J Math Chem (2008) 44:228–234
DOI 10.1007/s10910-007-9305-z

ORIGINAL PAPER

A quantum similarity matrix (QSM) Aufbau procedure

Ramon Carbó-Dorca

Received: 15 February 2007 / Accepted: 10 July 2007 / Published online: 21 September 2007

Abstract An Aufbau recursive algorithm, leading to the construction of molecular Quantum similarity matrices (QSM) with positive definite structure is described. As a consequence, Molecular Quantum Similarity measures optimization has to be restricted by a recursive constraint, related to the Euclidian norm of the QSM column elements in Quantum Object density tag reciprocal space.

Keywords Quantum similarity · Quantum similarity matrices · Quantum similarity Aufbau condition · Quantum similarity matrices Aufbau recursive construction

Introduction

Suppose a known given Quantum Object Set (QOS) [1] formed by N molecules, with density tags described as: $\{\rho_I(\mathbf{r})\}$. The usual procedure for constructing the symmetric $(N \times N)$ Quantum Similarity Matrix (QSM), $\mathbf{Z} = \{z_{IJ}\}$, using overlap similarity measures, for example, has been described as [2]:

$$\forall I > J : z_{IJ} = \langle \rho_I \rho_J \rangle = \int_D \rho_I(\mathbf{r})\, \rho_J(\mathbf{r})\, d\mathbf{r} = z_{JI}, \tag{1}$$

However, as it is well know since the first paper on the subject [3] the set of quantum similarity measures $\{z_{IJ}\}$ depend on the relative position in 3D space of the implied

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

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

Quantum Objects (QOs). As the QO density function labels are positive definite functions, the integrals of type (1) can be considered as measures; thus, they are positive definite too. Since now the usual procedure to construct the QSM has been to maximize each of the integrals of type (1) with respect the translations and rotations of one of the implied QO with respect to the other. This can be expressed formally, for instance, as:

$$\forall I > J : z_{IJ} = \max_{\mathbf{t};\boldsymbol{\Omega}} \int_D \rho_I(\mathbf{r})\, \rho_J(\mathbf{r}\,|\mathbf{t};\boldsymbol{\Omega})\, d\mathbf{r} \tag{2}$$

where the pairs: $\{\mathbf{t};\boldsymbol{\Omega}\}$ are translations and rotations respectively, performed on the center of coordinates of the J-th QO [4]. It is irrelevant which one of the QO pair is chosen in order to optimize the integral (1) by means of the algorithm (2), the same result shall be obtained choosing the I-th QO for undertaking translations and rotations.

Apparently, such a procedure, repeated for every non redundant couple of QO's, shall provide a QSM $\mathbf{Z}$ with appropriate characteristics [5] associated to a metric matrix. The most important one consists into that the attached QSM has the property to be positive definite; as the density tag set is linearly independent, if the QOS is made of different QOs, then $\mathbf{Z}$ has to be a metric matrix of a pre-Hilbert space [6]. However, in many cases the use of algorithm (2) does not provide a QSM whose whole spectrum is positive definite, but a small amount of the $\mathbf{Z}$ eigenvalues may appear to be negative. This non-definite behavior of the metric matrix $\mathbf{Z}$ can be attributed to the fact that following algorithm (2), when facing the J-th QO to the rest of the QOS elements, then for every distinct QO a different relative position of the J-th QO is found, while reaching the optimal value of the similarity measure (1) for every pair of QOs; that is: the relative position of the J-th QO with respect to the I-th QO, $\forall I : J \neq I$, in order to optimize every element z_{IJ}, becomes different, and therefore when optimizing Eq. 2 one will obtain a set of different optimal translations-rotations: $\{\mathbf{t}_I;\boldsymbol{\Omega}_I\}\forall I \neq J$.

