Chemical algebra.
III: Thermochemical approach to completely G-invariant distances

Remi Chauvin

Laboratoire de Chimie de Coordination du C.N.R.S., Unité 8241, liée par convention à l'Université Paul Sabatier, 205 Route de Narbonne, 31077 Toulouse Cedex, France

Received 7 July 1994

The search for a definition of distances over sets of skeletal analogs (identified to G-Hilbert spaces of vector ligand parameters) is initiated from the algebraic formulation of the constant of stereogenic pairing equilibria (pairing product). A basic definition equation is devised from thermodynamical speculations. The equation is proved to have always a single potential distance solution D_p as soon as the pairing product is discriminating. The equation of D_p is constructed in order to satisfy three consistency requirements: complete G-invariance (arbitrary orientations selected to describe skeletal analogs do not affect the value of D_p); extension properties (D_p coincides with two standard completely G-invariant distances or with the Euclidean distance in borderline cases); all the distance properties except, perhaps, the triangular inequality. The latter point remains challenging in general, and is computationally verified in some examples.

1. Introduction

In current molecular representations, constituting atoms are located at sites of a symmetrized skeleton. Chemical transformations take place either at the level of the skeletal geometry (rearrangement, condensation, etc.) or at the level of the ligand distribution (isomerization, substitution, etc.). In quantitative approaches, the constituting atoms are characterized by ligand parameters, and the structural modification brought about by the assignment of these parameters can be formally quantified through the notion of "fuzzy symmetry subgroup" [1]. Treatment of stereogenic pairing equilibria between skeletal analogs has received a fairly consistent algebraic formulation [2,3]: the trend of the results proved hitherto prompts us to explore further the mathematical scope addressed by the basic hypotheses below:

(1) *Skeleton symmetrization*: skeleton of interacting molecules are identically symmetrized in a realistic manner.

(2) *Skeleton overlap pairing*: the geometry of the paired species is the juxtaposition of the two skeletons, in such a manner that they are parallel and close to each other.

(3) *Scalar product form of the ligand interactions*: only one kind of pairwise ligand interactions occurs, and the corresponding energy is proportional to a scalar product between real or vector ligand parameters.

When two molecules are supposed to pair under such conditions, the so-called chemical pairing (constant K) is shifted towards homo-pairing for attractive-type interactions in several general situations, and hetero-pairing has never been found to be favoured in any particular situation hitherto considered. In addition, the shift vanishes only if the molecules are chemically equivalent. On the very outset, the study is restricted to pairing equilibria where K satisfies

(a) $K \geqslant 1$.

(b) $K = 1$ only if the paired species are chemically equivalent.

The corresponding function $K(\mathbf{u}, \mathbf{v})$ is then called a "discriminating pairing product" [4]. These properties of a discriminating pairing product call to mind two of the properties of a distance D, namely:

(a′) $0 \leqslant D(\mathbf{u}, \mathbf{v})$,

(b′) $D(\mathbf{u}, \mathbf{v}) = 0 \Rightarrow \mathbf{u} = \mathbf{v}$.

Efforts are now made to exhibit a distance D_p on E/G (set of skeletal analogs with a given skeleton symmetry group G) from a discriminating pairing product defined on $(E, d : G)$ (E = set of skeletal analogs with all possible orientations generated by the skeletal symmetry group G; d = Euclidean distance of the ligand parameter space). D_p might also be considered as a completely G-invariant distance on E [5].

Standard completely G-invariant distances on metric spaces (E, d) are defined by

$$\forall(\mathbf{u}, \mathbf{v}) \in E^2, \quad D_\infty(\mathbf{u}, \mathbf{v}) = \operatorname*{Inf}_{g \in G, h \in G} d(g\mathbf{u}, h\mathbf{v})$$

and this notion is related to the notion of discriminating pairing product in the limit $p = -a/kT \rightarrow \infty$. Indeed, proposition 1 has been proved previously: [4].

PROPOSITION 1

When $p \rightarrow \infty$, K_p converges to a discriminating pairing product K_∞:

$$K_\infty(\mathbf{u}, \mathbf{v}) = \lim_{p \rightarrow \infty} K_p(\mathbf{u}, \mathbf{v}) = \exp\left[\operatorname*{Inf}_{g \in G} d^2(g\mathbf{u}, \mathbf{v})\right],$$

where

$$K_p(\mathbf{u},\mathbf{v}) = \frac{\left(\int_G \exp\left[-\frac{p}{2}d^2(g\mathbf{u},\mathbf{u})\right] dg\right)^{1/p} \left(\int_G \exp\left[-\frac{p}{2}d^2(g\mathbf{v},\mathbf{v})\right] dg\right)^{1/p}}{\left(\int_G \exp\left[-\frac{p}{2}d^2(g\mathbf{u},\mathbf{v})\right] dg\right)^{2/p}}.$$

Moreover, the limit of $\sqrt{\ln K_p}$ when $p \to +\infty$, i.e. the function D_∞: $E \times E \to R$, $(\mathbf{u},\mathbf{v}) \to \sqrt{\ln K_\infty(\mathbf{u},\mathbf{v})} = \mathrm{Inf}_{g\in G}\, d(g\mathbf{u},\mathbf{v})$ is a completely G-invariant distance on E, i.e. a distance on the set E/G of all the orbits.

