

Quantum polyhedra, definitions, statistics and the construction of a collective quantum similarity index

Ramon Carbó-Dorca

Received: 30 July 2014 / Accepted: 23 September 2014 / Published online: 2 October 2014
© Springer International Publishing Switzerland 2014

Abstract Quantum polyhedra (QP) are geometrical constructs whose vertices are made by quantum mechanical density functions (DF). In this paper a QP centroid and a variance are defined at two levels: functional and numerical. The numerical QP variance can be shown associated to a collective QP squared distance involving the whole DF set composing the QP vertices. In this manner, a global dissimilarity index corresponding to the set of QP vertices can be defined. Extension of the mathematical and computational techniques developed on QP to shape functions polyhedra and to classical descriptor N-dimensional multimolecular polyhedra, are also discussed.

Keywords Quantum object sets (QOS) · Quantum polyhedra (QP) · Quantum similarity · Density functions (DF) · DF Origin Shifts · QP variance · Generalized collective distances · Shape functions (ShF) · Multimolecular polyhedra (MP)

1 Introduction

A quantum object set (QOS), see for example reference [1], can be constructed by the Cartesian product: $Q = M \times P$ of a well-defined submicroscopic object set: $M = \{m_I | I = 1, N\}$ and a tag set made by quantum density functions (DF): $P = \{\rho_I | I = 1, N\}$.

Then, a quantum polyhedron (QP) can be defined in some functional vector space as a geometrical construct, made by means of the tag set P of some QOS. The corresponding cardinality of the QOS : N , is coincident with the number of vertices of the QP.

R. Carbó-Dorca (✉)

Institut de Química Computacional i Catàlisi, Universitat de Girona, 17071 Girona, Catalonia, Spain
e-mail: quantumqsar@hotmail.com

Recently, collective Euclidian distances have been studied [2,3] in connection with quantum similarity theoretical foundations. Also, recent studies performed on the similarity relations between the vertices of QP, provide the possibility to uniformize any of such DF sets, acting as QP vertices, via an Origin Shift (OS) with respect any vertex or convex linear combination of them [4]. Several publications describe, the general features and properties of such an operation [5,6].

On the other hand, preliminary studies [8,9] have conducted towards the application of the QP OS, which has been useful to set up an efficient quantum QSPR (QQSPR) procedure [10]. Some interesting properties about QP OS have been obtained and recently published [7].

Based on the above definitions and the previous literature, the aim of the present work is the development of another set of characteristic properties related to QP and their connection with straightforward statistical algorithms.

The main result of the present discussion can be resumed with a surprisingly simple definition of a global dissimilarity index, related to a collective distance feature existing within QP. A similar question constituted a must, which was initially discussed in several previous papers [11–13], but taking into account a complete different perspective than the one proposed here in the present study.

Owing to the above proposals, QP computational structures will be first studied over origin shifted DF vertices, to be subsequently extended over shape functions and finally to classically N-dimensional descriptor defined multimolecular polyhedra (MP).

2 Quantum polyhedra characteristic functions

2.1 Centroid

The simpler and general convex linear combination of the vertices belonging to any QP is the centroid, which can be directly expressed as:

$$\rho_C = N^{-1} \sum_I \rho_I. \quad (1)$$

Such a trivial definition acts as an arithmetic mean of the involved DF set forming the QP. As it has been said before, the centroid, among other applications, has been employed to origin shifting QP for QQSPR purposes [10].

One must note now that the primary mathematical characteristic of the DF tag set elements, acting as QP vertices, is essentially their positive definite nature. The centroid definition (1) can be taken as a convex linear combination of the QP vertices, where all the coefficients are equal to the inverse of the number of vertices. Therefore, the centroid function inherits in this way the character of being positive definite. Besides, the centroid function can be considered as the point in infinite dimensional space, for which the sum of squared Euclidian distances to all the vertices is minimal.

2.2 Variance

Now, from such intuitive construction of the QP centroid, it is easy to design the function equivalent to the variance attached to the elements of the QP, by means of the expression:

$$v_C = N^{-1} \sum_I (\rho_I - \rho_C)^2 = \rho_C^{(2)} - \rho_C^2, \tag{2}$$

where the function,

$$\rho_C^{(2)} = N^{-1} \sum_I \rho_I^2$$

represents the DF set squared elements centroid.

