Fundamental Quantum QSAR (Q²SAR) Equation: Extensions, Nonlinear Terms, and Generalizations within Extended Hilbert–Sobolev Spaces

RAMON CARBÓ-DORCA, EMILI BESALÚ
Institut de Química Computacional, Universitat de Girona, Girona 17071 Catalonia, Spain

Received 13 September 2001; accepted 16 October 2001
DOI 10.1002/qua.10081


Key words: extended wave functions; extended Hilbert spaces; extended Sobolev spaces; extended quantum similarity measures; extended expectation values; quantum QSAR

Introduction

S
ome years ago a paper was published [1] about a remarkable feature of Quantitative Structure-Activity Relationships (QSAR), that is, the linear equations relating properties and structure in microscopic systems. The noteworthy characteristic consisted in the fact that a QSAR-like equation could be easily deduced from a framework of Quantum Similarity Measures (QSM) [2–20]. QSM appear to be an old set of concepts related to the geometric interpretation of some aspects of quantum theory. Since then, a large number of applications of this newly defined quantum QSAR (Q²SAR) scheme has appeared in the current literature, dealing with quite a large variety of problems from nuclear structure [21–23], up to molecular properties like physical parameters and biological activity [24–35] as well as ecotoxicity [16–20, 34, 36–43]. Several QSM-related new algorithms [44–50], its general theoretical foundation [24, 25, 36, 51, 52], and some practical simplifications [53] have also been published. Even the possibility of connecting QSM with a new breed of topological indices has been described [51] and also the transformation of similarity matrices (SM) [10], composed by ordering into ar-
rays the QSM evaluated among the elements of a set of microscopic systems, into stochastic matrices has been also discussed [19]. The theoretical basis of QSM has been deepened accordingly [20], producing new concepts such as tagged sets [25, 36], vector semispaces [24, 25, 36], and Quantum Similarity Indices, which have been studied from a completely general point of view [10, 11], as by-products of the theory constructed by freely transforming QSM in multiple ways.

The general connection of the quantum mechanical expectation value concept with the QSAR problem has posed several problems, which have been solved step-by-step, describing a new quantum mechanical related formalism, where the typical scalar Schrödinger wave function is extended with the aid of its gradient [54 – 57]. This is made in such a way as to form a column vector or a diagonal matrix containing both elements. This new way of describing the wave function has permitted several interesting applications in quantum chemistry [57], SCF theory [58], and density function structure [54, 55, 59].

The extended wave function formalism puts forward, as its most distinguishing feature, the idea that the wave function and its gradient could be collected together to form a new mathematical entity. This possibility has led, as a natural result, to the description of extended density functions, where a superposition of position and kinetic energy probability density functions become a unique composite function, associated to the uncertainty principle. Within the quantum mechanical setup related to Schrödinger, Levy-Leblond, and Dirac equations, the implications of the extended wave function concept have been recently analyzed [55]. This has been done in parallel with the study of the interrelation of the extended wave function resulting norm with Sobolev spaces [60]. In turn, Sobolev spaces have also been generalized accordingly, in order to provide the appropriate tools to naturally connect extended wave functions formalism with the theoretical structure of nonlinear Schrödinger equations [59] and with QSM [61].

Extended Hilbert or Sobolev spaces and semispaces, respectively, have been primarily developed in order to contain extended wave functions and the resultant composite densities, as a way to develop the mathematical basis of QSM. Otherwise, QSM theory lacked a robust background mathematical structure. Consequently, it must be said that the main purpose for searching into these new mathematical structures was essentially to prove that, with an appropriate choice of the theoretical frame-
the tag set, made in turn by some coherent organized information about the objects, which must be previously known in order to construct the corresponding tagged set. This previously defined tagged set structure contains as a particular case the description of fuzzy sets [36, 51, 52]. Then it can be said that QOS are tagged sets, adapted to the quantum mechanical framework, considering the object set made of systems bearing a microscopic nature. QOS then can be defined as being formed by quantum systems with their attached probability DF acting as elements of the tag set.

Quantum Similarity Measures

The first precise definition of QSM was proposed several years ago [5–7, 63]. Actually, any QSM can be considered made by an application of the direct product of two, or more, QO DF tags, belonging to some QOS, into the set of the positive definite real numbers, \( \mathbb{R}^+ \) [62]. In the simplest way, a QSM is defined by means of two QO: \( \{\rho_A;\rho_B\} \), using their tags in terms of quantum mechanical DF: \( \{\rho_A;\rho_B\} \), and connecting them by means of an arbitrarily chosen positive definite operator, \( \Omega \), through the volume integral,

\[
z_{AB}(\Omega) = \int\int \rho_A(r_1)\Omega(r_1, r_2)\rho_B(r_2) \, dr_1 \, dr_2.
\]

Substituting, for instance, the positive definite operator, \( \Omega(r_1; r_2) \), in Eq. (1), by the Dirac’s delta function: \( \delta(r_1 - r_2) \), then an overlap-like QSM is obtained:

\[
z_{AB} = \int \rho_A(r)\rho_B(r) \, dr.
\]

Overlap-like QSM were the kind of QSM used from the early times of the definition of quantum similarity [64]. In the same way, when a Coulomb operator is employed in Eq. (1): \( \Omega(r_1; r_2) = |r_1 - r_2|^{-1} \), a Coulomb-like QSM appears connected with the integral form,

\[
z_{AB}(r^{-1}) = \int\int \rho_A(r_1)|r_1 - r_2|^{-1}\rho_B(r_2) \, dr_1 \, dr_2.
\]

A QSM \( z_{AA} \), involving only a unique QO DF tag, is usually called a quantum self-similarity measure (QSSM):

\[
z_{AA}(\Omega) = \int\int \rho_A(r_1)\Omega(r_1; r_2)\rho_A(r_2) \, dr_1 \, dr_2.
\]

QSM, defined in this way, can be interpreted as generalized volumes and they could be easily related to quantum mechanical expectation values [62]. Many alternative QSM definitions are also possible [6, 20, 65], and a general description has also been advanced [52, 62]. Throughout the historical development of these basic ideas, quantum similarity has become an extremely flexible and general theoretical tool; see for example Ref. [6]. Besides QSM integrals like those of Eqs. (2) and (3), there can be the possibility that the operator \( \Omega(r_1; r_2) \) becomes the DF tag associated to another QO, for instance \( \rho_C(r_1; r_2) \) or simply \( \rho_C(r) \); then one of the many forms [65] of a triple density QSM is readily defined:

\[
z_{ABC} = \int \rho_A(r)\rho_C(r)\rho_B(r) \, dr,
\]

and, in the same way, multiple density QSM can be structured, see Refs. [65, 66] and the Appendix for more details on a possible general definition of QSM.

