Journal of Mathematical Chemistry 23 (1998) 353–364

Tagged sets, convex sets and quantum similarity measures

Ramon Carbó-Dorca

Institute of Computational Chemistry, University of Girona, Girona, Spain

Received 19 June 1997

Tagged and convex sets concepts and definitions are applied with the aim to discover a general mathematical pattern enveloping the quantum similarity measures framework. As a consequence, several aspects of the quantum similarity theoretical structure become beautifully related to a mathematical construction, adopting the form of some interwoven essential formalism, connecting quantum theory, molecular similarity, convexity and tagged sets.

1. Introduction

Quantum Similarity Measures (QSM) in the first place [4] and *Tagged and Convex Sets* (CS) in the second place [23], constitute two interestingly interconnected fields. The relationship has to be found in the characteristic definitions, which have been developed with time in the realm of QSM [6,7,10].

Throughout the recent history, the structure of such QSM theory is always associated to integrals involving the description of *Quantum Objects* (QO).

By a QO must be understood a microscopic system, which possesses *all* its information included into an associated *Positive Definite* (PD) function. Such functional descriptive power lies in the assignation to the attached PD function of some statistical probability distribution formalism, the so-called *Density Function* (DF) [21,22], which, in turn, is nothing else but a result of the *system wavefunction* squared module manipulation. Such initial function information, in form of wavefunctions, can be obtained as a solution of the system's Schrödinger equation. The whole conceptual structure may be cast as a quantum mechanical postulate.

Also, besides the PD DF dependence of any QSM, they can be easily attached to PD operators too, as well as to the functional spaces where PD DF belongs. This apparent PD mathematical background makes of QSM theory a good candidate to be related, somehow, to the structure and definitions of convex sets, because the particular structure of this kind of sets deals, preferentially, with PD linear combinations of vector space elements.

Here, the role, which could be played by tagged and convex sets definitions, will be studied within the framework of quantum similarity measures and the related techniques and algorithms. Therefore, the present work will be structured in the following

manner. In the first place, the appropriate definitions to construct the general theoretical background of QSM theory will be given. Then, afterwards, some necessary notions associated to tagged and convex sets will be provided. The connection between both subjects will be discussed next. Finally, an application example involving Quantitative Structure–Activity or Structure–Properties Relationships (QSAR-QSPR) will be discussed.

2. Quantum Similarity Measures (QSM)

Quantum similarity measures have been defined in several previous papers [8,12], along with the basic concepts associated to the theoretical body, involving the ideas underlying the QSM basic integral structure. In order to keep these primary ideas present at the time to deepen in the QSM theory, several definitions will be given first.

2.1. Definitions

2.1.1. Tagged sets and Quantum Objects (QO)

A parallel definition to the one, associated to *Boolean tagged sets*, which have been recently discussed [3], as an alternative to fuzzy sets, will be employed to construct QO Sets (QOS).

A Boolean tagged set, T_n can be defined as the direct product between a given arbitrary set, S, the *background set*, and the elements of the 2^n bit strings associated to the integers $\{0, 1, 2, \ldots, 2^n - 1\}$. These bit strings form the vertex vectors of an n-dimensional unit cube, pointing towards the PD space directions. They constitute the *tag set* K_n. Then

$$T_n = S \times K_n = \left\{\tau \mid \forall s \in S,\ \exists \nu \in K_n\colon\ \tau = (s, \nu)\right\}. \tag{D.1}$$

An interesting example of this Boolean tagged set form, connected to a particular molecular discrete description, may be briefly discussed. Consider a given set of molecular structures. Take it as a background set. Construct the attached topological matrix to every molecule in the background set and order it as a column vector. When this task is finished, homogenise the dimension of these column topological vectors by, for instance, adding the appropriate number of zeroes. Take the homogenised column topological vectors as the tag set. A Boolean tagged set has been defined, in this manner, over the molecular set.

The Boolean tagged set definition can be easily generalised, considering various possible tag set extensions. Suppose now known a set of microscopic systems, S, and let us associate it with a *background set*. Suppose also a set made of PD DF, P, and let us call it the *tag set*. According to quantum mechanics, to every element of S there can be found a one-to-one correspondence with a DF of P. The situation may be cast into a new set, a *function tagged set*, T, defined in the following way:

$$T = S \times P = \left\{\tau \mid \forall s \in S,\ \exists \rho \in P\colon\ \tau = (s, p)\right\}. \tag{D.2}$$

Within this definition, a QOS may be associated to a function tagged set T, and a QO can be considered in consequence as an element of T.