When computing any optimal quantum similarity measure by means of algorithm (2), one also must be aware that the final result, can be used to construct the symmetric (2×2) matrix:

$$\mathbf{Z}^{IJ} = \begin{pmatrix} z_{II} & z_{IJ} \\ z_{JI} & z_{JJ} \end{pmatrix} \wedge z_{IJ} = z_{JI}, \tag{3}$$

and also has to provide at least a positive definite matrix (3), which is the same as to consider the following property has to be fulfilled:

$$Det\left|\mathbf{Z}^{IJ}\right| = z_{II}z_{JJ} - z_{IJ}^2 > 0 \rightarrow z_{II}z_{JJ} > z_{IJ}^2. \tag{4}$$

The restriction (4) can also be written as:

$$z_{JJ} > z_{IJ}^2 z_{II}^{-1}, \tag{5}$$

and this will provide a form of the (2×2) positive definite restrictions to be easily

related to the general analysis which follows. Therefore, the algorithm (2) has to be modified accordingly incorporating the inequality (4) as a restriction:

$$\forall I > J : z_{IJ} = \max_{\mathbf{t};\boldsymbol{\Omega}} \int_D \rho_I(\mathbf{r})\, \rho_J(\mathbf{r}\,|\mathbf{t};\boldsymbol{\Omega})\, d\mathbf{r} \wedge z_{IJ}^2 < z_{II} z_{JJ} \tag{6}$$

and one can expect that the general QSM **Z**, can approach in this way the required complete positive definiteness, although this cannot be completely assured. In fact, this (2×2) restrictions constitute an incomplete point of view, as nothing can be said about the positive definiteness of higher the dimensional submatrices of the QSM **Z**. In this sense, the restricted algorithm (6) is more or less similar to the triangle distance relationship coherence, sought by an already published procedure [7].

Gershgorin circles and positive definiteness of QSM

In general, the positive definiteness of the similarity matrices of QSM type **Z** could be assessed by means of the Gershgorin theorem, see for more details reference [8]. Indeed, the QSM: $\mathbf{Z} = \{z_{IJ}\}$ are strictly positive, that is it can be written:

$$\mathbf{Z}* > 0 \rightarrow \forall z_{IJ} \in \mathbf{R}^+;$$

thus, the attached Gershgorin circles become intervals in the real line $\{\Gamma_I\}$, defined by means of the Gershgorin radius:

$$\forall I : r_I = \sum_{J \neq I} z_{IJ} \rightarrow \Gamma_I = z_{II} \pm r_I \tag{7}$$

and then all the eigenvalues, the spectrum: $Sp\,[\mathbf{Z}]$, of the QSM must be included within the union of the intervals (7). Therefore, defining the upper and lower bounds of each Gershgorin interval as:

$$\forall I : \Gamma_I^+ = z_{II} + r_I \wedge \Gamma_I^- = z_{II} - r_I$$

respectively. Then within a coarse rule, it can be said that the spectrum of the similarity matrix shall be necessarily contained into the line segment defined by the extreme values of the Gershgorin intervals:

$$Sp\,[\mathbf{Z}] \subset \left(\min_I \left\{\Gamma_I^-\right\} ;\ \max_I \left\{\Gamma_I^+\right\} \right).$$

Because of this inclusion due to Gershgorin theorem, a QSM is positive definite: $\mathbf{Z} > 0$, if the lower Gershgorin limit is positive:

$$\min_I \left\{\Gamma_I^-\right\} \in \mathbf{R}^+ \rightarrow Sp\,[\mathbf{Z}] \subset \mathbf{R}^+ \rightarrow \mathbf{Z} > 0.$$

This will be equivalent to say that whenever a QSM is diagonally dominant it is also positive definite [8], that is:

$$\forall I : z_{II} > r_I \rightarrow \forall I : z_{II} - r_I \in \mathbf{R}^+ \rightarrow \min_I \left\{\Gamma_I^-\right\} \in \mathbf{R}^+ \rightarrow Sp\,[\mathbf{Z}] \subset \mathbf{R}^+ \rightarrow \mathbf{Z} > 0. \quad (8)$$

This last property becomes a straightforward test to perform on QSM, as a result of the test, if the matrix diagonal dominance is fulfilled it can be assured that: $\mathbf{Z} > 0$. However, the QSM are usually not diagonally dominant, although they still are positive definite if properly constructed.

