

# Mathematical aspects of the LCAO MO first order density function (1): atomic partition, metric structure and practical applications

Ramon Carbó-Dorca

*Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281, B-9000 Gent, Belgium*

*Institut de Química Computacional, Universitat de Girona, Girona, 17071 Catalonia, Spain*  
E-mail: quantumqsar@hotmail.com

Received 9 October 2006; Revised 20 January 2007

Partition of the first order density function within LCAO MO theory permits the definition of Mulliken and Minkowski metric spaces. The metric matrices obtained become useful to connect density atomic partition with quantum similarity measures, Mulliken populations, EHT, expectation values of Hermitian operators and the comparison of two density distributions.

**KEY WORDS:** First order density function, Mulliken populations, atomic basis set, Minkowski density function partition, quantum similarity measures, comparison of two density functions on the same molecule

## 1. Introduction

The analysis of the first order density function in LCAO MO theoretical framework is far from being completed in all its aspects and nuances. The present paper, the first of a self-contained series on density function mathematical aspects, pretends to study the atomic partition of the density function aimed to constructing a functional basis set able to generate a metric vector space. Two possible atomic partitions will be presented, leading to alternative Euclidean and Minkowski metrics. The obtained results show how density function partition analysis permits to connect metric matrices with atomic projection operators, quantum similarity measures, Mulliken populations, EHT structure, expectation values of Hermitian operators and other subjects. It must be noted finally that such partition is not unique at all, as other possible basis sets can be devised, implying groups of atoms or mixtures of atomic diatomic and polyatomic moieties. The generalization of the framework studied in this paper into these alternative possibilities is straightforward and will not be further commented.

## 1.1. LCAO MO first order density function and atomic Euclidian metric

### 1.1.1. First order density function

It is well known that first order molecular density can be expressed, without loss of generality, within LCAO MO theory in a very general manner using a coefficient diagonal form [1]:

$$\rho_M(\mathbf{r}) = \sum_i \omega_i |i\rangle \langle i|, \quad (1)$$

where  $\{\omega_i\}$  are occupation numbers and  $\{|i\rangle\}$  is the MO set expressed as a collection of kets.

The first order density function (1) has a Minkowski norm equal to the number of electrons,  $N_M$ , of the molecule  $M$  considered; that is, one can write:

$$\langle \rho_M \rangle = \int_D \rho_M(\mathbf{r}) \, d\mathbf{r} = N_M. \quad (2)$$

Now it can be considered that the MO set as used in equation (1) is usually expressed within the LCAO formalism, that is: knowing a one-electron basis set, which can be expressed in a column vector form:  $|\mathbf{x}\rangle = \{|\mu\rangle\}$ , then the MO set is expressed as a linear combination of the basis functions; for instance, it can be written:

$$\forall i : |i\rangle = \langle \mathbf{c}_i * |\mathbf{x}\rangle \rangle = \sum_{\mu} c_{\mu i} |\mu\rangle, \quad (3)$$

where:  $\mathbf{c}_i = \{c_{\mu i}\}$  are the coordinates of the  $i$ th MO with respect to the basis set  $|\mathbf{x}\rangle$ . The symbols of the *inward matrix product*<sup>1</sup> [2, 3, 11], also called Hadamard product in a restricted definition, and also the *complete sum of the elements of a matrix*<sup>2</sup> [4] are also introduced in equation (3), as providing a compact notation for the MO LCAO expression and will be also employed afterwards in other forthcoming formulae and related papers. Moreover, both operations are intrinsic procedures in modern Fortran 90 or 95 compilers, see for example [5]; thus, the formulae written in that way are easily translated to Fortran language compact sentences.

Substituting the LCAO MO expansion (3) into the density function (1), one arrives to the well-known expression of the first order density function in the

<sup>1</sup>The *inward matrix product* between two matrices:  $\mathbf{A} = \{a_{ij}\} \wedge \mathbf{B} = \{b_{ij}\}$  of dimension  $(m \times n)$ , is defined as a matrix of the same dimension by means of the algorithm:  $\mathbf{A} * \mathbf{B} = \mathbf{P} = \{p_{ij}\} \rightarrow \forall i, j : p_{ij} = a_{ij} b_{ij}$ .

<sup>2</sup>The *complete sum* of the elements of arbitrary dimension  $(m \times n)$ :  $\mathbf{A} = \{a_{ij}\}$ , is described according to the algorithm:  $\langle \mathbf{A} \rangle = \sum_{i=1}^m \sum_{j=1}^n a_{ij}$ .

LCAO framework:

$$\rho_M(\mathbf{r}) = \langle \mathbf{D}^* | \mathbf{x} \rangle \langle \mathbf{x} | \rangle = \sum_{\mu} \sum_{\nu} D_{\mu\nu} |\mu\rangle \langle \nu|, \quad (4)$$

with the so-called charge and bond order matrix  $\mathbf{D}$ , defined as follows:

$$\mathbf{D} = \{D_{\mu\nu}\} = \sum_i \omega_i \mathbf{c}_i \mathbf{c}_i^+. \quad (5)$$

Then, one can take into account that the AO basis set  $|\mathbf{x}\rangle$  is customarily attached to the atomic centres,  $I$  say, of the initially chosen molecule,  $M$ , so formally one can write:

$$\forall \mu \in I \wedge I \in M : |\mu\rangle \equiv \chi_{\mu}(\mathbf{r} | \mathbf{R}_I) \rightarrow |\mu\rangle \in I,$$

where  $\mathbf{R}_I$  are the position coordinates of the  $I$ th atomic centre. Thus, there can be always computed a positive definite metric or overlap matrix for the AO basis set, which can be formally written as:

$$\mathbf{S} = \int_D |\mathbf{x}\rangle \langle \mathbf{x}| d\mathbf{r} = \{S_{\mu\nu}\}. \quad (6)$$

### 1.1.2. Atomic density basis set

For a given molecular structure, a set of projectors can be constructed [6] in such a way that, applied to the molecular first order density function, yield a density function, associated to any AO belonging to a given atomic centre; for example, it can be written:

$$\Pi_{\alpha} = \sum_J \sum_{\beta \in J} S_{\alpha\beta}^{(-1)} |\alpha\rangle \langle \beta|, \quad (7)$$

that is:

$$\begin{aligned} \Pi_{\alpha} [\rho_M] &= \sum_J \sum_{\beta \in J} S_{\alpha\beta}^{(-1)} |\alpha\rangle \langle \beta| \left[ \sum_{\mu} \sum_{\nu} D_{\mu\nu} |\mu\rangle \langle \nu| \right] \\ &= \sum_J \sum_{\beta \in J} \sum_{\mu} \sum_{\nu} D_{\mu\nu} S_{\alpha\beta}^{(-1)} S_{\mu\beta} |\alpha\rangle \langle \nu| \\ &= \sum_{\mu} \sum_{\nu} D_{\mu\nu} \delta_{\alpha\mu} |\alpha\rangle \langle \nu| = \sum_{\nu} D_{\alpha\nu} |\alpha\rangle \langle \nu| = \rho_{\alpha} \end{aligned}$$

therefore, as the following sum property must necessarily hold:

$$\sum_{I \in M} \left( \sum_{\alpha \in I} \Pi_{\alpha} \right) = \sum_{I \in M} \Pi_I = \hat{I},$$

being  $\hat{I}$  the unit operator, then it results that the density function can be decomposed as follows:

$$\rho_M(\mathbf{r}) = \sum_{I \in M} \left( \sum_{\alpha \in I} \rho_\alpha(\mathbf{r}) \right). \quad (8)$$

### 1.2. Atomic density basis set, Mulliken atomic populations and density partition

The set of AO density functions obtained in this way:  $\{\rho_\alpha(\mathbf{r})\}$  necessarily has to structure a basis set, as it can always be condensed to atomic centres; that is, calling:

$$\forall I \in M : \rho_I(\mathbf{r}) = \sum_{\alpha \in I} \rho_\alpha(\mathbf{r}). \quad (9)$$

The basis set:  $\{\rho_I(\mathbf{r})\}$ , will be logically called from now on: *atomic density basis set*. The set of functions (9) forming the atomic density basis set is necessarily dependent on the computational level and on the completeness of the AO basis set employed in order to obtain the molecular wave function. They are positive definite functions and the dimension of the subtended semispace<sup>3</sup> will be always the same:  $n_M$ , equal to the number of atomic centres present in molecule  $M$ .