This structure may be considered as the simplest case of a possible general tagged set form, which, for the moment, is not relevant here, where definition (D.2) will be used throughout. In fact, in this possible general definition, QOS may be considered tagged sets, with the tag set part, formed not only employing a unique function, but the collection of all the possible DF of the system's quantum states. The tag, ρ, will have, from this point of view, necessarily a vector structure, whose elements will be the state densities. An alternative way could be associated to the tagged set formed with the elements of the microscopic system background set and one quantum-state density as the tag part: in this case, the tagged set elements will possess the same background part and diverse tags.

2.1.2. QSM

Let us define primarily a QSM as a composition involving two QO, constructed with the rule

$$\forall a, b \in T \wedge \{a = (s_a, \rho_a);\ b = (s_b, \rho_b)\}:$$
$$Z_{ab}(\omega) \equiv \langle a|\omega|b\rangle = \iint \rho_a(\mathbf{r}_1)\omega(\mathbf{r}_1, \mathbf{r}_2)\rho_b(\mathbf{r}_2)\,\mathrm{d}\mathbf{r}_1\,\mathrm{d}\mathbf{r}_2, \tag{1}$$

where ω is a PD operator, whose dependence of the variable set $\{\mathbf{r}_1, \mathbf{r}_2\}$ must be coherent with the ones associated to the tag functions of the involved QO. The case of computing a QSM when both involved systems are the same, $a = b$, produces a *Quantum Self-Similarity Measure* (QS-SM).

Due to the PD nature of all the involved elements in the QSM, the values of the measure integrals, described in equation (1), are always positive. Thus, a QSM, as previously defined, can be considered an operation such as to transform the ordered pairs of tagged set elements into the set of positive real numbers: $Z : (T \times T) \rightarrow \mathbf{R}^+$. Integral defined scalar products with weights, associated to the PD operators ω, are good candidates to be connected to QSM too.

The most usual operator used up to date corresponds to the Dirac's delta function $\delta(\mathbf{r}_1 - \mathbf{r}_2)$, and in this case, integral (1) transforms into a so-called *overlap-like* QSM. A third element of the QOS can be used instead, producing several possible forms of *triple* QSM [11]. Multiple products of QO substituting the operator will yield a *multiple* QSM.

2.2. Similarity matrices and discrete representations of QO

In any case, given a QOS, the computation of QSM involving, at least, element pairs produces a new set, which has been discussed in the literature in many ways [9]. In the simplest situation, every QO can be connected with the rest of the QOS elements, including itself. When all the QOS elements are involved in the QSM calculation and ordered, this gives as a result a symmetric matrix – the Similarity Matrix (SM) **Z**. The

dimension of the SM will depend on the cardinality of the QOS: if #(T) $= n$, then Dim($\mathbf{Z}$) $= (n \times n)$. The SM can be considered a row hypermatrix whose elements are n-dimensional column vectors, collecting all the matrix elements involving a given QO. That is,

$$\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n) \equiv \{\mathbf{z}_i\} \subset C_n\big(\mathbf{R}^+\big). \tag{2}$$

As it has already been commented, the SM elements are computed within some kind of scalar product formalism. Owing to the fact that the DF tag part, if the background set elements are chosen really different, could be considered a linearly independent function set. Then, the collection of QSM can be easily seen as forming a metric matrix, computed over the DF tag set, and thus the SM, $\mathbf{Z}$, may be considered a PD matrix too. This is so, whenever the QSM are calculated with DF bearing the *same* orientation in the particle coordinates space, for all matrix elements.