2.3 Some remarks

The expression (2) above coincides with the well-known algorithm attached to the variance of a set of discrete scalar values, except that in this QP case, the involved elements in the variance computation are themselves DF. Thus, defined in this way, the QP variance also appears as a positive definite function.

As will be discussed later on, centroid and variance functions as previously defined here, can produce at the end numerical values, which are obviously related with the usual mean and variance of a set of values of a random variable, as they are usually defined in statistical lore, but have not to be confused with them as in the present paper the source are functions.

While the centroid can be associated to a central point within the QP, thus to some kind of mean value, the numerical QP variance described below acquires the general definition of some squared Euclidian distance involving an indefinite number of QO.

Both QP centroid and variance functions are attached to the same number of variables which forms the QP vertex DF set. They belong in this way to the same functional vector space subtended by the DF set.

3 Numerical QP arithmetic mean and variance

In the same way as the pair of QP characteristic functions: $\{\rho_C; v_C\}$ as above defined, there seems also interesting to know the positive scalars provided by the Minkowski norms of both.

3.1 Centroid Minkowski norm as an arithmetic mean

That is, in one hand the Minkowski norm of the centroid DF yields:

$$\langle \rho_C \rangle = N^{-1} \sum_I \langle \rho_I \rangle = N^{-1} \sum_I v_I = v_C, \tag{3}$$

provided that the number of particles of every QO is described by the number set: $\{v_I | I = 1, N\}$. Thus, v_C is the average number of particles involved in the DF of the QO's contained in the QOS, associated in turn with the QP vertices.

When the QO's are molecules, then ν_C corresponds to the average number of electrons of all the involved structures attached to the DF set of vertices of the QP.

3.2 Minkowski norm of the variance as a collective QP index

On the other hand, the Minkowski norm of the QP variance, provides the sequence of equalities:

$$\begin{aligned}
 \langle \nu_C \rangle &= \langle \rho_C^{(2)} \rangle - \langle \rho_C^2 \rangle \\
 &= N^{-1} \sum_I \langle \rho_I^2 \rangle - N^{-2} \left\langle \left(\sum_I \rho_I \right)^2 \right\rangle \\
 &= N^{-1} \sum_I \langle \rho_I \rho_I \rangle - N^{-2} \sum_I \sum_J \langle \rho_I \rho_J \rangle \\
 &= N^{-1} \sum_I Z_{II} - N^{-2} \sum_I \sum_J Z_{IJ} \\
 &= N^{-1} \text{Tr}(\mathbf{Z}) - N^{-2} \langle \mathbf{Z} \rangle
 \end{aligned} \tag{4}$$

where use has been made of the similarity matrix \mathbf{Z} definition, see for example [14–16], attached to the DF vertex set forming the QP:

$$\mathbf{Z} = \{Z_{IJ} | I, J = 1, N\} \leftarrow \forall I, J : Z_{IJ} = \langle \rho_I \rho_J \rangle = \int_D \rho_I(\mathbf{r}) \rho_J(\mathbf{r}) d\mathbf{r}.$$

Moreover, as it is well-known, one can write:

$$\text{Tr}(\mathbf{Z}) = \sum_I Z_{II} = \langle \text{Diag}(\mathbf{Z}) \rangle.$$

Then, $\langle \nu_C \rangle$ the numerical variance of the QP, corresponds to the arithmetic mean of the QOS self-similarities minus the arithmetic mean of the similarity matrix. By definition one can certainly write: $\langle \nu_C \rangle \geq 0$.

The magnitude of this QP numeric variance will indicate the generic similarity-dissimilarity between the whole QP vertices. In fact, taking into account that it can be also defined: $\text{Offdiag}(\mathbf{Z}) = \mathbf{Z} - \text{Diag}(\mathbf{Z})$, as the original similarity matrix but provided with a zero diagonal. Also, due to the symmetric nature of the similarity matrix, it can be used the algorithm:

$$\langle \text{Offdiag}(\mathbf{Z}) \rangle = 2 \sum_I \sum_{J>I} Z_{IJ}.$$

Therefore, the QP numerical variance can be also written as:

$$\begin{aligned} \langle v_C \rangle &= N^{-1} \langle \text{Diag}(\mathbf{Z}) \rangle - N^{-2} \langle \text{Offdiag}(\mathbf{Z}) + \text{Diag}(\mathbf{Z}) \rangle \\ &= N^{-1} (1 - N^{-1}) \langle \text{Diag}(\mathbf{Z}) \rangle - N^{-2} \langle \text{Offdiag}(\mathbf{Z}) \rangle \end{aligned}$$

From this result one can argue that the numerical value of the QP variance measures the balance between self-similarities: $\{Z_{II}\}$ and pair similarities: $\{Z_{IJ} | I \neq J\}$, associated with the different vertices of the QP.