Thus, it is obvious that at any computational level, the set (9) of atomic density functions  $\{\rho_I\}$  can be ordered into a  $n_M$  - dimensional column vector as:

$$|\rho\rangle = |\rho(\mathbf{r})\rangle = \{\rho_I(\mathbf{r})\}. \quad (10)$$

With such a basis set one has to be able to reconstruct the molecular first order density function, by using coordinates equal to the *unity vector*:  $|\mathbf{1}\rangle = \{1_k = 1\}$ ; that is: a vector whose elements are the real unit element. So, in this context one can write the molecular density function (8) as:

$$\rho_M(\mathbf{r}) = \langle \rho | \mathbf{1} \rangle. \quad (11)$$

Inward integration of the atomic density basis set vector provides the gross atomic populations in the Mulliken sense [8] ordered as a column vector:  $|\mathbf{q}\rangle = \{q_I\}$ :

$$\int_D |\rho(\mathbf{r})\rangle d\mathbf{r} = |\mathbf{q}\rangle \rightarrow |\mathbf{q}\rangle = \left\{ q_I = \int_D \rho_I(\mathbf{r}) d\mathbf{r} = \sum_{\alpha \in I} \sum_{\mu} D_{\alpha\mu} S_{\alpha\mu} \right\}, \quad (12)$$

<sup>3</sup>A *vector semispace* is a vector space defined over the positive real numbers, where the additive group has been substituted by a semigroup. See for example: [7].

in this way, the complete sum of the vector  $|\mathbf{q}\rangle$  will provide the total number of electrons in the attached molecular structure  $M$ :

$$\langle |\mathbf{q}\rangle \rangle = \sum_I q_I = N_M. \quad (13)$$

Therefore, it can be easily said that the atomic density basis set described so far is coincident with the earlier well-known Mulliken ideas about the partition of the density function and the molecular density condensation to form atomic charges and bond orders [8].

### 1.3. Atomic basis set Euclidean metric

The atomic density basis set, as such, can be directly attached to a positive definite metric matrix of dimension  $(n_M \times n_M)$  in the Euclidian scalar product sense. It is easy to find out that the *atomic similarity matrix*:

$$\mathbf{Z} = \{z_{IJ} | \forall I, J \in M\}, \quad (14)$$

where the overlap similarity measures [10] have been used:

$$z_{IJ} = \int_D \rho_I(\mathbf{r}) \rho_J(\mathbf{r}) d\mathbf{r} = \langle \rho_I | \rho_J \rangle \quad (15)$$

has to be positive definite, due to the linear independence of the atomic density basis set, supposedly employed in this way by construction. There is also interesting to note that the integrals intervening in equation (15) above, possess a lower bound defined by the inequality:

$$\forall I, J : \int_D \rho_I(\mathbf{r}) \rho_J(\mathbf{r}) d\mathbf{r} \geq (\rho_I(\langle \mathbf{r} \rangle_J) \rho_J(\langle \mathbf{r} \rangle_I))^{+\frac{1}{2}}, \quad (16)$$

where:

$$P = I, J : \langle \mathbf{r} \rangle_P = \int_D \rho_P(\mathbf{r}) \mathbf{r} d\mathbf{r}$$

are the position vector expectation values with respect to the involved density functions. The proof of the relationship (16) follows straightforwardly from Jensen's inequality [9]:

$$\langle \varphi(\mathbf{r}) \rangle \geq \varphi(\langle \mathbf{r} \rangle),$$

after taking into account that the integrals (15) define two equivalent mean values.

As defined in equation (14), the atomic similarity matrix  $\mathbf{Z}$  constitute an adequate Euclidean metric, associated to the atomic density basis set defined

previously and coincident with the atomic Mulliken partition of the density function.

The Euclidean metric can be also computed weighted by a positive definite operator  $W$ , providing in this case the positive definite metric matrix elements:

$$\begin{aligned} \mathbf{Z}[W] &= \left\{ z_{IJ} [W] \right. \\ &= \left. \int_D \int_D \rho_I(\mathbf{r}_1) W(\mathbf{r}_1, \mathbf{r}_2) \rho_J(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2 = \langle \rho_I | W | \rho_J \rangle \right\}, \quad (17) \end{aligned}$$

constituting a general expression, which appears reminiscent of the usual definition of *quantum similarity measures* involving two densities attached to a pair of *quantum objects*.

The complete sum of any weighted metric matrix  $\mathbf{Z}[W]$  produces the corresponding weighted selfsimilarity measure, as:

$$\langle \mathbf{Z}[W] \rangle = \sum_I \sum_J \langle \rho_I | W | \rho_J \rangle = \langle \rho | W | \rho \rangle$$

and the so-called overlap selfsimilarity of the associated molecular structure is recovered when the weight operator becomes a Dirac delta function [10].

#### 1.4. Stochastic matrices

As it has been done in previous studies [11], this kind of similarity matrices, either the simple overlap (15) or the general form (17), representing any Euclidian metric, can be transformed into row or column Stochastic matrices. In fact this is so for any vector space with such a metric. It is sufficient to represent the matrix into a row or column partition and proceed obtaining the sums of row or column elements. Let us suppose the following column partition of the overlap Euclidian metric  $\mathbf{Z}$  as defined in equations (14) and (15):

$$\mathbf{Z} = (|\mathbf{z}_1\rangle; |\mathbf{z}_2\rangle \dots |\mathbf{z}_{n_M}\rangle),$$

where:

$$\forall I = 1, n_M : |\mathbf{z}_I\rangle = \{z_{JI} | J = 1, n_M\}.$$

Representing the complete sums of each column as:

$$\theta_I = \langle |\mathbf{z}_I\rangle \rangle = \sum_J z_{JI} > 0,$$

one can construct a positive definite diagonal matrix:

$$\Theta = \text{Diag}(\theta_1; \theta_2; \dots \theta_{n_M}),$$

whose inverse transforms the original Euclidian metric into a column stochastic matrix:

$$\mathbf{S}_C = \mathbf{Z}\Theta^{-1}, \quad (18)$$

whose columns have complete sum equal to the unit. Every column of the stochastic matrix (18) becomes a  $n_M$ -dimensional discrete probability distribution. The associate row stochastic matrix is just the transpose of the matrix  $\mathbf{S}_C$ , or:

$$\mathbf{S}_R = \mathbf{S}_C^T = \Theta^{-1}\mathbf{Z}$$

as the metric matrix is symmetric.

The columns or rows of the stochastic matrices described above can be interpreted as discrete statistical representations of the attached atomic centres present in the molecule  $M$ . For a given molecular configuration and a chosen atomic basis set, these atomic representations are unique and linearly independent.

### 1.5. Alternative Minkowski metric associated to the density function

So far, the discussion on the possible description of the atomic density basis set has been associated to the Euclidian metric structure, yielding the matrix  $\mathbf{Z}$  as defined in equations (14) and (15). However, the density expression (4) can be also associated to the complete sum of a paired atomic contributions matrix which can be defined in the following manner:

$$\mathbf{P}(\mathbf{r}) = \left\{ p_{IJ}(\mathbf{r}) = \sum_{\mu \in I} \sum_{v \in J} D_{\mu v} |\mu\rangle \langle v| \right\} \rightarrow \rho(\mathbf{r}) = \langle \mathbf{P}(\mathbf{r}) \rangle. \quad (19)$$

The matrix defined in equation (19) can be named *atomic density matrix*.

The sum of the rows or columns of the atomic density matrix yields the atomic density basis set vector (10) as described above.