An interesting feature, derived from the PD nature of the DF tag set and from the resulting QSM, corresponds to the elements of such vectors and matrices: the column vectors, $\{\mathbf{z}_i\}$, elements are positive too, as represented schematically in equation (2). Given a QOS, constructed as a function tagged set, as the one previously defined for QO, and provided the QSM among the elements of the QOS, as defined in equation (1), equation (2) introduces the possibility to construct another tagged set. This can be performed employing the following definition, which has a parallel form to the previous definition (D.2):

$$\Theta = S \times \mathbf{Z} = \big\{\theta \mid \forall s \in S,\ \exists \mathbf{z} \in \mathbf{Z}\colon\ \theta = (s, \mathbf{z})\big\}, \tag{D.3}$$

where, the symbol for the tag set, $\mathbf{Z}$, has been chosen to be the same as the one used for the SM. The new *vector tagged set* is another possible representation of the QOS. The vector tagged set constructed as in definition (D.3) is immediately connectable to concepts defined early in the QSM context [9]. Indeed, a QOS in the form of a vector tagged set has been called, when molecular QOS were studied, a *molecular point cloud*, and the elements contained in it, *point-molecules*.

Vector tagged sets or *QO point clouds* can be derived from the original function tagged sets, T, by projecting the PD DF tagged set into a PD vector tagged set:

$$\begin{aligned} &\mathcal{P}(T) = \Theta \quad \Rightarrow \\ &\quad \forall \tau \in T\colon\ \mathcal{P}(\tau) = \mathcal{P}\big((s, \rho)\big) = \big(s, \mathcal{P}(\rho)\big) = (s, \mathbf{z}) = \theta \in \Theta. \end{aligned} \tag{D.4}$$

The new projected QOS corresponds to a discrete QO description, where every system from the background set is no more tagged by a continuous DF, but by an n-dimensional PD vector. As such, it can be easily transformed into a Boolean tagged set, according to the comments of previous work [3], based on the evidence of the usual computational practice. According to this, computations are made within the field of rational numbers, and as such the numerical QSM involved in the elements of the tag set part of the QO point cloud, can be associated to this numerical form. Rational numbers are expressed within a usual computational structure in form of bit

strings. Already in this numerical form, they can enter as the elements of the tag part of vector tagged sets, transforming them into the equivalent Boolean tagged sets.

The sets P and the projection $Z = \mathcal{P}(P)$ constitute very peculiar ensembles. In fact, they are properly defined by elements, which have values only in the PD set of real numbers. Both correspond to sets whose elements belong to some vector space with operations defined in $\mathbf{R}^+$. The vectorial addition cannot possess a complete group structure, but a *semigroup* one instead [16], lacking of reciprocal elements. Everything else can be considered preserved. One can refer to this kind of structures as *vector semispaces*.

3. Convex Sets (CS)

CS can be referred to as vector semispaces for our purposes, although there are interesting problems related to this concept, which can be easily found in the literature [23]. A CS is a collection of linear combinations of vectors in a semispace, thus having positive coefficients, which fulfil an extra constraint such that the coefficient addition yields the unity. That is, a convex set is a subset K of a semispace H, which is defined by means of

$$\begin{aligned} &\forall \mathbf{x} \in K \subseteq H \quad \Rightarrow \\ &\quad \exists \{\mathbf{c}_i\} \subset K \wedge \exists \{\alpha_i\} \subset R^+ \colon \mathbf{x} = \sum_i \alpha_i \mathbf{c}_i \wedge \sum_i \alpha_i = 1. \end{aligned} \tag{D.5}$$

A curious question can be formulated now about how CS, as defined above, can be considered made using vectors of a general vector space. Because, whenever a set of complex numbers is defined such as to form a normalised vector:

$$\begin{aligned} &\{\gamma_i\} \subset C \wedge \mathbf{g} = (\gamma_1, \gamma_2, \ldots, \gamma_n) \in V_n(C) \wedge \mathbf{g}^+\mathbf{g} = \sum_i |\gamma_i|^2 = 1 \wedge \\ &\text{if } \alpha_i = |\gamma_i|^2 \ \forall i \quad \Rightarrow \quad \sum_i \alpha_i = 1. \end{aligned} \tag{3}$$

Then, CS may be constructed from an n-dimensional vector space over the complex field, $V_n(C)$, associating normalised vectors of this general space to the positive coefficients summing unity of definition (D.5). The squared modules of the general vector elements are the needed positive coefficients of the CS. Such a vector space, as $V_n(C)$, may be called a *generating vector space* of the CS, K. By extension, any vector $\mathbf{g} \in V_n(C)$, will be called the *generating vector* of the attached CS element.