Besides, as a consequence of the fact that the resultant numerical QP variance value must be non-negative definite, a relationship can be written, which must be fulfilled by any similarity matrix and which can be written as:

$$\begin{aligned} (1 - N^{-1}) \langle \text{Diag}(\mathbf{Z}) \rangle &\geq N^{-1} \langle \text{Offdiag}(\mathbf{Z}) \rangle \\ \rightarrow (N - 1) \langle \text{Diag}(\mathbf{Z}) \rangle &\geq \langle \text{Offdiag}(\mathbf{Z}) \rangle \end{aligned}$$

A value of $\langle v_C \rangle$ nearby zero will indicate a large similarity between the whole QP vertex pairs. While larger the variance, greater the dissimilarity between the vertices. That is: between the QO DF forming the QP.

4 Examples

It is worthwhile to discuss several simple examples in order to grasp the nature and interest of the QP numerical variance.

4.1 Two QO case

In a QP just made by two QO, the similarity matrix and the variance can be written with the general formalism:

$$\mathbf{Z} = \begin{pmatrix} \alpha & \gamma \\ \gamma & \beta \end{pmatrix} \rightarrow \langle v_C \rangle = \frac{1}{2} \left[\frac{1}{2} (\alpha + \beta) - \gamma \right] \leftarrow N = 2$$

It is obvious that a null variance value will coincide with the fact that both involved QO are exactly the same and because in this case: $\alpha = \beta = \gamma$. Moreover, in this simple case nullity of the variance might also happen when: $\alpha + \beta = 2\gamma$, though. Also a greater value of the variance will be associated to the fact that the arithmetic average of self-similarities becomes greater than the similarity between both QO, that is, when the inequality: $\frac{1}{2}(\alpha + \beta) > \gamma$, holds.

In general, this also precludes a new distance-like similarity index between any pair of QO: $\{m_I, m_J\}$, say, as one can write in this circumstance the following expression:

$$\langle v_{IJ} \rangle = \frac{1}{2} \left[\frac{1}{2} (Z_{II} + Z_{JJ}) - Z_{IJ} \right].$$

Written as above, this index can be compared with the squared Euclidian Distance between a pair of QO, which could be written as:

$$D_{IJ}^2 = Z_{II} + Z_{JJ} - 2Z_{IJ}.$$

Therefore, comparing both expressions it might be also written:

$$\langle v_{IJ} \rangle = \frac{1}{4} D_{IJ}^2.$$

Resulting into that the generalized QP numeric variance will become a trivial scaled Euclidian distance definition, when considering two QO only.

4.2 Three QO case

After the previous result for two QO, one must take into account that the numerical variance of a QP structure can be extended to any number of QO. For instance, a set of three QO will have the following numerical variance:

$$\langle v_{IJK} \rangle = \frac{2}{9} [(Z_{II} + Z_{JJ} + Z_{KK}) - (Z_{IJ} + Z_{IK} + Z_{JK})],$$

owing to the symmetric structure of the similarity matrix \mathbf{Z} .

To have another particular point of view on how the multiple distance QP index will behave, just suppose that the three QO are the same, then all the similarity integrals will become equal and obviously $\langle v_{IJK} \rangle = 0$.

Moreover, imagine now that only two QO are the same, for instance:

$$m_I = m_J \rightarrow Z_{II} = Z_{JJ} = Z_{IJ} \wedge Z_{IK} = Z_{JK}$$

then

$$\begin{aligned} \langle v_{IJK} \rangle &= \frac{2}{9} [(2Z_{II} + Z_{KK}) - (Z_{II} + 2Z_{IK})] \\ &= \frac{2}{9} [(Z_{II} + Z_{KK}) - 2Z_{IK}] = \frac{2}{9} D_{IK}^2 = \frac{2}{9} D_{JK}^2 \end{aligned}$$

providing a result coherent with the fact that only two different QO are present, and obtaining an expression associated to the previous two QO case, where the numerical variance corresponds to a scaled Euclidian distance.

