J Math Chem (2008) 44:628–636
DOI 10.1007/s10910-008-9442-z

ORIGINAL PAPER

Molecular quantum similarity measures in Minkowski metric vector semispaces

Ramon Carbó-Dorca

Received: 5 November 2005 / Accepted: 16 November 2007 / Published online: 4 September 2008

Abstract Minkowski metric vector semispaces can be chosen as the natural mathematical framework, where quantum similarity measures are described and evaluated. The obtained results in this study show that the Minkowski metric option is easily feasible, providing a new set of computationally simpler expressions, computationally faster when compared with Euclidian based quantum similarity measures.

Keywords Quantum similarity measures · Vector semispaces · Unit shell · Minkowski metric · General weighted Minkowski metric · Shape functions

1 Introduction

Quantum molecular similarity measures have been applied in many chemical fields [1,2], since its first appearance in the literature [3] they have been mainly employed in a computational structure provided with a metric based on Euclidian norms and scalar products. The development of the quantum similarity theoretical background on the other hand, has brought essentially to the definition and properties of semispaces [4] within their natural Minkowski metric [5]. These findings had raised the expectative to use such a metric in the quantum similarity field; thus some of the related questions will be tentatively answered in a first instance within the present work.

This paper will be structured by the following scheme: First a short résumé about semispaces and Minkowski metric will be given. Then, the cosine of the subtended

R. Carbó-Dorca (✉)
Institut de Química Computacional, Universitat de Girona, Girona, 17071 Catalonia, Spain
e-mail: quantumqsar@hotmail.com

R. Carbó-Dorca
Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281,
B-9000 Gent, Belgium

angle will be studied using the construction of a pseudo-wave function attached to a known density function. The overlap similarity measure between the density functions of two systems will be studied afterwards, arriving to a simple algebraic form, based on the overlap matrix between the basis functions of both systems. Finally, the description of the triple density and Coulomb similarity measures are used as first steps to obtain a general expression for this new kind of weighted Minkowski metric quantum similarity measures.

2 Minkowski metric vector semispaces

A vector semispace (VSS) is a subset of any vector space, which is completely defined over the positive real numbers [4], in such a manner that the additive group of the usual vector space can be substituted in a VSS by a semigroup [5], a kind of group where no reciprocal elements are allowed. Metric VSS can be naturally attached to Minkowski norms [6].

A Minkowski norm is defined within a VSS in the following way:

(a) In infinite dimensional functional VSS, Minkowski norms are described by an integral:

$$\forall f(t) \in V_\infty\left(\mathbf{R}^+\right) \rightarrow \langle f \rangle = \int_D f(t)\, dt,$$

(b) In finite dimensional column or row VSS, the Minkowski norm is simply evaluated by summing up their elements, for example:

$$\forall \mathbf{x} = \{x_i\} \in V_n\left(\mathbf{R}^+\right) \rightarrow \langle \mathbf{x} \rangle = \sum_i x_i .$$

Metric VSS can be organized in shells [6]. A shell $S(\mu)$ is a set of vectors belonging to a VSS whose Minkowski norms are coincident. Shells are constructed by taking into account the Minkowski norms of their elements as follows:

$$\forall x \in V\left(\mathbf{R}^+\right) : \langle x \rangle = \mu \in \mathbf{R}^+ \rightarrow x \in S(\mu) \subset V\left(\mathbf{R}^+\right).$$

Among the infinite set of shells in a VSS, the unit shell $S(1)$ is most important, due that it can be considered the VSS subset made of all probability distributions; also, an element of any shell can be transformed into a unit shell one by a simple homothecy. This property has been discussed recently in order to set the relationship and associated properties between first order density functions and shape functions [7].

Scalar products in VSS shall be performed taking into account the square roots of the involved vectors' elements, in order they yield the Minkowski norm, when both vectors appearing into the scalar product are the same. That is, it can be written:

$$\langle f|g \rangle = \int_D \left(f(t)\, g(t)\right)^{\frac{1}{2}} dt \wedge \langle \mathbf{x}|\mathbf{y} \rangle = \sum_i \left(x_i\, y_i\right)^{\frac{1}{2}}$$

for the two kinds of possible Minkowski norms in VSS as commented before.