This bears, no doubt, a discrete quantum mechanical flavour. Quantum mechanical DF are computed as manipulated squared modules of some complex valued wavefunction, which can act as a generating ∞-dimensional vector. The generating vector space of DF is the Hilbert space whose elements are the quantum system's wavefunctions.

It must be remarked now that DF constitute some kind of PD functions, admitting only positive coefficients for linear combination purposes. More than this, the volume

subtended by any DF shall be finite and constantly equal to some function of the particle number. In consequence, DF can be, without generalisation loss, considered normalised to unity possessing a unit volume. Thus, DF could be considered the elements of a CS defined in a ∞-dimensional space. Indeed, suppose a set of DF, possessing coherent particle co-ordinates $\{\rho_i(\mathbf{r})\} \subset P$, such that $\forall i$,

$$\int \rho_i(\mathbf{r})\, \mathrm{d}\mathbf{r} = 1, \tag{4}$$

and construct a new DF, $\vartheta(\mathbf{r})$, using a set of PD coefficients $\{c_i\} \subset \mathbf{R}^+$:

$$\vartheta(\mathbf{r}) = \sum_i c_i \rho_i(\mathbf{r}) \quad \Rightarrow \quad 1 = \int \vartheta(\mathbf{r})\, \mathrm{d}\mathbf{r} = \sum_i c_i \int \rho_i(\mathbf{r})\, \mathrm{d}\mathbf{r} = \sum_i c_i. \tag{5}$$

As a consequence, forcing new linear combinations of DF to have constant unit volumes is the same as embedding the DF set into a CS structure.

4. Atomic Shell Approximation (ASA) within a CS

Recently, various papers have been devoted to the problem of constructing properly defined first-order DF, by using spherical function basis set superpositions [2,14,15]. The so-called ASA has been discussed in various molecular and atomic environments as well as has been practically studied through several methodological algorithms. According to the previous generalisation and analysis, the ASA becomes nothing else than a consequence of the nature of the DF and of the vector semispaces where they belong.

Indeed, suppose a PD basis set of spherical functions $\mathbf{\Sigma} = \{\sigma_i\}$, which can be built up from a normalised function set $\mathbf{\Phi} = \{\varphi_i\}$. This can be obtained using simply the equivalence $\sigma_i = |\varphi_i|^2$, $\forall i$. Taking into account the normalisation conditions of the functions belonging to the set $\mathbf{\Phi}$, the set $\mathbf{\Sigma}$ acquires automatically the property of unit volume, as commented earlier, when dealing with the CS structure of DF:

$$1 = \int \left|\varphi_i(\mathbf{r})\right|^2 \mathrm{d}\mathbf{r} = \int \sigma_i(\mathbf{r})\, \mathrm{d}\mathbf{r}. \tag{6}$$

Suppose a DF, $\rho(\mathbf{r})$, with an appropriate coherent variable set, $\mathbf{r}$, as the one associated to the basis function set $\mathbf{\Sigma}$. The ASA approach consists in expressing the DF as a CS element generated from the basis set $\mathbf{\Sigma}$:

$$\forall \rho(\mathbf{r}),\ \exists \{c_i\} \subset \mathbf{R}^+: \quad \rho(\mathbf{r}) \approx \sum_i c_i \sigma_i \wedge \sum_i c_i = 1. \tag{7}$$

The coefficients $\{c_i\}$ of equation (7), are to be estimated using a constrained least squares technique. Recently [1], it has been shown how elementary Jacobi rotations [18] can be employed efficiently to deal with this kind of problems. For atoms, the ASA procedure produces first-order DF with almost negligible quadratic error integral measures. It has been also investigated whether a *promolecular approach* can

be constructed as an accurate ersatz for molecular first-order DF. The promolecular formalism states that the first-order DF tag part for any molecule can be approximated by a simple sum of atomic ASA functions, computed by means of a constrained least squares procedure as commented before. Results have shown accurate enough values for QSM computational purposes [13]. More refined approaches can use a completely equivalent formalism as in equation (7), using atomic ASA optimal functions as the components of the basis set Σ.

