion is to inquire about the occurrence of comparable features as the ones found here, when constrained calculations are made within a STO basis set framework, for example in atomic calculations. Further research in these directions will be published elsewhere.
Acknowledgments

The referees’ comments are warmly acknowledged: they have definitely improved this article, helping to shed light on certain crucial aspects of the initial versions.

References

73. Available at: http://iqc.udg.edu/~quantum/software/Constrained_QSS/Atomic.
74. Available at: http://iqc.udg.es/cat/similarity/ASA/.
77. Available at: http://iqc.udg.edu/~quantum/software/Constrained_QSS/Molecular.
80. Spartan ’06; Wavefunction, Inc.; Irvine, CA.
84. Available at: http://iqc.udg.edu/~quantum/software/Constrained_QSS/Atomic_triple_density.