It must be stressed here that, given two or more QO and choosing the operator \( \Omega(r_1; r_2) \), then the QSM (1) becomes uniquely determined by the DF tags, attached by quantum mechanical links to the considered active QO within the QSM. That is: the values of QSM among several QO cannot be chosen beforehand, in the same way that a QO wave function cannot be chosen, but just uniquely computed as a solution of the Schrödinger equation. This remark must be stressed, because it establishes a major difference along with QSM, considered as QO descriptors, and the usual arbitrary way that such parameters, assigned to describe molecular structure, are chosen in classical QSAR procedures. The same consideration can be extended to the so-called quantum similarity indices (QSI), which are just subjective transformations of QSM [10, 11], and thus can be chosen in many ways and forms. In this sense, QSM shall be preferred to QSI as QO descriptor working tools.

Similarity Matrices and Discrete Quantum Objects Sets

Given a well-defined QOS, the collection of QSM of type (1), involving all the QO associated pairs, can be ordered in the form of a symmetric matrix: \( Z = [z_{ij}] \). Such an ordered array of QSM is consequently called a similarity matrix (SM). Any SM may be partitioned in terms of its columns (or rows);

*Usually a column matrix partition is performed, although a row transformation can be also used. When it would seem important to make relevant which one of the two possible vector forms is adopted, then a bra-ket formalism will be employed to stress the difference between row and column vectors.
that is,

$$Z = (z_1, z_2, ..., z_N) = [z_I],$$

where every column corresponds to the collection of QSM, associated to a precise QO, belonging to the studied QOS [6, 10, 24, 52, 62], and computed with the rest of QOS elements, including itself. The same can be applied to multiple DF QSM. Then, the SM becomes a similarity hypermatrix, and the above-mentioned partition will be modified accordingly. This practical aspect has been studied recently [65, 66] and will not be repeated here in full detail.

The meaning of a SM definition will be discussed as follows. If $\omega_A$ is a given QO and $\rho_A$ is associated to its DF tag, then the corresponding column of the SM, $z_A$, may be considered in correspondence to the $\omega_A$ QO DF tag, and could consequently be associated to a discrete, finite-dimensional, QO representation. This situation can be written symbolically as $\omega_A \leftrightarrow \rho_A \leftrightarrow z_A$, or, in a better way as $(\omega_A; \rho_A) \leftrightarrow (z_A)$, taking into account the QOS elements structure, made by the ordered pairs like: (quantum system; DF), which may be associated, in turn, to an alternative ordered pair constructed as: (quantum system; SM column), obtained by computing the corresponding QSM between QO. In this manner, the QSM collection of any chosen QO, belonging to a given QOS, with respect to all the elements of the QOS, defines an ordered discrete set of N-dimensional tags $\{z_I\}$, $N$ being the cardinality of the QOS. The columns of any SM can, thus, substitute for the former QO DF tags. This can be even better set in case a stochastic transformation is performed over the columns or rows of the SM [19]. A schematic discussion of this possibility is also given below in the next section.

As a consequence, a new discrete QOS (DQOS), of the same cardinality, $N$, as the original one of the QOS, can be constructed in this way. The tagged set objects are the same quantum systems of the QOS as before, but the columns or rows of the SM, substitute for the former DF tag set, infinite-dimensional, elements. The set of DQOS discrete tags naturally define a polyhedron in N-dimensional space. The DQOS tag set, for obvious reasons, can thus be called a QO point cloud [7–9].

A final remark on SM and, thus, also relative to DQOS construction, is necessary in order to be aware that diverse positive definite operators or the corresponding SM over some chosen QOS, can be superposed by means of a convex set of coefficients; see the Appendix, without the resulting operator or SM losing, by such a manipulation, the positive definiteness of the superposed elements.

**STOCHASTIC SM**

The computation of SM over a QOS, providing a new DQOS structure, as discussed above, produces a set of N-dimensional tags, which can be associated to the original infinite-dimensional DF tags. Despite the strict positive definiteness of the SM column set elements, $\{z_I\}$, which appear as a consequence of the QSM definition presented in Eq. (1), the connection between the N-dimensional tags and the DF is not immediately evident. However, it can be deduced after taking into account the nature of the involved DF tags [67–72], which can be considered, in turn, either as positive definite functions or projection operators, due to the quantum mechanical origin of its mathematical structure definition. The present section will discuss a SM transformation, producing a new natural collection of N-dimensional tags, which are provided with a structure, such that their elements can be considered as forming a discrete probability distribution. This possibility will be expected as a plausible outcome of QSM theory, due to the quantum mechanical origin of all the QO tags employed so far. This may be associated to the fact that, if some probability DF set is substituted by a discrete set of finite-dimensional vectors within a one-to-one rule, then the finite-dimensional substitutes will more likely be nearest to the infinite-dimensional DF tags, if they in addition possess a probability distribution character.

**Stochastic Transformation of Quantum Similarity Matrices**

N-dimensional quantum SM columns or rows, $\{z_I\}$, or, simply, their elements, are made, by construction, of positive definite real numbers. Also, from a practical computational point of view, they can safely be supposed positive definite rational numbers. This characteristic property can be summarized by saying that the set of the SM columns or rows, the QO point cloud, belongs to a vector semi-space (VSS): $\{z_I\} \subset \mathcal{V}_N(\mathbb{R}^+)$. VSS are vector spaces that lack reciprocal elements or negative scalars; see for more details Refs. [24, 51] and also the Appendix. It must be also noted, for example, that the QO point cloud, defined by the DQOS tag set elements, commented on a few paragraphs earlier, delimits a distinctive set of points in some N-dimensional VSS, as, due to the nature of the QSM definition, all their components are made of strictly positive numbers.
Obviously, they are rigorously oriented toward the positive definite space hyperquadrant.

The VSS structure in general and, in particular, the construction of SM precludes that, in any case, the sum of every SM row (or column, due to the SM symmetry) elements is a positive real number, for example, \(( z_i ) = \sum_j z_{ij} \in \mathbb{R}^+\) (see Definition 5 in the Appendix). These sums can be used as row (or column) scale factors in order to trivially obtain a new row (or column) set belonging to the same VSS, that is, \( s_i = ( z_i )^{-1} z_{ij} \), but with their elements possessing the imposed form of a discrete probability distribution. That is, the following equalities can be easily written:

\[
\langle s_i \rangle = \left( \langle z_i \rangle^{-1} z_i \right) = \langle z_i \rangle^{-1} \langle z_i \rangle = 1.
\]

The set of \( N \) rows: \( S = \{ s_i \} \), ordered forming a square \(( N \times N )\) matrix \( S \), produces a nonsymmetric stochastic matrix [19] as a result.

A trivial compact way to produce the stochastic row matrix \( S \) may be described by first constructing a diagonal matrix, \( D \), with elements made by the sums of row (or column) SM elements,

\[
D = \text{Diag}(\langle z_1 \rangle, \langle z_2 \rangle, \ldots, \langle z_N \rangle), \tag{7}
\]

and producing afterwards the matrix product \( S = D^{-1} Z \). In the same manner, a column stochastic matrix will be defined straightforwardly by the transpose of the previous definition, \( S^T = ZD^{-1} \). The symmetrical structure of the original SM must obviously be taken into account.