5. Convex operators

In the previous discussion, related to the ASA approach, the CS nature of the DF has been fruitful to design an efficient approximate algorithm, so as to obtain reasonable accurate approaches to first-order DF. From there and from the relationship between CS and DF, as stated in equation (5), one can try to go beyond the functional CS and, using the PD operator nature of DF [21,22], extend the previous ideas to PD sets of operators.

In doing such an extension, the QSM themselves will be affected, being based on PD operator weighted integrals. The present discussion, from now on, will be devoted to discuss this extension of the CS framework and, finally, as a corollary, it will be tried to find out if there is some possible application of the resulting formalism.

5.1. Convex linear combinations of PD operators

Suppose a set of PD operators, $\mathbf{\Omega} = \{\omega_\alpha\}$. They fulfil the following relationship, when studied over the coherently defined DF vector semispace P:

$$\mathbf{\Omega} = \left\{\omega \mid \omega : P \to \mathbf{R}^+\right\} \wedge \forall \omega \in \mathbf{\Omega},\ \forall \rho \in P\colon \quad \int \omega(\mathbf{r})\rho(\mathbf{r})\,\mathrm{d}\mathbf{r} \in \mathbf{R}^+. \tag{8}$$

Nothing opposes to consider the following situation:

$$\mathbf{\Omega} \subset P \quad \Rightarrow \quad \forall \omega, \rho\colon \int \omega(\mathbf{r})\rho(\mathbf{r})\,\mathrm{d}\mathbf{r} = \langle \omega | \rho \rangle \in \mathbf{R}^+, \tag{9}$$

where the application of the PD operator set over the PD DF set can be interpreted as a non-commutative scalar product, defined over the vector semispace, where both PD operators and DF sets belong. Moreover, this scalar product can be interpreted according to the usual quantum mechanical interpretation as the expectation value, $\langle \omega \rangle$, of the system observable, represented by the particular operator ω in terms of the QO particular state DF tag part, ρ.

Being possible to consider the operator set, $\mathbf{\Omega}$, as forming part of the vector semispace, P, the situation applied on DF as stated in equation (5) can be used over the elements of the operator set. In such a manner, if a set of coefficients $W = \{w_\alpha\}$ exists, and the following constraints, similar to definition (D.5) and equation (7), are

set upon it, then a linear combination of the operator set Ω will yield a PD operator, γ, such that

$$W = \{w_\alpha\} \subset \mathbf{R}^+ \wedge \sum_\alpha w_\alpha = 1 \quad \Rightarrow$$
$$\gamma = \sum_\alpha w_\alpha \omega_\alpha \colon \ \langle\gamma|\rho\rangle = \sum_\alpha w_\alpha \langle\omega_\alpha|\rho\rangle \in \mathbf{R}^+. \tag{10}$$

The second constraint has been introduced so as to obtain a pattern comparable to the CS structure of the semispace P and transfer it to Ω, but it is not strictly necessary to keep this unit coefficient sum, if not needed. The most interesting thing is the obvious result that PD operators can yield, in a CS environment, new PD operators.

5.2. Tuned QSM, SM and QO descriptors

The previous conservation of the PD property upon linear combinations of PD operators in a CS environment can be employed in the evaluation of new kinds of QSM, by constructing a new breed of PD operator weights. The γ-type operators, appearing in equation (10), can be tuned up, while maintaining the identity of the operator set, just by changing the values of the CS coefficient set, W, conserving the initial chosen constraints. A QSM, following the definition provided in equation (1), can be built up under the circumstances as

$$Z_{ab}(\gamma) = \sum_\alpha w_\alpha Z_{ab}(\omega_\alpha). \tag{11}$$

The resulting tuned-up SM elements $Z_{ab}(\gamma)$ produce another obvious result when looking at the SM set, $\{\mathbf{Z}(\omega_\alpha)\}$, associated to every operator in Ω. Each SM being attached to a PD operator, then they can be considered some discrete matrix representation of the associated operator in the corresponding basis set of the involved DF. These matrices, as it has been commented before, can be considered as PD matrices. Thus, equation (11) can be written in whole matrix form as

$$\mathbf{Z}(\gamma) = \sum_\alpha w_\alpha \mathbf{Z}(\omega_\alpha). \tag{12}$$