The matrix elements in equation (19) can be obtained using the projector (7) summed up to a given atomic centre:

$$\Pi_I = \sum_{\alpha \in I} \Pi_{\alpha},$$

and multiplying the density expression (4) on the left- and right-hand sides, respectively, as:

$$\begin{aligned}
 \Pi_I \rho(\mathbf{r}) \Pi_J &= \sum_{\alpha \in I} \sum_{\beta \in J} \sum_K \sum_L \sum_{\kappa \in K} \sum_{\lambda \in L} \sum_{\mu} \sum_{\nu} S_{\alpha\kappa}^{(-1)} S_{\beta\lambda}^{(-1)} |\alpha\rangle S_{\kappa\mu} S_{\nu\lambda} \langle\beta| D_{\mu\nu} \\
 &= \sum_{\alpha \in I} \sum_{\beta \in J} \sum_{\mu} \sum_{\nu} \delta_{\alpha\mu} \delta_{\beta\nu} |\alpha\rangle \langle\beta| D_{\mu\nu} \\
 &= \sum_{\alpha \in I} \sum_{\beta \in J} D_{\alpha\beta} |\alpha\rangle \langle\beta| \\
 &= p_{IJ}(\mathbf{r})
 \end{aligned}$$

Thus, the matrix (19) is not at all arbitrarily build, but their elements can be considered projected parts, made by pairs of atomic contributions, of the whole density function.

### 1.6. Integration of the atomic density matrix: Minkowski metric matrix

Inward integration<sup>4</sup> of the atomic density matrix provides an alternative metric matrix with the form:

$$\mathbf{Q} = \int_D \mathbf{P}(\mathbf{r}) d\mathbf{r} = \left\{ q_{IJ} = \int_D p_{IJ}(\mathbf{r}) d\mathbf{r} \right\}, \tag{20}$$

which has the property that its complete sum provides the number of electrons:

$$\langle \mathbf{Q} \rangle = N_M, \tag{21}$$

and the sum of every row or column yield the corresponding Mulliken gross atomic populations. Equation (21) proves that the complete sum of the Minkowski metric matrix behaves as a Cioslowski [12] selfsimilarity measure, and in this way the previous Euclidean metric and the present Minkowski one can be considered different aspects of the same ideas.

### 1.7. Cioslowski similarity measures on the atomic density basis set as a Minkowski metric

The atomic density basis set (9), can be also described using it as elements of the off-diagonal first order density matrix [12]. Rewriting the density function

<sup>4</sup>The *inward integration* over an appropriate integration domain  $D$  of some matrix  $\mathbf{A}(\mathbf{r})$ , whose elements  $\{a_{IJ}(\mathbf{r})\}$  are functions of some variable vector  $\mathbf{r}$ , is defined as a matrix of the same dimension by means of the algorithm:  $\mathbf{Z} = \int_D \mathbf{A}(\mathbf{r}) d\mathbf{r} \rightarrow \forall I, J: z_{IJ} = \int_D a_{IJ}(\mathbf{r}) d\mathbf{r}$

in the following way:

$$\rho_M(\mathbf{r}, \mathbf{r}') = \sum_{I \in M} \left( \sum_{\alpha \in I} \rho_\alpha(\mathbf{r}, \mathbf{r}') \right) = \sum_{I \in M} \left( \sum_{\alpha \in I} \sum_{\mu} D_{\alpha\mu} \chi_\alpha(\mathbf{r}) \chi_\mu^*(\mathbf{r}') \right),$$

then the atomic density basis set may be expressed as:

$$\forall I \in M : \rho_I(\mathbf{r}, \mathbf{r}') = \sum_{\alpha \in I} \rho_\alpha(\mathbf{r}, \mathbf{r}')$$

and the basis set rewritten in this way can be employed to construct an Euclidean metric matrix, in the way an alternative overlap similarity measure can be constructed, as proposed by Cioslowski [12], in the form like:

$$\begin{aligned} \forall I, J \in M : c_{IJ} &= \int_D \int_D \rho_I(\mathbf{r}, \mathbf{r}') \rho_J(\mathbf{r}', \mathbf{r}) \, d\mathbf{r} \, d\mathbf{r}' \\ &= \sum_{\alpha \in I} \sum_{\mu} \sum_{\beta \in I} \sum_{\nu} D_{\alpha\mu} D_{\nu\beta} \int_D \chi_\beta^*(\mathbf{r}) \chi_\alpha(\mathbf{r}) \, d\mathbf{r} \int_D \chi_\mu^*(\mathbf{r}') \chi_\nu(\mathbf{r}') \, d\mathbf{r}' \\ &= \sum_{\alpha \in I} \sum_{\mu} \sum_{\beta \in I} \sum_{\nu} D_{\alpha\mu} D_{\nu\beta} S_{\beta\alpha} S_{\mu\nu} \\ &= \sum_{\alpha \in I} \sum_{\beta \in J} \sum_{\mu} \sum_{\nu} (D_{\alpha\mu} S_{\mu\nu} D_{\nu\beta}) S_{\beta\alpha} \end{aligned}$$

the last term simplifies due to the definition of the charge and bond order matrix (5), which permits to write in general:

$$\begin{aligned} \mathbf{DSD} &= \sum_i \sum_j \omega_i \omega_j \mathbf{c}_i \mathbf{c}_i^+ \mathbf{S} \mathbf{c}_j \mathbf{c}_j^+ \\ &= \sum_i \sum_j \omega_i \omega_j (\mathbf{c}_i^+ \mathbf{S} \mathbf{c}_j) \mathbf{c}_i \mathbf{c}_j^+ \\ &= \sum_i \sum_j \omega_i \omega_j \delta_{ij} \mathbf{c}_i \mathbf{c}_j^+ \\ &= \sum_i \omega_i^2 \mathbf{c}_i \mathbf{c}_i^+ \end{aligned}$$

and therefore, for closed shell configurations:

$$[\mathbf{DSD}]_{\alpha\beta} = 2D_{\alpha\beta} = 2 \sum_{\mu} \sum_{\nu} D_{\alpha\mu} S_{\mu\nu} D_{\nu\beta}$$

thus:

$$\forall I, J \in M : c_{IJ} = 2 \sum_{\alpha \in I} \sum_{\beta \in J} D_{\alpha\beta} S_{\beta\alpha} = 2 \sum_{\alpha \in I} \sum_{\beta \in J} D_{\alpha\beta} S_{\alpha\beta} = 2q_{IJ},$$

an expression which is, except for a multiplicative factor, the same as the Minkowski metric elements defined in equation (20). Therefore, the Minkowski partition of the density function and the definition of the Minkowski density basis set matrix permit to construct an associated Minkowski metric, which becomes equivalent in turn to an Euclidian metric in terms of the atomic density basis set, but computed taking into account the off-diagonal elements of the first order density matrix within a Mulliken partition scheme.

### 1.8. Minkowski metric and Mulliken populations

The elements of the Minkowski metric matrix can be particularly expressed as:

$$\forall I, J : q_{IJ} = \langle p_{IJ}(\mathbf{r}) \rangle = \sum_{\alpha \in I} \sum_{\beta \in J} D_{\alpha\beta} \int_D |\alpha\rangle \langle \beta| d\mathbf{r} = \sum_{\alpha \in I} \sum_{\beta \in J} D_{\alpha\beta} S_{\alpha\beta}, \quad (22)$$

and, obviously enough, the atomic Mulliken charges as defined in equation (12), can be also described with the Minkowski metric elements, as:

$$\forall I : q_I = \sum_J q_{IJ}.$$

Expression (22) indicates that an *uncontracted* Minkowski metric matrix,  $\mathbf{M}$ , can be defined as the inward product of the charge and bond order matrix as defined in equation (5) by the overlap matrix, as described in equation (6):

$$\mathbf{M} = \mathbf{D} * \mathbf{S} \rightarrow m_{\alpha\beta} = D_{\alpha\beta} S_{\alpha\beta} \rightarrow q_{IJ} = \sum_{\alpha \in I} \sum_{\beta \in J} m_{\alpha\beta}. \quad (23)$$

Comparison of Mulliken and Minkowski metric matrices in diatomic molecules

In fact, the structure, but not the numerical values of the contracted Minkowski metric matrix  $\mathbf{Q}$  in equation (20) appear to be the same as the Euclidian metric  $\mathbf{Z}$  for some naïve molecular constructs: as it can be easily seen when considering diatomic and simplex molecular structures. This property can be extended in other cases as the atomic clusters of type  $A_n$ . For example, homonuclear diatomic molecular structures possess the canonical  $(2 \times 2)$  matrix form in both metric frames:

$$\mathbf{Z} \vee \mathbf{Q} = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix} \quad (24)$$

with:  $\alpha \geq \beta$ .