The quantum similarity matrix Aufbau recursive algorithm

Although one can use the Gershgorin theorem to test the positive definiteness of any QSM, a complete QSM calculation algorithm, based on the generalization of property (5) for (2×2) matrices, in order to assure the QSM $\mathbf{Z}$ positive definiteness, shall be based in an aufbau manner; that is: starting from any pair of QO, algorithm (6) is put forward. The result will be a positive definite matrix, $\mathbf{Z_0}$ say, with a structure like the matrix (3) above defined. A simple recursive aufbau algorithm can be described in order to obtain a final positive definite QSM.

Suppose that for some index $P < N$, a $(P \times P)$ positive definite QSM $\mathbf{Z_0}$ has been obtained, using the QOs sequence:$\{I_K; K = 1, P\}$. One can add a new QO to the aufbau procedure, the Q-th QO, say, in such a way that an augmented QSM, $\mathbf{Z_1}$, is obtained possessing the partitioned structure:

$$\mathbf{Z_1} = \begin{pmatrix} \mathbf{Z_0} & |\mathbf{z}\rangle \\ \langle \mathbf{z}| & \theta \end{pmatrix},$$

with the $(1 \times P)$ row vector defined as: $\langle \mathbf{z}| = \left(z_{I_1 Q}; z_{I_2 Q}; \ldots z_{I_P Q}\right)$, and the column vector $|\mathbf{z}\rangle$, being just the transpose of the former; finally, $\theta \equiv z_{QQ}$ is the selfsimilarity of the added QO.

The sufficient relationship (8), which can be written here as the set of conditions:

$$\theta > \langle |\mathbf{z}\rangle \rangle \wedge \forall K = 1, P : z_{I_K I_K} > \sum_{L \neq K} z_{I_K I_L} + z_{I_K Q}, \quad (9)$$

assuring that the augmented matrix $\mathbf{Z_1}$ has a positive definite structure, can be alternatively rewritten via a recursive Cholesky decomposition algorithm, described in several places [9].

The necessary and sufficient condition for the positive definiteness of the augmented QSM $\mathbf{Z_1}$ can be stated as:

$$\theta - \langle \mathbf{z}| \mathbf{Z_0^{-1}} |\mathbf{z}\rangle > 0 \rightarrow z_{QQ} > \sum_K \sum_L z_{I_K Q} z_{I_L Q} Z^{(-1)}_{0;I_K I_L}. \quad (10)$$

The Cholesky decomposition condition, which can be called *Quantum Similarity Aufbau Condition* (QSAC), means that it cannot be reliable to use a pair of QOs every time that a new element of the QSM has to be computed, but that the added QO density function: $\rho_Q\,(\mathbf{r}\,|\mathbf{t};\,\boldsymbol{\Omega}\,)$ has to be translated-rotated with the same values of the pair: $\{\mathbf{t};\,\boldsymbol{\Omega}\}$, for every computed element of the vector $|\mathbf{z}\rangle$, connecting recursively the QO Q with all the ones previously employed in constructing the QS submatrix $\mathbf{Z_0}$. When the QS submatrix $\mathbf{Z_0}$ has scalar (1×1) dimension as occurs in the submatrix (3) case, then the QSAC (10) becomes the relationship (5). Moreover the QSAC condition is a stronger positive definiteness condition than the diagonal dominance, as QSAC becomes the necessary and sufficient condition for constructing a positive definite augmented matrix.

The maximal pair condition (2) can be substituted in the general $(P \times P)$ case, for instance, by maximizing the sum of the whole vector$|\mathbf{z}\rangle$, which due to the positive definiteness of its elements is coincident with the search of a maximal Minkowski norm:

$$\max_{\mathbf{t};\boldsymbol{\Omega}}\left[\langle|\mathbf{z}\rangle\rangle\right] = \max_{\mathbf{t};\boldsymbol{\Omega}}\left[\sum_{K=1}^{P}\int_{D}\rho_{I_K}\,(\mathbf{r})\,\rho_Q\,(\mathbf{r}\,|\mathbf{t};\,\boldsymbol{\Omega}\,)d\mathbf{r}\right] = \max_{\mathbf{t};\boldsymbol{\Omega}}\left[\sum_{K=1}^{P} z_{I_K\,Q}\right]. \quad (11)$$

This can be done admitting the same translation–rotation sequence performed on every term of the vector $|\mathbf{z}\rangle$ in Eq. 11, whenever such transformation increases the Minkowski norm.