The resultant matrix being PD, because if in the SM set, $\Theta = \{\mathbf{Z}(\omega_\alpha)\}$, all the SM elements are PD, then the following property will hold:

$$\forall \mathbf{x} \in \mathbf{C}_n \wedge \forall \mathbf{Z}(\omega_\alpha) \in \Theta \colon \ \mathbf{x}^+ \mathbf{Z}(\omega_\alpha)\mathbf{x} \in \mathbf{R}^+ \quad \Rightarrow$$
$$\mathbf{x}^+ \mathbf{Z}(\gamma)\mathbf{x} = \sum_\alpha w_\alpha \mathbf{x}^+ \mathbf{Z}(\omega_\alpha)\mathbf{x} \in \mathbf{R}^+ \text{ if } \forall w_\alpha \in \mathbf{R}^+. \tag{13}$$

From the results above, one can see that a finely tuned set of QO descriptors can be obtained in this way. This is so due to that equation (12) holds for the SM columns too, in such a way as

$$\mathbf{z}_i(\gamma) \in \mathbf{Z}(\gamma) \wedge \mathbf{z}_i(\omega_\alpha) \in \mathbf{Z}(\omega_\alpha): \qquad \mathbf{z}_i(\gamma) = \sum_\alpha w_\alpha \mathbf{z}_i(\omega_\alpha). \tag{14}$$

5.3. *Résumé*

So, one can shortly describe all the findings and definitions up to now in the following way: A QOS is chosen in the form of a DF tagged set. A PD set of suitable operators is used, as a set of weights, in the evaluation of QSM between QO. A set of SM is thus computed for each operator. A CS with suitable coefficients is chosen to combine the elements of the SM set. The resultant SM columns are convex descriptors of the corresponding QO, and provide a discrete vector tagged set representation of the QOS.

5.4. *Finely tuned QSAR*

If an immediate application of all the previous development has to be chosen, *Quantitative Structure–Activity*, or *Structure–Property, Relationships* (QSAR or QSPR), is a good candidate field. In our laboratory, the basic theory connecting QSM and QSAR or QSPR has been developed [5] some time ago and various practical applications have been reported [17,19,20] recently.

It has been deduced that molecular properties have to be, in some manner, related with the discrete representation of molecular descriptors furnished by the columns of SM, constructed from QSM over the molecular QO. As a consequence of equations (9) and (14), a given property value, π, for a particular molecular QO, described in turn by a discrete descriptor, $\mathbf{z}(\gamma)$, can be related by means of

$$\pi = \mathbf{u}^{\mathrm{T}} \mathbf{z}(\gamma), \tag{15}$$

where the vector $\mathbf{u}$ corresponds to an unknown discrete representation of some operator over the same PD DF basis set, used to construct the convex discrete molecular descriptor $\mathbf{z}(\gamma)$.

The usual procedure is to use a least-squares algorithm in order that, knowing the pairs $\{\pi, \mathbf{z}(\gamma)\}$ for a molecular QOS, the values of $\mathbf{u}$ can be obtained. Taking into account the tuned construction of the vectors $\mathbf{z}(\gamma)$, it can be easily seen that the vector $\mathbf{u}$ will depend on the tuning parameter set. Using the least-squares solution of the problem

$$\mathbf{u} = \left(\mathbf{Z}(\gamma)^{\mathrm{T}} \mathbf{Z}(\gamma)\right)^{-1} \mathbf{Z}(\gamma)^{\mathrm{T}} \mathbf{p}, \tag{16}$$

where the vector $\mathbf{p} = \{\pi_k\}$ contains the values of the property for each molecular QO, and $\mathbf{Z}(\gamma)$ is the SM of the QOS computed according to equation (12). Equation (16), however, has been written taking into account the possibility that the SM may be no

longer symmetric, but rectangular. This will constitute the more general case, where instead of a unique tagged set, two QOS with different cardinalities, m and n, are used to compute the QSM. The resultant SM will be of dimension $m \times n$.