The row \(| \{ s_i \} \rangle\) or column \(| \{ s_i \} \rangle\) stochastic sets, being associable to a collection of discrete probability distributions, may be even better connected to the original QOS DF tag set, \(| s_i \rangle\), than the rows or columns of the attached nontransformed SM. In the same way as was discussed when SM were analyzed, the rows or columns of the corresponding stochastic SM can be used as a tag set to produce a new tagged set, which can be called a discrete stochastic QOS (DSQOS), when combining the stochastic rows (or columns) with the microscopic systems belonging to the object set \(| \omega_i \rangle\) of the original QOS. For example, taking into account the same considerations as those used before when previously discussing the nature of the SM rows, the connection between the original QOS elements with the stochastic matrix rows: \( \langle \omega_i ; s_i \rangle \leftrightarrow ( \omega_i ; | s_i \rangle ) \); \( \forall I \), defines the elements of a DSQOS.

### Inward Symmetrization of Stochastic Quantum Similarity Matrices

The elements of the DSQOS tag set can be directly used as QO discrete descriptors, admitting the actual implications \( \forall I ; \omega_i \leftrightarrow \rho_i \leftrightarrow ( s_i ) \), in a similar manner, as has been previously discussed when dealing with the significance of the SM elements. The unique problem lies in the fact that, to the contrary of the SM, \( Z \), the row stochastic matrix, \( S \), is not symmetric. However, this does not constitute a restrictive problem when alternative manipulations of the DSQOS are envisaged and even has other possible uses, not contained in symmetric QO descriptor structures; see, for example, Ref. [19].

In addition to the classical symmetrization techniques, which customarily use the sum or classical product of the matrix and its transpose, there can be also described a simpler algorithm, involving an inward matrix product [37, 38, 62]. The inward matrix product between two matrices, \( A \) and \( B \), bearing the same arbitrary dimension, is defined without problems as another matrix, \( C \), with the same dimension form. Using \(( n \times m )\) matrices as a typical quite general example, the following straightforward algorithm can be designed for the inward matrix product definition:

\[
C = A \ast B \rightarrow \forall i,j : c_{ij} = a_{ij}b_{ij}. \tag{8}
\]

Inward matrix products possess a standard set of properties, which have already been described in detail [37, 38, 62] and will not be repeated here. It is trivial to see that the commutative inward matrix product,

\[
R = S^T \ast S = S \ast S^T \tag{9}
\]

of the stochastic matrix \( S \) with its transpose, \( S^T \), produces a symmetric matrix as a result.

### Stochastic Quantum Similarity Indices

As an example of QSI, now will be given a brief discussion on how stochastic transformations can produce similarity indices and how these new QSI forms can be related to previous definitions. To start, one must realize that the symmetric matrix \( R \) in Eq. (9) could be used in the same way as the original SM, but for unit homogenization purposes with respect to the initial SM, \( Z \), it is preferable to employ the inward square root of the symmetric inward
product (9), defined by the algorithm,*

\[ Q = \mathbb{R}^{[1/2]} \rightarrow \forall i, j : q_{ij} = \sqrt{y_{ij}}. \]  

(10)

One can consider the new symmetric matrix Q as holding stochastic quantum similarity indices (SQSI). It is worthwhile to consider, at this point, how the new kind of SQSI may be related to previous well-defined QSI, which, as the so-called Carbó index [10, 11, 64], are present in the literature since twenty years. An expression of the elements of matrix Q in terms of the original symmetric SM, leads to the equality sequence,

\[ q_{ij} = \sqrt{y_{ij}} = \sqrt{z_{ij}^2y_{ij}} = \frac{z_{ij}}{\sqrt{(z_i^*)z_j}}, \]  

(11)

which resembles the Carbó similarity index [64] continuous form, that is: the index defined over the involved QO DF tag couples, using the associated SM elements, which can be written as

\[ k_{ij} = \frac{z_{ij}}{\sqrt{z_i^*z_j^2}}. \]  

(12)

**STATISTICAL EXPECTATION VALUES**

When studying a quantum mechanical system, the expectation value of some observable \( \omega_i \) in a well-defined QO system state, can be written by means of the integral [54] equivalent to a statistical measure,

\[ \langle \omega \rangle = \int W(r)\rho(r)\,dr, \]  

(13)

where \( W(r) \) is an associated hermitian operator to be determined for a given QOS, and \( \rho(r) \) is any system state DF acting, as is customary, as a QO tag.

On the other hand, Eq. (13), from both a mathematical and a Q2SAR point of view, can be interpreted as a scalar product, that is

\[ \langle \omega \rangle = \langle W|\rho \rangle. \]  

(14)

Taking into account the a priori unknown nature of the Q2SAR operator \( W(r) \), one can consider that it can be decomposed as a product of some operator, \( W_0(r) \), leading to the expectation value, still unknown and to be determined, by a known positive definite weight operator, \( \Omega(r, r_0) \), that is,

\[ W(r, r_0) = W_0(r)\Omega(r, r_0). \]  

(15)

Thus, the expectation value as presented in expression (14) could be written now as

\[ \langle \omega \rangle = \langle W_0|\Omega|\rho \rangle. \]  

(16)

This is the same as to transform the previous expectation value, as expressed in Eq. (13), into the equivalent, albeit more general, integral,

\[ \langle \omega \rangle = \int \int W_0(r)\Omega(r, r_0)\rho_0(r_0)\,dr_0\,dr. \]  

(17)

It is easy to see that the positive definite weight operator \( \Omega(r, r_0) \) can particularly be chosen as the Dirac’s delta function: \( \delta(r - r_0) \), and, thus, in doing so, Eq. (13) is recovered. In order to distinguish the expectation value general definition, as presented in Eq. (17) from the usual choice in Eq. (13), where the weight can be considered as a unit operator, the general integral (17) can be named a weighted expectation value integral expression.

**FUNDAMENTAL Q2SAR EQUATION**

When studying a known QO attached DF set, \( \{\rho_i(r)\} \), forming the QOS tag set [51], and if a particular QOS element, \( A \), is chosen, then Eq. (13) can be particularly written as

\[ \langle \omega_A \rangle = \int W(r)\rho_A(r)\,dr. \]  

(18)

Now, whenever both operator and density function in the integral (13) can be considered as belonging to the same vector space and expressing approximately the unknown operator, \( W_0(r) \) as defined in Eq. (15), by means of a linear combination of the QOS tag set, acting as a basis:

\[ W_0(r) \approx \sum_i w_i\rho_i(r). \]  

(19)

A new relationship is obtained after substituting this last expression in Eq. (17), while taking into account the association of the involved integral with a QSM and, thus, with the elements of a SM,

\[ \langle \omega_A \rangle = \sum_i w_i\int \rho_i(r)\Omega(r, r_0)\rho_A(r_0)\,dr_0\,dr \]

\[ = \sum_i w_i z_{ij}^\omega(\Omega). \]  

(20)

This constitutes a result, which is also equivalent to the usually proposed equation in classical QSAR. To see this, it just must be admitted that Eq. (20) holds for every QO of the chosen QOS, and that the SM, \( Z = \{z_{ij}\} \), obtained employing the weight operator \( \Omega(r, r_0) \), is necessarily symmetric.
Then an equivalent linear equation can be written in matrix form as

\[ |\tau| = Zw, \quad (21) \]

where: \(|\tau|\) is a column vector containing the expectation values or the QO property of interest and \(W\) is another column vector, whose elements are made by the unknown coefficient set \(|w|\) of the operator expression (19). In light of all the discussed aspects of the problem, it seems that Eq. (21) can be logically named, from now on, the fundamental Q^2SAR equation.