4.3 Conclusion

At the light of the previous simpler examples it must be stressed again the fact that such an index, defined over a similarity matrix \mathbf{Z} , attached to the tag set of a QOS

forming a QP, by means of the difference of two mean values involving the similarity matrix elements:

$$\langle v_C \rangle = N^{-1} \text{Tr}(\mathbf{Z}) - N^{-2} \langle \mathbf{Z} \rangle, \quad (5)$$

corresponds to a *generalized squared Euclidian collective distance* between any arbitrarily large number of QO.

5 Origin Shifted (OS) DF set and variance

The QP variance function definition can be also constructed in the following manner. Take into account that in any QP it can be constructed an OS DF set in the following way [4,5]:

$$\forall I : \xi_I = \rho_I - \rho_C \rightarrow S = \{\xi_I | I = 1, N\}.$$

Therefore, the QP variance can be now easily rewritten as the mean function of the squares of the OS DF:

$$v_C = N^{-1} \sum_I |\xi_I|^2$$

and the previously performed numerical analysis can be effortlessly described as:

$$\langle v_C \rangle = N^{-1} \sum_I \langle |\xi_I|^2 \rangle.$$

Besides that, the Minkowski norms of the squared OS DF integrals averaged can be interpreted as the Euclidian norms of the OS DF set S , as one can write:

$$\forall I : \langle |\xi_I|^2 \rangle = \int_D |\xi_I(\mathbf{r})|^2 d\mathbf{r} = \langle \xi_I | \xi_I \rangle.$$

However, taking into account the definition of the OS DF set, the resultant final expression can be transformed into the already deduced one.

Contrarily to the non-negative definite structure of the DF set, the OS DF elements possess a non-definite structure, but obviously enough their squared modules behave as a non-negative set of functions.

The role of the OS DF set S is important in the definition of an efficient algorithm for QQSPR purposes [10].

6 Positive definite weighted variance

When variance is integrated in order to obtain a collective distance index, involving the whole set of a QP vertices, as previously commented, its definition can be generalized just weighting the involved integrals with a positive definite operator, see for example [14–16]. This prospect might be easily made, for instance using the generalized definition of the involved integrals employing some positive definite operator: $\Omega(\mathbf{r}_1, \mathbf{r}_2)$,

$$\forall I : \langle |\xi_I|^2 \rangle \rightarrow \langle \Omega |\xi_I|^2 \rangle \equiv \int_D \int_D \xi_I(\mathbf{r}_1) \Omega(\mathbf{r}_1, \mathbf{r}_2) \xi_I(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2.$$

Using this generalized definition, the variance of a QP and thus the generalized distance between their vertices are easily set up.

7 Shape functions and QP variance

Another question might be the construction of a Shape QP (SQP). It can be easily made starting with a set of shape functions (ShF) as vertices, substituting the QP DF set of vertices.

The set of SQP vertices can be now described as the ShF set: $\Sigma = \{\sigma_I | I = 1, N\}$. The ShF set can be straightforwardly obtained taking into account the number of particles attached to every DF: $\{v_I | I = 1, N\}$ as defined in Eq. (3). Then, one can easily write the ShF set as the set of function elements everyone having a unit Minkowski norm. Such a set can be obtained from the DF tag set: $P = \{\rho_I | I = 1, N\}$, defined at the beginning, in the following way:

$$\forall I : \sigma_I = v_I^{-1} \rho_I \rightarrow \langle \sigma_I \rangle = v_I^{-1} \langle \rho_I \rangle = v_I^{-1} v_I = 1$$

Obviously enough, the centroid of the SQP will be defined in turn as the shape function:

$$\sigma_C = N^{-1} \sum_I \sigma_I \rightarrow \langle \sigma_C \rangle = N^{-1} \sum_I \langle \sigma_I \rangle = 1.$$