In order to distinguish the Minkowski metric from the Euclidian one, the symbol $(\mathbf{a}|\mathbf{b})$ will be used to write an Euclidian scalar product; for instance, if: $\mathbf{a} = \{a_i\} \wedge \mathbf{b} = \{b_i\}$, are two n-tuple arrays, then the Euclidian scalar product symbol is computed as: $(\mathbf{a}|\mathbf{b}) = \sum_i a_i b_i$.

3 Cosine of the subtended angle between two vectors in a Minkowski metric VSS

The cosine of the subtended angle between two vectors in a Minkowski metric VSS can, thus, be computed as in the usual vector algebra:

$$\cos(\alpha) = (\langle \mathbf{x} \rangle \langle \mathbf{y} \rangle)^{-\frac{1}{2}} \langle \mathbf{x}|\mathbf{y} \rangle \wedge \mathbf{x} = \mathbf{y} \rightarrow \cos(\alpha) = 1,$$

because employing the previous Minkowski metric definitions: $\langle \mathbf{x}|\mathbf{x} \rangle = \langle \mathbf{x} \rangle$. Suppose now that the following generating rules [4] are set for two vectors in a known Minkowski metric VSS:

$$R(\mathbf{a} \rightarrow \mathbf{x}) \wedge R(\mathbf{b} \rightarrow \mathbf{y}),$$

meaning that:

$$\mathbf{x} = \mathbf{a}^* * \mathbf{a} \rightarrow \forall x_i = |a_i|^2 \wedge \mathbf{y} = \mathbf{b}^* * \mathbf{b} \rightarrow \forall y_i = |b_i|^2,$$

where an inward matrix product [8,9] is employed. This construction, if both generated VSS vectors are elements of the unit shell, implies that the generating vectors are normalized in the Euclidean norm sense, and vice versa: whenever the generating vectors are normalized in the Euclidian sense, the generated vectors belong to the unit shell. The interesting fact in this circumstance is that, then, the cosine of the VSS vectors and the one of their generators is the same, that is:

$$\cos(\alpha) = \langle \mathbf{x}|\mathbf{y} \rangle = \sum_i (x_i y_i)^{\frac{1}{2}} = \sum_i a_i b_i = (\mathbf{a}|\mathbf{b}),$$

moreover, the cosine is coincident with the Minkowski product of the generated vectors and the Euclidian product of the generating vectors.

4 Overlap quantum similarity measures

The equivalence between unit shell Minkowski metric and generating Euclidian scalar products has the most important consequence in evaluating the, so-called, overlap similarity measures (OSM) [10] and, thus, by extension the so-called Carbó index [11] is obtained, involving infinite dimensional VSS made of quantum mechanical density functions as elements.

Indeed, suppose two density functions, scaled by the respective number of electrons, that is, the so-called shape functions attached to a pair of molecular structures: $\{\rho_A, \rho_B\}$, the OSM according to a Minkowski metric can be written:

$$\langle \rho_A | \rho_B \rangle = \int_D (\rho_A \rho_B)^{\frac{1}{2}} dV.$$

On the other hand, it has been described that to any density function, ρ, a pseudo-wave function, ψ, which generates it in the previous sense:

$$R(\psi \rightarrow \rho) \equiv \rho(\mathbf{r}) = \psi^*(\mathbf{r}) * \psi(\mathbf{r})$$

can be easily described [6,12] under some specific circumstances. To gain insight into this generating process, suppose that for the molecule A, a MO description for the density function is known and that a pseudo-wave function is constructed as:

$$\rho_A = \sum_i \omega_i^A \left| \phi_i^A \right|^2 \wedge \psi_A = \sum_i \left(\omega_i^A \right)^{\frac{1}{2}} \phi_i^A \rightarrow \rho_A = \psi_A^* * \psi_A, \tag{1}$$

also the same holds for molecule B. The signs of the square roots of the scaled MO occupation numbers $\left\{ \omega_i^A = N_A^{-1} \nu_i^A \right\}$ are taken to be positive, in order to obviate the undetermined signature of the pseudo-wave function [6–12], which acts as an undetermined phase factor. It must be now commented that this set up is a generalized one in MO theory, as both the density expression and the attached pseudo-wave function, conveniently written in a similar manner as in Eq. 1, can be taken as feasible formal expressions valid up to any density function order [13]. This possibility will be deeply studied elsewhere.