However, while the maximal value of the sum leading to the Gershgorin radius is searched as in the condition (11) of the previous sentence, the QSAC relationship (10) has to be equally tested and if not fulfilled the pair $\{\mathbf{t};\,\boldsymbol{\Omega}\}$ rejected.

Such a procedure will assure the positive definiteness of the QSM $\mathbf{Z}$ at the final step of the recursion and will provide the same relative position in the calculation of the quantum similarity measures for every recursively added QO.

Geometrical interpretation of the QSAC

Leaving apart the linear algebra concept of diagonal dominance, expressed as equation (8), which similarity matrices usually do not fulfill, the alternative Cholesky decomposition condition property leading to the QSAC, assuring in this manner the positive definite structure of the final QSM form and written as in equation (10), has a clear geometrical meaning. The positive definite quadratic form:$\langle\mathbf{z}|\,\mathbf{Z_0^{-1}}\,|\mathbf{z}\rangle \in \mathbf{R}^+$, is nothing else than the Euclidian norm of the vector $|\mathbf{z}\rangle$ in the reciprocal metric space defined by the density tags: $\left\{\rho_{I_K}\,(\mathbf{r})\,;\,K=1,\,P\right\}$. The QOs tags are employed to form the QSM $\mathbf{Z_0}$, which because of the QSAC construction has been structured positive definite and acts accordingly as a metric matrix of a P-dimensional pre-Hilbert space. Since in the quadratic form appearing in equation (10), the inverse of the metric $\mathbf{Z_0}$ appears, the implicit Euclidian norm equivalent to the aforementioned quadratic form is computed in the metric reciprocal space with the matrix $\mathbf{Z}_0^{-1}$, acting as a positive definite metric matrix, because: $Sp\,[\mathbf{Z_0}] \in \mathbf{R}^+ \rightarrow Sp\left[\mathbf{Z}_0^{-1}\right] \in \mathbf{R}^+$. Accordingly, the QSAC forces

this Euclidian norm in the reciprocal P-dimensional pre-Hilbert space to be less than the self-similarity of the recursively added Q-th QO.

This permits to associate the described Quantum Similarity Aufbau procedure as an algorithm maximizing the Minkowski norm of each recursive column $|\mathbf{z}\rangle$ of the QSM, submitted to the QSAC restriction consisting in that its Euclidian norm, computed in the recursive reciprocal pre-Hilbert space, remains less than the recursive QSM diagonal self-similarity elements.

Final remarks

1) Due that it is not necessary to start the recursive QSAC with any *a priori* chosen QO, the final QSM will certainly depend on the QO recursive order chosen. Thus, there are just $N!$ possible choices, producing each one an equally positive definite QSM. However, the ordering imposed by the self-similarity measures can be chosen as a way to reach a systematic QSM Aufbau. That is, if one calls the QSM diagonal selfsimilarity measures set computed on the QOS elements: $D(\mathbf{Z}) = \{z_{II}\}$, then the obvious choices are defined by the maximal ordering:

$$z_{11} = \max_I [D(\mathbf{Z})] \rightarrow z_{22} = \max_I [D(\mathbf{Z}) - z_{11}] \ldots$$

or by the minimal:

$$z_{11} = \min_I [D(\mathbf{Z})] \rightarrow z_{22} = \min_I [D(\mathbf{Z}) - z_{11}] \ldots$$

This ensures that the QOs will be ordered in decreasing or increasing complexity, while providing with a generic reproducible way of computing QSM under QSAC premises.