The distance D_∞ is the limit for $q \to +\infty$ of the series

$$D'_q(\mathbf{u},\mathbf{v}) = \frac{1}{\left(\int_G \left[\frac{1}{d(g\mathbf{u},\mathbf{v})}\right]^q dg\right)^{1/q}}, \quad q > 1.$$

Any positive function D'_q satisfies all the properties of a distance on E/G but the triangular inequality. However, not only the limit $D'_\infty (q \to +\infty)$, but also the limit $D'_1 (q \to 1)$ satisfy the triangular inequality and are completely G-invariant distances

PROPOSITION 2

Under the preceding notations, D'_1 is a completely G-invariant distance on (E, d).

Proof of the triangular inequality:

$$\forall (\mathbf{u},\mathbf{v},\mathbf{w}) \in E^2, \quad D'_q(\mathbf{u},\mathbf{v}) + D'_q(\mathbf{w},\mathbf{v})$$
$$= \frac{1}{\left(\int_G \left[\frac{1}{d(g\mathbf{u},\mathbf{v})}\right]^q dg\right)^{1/q}} + \frac{1}{\left(\int_G \left[\frac{1}{d(g\mathbf{w},\mathbf{v})}\right]^q dg\right)^{1/q}}.$$

Thus,

$$\forall h \in G, \quad D'_q(\mathbf{u},\mathbf{v}) + D'_q(\mathbf{w},\mathbf{v})$$
$$= \frac{1}{\left(\int_G \left[\frac{1}{d(g\mathbf{u},\mathbf{v})}\right]^q dg\right)^{1/q}} + \frac{1}{\left(\int_G \left[\frac{1}{d(gh\mathbf{w},\mathbf{v})}\right]^q dg\right)^{1/q}}.$$

The triangular inequality of the q-norm of the $1/d$ function produces

$$D'_q(\mathbf{u},\mathbf{v}) + D'_q(\mathbf{w},\mathbf{v}) \geqslant \frac{1}{\rho}\left(\int_G \left[\frac{1}{d(g\mathbf{u},\mathbf{v})} + \frac{1}{d(gh\mathbf{w},\mathbf{v})}\right]^q dg\right)^{1/q},$$

where

$$\rho = \left(\int_G \left[\frac{1}{d(g\mathbf{u},\mathbf{v})}\right]^q dg\right)^{1/q} \cdot \int_G \left[\frac{1}{d(gh\mathbf{w},\mathbf{v})}\right]^q dg)^{1/q}.$$

Since d is a distance, $d(g\mathbf{u},\mathbf{v})+d(gh\mathbf{w},\mathbf{v})\geqslant d(g\mathbf{u},gh\mathbf{w})=d(\mathbf{u},h\mathbf{w})$:

$$D'_q(\mathbf{u},\mathbf{v})+D'_q(\mathbf{w},\mathbf{v})\geqslant\frac{1}{\rho}\left(\int_G\left[\frac{d(\mathbf{u},h\mathbf{w})}{d(g\mathbf{u},\mathbf{v})\cdot d(gh\mathbf{w},\mathbf{v})}\right]^q dg\right)^{1/q}.$$

Let us calculate the ratio R of the right term to $D'_q(\mathbf{u},\mathbf{w})$:

$$R=\frac{1}{D'_q(\mathbf{u},\mathbf{w})}\cdot\frac{1}{\rho}\cdot\left(\int_G\left[\frac{d(\mathbf{u},h\mathbf{w})}{d(g\mathbf{u},\mathbf{v})\cdot d(gh\mathbf{w},\mathbf{v})}\right]^q dg\right)^{1/q}$$

$$=\left(\int_G\left[\frac{1}{d(g\mathbf{u},\mathbf{w})}\right]^q dg\right)^{1/q}\cdot\frac{1}{\rho}\cdot\left(\int_G\left[\frac{d(\mathbf{u},h\mathbf{w})}{d(g\mathbf{u},\mathbf{v})\cdot d(gh\mathbf{w},\mathbf{v})}\right]^q dg\right)^{1/q}.$$