If the densities in systems A and B are normalized in the Minkowski sense, then due to the property already studied, the cosine of the subtended angle between both functions is written:

$$\langle \rho_A | \rho_B \rangle = (\psi_A | \psi_B) = \sum_i \sum_j \left(\omega_i^A \omega_j^B \right)^{\frac{1}{2}} \int_D \left(\phi_i^A \right)^* \phi_j^B dV, \tag{2}$$

but, as in molecule A, for instance, the MO set $\left\{ \phi_i^A \right\}$ within the LCAO theory can be expressed as a superposition of atomic centered functions $\left\{ \chi_\mu^A \right\}$, then:

$$\phi_i^A = \sum_\mu c_{\mu i}^A \chi_\mu^A$$

and the same holds for molecule B. Therefore, one can write the overlap between two MO belonging to systems A and B as:

$$\int_D \left(\phi_i^A\right)^* \phi_j^B dV = \sum_\mu \sum_\nu \left(c_{\mu i}^A\right)^* c_{\nu j}^B \int_D \left(\chi_\mu^A\right)^* \chi_\nu^B dV = \mathbf{c}_{A;i}^+ \mathbf{S}_{AB} \mathbf{c}_{B;j},$$

with the overlap matrix between basis functions of systems A and B defined by:

$$\mathbf{S}_{AB} = \left\{ S_{AB;\mu\nu} = \int_D \left(\chi_\mu^A\right)^* \chi_\nu^B dV = \left(\chi_\mu^A \middle| \chi_\nu^B\right) \right\},$$

and the MO coordinate coefficients are collected into the column vector: $\mathbf{c}_{A;i} = \left\{c_{\mu i}^A\right\}$ with a similar form for molecule B.

Consequently, the OSM conforming to a VSS Minkowski metric can be written as:

$$\langle \rho_A | \rho_B \rangle = \sum_i \sum_j \left(\omega_i^A \omega_j^B\right)^{\frac{1}{2}} \mathbf{c}_{A;i}^+ \mathbf{S}_{AB} \mathbf{c}_{B;j}.$$

In order to evaluate the OSM above described in an efficient computational manner, one can define the auxiliary column vectors, which are nothing else than the MO coordinates scaled by the square root of the occupation numbers:

$$\mathbf{d}_{A;i} = \left(\omega_i^A\right)^{\frac{1}{2}} \mathbf{c}_{A;i},$$

and the same for molecule B. With the aid of the scaled MO coordinate vectors, the following matrix:

$$\mathbf{T}_{AB} = \left\{ T_{AB;ij} = \mathbf{d}_{A;i}^+ \mathbf{S}_{AB} \mathbf{d}_{B;j} \right\}$$

can be easily constructed. Finally, the Minkowski OSM is expressible as a complete sum [13] of the elements of matrix $\mathbf{T}_{AB}$:

$$\langle \rho_A | \rho_B \rangle = \langle \mathbf{T}_{AB} \rangle = \sum_i \sum_j T_{AB;ij}. \tag{3}$$

The OSM obtained in this way still will depend on the relative positions of both molecules, as the associated Euclidean expression is, see for example [2], but the Minkowski MO algorithm is connected to integrals with a pair of orbital basis functions only, instead of a four-function dependence as encountered in the usual Euclidean representation [10]. A similar result has been described when comparing matrix representations of density functions acting as operators [14].