The Q^2SAR problems of type (21) are associated to the unknown operator, \(W_{Q}(r)\), supposedly formed by many entangled terms, whose coefficient coordinates, computed with respect to the QOS DF tags acting as a basis set, should be determined from the knowledge of the associated problem expectation values. The usual classical procedure to obtain the vector \(w\) could be based on some algorithm related to the well-known least-squares procedure or some variant of it; see the next section for a sketchy discussion on this subject. Also remembering the stochastic scaling of the SM, it can be admitted that the fundamental Q^2SAR equation can be transformed according to the computational needs in such a way or any alternative form. Some new insight on a possible appropriate procedure to solve the constrained fundamental Q^2SAR equation has been proposed recently [19].

This close relationship, between expectation values in continuous and discrete quantum frameworks, cannot be so easily deduced as a theoretical consequence outside of the QSM models of QSAR. Therefore, the term Q^2SAR is completely justified and could be applied from now on to QSAR models, obtained by means of QSM or QSI considerations such as these presented up to now. Several application papers have recently appeared in the literature [26–35], where it is shown how QSM can be successfully used within this Q^2SAR formalism. The SM appearing in the fundamental Q^2SAR Eq. (21) can be constructed in several ways, even using diverse weighting operators when computing the QSM, transforming the SM into similarity indices (SI) [11] or stochastic SM [19], even gathering several different SM in a convex combination [40], providing the methodological basis of tuned Q^2SAR.

A corollary, deducible from the above discussion, can now be proposed. If a Q^2SAR equation can be built up solely employing quantum mechanical concepts, then the far-reaching consequences of this situation obviously encompass the (let us call it) classical QSAR, which can be viewed now, as a convenient simplification of the geometrical and mathematical setup associated with the QSM theory. It can be shown that the unique statistical tools needed in Q^2SAR [19] are those coming from the quantum mechanical background, dating from the initial construction of quantum mechanics [73, 74]. Then, in the context of Q^2SAR, the fundamental equation can be solved and the associated Q^2SAR models later obtained employing simple geometrical procedures [19], transformations like stochastic scaling as discussed before, or any other coherent plausible procedure. Moreover, because of the need of testing the statistical significance of the obtained classical QSAR models, as a consequence of the empirical nature of the linear hypothesis in the classical case, within classical QSAR modeling premises there cannot be established any causal relationship between the molecular descriptors, appearing as model parameters connected to the structure and the modeled QO property. Instead, Q^2SAR models, which in any case are obtained by solving the fundamental Q^2SAR equation, produce a causal link, as tenuous as it may be, between SM elements, representing the QO structure and the studied property in a user bias-free context. Generality and uniqueness of the molecular descriptors in Q^2SAR can also be invoked here and stressed before classical QSAR procedures. This can be so because the SM elements, being computed from the quantum mechanical DF, which are associated in turn to the QO wave functions, are general in the sense that they can be employed to represent any QO structure: from nuclei to molecules. For the same reason, QSM descriptors are unique due to the associated wave functions uniqueness, as commented on above. This property produces, therefore, the consequent QSM parameter property of being user bias-free. Parameters in Q^2SAR cannot be freely chosen but are computed instead within the framework of quantum theoretical reasoning and employing the subsequent algorithms. Finally, on the contrary to what happens in the classical QSAR framework, there is a linear equation relying QO descriptors and QO observable properties, the fundamental Q^2SAR equation, which can be deduced from quantum mechanical principles only and some straightforward mathematical reasoning.

Thus, after these considerations, it might be assured that Q^2SAR can produce, by means of universal and unbiased QO descriptors, causal models, based on a fundamental Q^2SAR equation, which connect, in a discrete quantum framework, QO structure and observable properties.
APPROXIMATE SOLUTIONS OF THE FUNDAMENTAL Q²SAR EQUATION

The fundamental Q²SAR Eq. (21), can be solved in various ways as discussed in several preceding works [19, 38, 75]. These earlier studies have leaned on the fact that the fundamental Q²SAR equation solutions could be constrained to be positive definite, in the same way as the known SM and, in many cases, as the property vectors are positive definite, too. Here, a general solution-seeking formalism will be presented, reviewing a possible way to obtain approximate unconstrained solutions of the fundamental Q²SAR equation. The description of the plausible algorithm will be presented here schematically and as a possible example of the theoretical development, as the aim of the present study is not this subject anyway.

The SM \( Z \) or any of the possible stochastic transformations previously discussed can be diagonalized without problems; see a recent study, for example, Ref. [76], which analyzes the problem. Because it has also been proved that even the SM nonsymmetric stochastic transforms are diagonalizable, the original symmetric SM only will be used in this discussion.

Algorithm for Obtaining Approximate Solutions of the Fundamental Q²SAR Equation

Suppose the secular equation for a given SM is written as

\[
Z|\pi\rangle = \theta_i|\pi\rangle \quad (i = 1, N) \wedge \forall i, j : \langle i|j \rangle = \delta_{ij},
\]

\( \theta_i \) being the eigenvalue spectrum and \( \{|i\rangle\} \) the associated orthonormalized eigenvector set. Then the SM spectral decomposition can be written exactly as

\[
Z = \sum_i \theta_i |i\rangle \langle i|.
\]

However, due to possible redundancies in the structure of the QOS, some eigenvalues may appear to be lesser than a given threshold \( \epsilon \), and then an approximate SM, \( \approx Z \), can be reconstructed by means of a truncated spectral decomposition,*

\[
\approx Z = \sum_i \delta(\theta_i > \epsilon) \theta_i |i\rangle \langle i|,
\]

in the same way as principal components techniques use the decomposition of the variance-correlation matrix in classical QSAR model-building. An approximate SM inverse can be defined using the same technique as well, simply writing

\[
\approx Z^{-1} = \sum_i \delta(\theta_i > \epsilon) \theta_i^{-1} |i\rangle \langle i|.
\]

From Eq. (24) it is easy to obtain an approximate solution of the fundamental Q²SAR equation as

\[
\approx w \approx \approx Z^{-1}|\pi\rangle.
\]

The approximate solution (25) will produce an estimated property vector, which can be computed in turn as

\[
|\pi\rangle = Z(\approx w) = Z(\approx Z^{-1})|\pi\rangle.
\]

Owing to the orthonormalized structure of the eigenvector set, it is easy to see that the exact and approximate property vectors are related by a projector chosen by the threshold value, which projects the exact property vector into the eigenvector subspace, yielding the approximate one,

\[
|\approx\pi\rangle = \left[ \sum_i \delta(\theta_i > \epsilon) |i\rangle \langle i| \right]|\pi\rangle
\]

\[
= \sum_i \delta(\theta_i > \epsilon) |i\rangle \langle i| \pi \rangle |i\rangle.
\]

So, the approximate properties can be associated to a linear combination of the eigenvectors, associated to eigenvalues greater than the threshold, with coefficients computed by the scalar products of the exact property vector by the chosen eigenvectors.