Both metric matrices provide a set of parameters able to be used as molecular descriptors, bearing close information to the molecular structure; that is: both matrix structures are constructed in such a way that their elements vary with molecular coordinate conformation, and atomic charge distribution.

Comparison of the metric matrices in general will be discussed below, but it is interesting here to see how two matrices associated to homonuclear diatomic structures can be compared. Suppose two of such matrices are known:

$$\mathbf{Z} = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix} \wedge \mathbf{Q} = \begin{pmatrix} a & b \\ b & a \end{pmatrix}. \quad (25)$$

A cosine-like index can be easily defined as:

$$\begin{aligned} \cos(\mathbf{Z}; \mathbf{Q}) &= (\langle \mathbf{Z} * \mathbf{Z} \rangle \langle \mathbf{Q} * \mathbf{Q} \rangle)^{-\frac{1}{2}} \langle \mathbf{Z} * \mathbf{Q} \rangle \\ &= \left( [\alpha^2 + \beta^2] [a^2 + b^2] \right)^{-\frac{1}{2}} (\alpha a + \beta b) \end{aligned}$$

and a squared Euclidian distance:

$$D^2(\mathbf{Z}; \mathbf{Q}) = \langle (\mathbf{Z} - \mathbf{Q}) * (\mathbf{Z} - \mathbf{Q}) \rangle = 2 \left( [\alpha - a]^2 + [\beta - b]^2 \right).$$

### 1.9. EHT and Minkowski metric

The formalism associated to equation (23) can be also applied within the atomic condensed structure of extended Hückel theory (EHT), proposed many years ago by Hoffman [13]; as the elements of the Minkowski metric can be easily expressed as in equation (22), while in EHT framework the Hamiltonian matrix  $\mathbf{H}$  can be also expressed as an inward product of the overlap matrix  $\mathbf{S}$  by a parameter matrix  $\mathbf{G}$  as:

$$\mathbf{H} = \mathbf{G} * \mathbf{S},$$

with the elements of the parameter matrix computed by means of the atomic orbital ionization potentials  $\{I_\alpha\}$ :

$$g_{\alpha\beta} = \lambda_{\alpha\beta} (I_\alpha + I_\beta)$$

and the parameters  $\{\lambda_{\alpha\beta}\}$  depending of an adjustable constant  $\kappa$ :

$$\lambda_{\alpha\beta} = \frac{1}{2} (\delta_{\alpha\beta} + (1 - \delta_{\alpha\beta}) \kappa);$$

so, the atomic condensed EHT Hamiltonian elements can be obtained with the algorithm:

$$\forall I, J : h_{IJ} = \sum_{\alpha \in I} \sum_{\beta \in J} G_{\alpha\beta} S_{\alpha\beta},$$

which resembles the atomic density Minkowski metric elements expression (23) with the charge and bond order matrix elements substituted by the elements of the matrix  $\mathbf{G}$ .

Atomic condensed EHT for diatomic molecules will have the form of an extended secular equation like:

$$\begin{pmatrix} h_{AA} & h_{AB} \\ h_{AB} & h_{BB} \end{pmatrix} \begin{pmatrix} c_A \\ c_B \end{pmatrix} = \varepsilon \begin{pmatrix} s_{AA} & s_{AB} \\ s_{AB} & s_{BB} \end{pmatrix} \begin{pmatrix} c_A \\ c_B \end{pmatrix}.$$

And thus the eigenvalues can be obtained as the roots of the characteristic polynomial:

$$\begin{aligned} \text{Det } |\mathbf{H} - \varepsilon \mathbf{S}| &= \text{Det} \begin{vmatrix} h_{AA} - \varepsilon s_{AA} & h_{AB} - \varepsilon s_{AB} \\ h_{AB} - \varepsilon s_{AB} & h_{BB} - \varepsilon s_{BB} \end{vmatrix} \\ &= (h_{AA} - \varepsilon s_{AA})(h_{BB} - \varepsilon s_{BB}) - (h_{AB} - \varepsilon s_{AB})^2 \\ &= (s_{AA}s_{BB} - s_{AB}^2) \varepsilon^2 - (h_{AA}s_{BB} + h_{BB}s_{AA} + 2h_{AB}s_{AB}) \varepsilon \\ &\quad + (h_{AA}h_{BB} - h_{AB}^2) \\ &= \text{Det } |\mathbf{S}| \varepsilon^2 - \langle \mathbf{H} * \mathbf{S} \rangle \varepsilon + \text{Det } |\mathbf{H}| \end{aligned}$$

so, the eigenvalues can be easily calculated as:

$$\varepsilon = (2\text{Det } |\mathbf{S}|)^{-1} \left( \langle \mathbf{H} * \mathbf{S} \rangle \pm \left( \langle \mathbf{H} * \mathbf{S} \rangle^2 - 4\text{Det } |\mathbf{HS}| \right)^{\frac{1}{2}} \right).$$

### 1.10. Atomic partition of the expectation values of an operator

The condensation of the atomic density matrix using a Hermitian operator instead of the unit one, providing the construction of the Minkowski metric matrix (20), yields an atomic partition of the expectation values. For instance, suppose an one-electron Hermitian operator,  $W(\mathbf{r})$ ; then, the following matrix representation can be easily written:

$$\mathbf{Q}[W] = \int_D W(\mathbf{r})[\mathbf{P}(\mathbf{r})] \mathbf{d}\mathbf{r} = \left\{ q_{IJ}[W] = \int_D W(\mathbf{r})[p_{IJ}(\mathbf{r})] \mathbf{d}\mathbf{r} \right\}.$$

Or it can be also written an expression generalizing equation (23):

$$q_{IJ}[W] = \sum_{\alpha \in I} \sum_{\beta \in J} D_{\alpha\beta} w_{\alpha\beta} \rightarrow \mathbf{Q}[W] = \mathbf{D} * \mathbf{W},$$

with the matrix representation  $\mathbf{W}$  of the one-electron operator  $W$  defined as usual over the chosen basis set in LCAO framework:

$$\mathbf{W} = \left\{ w_{\alpha\beta} = \int_D \chi_{\alpha}^*(\mathbf{r}) W(\mathbf{r}) \chi_{\beta}(\mathbf{r}) \mathbf{d}\mathbf{r} \right\}.$$

The resultant weighted atomic condensed density matrix  $\mathbf{Q}[W]$  is associated with the additional elegant property, consisting into that the complete sum of the weighted atomic matrix representation, yields the expectation value of the associated operator:

$$\langle W \rangle = \langle \mathbf{Q}[W] \rangle.$$

### 1.11. Molecular atomic probability matrix distributions

In any case, the weighted atomic condensed density matrix representation:  $\mathbf{Q}[W]$ , can always be scaled with the corresponding operator expectation value in order to obtain a matrix whose complete sum yields the unity:

$$\mathbf{Q}_N[W] = \langle W \rangle^{-1} \mathbf{Q}[W] \rightarrow \langle \mathbf{Q}_N[W] \rangle = 1.$$

And in the case where the following property of the elements of the matrix is present:  $\forall I, J: q_{IJ}[W] \in \mathbf{R}^+$ , then the matrix  $\mathbf{Q}_N[W]$  can act as a discrete probability distribution, representing the molecular structure in a unique way associated to the chosen operator.

### 1.12. Comparison of atomic condensed matrices

In any case, for atomic condensed matrices, the corresponding structures will become easily comparable within isoatomic molecules or between different conformations or electronic states of the same molecule. Indeed, for molecular structures with the same number of atoms, let the matrices associated to a pair of molecular structures be named  $\mathbf{Q}_A$  and  $\mathbf{Q}_B$ . Let their inward product be expressed as:

$$\mathbf{R}_{AB} = \mathbf{Q}_A * \mathbf{Q}_B,$$

then the Euclidian product of both matrices can be written as the complete sum:

$$\langle \mathbf{Q}_A | \mathbf{Q}_B \rangle = \langle \mathbf{R}_{AB} \rangle. \quad (26)$$

Therefore, from this point of view, similarity or cosine-like and dissimilarity or distance-like indices can be computed, respectively, as:

$$c_{AB} = (\langle \mathbf{R}_{AA} \rangle \langle \mathbf{R}_{BB} \rangle)^{-\frac{1}{2}} \langle \mathbf{R}_{AB} \rangle \quad (27)$$

and

$$d_{AB} = (\langle \mathbf{R}_{AA} \rangle + \langle \mathbf{R}_{BB} \rangle - 2 \langle \mathbf{R}_{AB} \rangle)^{\frac{1}{2}}. \quad (28)$$

In both cases given a set of molecular structures the indices (27) and (28) can be used to order the set and construct in this way molecular periodic tables for diatomic, triatomic, ... molecules.