A comment about the form of the self-similarity measure, when computed in this way may be interesting. Taking Eq. 2, the selfsimilarity measure for system A will become the number of electrons:

$$\langle \rho_A | \rho_A \rangle = (\psi_A | \psi_A) = \sum_i \sum_j \left(\omega_i^A \omega_j^A \right)^{\frac{1}{2}} \int_D \left(\phi_i^A \right)^* \phi_j^A dV$$
$$= \sum_i \sum_j \left(\omega_i^A \omega_j^A \right)^{\frac{1}{2}} \delta_{ij} = \sum_i \omega_i^A = N_A.$$

For other kind of weighting operators than the unity, the intermolecular overlap matrix shall be substituted by the corresponding matrix representation. Some examples will be discussed next and a generalization proposal given at the end.

5 Triple density similarity measures

As a first example of the extension of Minkowski metric to similarity measures with weighting operators, the operator associated to a third molecular structure C will be considered, forming in this way the so-called triple density similarity measures [15]. The integral to be computed can be written within the weighted Minkowski metric structure as:

$$\langle \rho_A | \rho_C | \rho_B \rangle = \int_{D_1} \int_{D_2} (\rho_A (\mathbf{r}_1))^{\frac{1}{2}} \rho_C (\mathbf{r}_1; \mathbf{r}_2) (\rho_B (\mathbf{r}_2))^{\frac{1}{2}} d\mathbf{r}_1 d\mathbf{r}_2$$

so, applying the same pseudo-wave function decomposition as in the OSM case, it can be written:

$$\langle \rho_A | \rho_C | \rho_B \rangle = (\psi_A | \rho_C | \psi_B)$$
$$= \sum_i \sum_j \sum_k \left[\left(\omega_i^A \omega_j^B \right)^{\frac{1}{2}} \omega_k^C \right] \left[\int_{D_1} \left(\phi_i^A \right)^* \phi_k^C dV_1 \right]$$
$$\times \left[\int_{D_2} \left(\phi_k^C \right)^* \phi_j^B dV_2 \right]$$

which, upon defining, in the same way as in the OSM case, the matrices $\{\mathbf{T}_{AC}; \mathbf{T}_{CB}\}$, the Minkowski triple density measure can be finally written with the complete sum of the product of the involved matrices:

$$\langle \rho_A | \rho_C | \rho_B \rangle = \langle \mathbf{T}_{AC} \mathbf{T}_{CB} \rangle = \sum_i \sum_j \sum_k T_{AC;ik} T_{CB;kj}.$$

This result is computationally interesting, because there is not involved the superposition optimization of the three systems simultaneously, as it should be in the Euclidian norm case, but one can just separately optimize the involved systems two by two using an overlap expression of type (3); once every pair of molecules is optimally superposed, then the optimal triple similarity measures can be immediately evaluated.

Here, the selfsimilarity implying that all the involved densities are the same provides the sum of the scaled occupation numbers squared:

$$\begin{aligned}\langle \rho_A |\, \rho_A \,| \rho_A \rangle &= (\psi_A \,| \rho_A |\, \psi_A) \\ &= \sum_i \sum_j \sum_k \left[\left(\omega_i^A \omega_j^A \right)^{\frac{1}{2}} \omega_k^A \right] \delta_{ik} \delta_{kj} = \sum_k \left(\omega_k^A \right)^2 .\end{aligned}$$

6 Coulomb operator weighted similarity measures

The triple density similarity measure studied above indicates how Coulomb operators can act as weights in Minkowski metric OSM. That is, when computing the Coulomb weighted similarity measure, as:

$$\begin{aligned}\langle \rho_A |\, r_{12}^{-1} \,| \rho_B \rangle &= \int_{D_1} \int_{D_2} \left(\rho_A (\mathbf{r}_1)\, r_{12}^{-2} \rho_B (\mathbf{r}_2) \right)^{\frac{1}{2}} d\mathbf{r}_1 d\mathbf{r}_2 \\ &= (\psi_A \left| r_{12}^{-1} \right| \psi_B) = \int_{D_1} \int_{D_2} \psi_A^* (\mathbf{r}_1)\, r_{12}^{-1} \psi_B (\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2\end{aligned}$$