One can be interested on how different both QP and SQP centroid functions are. For instance, the cosine of the angle between them can be easily computed, provided that a N -dimensional (ND) column vector with the inverses of the number of particles, like: $|\mathbf{v}\rangle = \{v_I^{-1}\}$ is defined, then the squared cosine between both centroids can be written as:

$$r^2 = \frac{\langle \rho_C \sigma_C \rangle^2}{\langle \rho_C \rho_C \rangle \langle \sigma_C \sigma_C \rangle} = \frac{\left(\sum_I \sum_J v_J^{-1} Z_{IJ} \right)^2}{\left(\sum_I \sum_J Z_{IJ} \right) \left(\sum_I \sum_J v_I^{-1} v_J^{-1} Z_{IJ} \right)} = \frac{\langle \mathbf{Z} | \mathbf{v} \rangle^2}{\langle \mathbf{Z} | \langle \mathbf{v} | \mathbf{Z} | \mathbf{v} \rangle}.$$

From this result it is interesting to note that, as: $r^2 \leq 1$; then, the following inequality will always hold:

$$\langle \mathbf{Z} | \mathbf{v} \rangle^2 \leq \langle \mathbf{Z} | \langle \mathbf{v} | \mathbf{Z} | \mathbf{v} \rangle.$$

A squared Euclidian distance between both centroids can be also written as:

$$D^2 = \langle \rho_C \rho_C \rangle + \langle \sigma_C \sigma_C \rangle - 2 \langle \rho_C \sigma_C \rangle = N^{-2} (\langle \mathbf{Z} \rangle + \langle \mathbf{v} | \mathbf{Z} | \mathbf{v} \rangle - 2 \langle \mathbf{Z} | \mathbf{v} \rangle)$$

which provides another relationship which holds for any similarity matrix:

$$\langle \mathbf{Z} \rangle + \langle \mathbf{v} | \mathbf{Z} | \mathbf{v} \rangle \geq 2 \langle \mathbf{Z} | \mathbf{v} \rangle.$$

The associated variance function of any SQP can be written now in this context as:

$$v_S = N^{-1} \sum_I (\sigma_I - \sigma_C)^2 = N^{-1} \sum_I \sigma_I^2 - \sigma_C^2 = \sigma_C^{(2)} - \sigma_C^2 \tag{6}$$

where: $\sigma_C^{(2)} = N^{-1} \sum_I \sigma_I^2$.

The numerical SQP variance is easily written as the former DF one was, just obtaining the Minkowski norm of Eq. (6):

$$\langle v_S \rangle = N^{-1} \sum_I \langle \sigma_I^2 \rangle - \sigma_C^2 = N^{-1} \sum_I S_{II} - N^{-2} \sum_I \sum_J S_{IJ} \tag{7}$$

where use is now made of the definition of the similarity shape matrix $\mathbf{S} = \{S_{IJ}\}$:

$$\forall I, J : S_{IJ} = \langle \sigma_I \sigma_J \rangle = \int_D \sigma_I(\mathbf{r}) \sigma_J(\mathbf{r}) d\mathbf{r},$$

which can also be written as the matrix inward product [16]:

$$\mathbf{S} = |\mathbf{v}\rangle \langle \mathbf{v}| * \mathbf{Z}.$$

Within such a definition one constructs first the tensor product:

$$|\mathbf{v}\rangle \langle \mathbf{v}| = \left\{ [|\mathbf{v}\rangle \langle \mathbf{v}|]_{IJ} = v_I^{-1} v_J^{-1} \right\},$$

then the inward product is simply performed as the products of the elements of the involved matrices, in such a way that:

$$\forall I, J : S_{IJ} = [|\mathbf{v}\rangle \langle \mathbf{v}|]_{IJ} Z_{IJ} = v_I^{-1} v_J^{-1} Z_{IJ}.$$

Therefore, the expression of the Eq. (7) result can be also written compactly as:

$$\langle v_S \rangle = N^{-1} Tr(\mathbf{S}) - N^{-2} \langle \mathbf{S} \rangle, \tag{8}$$

which constitutes an equivalent result as the one obtained in the QP case. This equivalence can be made self-evident, when comparing the Eq. (8) above with formerly described Eq. (5).

The numerical squared Euclidian distance of the shape centroid and any vertex can be written as:

$$\begin{aligned} \forall I : D_{IC}^2 &= \langle (\sigma_I - \sigma_C)^2 \rangle = \langle \sigma_I^2 \rangle + \langle \sigma_C^2 \rangle - 2 \langle \sigma_I \sigma_C \rangle \\ &= S_{II} + N^{-2} \langle \mathbf{S} \rangle - N^{-1} \langle |s_I\rangle \rangle \end{aligned}$$

where $\{|s_I\rangle | I = 1, N\}$ is the set of the columns of the shape similarity matrix \mathbf{S} .