Minkowski products and indices can be also easily constructed. For this purpose it is necessary to describe the inward square root of the condensed matrices as:

$$\mathbf{X}_A = \mathbf{Q}_A^{\left[\frac{1}{2}\right]} \rightarrow x_{A;IJ} = q_{A;IJ}^{\frac{1}{2}}.$$

Then the Minkowski product may be defined using the complete sum of the inward product:

$$\mathbf{Y}_{AB} = \mathbf{X}_A^* * \mathbf{X}_B$$

so as:

$$\langle \mathbf{Q}_A | \mathbf{Q}_B \rangle = \langle \mathbf{Y}_{AB} \rangle.$$

This definition of the Minkowski scalar product yields the sum of the elements of the original matrix inward module<sup>5</sup>, as:

$$\langle \mathbf{Q}_A | \mathbf{Q}_A \rangle = \langle \mathbf{Y}_{AA} \rangle = \langle \mathbf{X}_A^* * \mathbf{X}_A \rangle = \langle * | \mathbf{Q}_A | \rangle.$$

It must be noted that, due to the appearance of possible negative elements in the atomic condensed Minkowski matrices, the inward square roots may be defined into the complex field, so the similarity-dissimilarity indices are constructed, respectively, in this case as a cosine index:

$$c_{AB}^M = (\langle \mathbf{Y}_{AA} \rangle \langle \mathbf{Y}_{BB} \rangle)^{-\frac{1}{2}} |\langle \mathbf{Y}_{AB} \rangle|$$

and a corresponding Minkowski distance:

$$d_{AB}^M = \langle * | \mathbf{X}_A - \mathbf{X}_B | \rangle.$$

### 1.13. Comparison of atomic density metric matrices in practice

However, the comparison of two matrices corresponding to two isoatomic molecules will not be completely defined until some consideration is made. Taking into account again the Euclidean product (26), for instance, it is easy to see that the result, in general, will depend on the arbitrary numbering of the

<sup>5</sup>The *inward matrix module* of a  $(n \times m)$  matrix  $\mathbf{A}$  is associated to the symbol:  $*|\mathbf{A}|$  and by it, it is meant another matrix  $\mathbf{Z}$  of the same dimension, computed by the following algorithm:  $\mathbf{Z} = *|\mathbf{A}| \rightarrow \forall I, J: z_{IJ} = |a_{IJ}|$ .

atoms of both molecules. This can be illustrated by considering the metric matrices of two diatomic heteronuclear molecules:

$$\mathbf{Q}_A^{(2)} = \begin{pmatrix} a_1 & a \\ a & a_2 \end{pmatrix} \wedge \mathbf{Q}_B^{(2)} = \begin{pmatrix} b_1 & b \\ b & b_2 \end{pmatrix}, \quad (29)$$

where the symmetry of both matrix representations is taken into account. If, for instance, one takes the ordering (29), it can be written:

$${}_1\mathbf{R}_{AB}^{(2)} = \begin{pmatrix} a_1b_1 & ab \\ ab & a_2b_2 \end{pmatrix} \rightarrow \langle {}_1\mathbf{R}_{AB}^{(2)} \rangle = a_1b_1 + a_2b_2 + 2ab,$$

but permuting the atomic numbering of one of the molecules, the molecule  $B$ , for instance, one has:

$${}_2\mathbf{R}_{AB}^{(2)} = \begin{pmatrix} a_1b_2 & ab \\ ab & a_2b_1 \end{pmatrix} \rightarrow \langle {}_2\mathbf{R}_{AB}^{(2)} \rangle = a_1b_2 + a_2b_1 + 2ab.$$

A criterion of maximal similarity can be now imposed stating in this case:

$$(\mathbf{Q}_A | \mathbf{Q}_B) = \max \left\{ \langle {}_1\mathbf{R}_{AB}^{(2)} \rangle; \langle {}_2\mathbf{R}_{AB}^{(2)} \rangle \right\}.$$

Expressing by the operator symbol:  $P \left[ \mathbf{Q}_B^{(n)} \right]$  any of the permutations of the numbering of the  $n$  atoms in some molecule  $B$ , then, the scalar product shall be possibly defined as:

$$(\mathbf{Q}_A | \mathbf{Q}_B) = \max \left\{ \langle \mathbf{Q}_A * P \left[ \mathbf{Q}_B^{(n)} \right] \rangle \right\}.$$

Distastefully, for usual up to date computers, this necessary computation permits to estimate that to easily compare molecules with more than 12 atoms appears out of reach.

However, a possible solution to this permutational problem could be to prepare the metric matrices to be compared, with their diagonal elements ordered from larger to lesser, with the following rows and columns situated according to the same order. This will be the equivalent as to order the atoms in a given molecule in the way the diagonal elements of their atomic density metric matrix is ordered. That is, for example, in  $n$ -atomic molecules, then atom 1 becomes the one with the largest diagonal element and the  $n$ th the smallest. As the out-diagonal elements are, in principle, expected to be smaller than the diagonal ones, this will insure the proper products of the elements with similar ordering in both compared structures.

The problem of comparing molecular metric matrices of different number of atoms, can be circumvent in some cases, taking into account the appropriate number of atoms forming a fragment or a cluster in both structures. For example, benzene can be compared, even without the appearance of the permutational problem previously discussed, with polycyclic hydrocarbons, just taking the

six carbon atom submatrix and comparing it with a similar submatrix formed with six carbon atoms in the hydrocarbon.

#### 1.14. Reciprocal metric

Among other interesting possibilities, the atomic density basis set, which has been used to define the Euclidian metric matrix:  $\mathbf{Z}$ , can be also employed to define a reciprocal space. The reciprocal space of any known vector space possessing a known metric  $\mathbf{Z}$  is a new vector space with metric:  $\mathbf{Z}^{-1} = \{z_{IJ}^{(-1)}\}$ . The inverse of a given Euclidian metric matrix shall always exist because of the positive definite nature of such metric matrices, resulting from their construction in pre-Hilbert spaces. The metric matrix defined by means of equations (14) and (15), can be also formally expressed as:

$$\mathbf{Z} = \int_D |\rho\rangle \langle \rho| \mathbf{d}\mathbf{r},$$

so, left and right multiplying by the inverse metric the previous formal definition, one obtains:

$$\mathbf{Z}^{-1} = \mathbf{Z}^{-1} \mathbf{Z} \mathbf{Z}^{-1} = \int_D \mathbf{Z}^{-1} |\rho\rangle \langle \rho| \mathbf{Z}^{-1} \mathbf{d}\mathbf{r} = \int_D |\rho^{(-1)}\rangle \langle \rho^{(-1)}| \mathbf{d}\mathbf{r},$$

where the vector:

$$|\rho^{(-1)}\rangle = \left\{ \rho_I^{(-1)} = \sum_J z_{IJ}^{(-1)} \rho_J \right\}$$

contains the basis set elements of the reciprocal space. The reciprocal space metric matrix can be used in the same way as the metric in original spaces. Thus, in diatomic homonuclear molecules, where the original metric can be associated to a form like the one in equation (25), possess a reciprocal metric with the same pattern, because it can be written:

$$\mathbf{Z}^{-1} = D^{-1} \begin{bmatrix} \alpha & -\beta \\ -\beta & \alpha \end{bmatrix} \wedge D = \alpha^2 - \beta^2.$$

Reciprocal atomic charges can be defined in the same way as:

$$q_I^{(-1)} = \langle \rho_I^{(-1)} \rangle = \sum_J z_{IJ}^{(-1)} \langle \rho_J \rangle = \sum_J z_{IJ}^{(-1)} q_J,$$