Error Analysis of the Approximate Solutions to Fundamental Q²SAR Equation

The residual error of the property vector will also be related to the above discussed mathematical peculiarity and can be set in the following way:

\[
|\epsilon\pi\rangle = |\pi\rangle - |\approx\pi\rangle
\]

\[
= \left[ I - \sum_i \delta(\theta_i > \epsilon) |i\rangle \langle i| \right]|\pi\rangle
\]

\[
= \sum_i \delta(\theta_i \leq \epsilon) |i\rangle \langle i| \pi \rangle |i\rangle,
\]

being the operator between brackets, at the end of the equality sequence, the projector over the eigensubspace complementary to the chosen one in order to construct the approximate matrices.

Finally, the quadratic error appears as a positive definite expression. It can be simply computed as the positive definite norm of the residual error vec-
tor (26) as follows:

\[
e^{(2)} = \langle \pi |, \pi \rangle = \left( \pi \left[ \sum_i \delta(\theta_i \leq \epsilon) \left\langle i | i \right\rangle \right] | \pi \rangle = \sum_i \delta(\theta_i \leq \epsilon) \left\langle i | i \right\rangle \right| | \pi \rangle = \sum_i \delta(\theta_i \leq \epsilon) \left| \langle i | \pi \rangle \right|^2,
\]

(27)

where the idempotency of the projector over the eigenvalue subspace has been taken into account. This last result suggests an alternative way to obtain the corresponding approximate terms into the fundamental Q²SAR equation.

**Alternative Choice of the Approximate Solution Fundamental Q²SAR Equation Eigenvector Subspace**

The quadratic error structure as obtained in Eq. (27), permits seeking for a possible minimal value of this quantity. Of course, defined as it is, related to the eigenspace of the SM, when the complementary eigensubspace is void, the minimal quadratic error value will be zero. That will be the same as to consider an exact solution of the problem computed and, thus, the use of the whole available eigenspace. Approximate solutions with increasing quadratic error can be obtained by using as complementary eigenspace any of the mono-dimensional ones constructed by just an eigenvector, then by two of them and so on. This construction way is not equivalent, in general, to the choice of the eigenvalue threshold, except in the case where a decreasing eigenvalues sequence is in a one-to-one correspondence to a decreasing scalar product set \( \{\langle i | \pi \rangle\} \) absolute values sequence. If the minimal value of the quadratic error is to be obtained, after a choice of the eigensubspace dimension, as is the usual way, then the eigenvalue–eigenvector ordering must be modified in the sense that, irrespective of the eigenvalue ordering, it will correspond to the ordering of the quadratic error elementary terms, and then the employed threshold will be referred to the minimal absolute value or to the square module of the scalar product set \( \{\langle i | \pi \rangle\} \).

In this way, suppose that the following eigenvector ordering has been obtained through the rule

\[
|\langle 1 | \pi \rangle|^2 > |\langle 2 | \pi \rangle|^2 > \cdots > |\langle p | \pi \rangle|^2 > \cdots > |\langle N | \pi \rangle|^2,
\]

and that this order has no implication on the eigenvalue order whatsoever. The reordered eigenvector set, can be used with a given threshold \( \tau \), to obtain approximate SM inverses and thus approximate property vectors with minimal residual error. For example, the approximate SM inverse could now be written as

\[
\delta Z^{-1} = \sum_p \delta(\langle p | \pi \rangle|p\rangle^2 > \tau) \theta_p^{-1} | p \rangle | p \rangle,
\]

so the approximate property vector will now be expressed as

\[
\delta \pi = \sum_p \delta(\langle p | \pi \rangle|p\rangle^2 > \tau) | p \rangle | p \rangle.
\]

As the involved scalar products are ordered in decreasing absolute value, the approximated matrices always will yield a minimal quadratic error to the approximate fundamental Q²SAR equation, which can be now computed as

\[
e^{(21)} = \sum_p \delta(\langle | p \rangle | \pi \rangle|p\rangle^2 \leq \tau) | p \rangle | p \rangle.
\]

It seems that from the discussion of the above results, the corresponding eigenvector ordering and the subsequent algorithm to obtain approximate solutions, can shed some light on the fact, often encountered in principal component analysis, that SM eigenvalue ordering does not have a general application [77, 78].

**Extended Hilbert and Sobolev Spaces**

The main idea leading to the construction of extended Hilbert spaces has been latent in the usual quantum mechanical postulates, associated to the wave function definition and properties since the dawn of quantum mechanics [79 – 87]. Certainly, in this respect the quantum mechanical wave functions, as solutions of the Schrödinger equation, will behave properly and, essentially as such, they will form part of the set of square sumable functions [81 – 84]. However, the first derivatives of quantum mechanical wave functions should also possess the same properties as their original counterparts have [84, 88]. Thus, the wave functions and their gradients must be taken, in this sense, on the same footing, while being independent from each other.

**STRUCTURE OF EXTENDED HILBERT SPACES**

These wave function characteristics, when gradients are to be taken into account as well as the
bulk wave functions themselves, originated the idea of extended Hilbert spaces. The definition of an extended Hilbert space can be cast as follows. Suppose a stationary wave function, \( \Psi(r) \), of some state of a known \( n \)-particle QO; then the vector \( r \) has to contain the coordinates of the positions associated to the QO particles. Suppose the gradient of such a wave function is symbolized by \( \nabla \Psi(r) \). An element of the extended Hilbert space can be defined as the following ket vector of the appropriate dimensions:

\[
\Phi(r) = \left( \Psi(r) \right) = \Phi = |\Phi\rangle. \tag{28}
\]

The Hilbert norm [89] of the extended wave functions can be easily defined, and can be interpreted as follows:

\[
\langle \Phi | \Phi \rangle = \langle \Psi | \Psi \rangle + \langle \nabla \Psi | \nabla \Psi \rangle = 1 + 2(K),
\]

where the system wave function \( \Psi \) is supposed to be normalized as

\[
\langle \Psi | \Psi \rangle = 1,
\]

and the symbol \( \langle K \rangle \), which can be associated to the following integrals:

\[
\frac{1}{2} \langle \nabla \Psi | \nabla \Psi \rangle = -\frac{1}{2} \langle \Psi | \nabla^2 \Psi \rangle = \langle K \rangle
\]

is nothing else but the expectation value of the quantum mechanical kinetic energy of the QO under study.