This can be written for any operation h, and in particular for the element h giving rise to the supremum of the integral wherein it takes place:

$$R=\left(\int_G\left[\frac{1}{d(\mathbf{u},h\mathbf{w})}\right]^q dh\right)^{1/q}\cdot\frac{1}{\rho}\cdot\operatorname*{Sup}_{h\in G}\left(\int_G\left[\frac{d(\mathbf{u},h\mathbf{w})}{d(g\mathbf{u},\mathbf{v})\cdot d(gh\mathbf{w},\mathbf{v})}\right]^q dg\right)^{1/q}.$$

The function "$\mathrm{Sup}_{h\in G}$" is a p-norm limit for $p=\infty$. Therefore, the Hölder inequality for conjugate exponents (p,q) satisfying $1/p+1/q=1$ can be applied as soon as $q=1$:

$$R\geqslant\frac{1}{\rho}\cdot\int_G\frac{1}{d(\mathbf{u},h\mathbf{w})}\int_G\frac{d(\mathbf{u},h\mathbf{w})}{d(g\mathbf{u},\mathbf{v})\cdot d(gh\mathbf{w},\mathbf{v})}\,dg\,dh$$

$$=\frac{1}{\rho}\cdot\int_G\frac{dg}{d(g\mathbf{u},\mathbf{v})}\cdot\int_G\frac{dk}{d(k\mathbf{w},\mathbf{v})}=1.$$

Thus, if $q=1$:

$$R=\frac{1}{D'_q(\mathbf{u},\mathbf{w})}\cdot\frac{1}{\rho}\cdot\left(\int_G\left[\frac{d(\mathbf{u},h\mathbf{w})}{d(g\mathbf{u},\mathbf{v})\cdot d(gh\mathbf{w},\mathbf{v})}\right]^q dg\right)^{1/q}\geqslant 1.$$

Consequently,

$$\frac{1}{\rho}\cdot\int_G\frac{d(\mathbf{u},h\mathbf{w})}{d(g\mathbf{u},\mathbf{v})\cdot d(gh\mathbf{w},\mathbf{v})}dg\geqslant D'_1(\mathbf{u},\mathbf{w}).$$

The triangular inequality for $q=1$ and proposition 2 follow. D'_1 is opposite to D'_∞ in the following sense: whereas in D'_∞, the weights of the $d(g\mathbf{u},\mathbf{v})$ equal 0 or 1, in the distance D'_1, the relative weights of the $d(g\mathbf{u},\mathbf{v})$'s are smoothed off. The concern resides in the design of completely G-invariant distances such as D'_1 where all the "covering gaps" (i.e.: $d(g\mathbf{u},\mathbf{v})$, $d(g\mathbf{u},\mathbf{u})$ and $d(g\mathbf{v},\mathbf{v})$, $g\in G$) occur, without neglecting a priori all of those which are not the smallest ones. A possible distance $D_p(\mathbf{u},\mathbf{v})$ with $p<\infty$ is sought as a solution of some functional equation involving the discriminating pairing product K_p:

$$\Phi(D_p) = K_p^p .$$

Even if the distance d is not yet supposed to be Euclidean, the following properties are required in order to devise a consistent equation:

(i) G-invariance requirements:
(a) D_p is completely G-invariant, i.e. the functional equation linking K_p and D_p must preserve the complete G-invariance of K_p.
(b) $\forall \mathbf{u} \in \mathbf{E}, \forall g \in G, D_p(\mathbf{u}, g\mathbf{u}) = 0$.

(ii) Extension requirements:
(a) D_p is an extension of the metric distance d: if $\mathbf{u}_0$ is invariant to all the operations of G, then: $\forall \mathbf{u} \in E, \forall g \in G, D_p(\mathbf{u}, \mathbf{u}_0) = d(g\mathbf{u}, \mathbf{u}_0) = d(\mathbf{u}, \mathbf{u}_0)$. In particular, if E is a normed vector space: $D_p(\mathbf{u}, \mathbf{0}) = \|\mathbf{u}\|$.
(b) When $p \to \infty$, D_p tends to the standard completely G-invariant distance D_∞.
(c) When $p \to 0$, D_p tends to the "smooth" completely G-invariant distance D'_1.

(iii) Distance properties on E/G:
(a) $\forall (\mathbf{u}, \mathbf{v}) \in E^2, 0 \leqslant D_p(\mathbf{u}, \mathbf{v})$.
(b) $\forall (\mathbf{u}, \mathbf{v}) \in E^2, D_p(\mathbf{u}, \mathbf{v}) = 0 \Rightarrow \exists h \in G, \mathbf{v} = g\mathbf{u}$ (converse of (ib)).
(c) $\forall (\mathbf{u}, \mathbf{v}, \mathbf{w}) \in E^3, D_p(\mathbf{u}, \mathbf{w}) \leqslant D_p(\mathbf{u}, \mathbf{v}) + D_p(\mathbf{v}, \mathbf{w})$.