8 Discrete multimolecular polyhedra or point clouds

Recently, some discussion has been associated to Multimolecular Polyhedra (MP) [7]. They are geometrical constructs whose vertices are made not with quantum mechanical DF or the related ShF, but with discrete N-Dimensional (ND) vectors, constructed by N ordered parameters. In previous papers, see for example [1], MP were also named as point clouds.

The term MP is preferable than point cloud though, as it distinguishes quite clearly those polyhedra from the subject of the present paper: QP, while all of them can be considered as point clouds, as both MP and QP are related mathematical objects.

The relationship between them can be obtained easily from the similarity matrices directly obtained from the DF or ShF vertices of QP or SQP respectively. Indeed, it has been earlier discussed in several places [17, 18] the role of the columns of similarity matrices as projections associated to the QP functional vertex sets.

Even so, there is a large class of MP which do not need at all to have a quantum molecular origin. This is the case of MP whose vertices are just defined classically with ND vectors, whose elements are empirically and arbitrarily made with the so-called molecular descriptors, which are parameters extracted from diverse origins, even containing experimental values. In fact, any kind of MP, can be easily considered as the basis of QSPR or QSAR, see for example [8].

A classical MP can be defined by means of a set of ND vectors: $X = \{|\mathbf{x}_I\rangle | I = 1, N\}$ which act as vertices of the polyhedron. Then it is trivial to define the MP centroid vector as:

$$|\mathbf{x}_C\rangle = N^{-1} \sum_I |\mathbf{x}_I\rangle$$

but the variance vector, which could be equivalent to Eqs. (2) and (6) in the functional QP or SQP spaces respectively, it is not so trivial to define. To do so, first the inward product of two vectors, for more details see for example reference [16], shall be defined in a similar manner as it has been done before with two matrices when dealing with SQP variance:

$$|\mathbf{p}\rangle = |\mathbf{a}\rangle * |\mathbf{b}\rangle \rightarrow \forall I : p_I = a_I b_I.$$

It is easy to see that the inward product is symmetric. Also, when considering inward product of vectors, an interesting characteristic consists into the fact that the sum of the elements of the inward product vector is coincident with the scalar product of the involved vectors:

$$\langle |\mathbf{p}\rangle \rangle = \sum_I p_I = \langle \mathbf{a} * \mathbf{b} \rangle = \sum_I a_I b_I = \langle \mathbf{a} | \mathbf{b} \rangle.$$

With this in mind, one can write a variance vector of a MP, which will be equivalent to the variance functions formerly described in QP. It is as simple as to write the variance vector as a sum of inward products, by means of the expression:

$$|\mathbf{v}_X\rangle = N^{-1} \sum_I (|\mathbf{x}_I\rangle - |\mathbf{x}_C\rangle) * (|\mathbf{x}_I\rangle - |\mathbf{x}_C\rangle) = \left| \mathbf{x}^{[2]} \right\rangle - |\mathbf{x}_C\rangle^{[2]}. \quad (9)$$

Equation (9) is easily obtained taking into account that the inward product is also distributive with respect vector addition. The supplementary definition has been also employed:

$$|\mathbf{x}^{[2]}\rangle = N^{-1} \sum_I (|\mathbf{x}_I\rangle * |\mathbf{x}_I\rangle) = N^{-1} \sum_I |\mathbf{x}_I\rangle^{[2]}$$

and accordingly one can also write:

$$|\mathbf{x}_C\rangle^{[2]} = |\mathbf{x}_C\rangle * |\mathbf{x}_C\rangle.$$

Moreover, within MP the numerical variance might be written as the complete sum of the elements of both terms appearing into the vector variance, as defined in Eq. (9), that is:

$$\begin{aligned} v_X &= \langle |\mathbf{v}_X\rangle \rangle = \langle \langle |\mathbf{x}^{[2]}\rangle \rangle \rangle - \langle |\mathbf{x}_C\rangle^{[2]} \rangle = N^{-1} \sum_I \langle \mathbf{x}_I | \mathbf{x}_I \rangle - \langle \mathbf{x}_C | \mathbf{x}_C \rangle \\ &= N^{-1} \sum_I \langle \mathbf{x}_I | \mathbf{x}_I \rangle - N^{-2} \sum_I \sum_J \langle \mathbf{x}_I | \mathbf{x}_J \rangle. \end{aligned}$$