**SOBOLEV SPACES**

The resultant structure of the extended Hilbert space norms, as previously defined, can be associated to the usual form of a Sobolev space [59]. In Sobolev spaces [60], the norm of any of their elements, the function \( \Psi \) say, is defined within a general formulation as:

\[
\|\Psi\|_n^m = \sum_{p=0}^{m} \|\nabla^p \Psi\|_{n'} \tag{29}
\]

where the first term on the sum in Eq. (29) has to be considered as a way to write the bulk wave function

\[
\nabla^0 \Psi \equiv \Psi. \tag{30}
\]

It is obvious that the extended Hilbert space function from the point of view of the above-defined norms can be considered a Sobolev space with the norm defined as

\[
\langle \Phi | \Phi \rangle \equiv \|\Psi\|_1^2. \tag{31}
\]

**GENERALIZED SOBOLEV SPACES**

A trivial generalization of Sobolev spaces can be readily described, taking as a starting point the extended Hilbert spaces as defined above. Suppose the Sobolev definition (29) can be generalized in such a way that it can be written, among other possibilities, as

\[
\|\Psi\|_{n'}^s = \sum_{p=1}^{r} \|\Psi|^p\|_n + \sum_{q=1}^{s} \|\nabla^q \Psi\|_n. \tag{32}
\]

In a symmetrical way with the first term in Eq. (32), there can also be defined the following more general norm:

\[
\|\Psi\|_{n'}^s = \sum_{p=1}^{r} \|\Psi|^p\|_n + \sum_{k=1}^{t} \sum_{q=1}^{s} \|\nabla^q \Psi\|^k_{n'}. \tag{33}
\]

so, in fact, Eq. (33) transforms into Eq. (32), when \( t = 1 \), in the same way as Eq. (32) becomes the original Sobolev definition (29), when \( r = 1 \).

Taking these definitions into account and the conventional meaning of the zeroth order gradient symbol, as already treated in Eq. (30), one can first try, as an appropriate example, to reach an extended Hilbert space whose functions could be associated to some sort of a generalized Sobolev space (32) with the following norm structure:

\[
\langle \Phi | \Phi \rangle \equiv \|\Psi\|_2^2 = \|\Psi\|_2 + \|\Psi|^2\|_2 + \nabla \Psi \|_2. \]

So, there may be written anew the vector form of the extended wave functions, now expressed as a three-dimensional column matrix:

\[
|\Phi\rangle = \begin{pmatrix} \Psi \\ |\Psi|^2 \\ \nabla \Psi \end{pmatrix} = \begin{pmatrix} \rho \\ \Psi \\ \nabla \Psi \end{pmatrix}, \tag{34}
\]

taking into account that the density function \( \rho \) in the expression (34) is the squared module of the initial wave function form. The explicit extended function norm, associative to a corresponding Sobolev space, will be written now as

\[
\langle \Phi | \Phi \rangle = 1 + \langle \rho | \rho \rangle + 2(K). \tag{35}
\]

The Hilbert norm of the total density form appears in Eq. (35) as another term in the extended Hilbert space norm. The new positive definite term corresponds to an integral with the form

\[
\langle \rho | \rho \rangle = \int \rho^2 \, dr = \langle |\Psi|^2 |\Psi|^2 \rangle = \int |\Psi|^4 \, dr. \tag{36}
\]

Such integrals are well defined in the field of QSM and correspond to a QSSM overlap-like measure [10], involving the \( n \)th order density matrices.
An earlier definition of a QSSM integral can be seen in Eq. (4).

From now on, these kinds of Hilbert spaces, which can be in connection with generalized Sobolev space norms, will be referred to as extended Hilbert–Sobolev spaces.

**Expectation Values in Extended Hilbert–Sobolev Spaces**

Expectation values in extended Hilbert–Sobolev spaces are of extreme interest, because the differential parts of the involved quantum mechanical operators can be easily separated from the functional parts, as recently analyzed [54, 55]. To start with the use of this extended facility of writing wave functions, there is only the need to quote that, in extended Hilbert–Sobolev spaces, the energy expression equivalent to the usual scalar wave function can be easily obtained. For example, using the extended wave function (28) employed in the usual scalar operator case (15), the Hamiltonian operator, constructed as a diagonal matrix, in the form

\[
H = \begin{pmatrix} V(r) & 0 \\ 0 & \frac{1}{2}I \end{pmatrix},
\]

where \( V(r) \) is the potential operator; then from the extended wave function (28) employed in the usual way, the following well-known energy expectation value expression is obtained:

\[
E = \langle H \rangle = \langle \Phi | H | \Phi \rangle = \langle \Psi | V | \Psi \rangle + \frac{1}{2} \langle \nabla \Psi | \nabla \Psi \rangle = \langle V \rangle + \langle K \rangle.
\]

Considering this result carefully, the expectation value of any observable in extended Hilbert–Sobolev spaces can be written by means of an operator, constructed as a diagonal matrix, in the form

\[
Z = \begin{pmatrix} W & 0 \\ 0 & L \end{pmatrix} = \text{Diag}(W; L).
\]

Then a convenient expectation value expression is obtained using the extended wave function (28)

\[
Z = \langle Z \rangle = \langle \Phi | Z | \Phi \rangle = \langle \Psi | W | \Psi \rangle + \langle \nabla \Psi | L | \nabla \Psi \rangle,
\]

and provided that none of the composite operator components contain differential terms, Eq. (38) can also be written as

\[
Z = \langle W | \Psi|^2 \rangle + \langle L | \nabla \Psi|^2 \rangle = \langle W | \rho \rangle + \langle L | \kappa \rangle,
\]

where the symbols \( \rho \) and \( \kappa \) represent the position and kinetic energy DF, respectively.

In case the circumstance appear to be forcing higher order wave function derivatives in the expectation value expression, then a higher order extended Hilbert–Sobolev space can be used to deal with this problem within the extended wave function formalism [59]. In the same way as in the extended wave function (34), a term with the kinetic energy DF can appear, for instance, in the column vector representing the extended wave function, providing the needed derivative extra terms.

**Extended Fundamental \( Q^2 \text{SAR} \) Equation**

Then, the possibility to obtain extended expectation values, by using appropriate extended wave functions and operators, can be surely employed in the theoretical framework leading to an extended fundamental \( Q^2 \text{SAR} \) equation.

The reasoning can be made in parallel terms to those employed in the manner provided within the section on the fundamental \( Q^2 \text{SAR} \) equation. A QO property can also be associated with an expectation value expression in extended Hilbert–Sobolev spaces.