then, the integral: $\left(\psi_A \left| r_{12}^{-1} \right| \psi_B \right)$ appears to be a Coulomb repulsion integral involving two one-electron functions, which obviously can be associated to an overlap integral with the Coulomb operator acting as a weight. Within the LCAO theory there will be necessary to compute Coulomb integrals involving two basis functions only, obtaining a new Minkowski metric weighted overlap matrix with the definition:

$$\mathbf{R}_{AB} = \left\{ R_{AB;\mu\nu} = \int_{D_1} \int_{D_2} \left(\chi_\mu^A (\mathbf{r}_1) \right)^* r_{12}^{-1} \chi_\nu^B (\mathbf{r}_2)\, d\mathbf{r}_1 d\mathbf{r}_2 = \left(\chi_\mu^A \left| r_{12}^{-1} \right| \chi_\nu^B \right) \right\}. \quad (4)$$

Thus, employing the additional matrix:

$$\mathbf{Q}_{AB} = \left\{ Q_{AB;ij} = \mathbf{d}_{A;i}^{+} \mathbf{R}_{AB} \mathbf{d}_{B;j} \right\}$$

one arrives to obtain the Coulomb operator weighted Minkowski OSM as the complete sum:

$$\langle \rho_A |\, r_{12}^{-1} \,| \rho_B \rangle = \langle \mathbf{Q}_{AB} \rangle = \sum_i \sum_j Q_{AB;ij}.$$

The necessary basic two-center two-function Coulomb integrals: $\left\{\left(\chi_\mu^A \left| r_{12}^{-1} \right| \chi_\nu^B\right)\right\}$, can be taken as simplified forms of the general four-center four-function repulsion integrals. This can be easily seen, when inspecting old GTO integral expressions, like the ones provided by Saunders [16], for example. Interesting enough, the matrix (4), when associated to one molecular system only, has to behave as a Coulomb weighted metric positive definite matrix.

7 General operator weighted Minkowski metric OSM

The previous results on triple density and Coulomb weighted Minkowski OSM, preclude and made easy to write a general OMS weighted operator formalism within a Minkowski metric VSS. It is only necessary to consider a positive definite operator, which can be written in a module like form:

$$\Omega\left(\mathbf{r}_1;\mathbf{r}_2\right) = \left|\Theta\left(\mathbf{r}_1;\mathbf{r}_2\right)\right|,$$

then the weighted OSM can be immediately evaluated as:

$$\begin{aligned}\langle\rho_A|\,\Omega\,|\rho_B\rangle &= \int_{D_1}\int_{D_2}\left(\rho_A\left(\mathbf{r}_1\right)\left|\Theta\left(\mathbf{r}_1;\mathbf{r}_2\right)\right|^2\rho_B\left(\mathbf{r}_2\right)\right)^{\frac{1}{2}} d\mathbf{r}_1 d\mathbf{r}_2 \\ &= \left(\psi_A\,|\Omega|\,\psi_B\right) = \int_{D_1}\int_{D_2}\psi_A^*\left(\mathbf{r}_1\right)\Omega\left(\mathbf{r}_1;\mathbf{r}_2\right)\psi_B\left(\mathbf{r}_2\right) d\mathbf{r}_1 d\mathbf{r}_2\end{aligned}$$

so, as in the previous cases, a matrix representation over the basis functions shall be considered to be computed. Therefore employing the following matrix expression:

$$\mathbf{Z}_{AB} = \left\{Z_{AB;\mu\nu} = \int_{D_1}\int_{D_2}\left(\chi_\mu^A\left(\mathbf{r}_1\right)\right)^*\Omega\left(\mathbf{r}_1;\mathbf{r}_2\right)\chi_\nu^B\left(\mathbf{r}_2\right)d\mathbf{r}_1 d\mathbf{r}_2 = \left(\chi_\mu^A\,|\Omega|\,\chi_\nu^B\right)\right\} \tag{5}$$

a new matrix can be constructed by using the scaled MO coordinate vectors:

$$\mathbf{A}_{AB} = \left\{A_{AB;ij} = \mathbf{d}_{A;i}^{+}\mathbf{Z}_{AB}\mathbf{d}_{B;j}\right\},$$

in this way, the weighted Minkowski OSM can be written as the total sum:

$$\langle\rho_A|\,\Omega\,|\rho_B\rangle = \langle\mathbf{A}_{AB}\rangle = \sum_i\sum_j A_{AB;ij}.$$

The weighted Minkowski metric OSM in this easy manner present a generalized structure associated to positive definite operators. In the same manner as has been commented when the Coulomb operator was studied, the matrix (5) will behave as a positive definite metric matrix.