Which implies that using an atomic charge vector  $|\mathbf{q}\rangle$ , like the one constructed in equation (12), the corresponding reciprocal charge vector can be easily defined as:

$$|\mathbf{q}^{(-1)}\rangle = \mathbf{Z}^{-1} |\mathbf{q}\rangle.$$

The sum of the original charges provides the total number of electrons as shown in equation (13), however the sum of reciprocal charge vector:

$$\left\langle \left| \mathbf{q}^{(-1)} \right\rangle \right\rangle = N_M^{(-1)}$$

can be used to construct a reciprocal atomic probability distribution vector:

$$\left| p^{(-1)} \right\rangle = \left( \frac{1}{N_M^{(-1)}} \right) \left| \mathbf{q}^{(-1)} \right\rangle,$$

which scaled by the number of electrons could provide, in turn, with an alternative atomic charge distribution in reciprocal space:

$$\left| \mathbf{q}_N^{(-1)} \right\rangle = N_M \left| p^{(-1)} \right\rangle.$$

For the previous diatomic molecular example one can easily define the reciprocal atomic charge distribution in every one of both atoms as:

$$q_A^{(-1)} = D^{-1} (\alpha - \beta) q_A = (\alpha + \beta)^{-1} q_A;$$

thus, the sum of both charges become twice the above quantity or:

$$N_M^{(-1)} = 2q_A^{(-1)} = 2(\alpha + \beta)^{-1} q_A = (\alpha + \beta)^{-1} N_M$$

and the scaled reciprocal atomic charge will be:

$$q_{N:A}^{(-1)} = N_M \left( N_M^{(-1)} \right)^{-1} q_A^{(-1)} = N_M \left[ (N_M)^{-1} (\alpha + \beta) \right] \left[ (\alpha + \beta)^{-1} q_A \right] = q_A$$

Which is quite an interesting result: in homonuclear diatomic molecules both charge sets are coincident. Such behaviour can be also found in highly symmetric simplex structures made with one kind of atoms only, whose charges are all equal.

It is also instructive to analyse the diatomic heteronuclear case, so the Euclidean metric can be written as:

$$\mathbf{Z} = \begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} \rightarrow \mathbf{Z}^{(-1)} = D^{-1} \begin{pmatrix} \gamma & -\beta \\ -\beta & \alpha \end{pmatrix} \wedge D = \alpha\gamma - \beta^2 > 0.$$

The reciprocal charges are:

$$q_A^{(-1)} = D^{-1} (\gamma q_A - \beta q_B) \wedge q_B^{(-1)} = D^{-1} (-\beta q_A + \alpha q_B)$$

and it is readily obtained:

$$N_M^{(-1)} = D^{-1} [(\gamma - \beta) q_A + (\alpha - \beta) q_B];$$

then, defining the ratio:

$$r = (\alpha q_B - \beta q_A) (\gamma q_A - \beta q_B)^{-1}$$

the corresponding reciprocal probability distribution becomes:

$$p_A^{(-1)} = [1 + r]^{-1} \wedge p_B^{(-1)} = [1 + r^{-1}]^{-1}$$

and the reciprocal charges are readily obtained by scaling the probabilities by the number of electrons:  $N_M = q_A + q_B$ .

Therefore, one can easily predict the inequalities:

$$r > 1 \rightarrow q_A^{(-1)} < q_B^{(-1)} \wedge r < 1 \rightarrow q_A^{(-1)} > q_B^{(-1)}.$$

However, one can also write:

$$r > 1 \rightarrow \alpha q_B - \beta q_A > \gamma q_A - \beta q_B;$$

so, if in case it is chosen the rough approach:  $\alpha \approx q_A^2 \wedge \gamma \approx q_B^2$ , then one will obtain:

$$(q_A q_B - \beta) q_A > (q_B q_A - \beta) q_B \rightarrow q_A > q_B,$$

and one can conclude that the original and reciprocal atomic charges will possess opposite trends. A similar relationship will occur when the case  $r < 1$  is studied.

### 1.15. Shape functions and density functions

It is well-known the relationship between density and shape functions, see for a recent example [14], but how density and shape functions are related from the point of view of the present discussion is still to be developed. One can easily describe the shape function as the homothety [14]:

$$\sigma_M(\mathbf{r}) = N_M^{-1} \rho_M(\mathbf{r}),$$

which transforms the density function, belonging to the shell  $S(N_M)$ , into an element of the semispace unit shell,  $S(1)$ :

$$\langle \sigma_M(\mathbf{r}) \rangle = N_M^{-1} \langle \rho_M(\mathbf{r}) \rangle = N_M^{-1} N_M = 1.$$

According to Mulliken decomposition (8) of the density function, it can be written:

$$\sigma_M(\mathbf{r}) = \sum_{I \in M} N_M^{-1} \left( \sum_{\alpha \in I} \rho_\alpha(\mathbf{r}) \right) = \sum_{I \in M} N_M^{-1} \rho_I(\mathbf{r}), \quad (30)$$

which provides an unsatisfactory description of the shape function, as the set of atomic density functions (10):  $|\rho\rangle = \{\rho_I(\mathbf{r})\}$  from the semispace shell structure point of view constitute a non homogeneous atomic basis set. However, a much coherent pattern can be found describing the atomic shape basis functions as:

$$|\sigma\rangle = \left\{ \sigma_I(\mathbf{r}) = q_I^{-1} \rho_I(\mathbf{r}) \right\} = P^{-1} |\rho\rangle \wedge P = \text{Diag}(q_I) \quad (31)$$

with the elements of the diagonal matrix  $P$  borrowed from the Mulliken population vector  $|\mathbf{p}\rangle$  as defined in equation (12). The atomic shape functions defined in equation (31) above belong to the unit shell, as:

$$\forall I : \langle \sigma_I | \sigma_I \rangle = q_I^{-1} \langle \rho_I | \rho_I \rangle = q_I^{-1} q_I = 1 \rightarrow \{\sigma_I\} \subset S(1).$$

Thus, the homothetic shape function (30) can be written as:

$$\sigma_M(\mathbf{r}) = \sum_I w_I \sigma_I(\mathbf{r}) \wedge \forall I : w_I = N_M^{-1} q_I; \quad (32)$$

the coefficient set  $|\mathbf{w}\rangle = \{w_I\}$ , acting as coordinates of the shape function with respect of the atomic shape basis set  $|\sigma\rangle$  as defined in equation (31), possess the important property to be convex, that is:

$$\langle |\mathbf{w}\rangle \rangle = \sum_I w_I = N_M^{-1} \sum_I q_I = N_M^{-1} \langle |\mathbf{q}\rangle \rangle = N_M^{-1} N_M = 1 \wedge \forall I : w_I \in [0, 1].$$

In other words, the shape function coordinate vector  $|\mathbf{w}\rangle$  with respect to the basis set  $|\sigma\rangle$  can be also considered as a discrete probability distribution. Thus, to the construct of the density function in terms of a unity vector coordinates as discussed in equation (11), when a Mulliken atomic density function basis set is used, correspond a construct of a well-defined unit shell shape function and vice versa:

$$\rho = \langle \rho | \mathbf{1} \rangle \leftrightarrow \sigma = \langle \sigma | \mathbf{w} \rangle. \quad (33)$$

Curiously enough, the shape function constructed accordingly to equation (32) posses a homogeneous structure in both atomic shape basis set and coordinates, which both belong to the appropriate unit shell; a property which lacks on the original density function, being the Mulliken atomic basis set, in general, a set of linearly independent functions belonging to different shells (excepting the cases when symmetry imposes that two or more atoms had the same population):

$$\forall I : \langle \rho_I | \rangle = q_I \rightarrow \rho_I(\mathbf{r}) \in S(q_I).$$

From the statistical point of view, therefore it seems that shape functions can play a more coherent role than density functions.