From this previous definition, one can easily deduce that the vector **u** will depend on the tuning coefficients W. Optimisation of the tuning set coefficients W, can be done at the same time as the classical least-squares problem is solved, keeping in mind the associated CS constraints, which the tuning set W bears. A parallel non-linear constrained optimisation on a quadratic function of the W elements will appear. The interesting feature here is that CS constraints can be studied in the same way as the CS constraints are kept in the optimal ASA problem. Thus, to the usual least-squares problem, involving the operator associated vector, **u**, there appears another least-squares equation, which starts defining the residual vector

$$\Delta = \mathbf{p} - \mathbf{Z}(\gamma)\mathbf{u} = \mathbf{p} - \sum_\alpha w_\alpha \mathbf{Z}(\omega_\alpha)\mathbf{u} = \mathbf{p} - \sum_\alpha w_\alpha \mathbf{v}_\alpha = \mathbf{p} - \mathbf{V}\mathbf{w}, \tag{17}$$

where the previous definitions of the involved matrices have been employed. Also, the matrix **V** collects the vector set $\{\mathbf{v}_\alpha = \mathbf{Z}(\omega_\alpha)\mathbf{u}\}$, and the vector $\mathbf{w} = \{w_\alpha\}$ contains the coefficients of the tuning set W. The residual vector (17) is obviously dependent on the classical least squares solution **u** in equation (16). From the inspection of the residual vector Δ, it is easy to see that the quadratic error will depend on a generalised quadric function with a variable set formed by the new unknown vector **w**. This least-squares problem has to be solved under the constraints associated to the PD nature of the vector **w**. A CS constraint structure may be very convenient in order to normalise the problem form. Thus the quadratic function and the constraints may be written as

$$\varepsilon^{(2)} = \chi - 2\mathbf{q}^{\mathrm{T}}\mathbf{w} + \mathbf{w}^{\mathrm{T}}\mathbf{Q}\mathbf{w} \wedge \mathbf{w} \in C_n\left(\mathbf{R}^+\right) \wedge \sum_\alpha w_\alpha = 1, \tag{18}$$

where the following simplifications have been used:

$$\chi = \mathbf{p}^{\mathrm{T}}\mathbf{p} \wedge \mathbf{q} = \mathbf{V}^{\mathrm{T}}\mathbf{p} \wedge \mathbf{Q} = \mathbf{V}^{\mathrm{T}}\mathbf{V}. \tag{19}$$

It is important to remark here the importance to define the appropriate SM set, $\{\mathbf{Z}(\omega_\alpha)\}$, in order that the matrix $\mathbf{V} = \{\mathbf{v}_\alpha\}$ possesses elements linearly independent, to obtain a PD matrix Q. This is equivalent to saying that the SM set shall provide images of the least-squares solution **u**, which must be linearly independent.

The solution of the second optimisation problem for the vector **w** could be sought using a generating vector, for example, **x**, which will substitute the **w** elements in the following way: $w_\alpha = |x_\alpha|^2$, $\forall\alpha$. Function (18) will transform into a quadratic function in terms of the components of vector **x**. Optimisation under the unit norm of the generating vector, $\mathbf{x}^+\mathbf{x} = 1$, may be obtained by means of EJR, as in the ASA case [1].

The whole optimisation process shall be made in an iterative manner:

(1) Using a starting approximate tuning vector **w** obtain **u**, solving equation (16).

(2) Known **u**, compute a new **w**, minimising function (18).

(3) Go to (1) while the vector pair $\{\mathbf{u}, \mathbf{w}\}$ remains inconsistent with respect to the previous iteration.

Changing the number and nature of the SM composite $\mathbf{Z}(\gamma)$ will obviously produce different results, but within a given choice these can be coherently tuned up. This can add extraordinary possibilities to QSAR procedures.

6. Conclusions

A general framework, where quantum objects can be described in a systematic way, has been constructed. The concept of density function tagged set encompasses an earlier generalisation, which was proposed as a sound substitution of fuzzy set definitions, to describe molecular structures – the Boolean tagged sets. At the same time, the definition of quantum mechanical density functions has been used in order to put in evidence its essential positive definite nature. This fundamental property of density functions, often forgotten in the current literature, has been used too to connect quantum similarity measures, a simple concept, which compares two or more quantum objects with the spaces containing positive definite operators. Vector semispaces and, more conventional, convex set algebra has been put into the context of the computation of approximate density functions, as in the ASA framework. This kind of computational algorithmic experience has been extended to positive definite operators and their matrix representation, the similarity matrices, from the point of view of quantum similarity measures. Positive definite operators can be used to construct a convex set of new positive definite operators and consequently their matrix representations remain positive definite. In this way, a new window is opened to obtain discrete, finely tuned molecular descriptors in the form of positive definite vectors belonging to n-dimensional vector semispaces. The utility of the presented theoretical results in the context of quantitative structure–activity relationships is nothing more but one of the vast prospective application fields.