Defining now the Gram matrix of the ND vertex set of the MP, as the ordered scalar products of its vertices:

$$\mathbf{X} = \{X_{IJ} = \langle \mathbf{x}_I | \mathbf{x}_J \rangle\},$$

then the numerical equivalent of the QP and SQP variances, as defined in Eqs. (4) and (8), which can be now constructed in classical MP structures, can be written without problems as:

$$v_X = N^{-1} Tr(\mathbf{X}) - N^{-2} \langle \mathbf{X} \rangle.$$

Therefore, taking the Gram matrix of the MP vertices as a ND version of the similarity matrices in DF or ShF spaces, all the considerations developed in the two previous functional QP and SQP cases are applicable in discrete ND classical MP.

9 Discussion

A QP has been defined as a set of DF associated to a set of QO, taken for instance as molecules. A simple definition of the vertex variance function of a QP leads to the numerical evaluation of the QP numerical variance, which can be set up in terms of the elements associated to any similarity matrix computed using the DF set.

It has been shown that such a QP numerical variance measures a global degree of dissimilarity between the *whole* set of QP DF vertices. It can be considered in general as a collective squared Euclidian distance between these QP vertices.

The same concept is valid in SQP, that is: QP defined over ShF instead of DF. Finally, the coherence of the mathematical and computational procedures appears when one realizes that the developed definitions and the derived algorithms, previously described here for QP and SQP, are also easily extensible to MP. Taking into account that MP are classical polyhedral structures made of ND discrete vectors, whose elements are molecular descriptors of any origin.

The whole set of results indicates that a collective distance can be in general easily described in metric vector spaces.

It has not been discussed the QP nature of the MO components of the DF in LCAO MO theory. It might be worth of a separate discussion, which will be done elsewhere.

References

1. R. Carbó-Dorca, E. Besalú, J. Math. Chem. **50**, 210–219 (2012)
2. R. Carbó-Dorca, J. Math. Chem. **51**, 338–353 (2013)
3. R. Carbó-Dorca, J. Math. Chem. **49**, 2109–2115 (2011)
4. R. Carbó-Dorca, E. Besalú, J. Math. Chem. **50**, 1161–1178 (2012)
5. R. Carbó-Dorca, J. Math. Chem. **51**, 289–296 (2013)
6. R. Carbó-Dorca, J. Math. Chem. **51**, 1397–1409 (2013)
7. R. Carbó-Dorca, J. Math. Chem. **52**, 1848–1856 (2014)
8. R. Carbó-Dorca, A. Gallegos, Á.J. Sánchez, J. Comput. Chem. **30**, 1146–1159 (2008)
9. R. Carbó-Dorca, A. Gallegos, J. Comput. Chem. **30**, 2099–2104 (2009)
10. R. Carbó-Dorca, J. Comput. Chem. **34**, 766–779 (2013)
11. R. Carbó-Dorca, J. Math. Chem. **44**, 628–636 (2008)
12. R. Carbó-Dorca, J. Math. Chem. **47**, 331–334 (2010)
13. R. Carbó-Dorca, J. Math. Chem. **50**, 734–740 (2012)
14. R. Carbó-Dorca, Quantum similarity, Chapter 17, in *Concepts and Methods in Modern Theoretical Chemistry*, ed. by S.K. Ghosh, P.K. Chattaraj (CRC Press Taylor & Francis, Boca Raton, 2013), pp. 349–365
15. P. Bultinck, X. Gironés, R. Carbó-Dorca, Molecular quantum similarity: theory and applications, in *Reviews in Computational Chemistry*, ed. by K.B. Lipkowitz, R. Larter, T. Cundari (Wiley, Hoboken, 2005), pp. 127–207
16. R. Carbó, E. Besalú, Ll Amat, X. Fradera, Quantum molecular similarity measures: concepts, definitions and applications, in *Proceedings of the 2nd. Girona Seminar on Molecular Similarity. July 1995. Advances in Molecular Similarity*, ed. by R. Carbó, P.G. Mezey (JAI Press Inc., Greenwich (Conn.), 1996), pp. 1–42
17. L.D. Mercado, R. Carbó-Dorca, J. Math. Chem. **49**, 1558–1572 (2011)
18. R. Carbó-Dorca, J. Math. Chem. **50**, 2339–2341 (2012)