For instance, using Eq. (39) along with the wave function-operator context where it has been deduced, there can be written the extended form of Eq. (17) adapted to the structure of the extended expectation value (39),

\[
\langle \omega \rangle = \int \int [W_\omega \omega + L_\omega \kappa] d\mathbf{r}_0 d\mathbf{r},
\]

where the extended operator components, defined in Eq. (37), have been split into a product of two operators as in the scalar operator case (15). Then, the unknown operators \( W_\omega \) and \( L_\omega \) can be expressed in terms of the DF tags associated with the particular DF \( \rho \) and \( \kappa \), respectively, so that

\[
W_\omega \approx \sum_i w_i \rho_i \wedge L_\omega \approx \sum_i \lambda_{ij} \kappa_j.
\]

Then, if these approximate operator forms are substituted into the extended expectation value integral, in such a way that the expectation value of a given QO, \( A \), say, can be obtained by means of the expression

\[
\langle \omega_A \rangle = \sum_i w_i z_{iA} + \sum_j \lambda_{ij} u_{iA}
\]

where use has been made of the QSM over the electronic DF,

\[
z_{iA} = \int \int \rho_i(\mathbf{r}) \Omega_{ij}(\mathbf{r}; \mathbf{r}_0) \rho_A(\mathbf{r}_0) d\mathbf{r}_0 d\mathbf{r}
\]
In this way, the extended Eq. (45) becomes the fundamental Q2SAR equation. That is, the unknown extended Q2SAR problem. Provided there is a QOS, whose elements possess a property, attached to some expectation value, and with its values ordered as a vector: \(|\pi\rangle = \{\langle \omega_0 \rangle \} \equiv \{\pi_l\}\), then Eq. (40) can be written in matrix form as in the previous scalar framework

\[ |\pi\rangle = Zw + Uv, \]  

(43)

whenever the symmetric matrices \(Z = \{z_{ik}\}\) and \(U = \{u_{il}\}\) collect, respectively, the QSM integrals of types (41) and (42) between the QOS elements.

Equation (43) becomes equivalent to the hypervector equation

\[ |\pi\rangle = (Z \ U)\begin{pmatrix} w \\ v \end{pmatrix}, \]  

(44)

and can be brought into a square linear system by using the simple transformation consisting of left multiplying both sides by the extended SM transpose vector. Owing to the symmetry of both SM components,

\[ \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} Z^2 & ZU \\ UZ & U^2 \end{pmatrix}\begin{pmatrix} w \\ v \end{pmatrix}, \]  

(45)

with the left-hand side two-component hypervector defined by the property vector transform by means of both SM components. That is,

\[ \begin{pmatrix} p \\ q \end{pmatrix} = (Z \ U)|\pi\rangle = (Z|\pi\rangle |\pi\rangle). \]

In this way, the extended Eq. (45) becomes the fundamental Q2SAR equation for the extended Hilbert–Sobolev space expectation value discretization problem.

An important remark shall be made here. When expressing the components of the involved hypermatrices in Eq. (45) by means of a unique symbol, then it is straightforward to realize that such an extended form of the fundamental Q2SAR equation is formally equivalent to the former scalar wave function counterpart represented by Eq. (21). The most striking differences consist of the presence of a SM originated by the kinetic energy DF and the doubled dimensions of the linear system leading to the extended fundamental Q2SAR equation. Consequently, the transformations provided schematically in the scalar setup apply here in the same way as before. For example, stochastic transformations can be performed in this extended case with small irrelevant modifications.

### Nonlinear Fundamental Q2SAR Equation

In the manner that the extended Hilbert–Sobolev space structure allows the study of nonlinear Schrödinger equations [59], the same ideas can be used to include nonlinear terms in extended fundamental Q2SAR equations. This can be achieved in the simplest way by using extended wave functions like the one depicted in Eq. (34), associated to a Sobolev norm as in Eq. (35). In this particular case, the associated operators no longer can be a diagonal matrix with two terms as in the definition (37); an extra diagonal term will be taken into account, so the appropriate operator form will become in this case,

\[ Z = \begin{pmatrix} W & 0 & 0 \\ 0 & P & 0 \\ 0 & 0 & L \end{pmatrix} = \text{Diag}(W; P; L), \]

in order that the expectation values, provided that no component of the operator bears differential terms, can be expressed this time as a triple contribution,

\[ Z = \langle Z \rangle = \langle W \rho \rangle + \langle P \rho^2 \rangle + \langle L \kappa \rangle. \]  

(46)

Almost no differences seems to be present in Eq. (46) with respect to the initial extended expectation value written as in expression (39), except for the central term on the right side. This term precisely includes a nonlinear electronic DF structure in the associated Schrödinger equation and in the expectation value (46) as well.

In order to grasp the form of such nonlinear terms in the associated extended fundamental Q2SAR equations, a short discussion will be given as an example. From the knowledge of QOS, it is easy to see that both integrands in the nonlinear term: \(\langle P | \rho^2 \rangle\) can be approximately expressed as superpositions of the elements of the attached QOS DF tag set: \(P = \{\rho_l\}\), say [51, 52, 56]. This means, for example,

\[ P = \sum_l a_l \rho_l \wedge \rho = \sum_l \beta_l \rho_l, \]

with the linear combination of the DF being necessarily convex; see Definition 6 in the Appendix. Then, using these superpositions in the nonlinear
integral, one obtains
\[
\langle p | \rho^2 \rangle = \sum_I \sum_J \sum_K \alpha_I \beta_J \beta_K \langle \rho_I \rho_J \rho_K \rangle,
\]

where the symbol \( \langle \rho_I \rho_J \rho_K \rangle \) stands for a triple density QSM, already described in the section on quantum similarity measures, as well as in previous studies [65, 66], and even employed in QSAR models under QSM descriptors [14, 30]. Thus, a possible particular choice permits the definition of such a QSM, which can be formally written as
\[
\langle \rho_I \rho_J \rho_K \rangle = \int \rho_I(\mathbf{r}) \rho_J(\mathbf{r}) \rho_K(\mathbf{r}) d\mathbf{r}.
\]

So, triple density QSM, already defined several years ago [65], can now be also related to the introduction of nonlinear terms, within the extended Hilbert–Sobolev formalism, in the quantum mechanical expectation values expression.

To construct the associated extended fundamental Q²SAR equation from the extended expectation value expression (46) now becomes an elementary task. An equation similar to Eq. (44) can be easily built up, where instead of two-component hypermatrices both in the SM and unknown coefficient parts, there will appear a three-component submatrix structure. The new extended fundamental Q²SAR equation in matrix form is so straightforward that it seems useless to write here the corresponding form. In any case, the transformations put forward in the earlier occurrence can be exactly employed here, and all the remarks, previously made, apply in the present example.

Generalized Fundamental Q²SAR Equation

The previous discussion can lead to a natural generalization of the expectation value concept within the extended Hilbert–Sobolev space framework and hence of the buildup of an extended fundamental Q²SAR equation as general as possible.