### 1.16. Minkowski decomposition

Quite a parallel reasoning can be used for the Minkowski decomposition. In this case, however, one shall take into account that the Minkowski norms of the decomposition of the density into the atomic density matrix (19), as described in equations (20) and (22), may not be positive definite:

$$q_{IJ} = \langle p_{IJ} \rangle = \sum_{\alpha \in I} \sum_{\beta \in J} D_{\alpha\beta} S_{\alpha\beta} \in \mathbf{R} \quad (34)$$

but one can only assure that:

$$q_I = \sum_J q_{IJ} \in \mathbf{R}^+.$$

However, the diagonal elements of the Minkowski metric are positive definite, as they can be written:

$$q_{II} = \langle p_{II} \rangle = \sum_{\alpha \in I} \sum_{\beta \in I} D_{\alpha\beta} S_{\alpha\beta} = \sum_i \omega_i \sum_{\alpha \in I} \sum_{\beta \in I} c_{\alpha i}^* c_{\beta i} S_{\alpha\beta} = \sum_i \omega_i \mathbf{c}_{Ii}^+ \mathbf{S}^{II} \mathbf{c}_{Ii} \in \mathbf{R}^+,$$

as the diagonal elements of the atomic partition of the overlap matrix:  $\mathbf{S}^{II} = \{\forall \alpha, \beta \in I : S_{\alpha\beta}\}$  are positive definite matrices, being the whole overlap positive definite; so, given any arbitrary non-zero vector,  $\mathbf{x}$ , then:  $\forall I : \mathbf{x}^+ \mathbf{S}^{II} \mathbf{x} > 0$ .

The construction of the shape function in terms of Minkowski shape basis functions cannot follow exactly the same algorithm as in the Mulliken atomic partition and it must be taken into account the possibility of using the atomic matrix basis set without a coherent shape structure as in the former Mulliken case. However, one can define a shape function adapted construct:

$$\sigma_M(\mathbf{r}) = \sum_I \sum_J N_M^{-1} p_{IJ}(\mathbf{r}) = \sum_I \sum_J w_{IJ} s_{IJ}(\mathbf{r})$$

using the corresponding definitions:

$$\forall I, J : s_{IJ}(\mathbf{r}) = \delta(q_{IJ} \neq 0) q_{IJ}^{-1} p_{IJ}(\mathbf{r}) \wedge w_{IJ} = N_M^{-1} q_{IJ}. \quad (35)$$

The shape function matrix elements of the basis set (35) can be collected into a matrix:  $\mathbf{\Sigma} = \{s_{IJ}(\mathbf{r})\}$  and the matrix-type coordinates of the Minkowski shape function:  $\mathbf{W} = \{w_{IJ}\}$ , fulfil the condition to sum up to one:

$$\langle \mathbf{W} \rangle = \sum_I \sum_J w_{IJ} = N_M^{-1} \sum_I \sum_J q_{IJ} = N_M^{-1} N_M = 1,$$

but as they can be negative the whole coordinate set cannot be considered convex.

In this manner, to the density function construct there correspond only one shape function construct. In this case one can thus write:

$$\rho_M(\mathbf{r}) = \langle \mathbf{P} * \mathbf{1} \rangle \leftrightarrow \sigma_M(\mathbf{r}) = \langle \Sigma * \mathbf{W} \rangle$$

which constitutes the equivalent relationship as the one described in Mulliken partition (33) between density and shape function in atomic Minkowski basis set.

### 1.17. Comparison of two density functions

In many circumstances, there appears the following problem. Two density functions for the same molecule in the same spatial configuration are known under two different approaches. Call this different functions of the same structure:  $\{\rho^{(A)}; \rho^{(B)}\}$ . They admit, by construction, similar Mulliken decompositions of type (8), which can be written in this case as:

$$\forall M = A, B : \rho^{(M)} = \sum_I \rho_I^{(M)}.$$

So, two shape functions can be described in the way equation (32) has been built by means of the convex expressions:

$$\forall M = A, B : \sigma^{(M)} = \sum_I w_I^{(M)} \sigma_I^{(M)}.$$

A technique to transform one of the densities into a form as similar or dissimilar possible to the other can be devised as follows. By constructing the scalar product of both shape functions one arrives to:

$$\langle \sigma^{(A)} | \sigma^{(B)} \rangle = \sum_{I \in A} \sum_{J \in B} w_I^{(A)} w_J^{(B)} \langle \sigma_I^{(A)} | \sigma_J^{(B)} \rangle = \langle \mathbf{w}^A | \mathbf{Z}^{AB} | \mathbf{w}^B \rangle, \quad (36)$$

where:

$$\forall M = A, B : |\mathbf{w}^M\rangle = \left\{ w_I^{(M)} \right\} \wedge \mathbf{Z}^{AB} = \left\{ z_{IJ}^{AB} = \langle \sigma_I^{(A)} | \sigma_J^{(B)} \rangle \right\}.$$

The optimization of equation (36) with respect to any of the coefficient vectors can be done as follows. The problem can be associated to a rotation in the  $n_M$ -dimensional space of the molecular atoms, which in turn can be constructed by means of elementary Jacobi rotations [15] in the way that some previous problems have been solved. Choosing the vector  $|\mathbf{w}^B\rangle$  as the one to be used to optimize the scalar product, one shall note that there can be chosen an auxiliary vector  $|\mathbf{x}\rangle$ , such that:

$$|\mathbf{w}^B\rangle = |\mathbf{x}\rangle * |\mathbf{x}\rangle \rightarrow 1 = \langle |\mathbf{w}^B\rangle \rangle = \langle |\mathbf{x}\rangle * |\mathbf{x}\rangle \rangle = \sum_I x_I^2 = \langle \mathbf{x} | \mathbf{x} \rangle$$

Thus, constructing the vector  $|\mathbf{w}^B\rangle$  to be optimized in this way is the same as to optimize a normalized vector  $|\mathbf{x}\rangle$ . This procedure has been employed in various situations and mainly in the search of ASA density functions [16].

Performing an elementary Jacobi rotation, using the cosine and sine  $\{c, s\}$  of some rotation angle acting on the elements  $\{P, Q\}$  of the vector  $|\mathbf{x}\rangle$ , results in a vector with all the rest of elements constant, but:

$$\begin{aligned} x_P &\rightarrow cx_P - sx_Q \\ x_Q &\rightarrow sx_P + cx_Q. \end{aligned}$$

So, the corresponding elements of  $|\mathbf{w}^B\rangle$ , vary according to:

$$\begin{aligned} w_P^B &= x_P^2 \rightarrow c^2x_P^2 + s^2x_Q^2 - 2scx_Px_Q \\ w_Q^B &= x_Q^2 \rightarrow s^2x_P^2 + c^2x_Q^2 + 2scx_Px_Q \end{aligned} \tag{37}$$

Therefore, if the following definition is used:

$$\begin{aligned} \langle\theta| = \{\theta_J\} &= \langle\mathbf{w}^A| \mathbf{Z}^{AB} \rightarrow \forall J : \theta_J = \sum_{I \in A} w_I^A z_{IJ}^{AB} \\ &\rightarrow \langle\sigma_A|\sigma_B\rangle = \langle\theta|\mathbf{w}^B\rangle \end{aligned}$$

employing the expression (36) one can easily write:

$$\langle\sigma^{(A)}|\sigma^{(B)}\rangle = \sum_{J \neq P, Q \in B} \theta_J x_J^2 + \theta_P x_P^2 + \theta_Q x_Q^2, \tag{38}$$

and from this result, equation (37) provides the scalar product expressed in terms of the elementary Jacobi rotation elements:

$$\begin{aligned} \langle\sigma^{(A)}|\sigma^{(B)}\rangle &= \sum_{J \neq P, Q \in B} \theta_J x_J^2 + c^2 (\theta_P x_P^2 + \theta_Q x_Q^2) \\ &\quad + s^2 (\theta_P x_Q^2 + \theta_Q x_P^2) + 2scx_Px_Q (\theta_Q - \theta_P) \end{aligned} \tag{39}$$

after using:  $c^2 = 1 - s^2$  and rearranging terms, the scalar product (39) variation simply becomes:

$$\delta\langle\sigma^{(A)}|\sigma^{(B)}\rangle = \left[ s^2 (x_Q^2 - x_P^2) - 2scx_Px_Q \right] (\theta_P - \theta_Q). \tag{40}$$