2) When constructing the QSM according to the proposed Aufbau procedure, it is well known that the overlap quantum similarity measures, as defined in Eq. (1), can be substituted by a more general form [10], involving a positive definite operator: $\Omega(\mathbf{r_1}; \mathbf{r_2})$; so, in general, the similarity measures can be described as the integral:

$$z_{IJ}(\Omega) = \int \int_D \rho_I(\mathbf{r_1}) \, \Omega(\mathbf{r_1}; \mathbf{r}_2) \, \rho_J(\mathbf{r}_2) \, d\mathbf{r_1} d\mathbf{r}_2;$$

while the positive definite operator choice ensures that the QSM, when constructed according the equivalent QSAC, like the one depicted previously for overlap quantum similarity measures in Eq. (10), is positive definite. One just shall make the substitution: $z_{IJ} \leftarrow z_{IJ}(\Omega)$.

3) The QSAC is also valid for quantum similarity measures involving the off-diagonal terms of the density matrix [11].

Conclusions

An Aufbau procedure subject to a so called Quantum Similarity Aufbau Condition (QSAC) in order to construct positive definite QSM has been described. It consists in the maximization of the Minkowski norm of the recursively constructed QSM columns, while keeping the Cholesky decomposition diagonal elements real, as a restriction.

Acknowledgment The author expresses his acknowledgment to the Ministerio de Ciencia y Tecnología for the grants: BQU2003-07420-C05-01 and CTQ2006-04410/BQU, which have partially sponsored this work, and also for a Salvador de Madariaga fellowship reference: PR2004-0547, which has permitted his stage at Ghent University during the 2004–2006 period. Discussions with Prof. P. Bultinck, Ghent University, on the QQSPR problem have lead to the present algorithm description. The help of Ms. S. Van Damme in order to perform some elementary numerical tests is highly appreciated.

References

1. (A) 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. *Advances in Molecular Similarity*, vol. 2, ed. by R. Carbó-Dorca, P.G. Mezey (JAI Press, London, 1998) pp. 43–72; (B) R. Carbó-Dorca, Mathematical Elements of Quantum Electronic Density Functions. Adva. Quant. Chemi. **49**, 121–207 (2005)
2. R. Carbó, E. Besalú, A general survey of molecular quantum similarity. J. Mol. Struct. Theochem. **451**, 11–23 (1998)
3. R. Carbó, M. Arnau, L. Leyda, How similar is a molecule to another? Int. J. Quant. Chem. **17**, 1185–1189 (1980)
4. P. Bultinck, X. Gironès, R. Carbó-Dorca, Molecular quantum similarity: theory and applications. Reviews in Computational Chemistry vol. **21**, ed. by K.B. Lipkowitz, R. Larter, T. Cundari (John Wiley & Sons, Hoboken, USA, 2005), pp. 127–207
5. R. Carbó-Dorca, D. Robert, Ll. Amat, X. Gironés, E. Besalú, Molecular Quantum Similarity in QSAR and Drug Design. *Lecture Notes in Chemistry*, vol. 73. (Springer Verlag, Berlin, 2000)
6. S.K. Berberian, *Introduction to Hilbert Space* (Oxford University Press, New York, 1961)
7. P. Bultinck, R. Carbó-Dorca, Molecular quantum similarity matrix based in clustering of molecules using dendrograms. J. Chem. Inf. Comp. Sci. **43**, 170–177 (2003)
8. R.S. Varga, *Gershgorin and His Circles* in Springer Series in Computational Mathematics, 36 (Springer, Berlin, 2004)
9. R. Carbó, E. Besalú, Many Center AO Integral Evaluation Using Cartesian Exponential Type Orbitals (CETO's). Adv. Quan. Chem. **24**, 115–237 (1992)
10. R. Carbó, B. Calabuig, L. Vera, E. Besalú, Molecular Quantum Similarity: Theoretical Framework, Ordering Principles, and Visualization Techniques. Adv. Quan. Chem. **25**, 253–313 (1994)
11. J. Cioslowski, E.D. Fleishmann, Assessing molecular similarity from results of ab initio electronic structure calculation. J. Am. Chem. Soc. **113**, 64–67 (1991)