8 Conclusions

The basic procedure described here constitutes a correct and generalized process to compute quantum similarity measures within the LCAO MO theory, conforming to a VSS Minkowski metric. The deducible algorithms in the present theory appear to be, in general, computationally faster than the usual ones, appeared so far in the literature, based on Euclidian metric.

Acknowledgements 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. R. Carbó-Dorca, D. Robert, L. Amat, X. Gironés, E. Besalú, in *Molecular Quantum Similarity in QSAR and Drug Design*. Lecture Notes in Chemistry, vol. 73 (Editorial Springer Verlag, Berlin, 2000)
2. P. Bultinck, X. Girones, R. Carbó-Dorca, in *Molecular Quantum Similarity: Theory and Applications*, Rev. Comput. Chem. **21**, ed. by K.B. Lipkowitz, R. Larter, T. Cundari (Wiley, Hoboken, 2005), p. 127
3. R. Carbó, L. Leyda, M. Arnau, Int. J. Quantum Chem. **17**, 1185 (1980)
4. R. Carbó-Dorca, in *Fuzzy Sets and Boolean Tagged Sets, Vector Semispaces and Convex Sets, QSM and ASA Density Functions, Diagonal Vector Spaces and Quantum Chemistry*, Adv. Mol. Simil. **2**, ed. by R. Carbó-Dorca, P.G. Mezey (JAI Press, London, 1998), p. 43
5. I.M. Vinogradov (ed.), *Encyclopaedia of Mathematics* (Reidel-Kluwer Academic Publishers, Dordrecht, 1987)
6. R. Carbó-Dorca, J. Math. Chem. **32**, 201 (2002)
7. P. Bultinck, M. Rafat, R. Ponec, R. Carbó-Dorca, P. Popelier, J. Phys. Chem. A **110**, 7642 (2006)
8. R. Carbó-Dorca, J. Mol. Struct. (Theochem) **537**, 41 (2001)
9. 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*. 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) pp. 1–31
10. R. Carbó-Dorca, J. Math. Chem., doi:10.1007/s10910-008-9442-2
11. R. Carbó, E. Besalú, Ll. Amat, X. Fradera, J. Math. Chem. **19**, 47 (1996)
12. R. Carbó-Dorca, in *Density Functions and Generating Wave Functions*, ed by K.D. Sen. Reviews of Modern Quantum Chemistry. A Celebration of the Contributions of Robert G. Parr, vol.**I** (World Scientific, Singapore, 2002), p. 401
13. E. Besalú, R. Carbó, in *Applications of Nested Summation Symbols to Quantum Chemistry:Formalism and Programming Techniques*, ed. by M. Defranceschi, Y. Ellinger. Strategies and Applications in Quantum Chemistry: From Astrophysics to Molecular Engineering, An Hommage to Prof. G. Berthier (Kluwer Academic Publishers, Amsterdam, 1996), p. 229
14. R. Carbó-Dorca, Adv. Quantum Chem. **49**, 121 (2005)
15. R. Carbó, B. Calabuig, E. Besalú, A. Martínez, Mol. Eng. **2**, 43 (1992)
16. V.R. Saunders, in *An Introduction to Molecular Integral Evaluation*, ed. by G.H.F. Diercksen, B.T. Sutcliffe, A. Veillardin. Computational Techniques in Quantum Chemistry and Molecular Physics, Proceedings of the NATO Advanced Study Institute held at Ramsau, Germany, 4–21 September 1974 (D. Reidel Pub. Co., Dordrecht, 1975), p. 347