Acknowledgements

This work has been partially supported by a CICYT grant #SAF 96-0158. The author wishes to warmly thank Mr. Ll. Amat for conducting some preliminary calculation tests on some tuned QSAR problems. Enlightening, informal discussions with Dr. J. Mestres have been carried out in the previous stages of this work.

References

[1] Ll. Amat and R. Carbó-Dorca, QSM and expectation values under ASA: First-order density fitting using EJR, Technical report IT-IQC-2-97; J. Comput. Chem. 18 (1997) 2023–2039.

[2] Ll. Amat, R. Carbó and P. Constans, Sci. Gerund. 22 (1996) 109–121.
[3] R. Carbó-Dorca, Fuzzy sets and Boolean tagged sets, Technical report IT-IQC-5-97; J. Math. Chem. 22 (1997) 143–147.
[4] R. Carbó, M. Arnau and L. Leyda, Int. J. Quantum Chem. 17 (1980) 1185–1189.
[5] R. Carbó, E. Besalú, L. Amat and X. Fradera, J. Math. Chem. 18 (1995) 237–246.
[6] R. Carbó and B. Calabuig, Comput. Phys. Commun. 55 (1989) 117–126.
[7] R. Carbó and B. Calabuig, in: *Computational Chemistry: Structure, Interactions and Reactivity*, Vol. A, ed. S. Fraga (Elsevier, Amsterdam, 1992) pp. 300–324.
[8] R. Carbó and B. Calabuig, Int. J. Quantum Chem. 42 (1992) 1681–1693, 1695–1709.
[9] R. Carbó and B. Calabuig, J. Chem. Inf. Comput. Sci. 32 (1992) 600–606.
[10] R. Carbó and B. Calabuig, J. Mol. Struct. (Theochem) 254 (1992) 517–531.
[11] R. Carbó, B. Calabuig, E. Besalú and A. Martínez, Mol. Eng. 2 (1992) 43–64.
[12] R. Carbó, B. Calabuig, L. Vera and E. Besalú, Adv. Quantum Chem. 25 (1994) 253–313.
[13] P. Constans, Ll. Amat and R. Carbó-Dorca, J. Comput. Chem. 18 (1997) 826–846.
[14] P. Constans, Ll. Amat, X. Fradera and R. Carbó-Dorca, in: *Advances in Molecular Similarity*, Vol. 1, eds. R. Carbó-Dorca and P.G. Mezey (JAI Press, Greenwich, CT, 1996) pp. 187–211.
[15] P. Constans and R. Carbó, J. Chem. Inf. Comput. Sci. 35 (1995) 1046–1053.
[16] *Enciclopaedia of Mathematics* (Reidel–Kluwer, Dordrecht, 1987).
[17] X. Fradera, Ll. Amat, E. Besalú and R. Carbó-Dorca, Quant. Struct. Act. Relat. 16 (1997) 25–32.
[18] C.G.J. Jacobi, J. Reine Angew. Math. 30 (1846) 51–94.
[19] M. Lobato, Ll. Amat, E. Besalú and R. Carbó-Dorca, Estudi QSAR d'una familia de quinolones, Technical report IT-IQC-4-97.
[20] M. Lobato, Ll. Amat, E. Besalú and R. Carbó-Dorca, Structure–activity relationship of a steroid family using QSM and topological QS indices, Technical report IT-IQC-8-97; QSAR 16 (1997) 465–472.
[21] P.O. Löwdin, Phys. Rev. 97 (1955) 1474–1489.
[22] R. McWeeny, Rev. Mod. Phys. 32 (1960) 335–369.
[23] J. Stoer and Ch. Witzgall, *Convexity and Optimization in Finite Dimensions*, Die Grundlehren der matematischen Wissenchaften in Einzeldarstellungen, Vol. 163 (Springer, Berlin, 1970).