The results obtained so far show, without doubt, that not only first-order terms on the density function can play a role in the definition of nonlinear extended fundamental Q²SAR equations, but that nothing opposes considering more general nonlinear structures. In order to provide an example of this possibility for both electronic and kinetic DF, a schematic discussion will be developed.

Supposing that both QO DF: \( \{ \rho; \kappa \} \) are known, then the following structure of an extended nonlinear operator can be used within the extended wave function form, provided with Eq. (28),
\[
Z = \text{Diag}(W + a(\exp(\alpha \rho) - 1); L + b(\exp(\beta \kappa) - 1)),
\]

where \( \{ a; \alpha \} \) and \( \{ b; \beta \} \) are parameters associated to the nonlinear DF contributions. The zeroth order operator will be a classical one as provided in Eq. (37) and the present exponential terms will produce the nonlinear contributions in both DF classes.

It is easy to see, employing a trivial Taylor series expansion on every exponential, that the included \( p \)th order operator nonlinear terms could be written in this case as
\[
Z_p = \frac{1}{p!} \text{Diag}(a(\alpha \rho)^p; b(\beta \kappa)^p); \quad \forall p \geq 1. \quad (47)
\]

The corresponding expectation value can be obtained considering the \( p \)th order operator term (47) as an isomorphic two-component vector. In this case the following extended density vector can be also defined:
\[
|P_p\rangle = \left(\begin{array}{c} \rho \\ \kappa \end{array}\right),
\]

so the \( p \)th order contribution to the operator expectation value could be simply written as
\[
\langle Z_p \rangle = \langle P | Z_p \rangle = \frac{1}{p!} (a \alpha \rho | \rho^p \rangle + b \beta \kappa | \kappa^p \rangle) = \frac{1}{p!} (a \alpha \rho | \rho^p \rangle + b \beta \kappa | \kappa^p \rangle). \quad (48)
\]

A general framework containing nonlinear terms in the matrix form of the extended fundamental Q²SAR equation is set in this manner. It must finally be noted that the nature of these \( p \)th order correction terms can be associated to (\( p + 1 \))th order QSSM, involving both DF classes: \( \{ \rho^p; \kappa^p \} \).

The matrix form, associated to such a development becomes, apart from the internal term matrix structure complexity, formally equivalent to the linear system (44), and thus to the scalar fundamental Q²SAR equation as commented before. Then, although the quantum mechanical structure of the corrections, represented by the \( p \)th order terms (48), may become impressively complex from the computational point of view, the final form of the extended fundamental Q²SAR equation remains extremely simple.
Conclusions

A general framework where extended fundamental Q²SAR equations can be defined up to any complexity level has been developed. Extended Hilbert–Sobolev spaces as well as the previously defined mathematical elements, related to QSM as resulting integrals over the QO tags made of DF, have been shown to be a coherent sequence of logical elements resulting, as far as is now possible, in a complete general background of the fundamental Q²SAR equations, which connect QO structure and properties. The theoretical basis of classical QSAR procedures, even those including nonlinear terms, appear to be founded in quantum mechanical grounds.

Appendix: Definitions Related to QSM

Definition 1: Tagged Sets. Let us suppose a given set, the Object Set, \( S \), and another set, made of some chosen mathematical elements, which will be hereafter called tags, forming a Tag Set, \( T \). A Tagged Set, \( Z \), can be constructed by the ordered product, \( Z = \delta \times T \),

\[
Z = \{ \forall \theta \in Z \mid \exists s \in \delta \wedge \exists t \in T \rightarrow \theta = (s,t) \}.
\]

Definition 2: Quantum Object (QO). A QO can be defined as an element of a Tagged Set. Quantum Systems in well-defined states are taken as the Object Set part and the corresponding Density Functions constitute the Tag Set part.

Definition 3: Vector Semispace (VSS). A VSS over the positive definite real field \( \mathbb{R}^+ \), is a Vector Space with the vector sum part provided by a structure of an abelian semigroup.

An additive semigroup is understood here to mean an additive group without the presence of reciprocal elements. All VSS elements can be seen as directed toward the region of the positive axis hyperquadrant. It can be accepted if necessary that null elements are included in the scalar field as well as in the VSS structure.

Definition 4: First Order Electronic Density Function (eDF). The first-order eDF form, as expressed within MO theory, can be defined by means of the linear combination,

\[
\rho(r) = \sum_i w_i |\psi_i(r)|^2.
\]

This MO eDF can be written in a general way, as a double sum of products of function pairs, coupled with a set of matrix coefficients. However, a simple matrix digitalization, followed by a unitary MO basis set transformation, can revert DF to the formal expression in Eq. (A1). The coefficient set \( W = \{w_i\} \subset \mathbb{R}^+ \), interpreted as MO occupation indices, corresponds to a collection of positive real numbers. A unit norm convention can be adopted:

\[
\int |\psi|^2 \, dr = 1; \quad \forall i
\]

\[
\Rightarrow \int \rho(r) \, dr = \sum_i w_i \int |\psi|^2 \, dr = \sum_i w_i = 1
\]

and this results in considering the coefficient set \( W = \{w_i\} \), as a discrete probability distribution.

Definition 5: Sum of Elements of a Matrix. The symbol \( \langle A \rangle \) has been used in many applications [90], and represents the sum of all the elements of matrix \( A \). So, for \((n \times m)\) matrices: \( A = \{a_{ij}\} \), then:

\[
\langle A \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}.
\]

Definition 6: Convex Conditions. By the term convex conditions applied to a vector is understood,

\[
\mathcal{X}_n(w) \equiv \{ w \in \mathcal{V}_n(\mathbb{R}^+) \wedge \langle w \rangle = \sum_i w_i = 1 \}.
\]

The set of the vector elements, \( \mathbf{w} = \{w_i\} \), can be used instead in the convex conditions symbol; that is,

\[
\mathcal{X}_n(\{w_i\}) \equiv \{ \forall i: w_i \in \mathbb{R}^+ \wedge \sum_i w_i = 1 \}.
\]

The coefficients of a first-order eDF fulfill a set of convex conditions, and thus become an application example of the convex conditions.

Definition 7: General Quantum Similarity Measure (QSM). A General QSM, \( G(\Omega) \), can be considered a positive definite multiple scalar product defined by a contracted \( \nu \)-direct product of a QOS, \( T \),

\[
G(\Omega) : \bigotimes_{\nu=1}^{\nu} T \rightarrow \mathbb{R}^+.
\]

This allows mixing \( \nu \) DF: \( \rho_l(r), l = 1, \nu \), of the QOS with \( \omega \) PD operators, collected into a set, \( \Omega = \{\Omega_k(r), K = 1, \omega\} \), belonging to the same VSS, for
example,

\[ G(\Omega) = \int \left[ \prod_{k=1}^{\omega} \Omega_k(r) \right] \left[ \prod_{i=1}^{\nu} \rho_i(r) \right] dr, \]

where the coordinate vector, \( r \), shall be taken here as a general position vector.

References

60. Sobolev, S. L. Math Sc 1938, 4, 471.