So, after taking also into account that:

$$\frac{dc}{ds} = -\frac{s}{c} = -t,$$

then, the optimal value with respect to the sine becomes:

$$\begin{aligned} \frac{d}{ds} \left( \delta \langle \sigma^{(A)} | \sigma^{(B)} \rangle \right) &= 2 (\theta_P - \theta_Q) c \left[ t^2 (x_P x_Q) - t (x_P^2 - x_Q^2) - x_P x_Q \right] = 0 \\ &\rightarrow t^2 (x_P x_Q) - t (x_P^2 - x_Q^2) - x_P x_Q = 0, \end{aligned}$$

so, finally one has two possible roots:

$$t = (2x_P x_Q)^{-1} \left( (x_P^2 - x_Q^2) \pm (x_P^2 + x_Q^2) \right)$$

providing two possible values:

$$t_+ = \frac{x_P}{x_Q} \wedge t_- = -\frac{x_Q}{x_P}. \quad (41)$$

From this result, substituting in (40), after taking into account that:

$$c = (1 + t^2)^{-\frac{1}{2}} \wedge s = tc,$$

then choosing one of the two roots<sup>6</sup> as written in equation (41),

$$\delta \langle \sigma^{(A)} | \sigma^{(B)} \rangle = \frac{t_*}{1 + t_*^2} \left[ t_* (x_Q^2 - x_P^2) - 2x_P x_Q \right] (\theta_P - \theta_Q)$$

will be an optimum. However, as the present procedure has not exactly the same structure as in the original Jacobi method, one can take into account that:

$$\frac{dt}{ds} = c^{-1} + s(c^{-2}t) = c^{-1}(1 + t^2)$$

so the second derivative can be written as:

$$\frac{d^2}{ds^2} (\delta \langle \sigma^{(A)} | \sigma^{(B)} \rangle) = 2 (\theta_P - \theta_Q) (t_* (t_*^2 + 3) (x_P x_Q) + (x_Q^2 - x_P^2)), \quad (42)$$

thus, the adequate root,  $t_*$ , can be chosen as the one which makes positive the expression (42) in case one seeks for a minimal value of the scalar product (36).

<sup>6</sup>According to the Jacobi diagonalization procedure, choosing the minimal root as:  $t_* = \min(t_+, t_-)$  it assures that the rotation angle will fulfil:  $|\alpha| \leq \frac{\pi}{4}$ .

### 1.17.1. Diatomic molecules as an example

When comparing two diatomic molecules, naming by  $\{P; Q\}$  the involved atoms whose coefficients will be transformed, then the working expression (38) becomes:

$$\langle \sigma^{(A)} | \sigma^{(B)} \rangle = \theta_P x_P^2 + \theta_Q x_Q^2,$$

So equation (40) and the sequel are already applicable in one Jacobi step. Suppose a case of a homonuclear diatomic structure compared with an arbitrary one. This situation will provide a clear answer as:

$$x_P = x_Q = \frac{1}{\sqrt{2}} \rightarrow w_P = w_Q = \frac{1}{2}$$

thus:

$$t_+ = 1 \wedge t_- = -1$$

Such a result gives second derivative values like:

$$\begin{aligned} \frac{d^2}{ds^2} \left( \delta \langle \sigma^{(A)} | \sigma^{(B)} \rangle \right)_+ &= 4(\theta_P - \theta_Q) \\ \frac{d^2}{ds^2} \left( \delta \langle \sigma^{(A)} | \sigma^{(B)} \rangle \right)_- &= -4(\theta_P - \theta_Q) \end{aligned}$$

which will depend on the elements of the vector  $|\theta\rangle$  and undoubtedly of the nature of the molecule  $A$ .

As equation (40) can be also written as:

$$\delta \langle \sigma^{(A)} | \sigma^{(B)} \rangle = c^2 \left[ t^2 (x_Q^2 - x_P^2) - 2tx_P x_Q \right] (\theta_P - \theta_Q),$$

in this case one will have:

$$\delta \langle \sigma^{(A)} | \sigma^{(B)} \rangle = \frac{t}{1+t^2} (\theta_Q - \theta_P)$$

Thus:

$$\begin{aligned} \delta \langle \sigma^{(A)} | \sigma^{(B)} \rangle_+ &= -\frac{1}{2} (\theta_P - \theta_Q) \\ \delta \langle \sigma^{(A)} | \sigma^{(B)} \rangle_- &= \frac{1}{2} (\theta_P - \theta_Q). \end{aligned}$$

If molecule  $A$  is homonuclear too, then:  $\theta_P = \theta_Q$  and all the involved scalar product functions will become zero.

## 2. Conclusions

An exhaustive analysis of the geometric aspects of the first order density function, with emphasis into the atomic metric aspects, leads to the well known Mulliken atomic density description, but also to an alternative Minkowski structure. In both cases the applications are studied. The most interesting feature is the condensation of both density atomic metrics, which lead to vector and matrix representations of a given molecule, depending of the number of atoms, and possessing a similar structure for all molecular constructs with the same number of atomic centres. Comparison of these atomic density matrices, the introduction of reciprocal atomic metric spaces, the impact of the atomic decomposition on shape functions and the comparison of two density functions based on Mulliken atomic partition prepares the landscape for extensive insight on further density analysis.

## Acknowledgments

The author expresses his acknowledgement to the Ministerio de Ciencia y Tecnología for the grant: BQU2003-07420-C05-01; which has partially sponsored this work and for a Salvador de Madariaga fellowship reference: PR2004-0547, which has made possible his stage at the Ghent University. The author also wishes to thank Professor Patrick Bultinck of the Ghent University for the deep discussions on the problems met in an early stage of the construction of the present work, his warm hospitality is also heartily acknowledged.

## References

- [1] See, for example: R. Carbó-Dorca, *Density Functions and Generating Wave Functions*, in: *Reviews of modern Quantum Chemistry*, ed. K.D. Sen (World Scientific. New Jersey, 2002) 401.
- [2] R. Carbó-Dorca, *Quantum Quantitative Structure-Activity Relationships (QQSAR): A Comprehensive Discussion Based On Inward Matrix Products, Employed As A Tool To Find Approximate Solutions Of Strictly Positive Linear Systems And Providing A QSAR-Quantum Similarity Measures Connection*. in: *Proceedings of European Congress on Computational Methods in Applied Sciences and Engineering. ECCOMAS 2000. Barcelona, 11–14 September 2000* ISBN-84-89925-70-4 (2000) 1–31.
- [3] R. Carbó-Dorca, *J. Mol. Struct. Teochem.* 537 (2001) 41.
- [4] R. Carbó-Dorca, *Int. J. Quantum Chem.* 91 (2003) 607.
- [5] See, for example: Lahey-Fujitsu Fortran Professional Edition. Lahey Computing Systems. Incline Village (NV) USA (2004). See the web site for more details: <http://www.lahey.com>
- [6] R. Carbó-Dorca, P. Bultinck, (a) *J. Math. Chem.* 36 (2004) 231; (b) *J. Math. Chem.* 36 (2004) 201.
- [7] R. Carbó-Dorca, *Fuzzy sets and boolean tagget sets, vector semiespaces and convex sets, QSM and ASA density functions, diagonal vector spaces and quantum chemistry*. *Adv. Molec. Simil.* 2 43–72. (1998), JAI Press.
- [8] R.S. Mulliken, *J. Chem. Phys.* 23 (1955) 1833, 1841, 2338, 2343.

- [9] J.L.W.V. Jensen, *Acta Math.* 30 (1906) 175.
- [10] P. Bultinck, X. Gironès, R. Carbó-Dorca, *Molecular quantum similarity: theory and applications*, *Rev. Comput. Chem.* 21 (2005) 127, eds. K.B. Lipkowitz, R. Larter, T. Cundari (Wiley, Hoboken, USA, 2005).
- [11] R. Carbó-Dorca; *J. Math. Chem.* 27 (2000) 357.
- [12] J. Cioslowski, E.D. Fleishmann, *J. Am. Chem. Soc.* 113 (1991) 64.
- [13] R. Hoffman, *J. Chem. Phys.* 39 (1969) 1397.
- [14] P. Bultinck, R. Carbó-Dorca, *J. Math. Chem.* 36 (2004) 191.
- [15] C.J.G. Jacobi, *J. Reine Angew. Math.* 30 (1846) 51.
- [16] L. Amat, R. Carbó-Dorca, *J. Comput. Chem.* 20 (1999) 911.