2. Thermochemical speculations

Consider the chemical equilibrium (E) occurring under the conditions mentioned above:

$$2AB \rightleftarrows AA + BB, \quad K = \frac{[BB][AA]}{[AB]^2}, \tag{E}$$

each canonical set of copies of an observable species, say AB, is a set of molecular states $\{A_iB_j\}$. The equilibrium is now formally regarded as a set of elemental equilibria between states of the involved paired species:

$$(A_iB_l + A_kB_j \rightleftarrows)\quad A_iB_j + A_kB_l \rightleftarrows A_iA_k + B_jB_l : \quad K_{ij,kl} .$$

The search for a quantity D allowing the equilibrium (E) to be defined as a "superposition" of these elemental equilibria, is undertaken from the equation:

$$\begin{aligned} K = \exp\left[-\frac{\Delta G^0}{RT}\right] &= \frac{1}{N}\sum_{i,j,k,l} [K_{ij,kl}]^{D^2/a_{ij,kl}} \quad (N: \text{ normalizing factor}) \\ &= \frac{1}{N}\sum_{i,j,k,l} \exp\left[-\frac{1}{kT}\{E(A_iA_k) + E(B_jB_l) \right. \\ &\qquad \left. -E(A_iB_j) - E(A_kB_l)\}\frac{D^2}{a_{ij,kl}}\right] . \end{aligned}$$

The $K_{ij,kl}$'s express the relative abundances of A_iA_k, B_jB_l, A_iB_j, and A_kB_l in the whole equilibrium (E). D^2 represents an energy defined with respect to the lowest energy state on one side of the equilibrium, say the right-hand side $AA + BB$, and takes into account the competitive cross-pairing process: $A_iB_l + A_kB_j \rightleftarrows A_iB_j + A_kB_l$. Coefficients $a_{ij,kl}$ are thus required to weight the $K_{ij,kl}$'s in K. They are defined by

$$a_{ij,kl} = (\alpha_{il}\alpha_{kj})^{1/2}$$

with $\alpha_{il} = E_0(AA) + E_0(BB) - 2E(A_iB_l)$, $\alpha_{kj} = E_0(AA) + E_0(BB) - 2E(A_kB_j)$, where $E_0(AA)$ and $E_0(BB)$ denote the ground states of the homo-pairs.

Although D is related to the standard free energy of the equilibrium, no clear macroscopical interpretation is claimed: D^2 represents a mean cohesion energy, averaged over molecular states, of a species AB with respect to disproportionation products AA and BB. These definitions are applied to pairing equilibria of skeletal analogs represented by vectors **u** and **v**:

$$2\mathbf{u}/\mathbf{v} \rightleftarrows \mathbf{u}/\mathbf{u} + \mathbf{v}/\mathbf{v}\,.$$

The states $(\mathbf{u}/\mathbf{v})_{ij}$, $(\mathbf{u}/\mathbf{v})_{kl}$, $(\mathbf{u}/\mathbf{u})_{ik}$, $(\mathbf{v}/\mathbf{v})_{jl}$ correspond to "stereoisomeric states" of observable pairs $AB = \mathbf{u}/\mathbf{v}$, $AA = \mathbf{u}/\mathbf{u}$, $BB = \mathbf{v}/\mathbf{v}$. Under the three basic hypotheses used in the algebraic treatment of stereogenic equilibria, the energy of paired species is expressed as a function of vector ligand parameters. It follows that the thermodynamic equation of D takes a form (given in theorem 1 below) suitable for a systematic study.

3. Definition equation

In what follows, p is a positive number which is supposed to have the thermodynamic significance: $p = -a/kT$, with $a < 0$.

THEOREM 1

Let K_p be a discriminating pairing product on a metric space $(E, d : G)$, and let us consider the equation of an unknown function D_p: $E \times E \to R_+$:

$$\Phi_{\mathbf{u},\mathbf{v}}(D_p(\mathbf{u},\mathbf{v})) = [K_p(\mathbf{u},\mathbf{v})]^p \qquad (\mathbb{E})$$

with

$$\Phi_{\mathbf{u},\mathbf{v}}(x) = \iiiint_{G^4} \times \exp\left[\frac{p}{2}\frac{d^2(g\mathbf{u}, l\mathbf{v}) + d^2(k\mathbf{u}, h\mathbf{v}) - d^2(g\mathbf{u}, k\mathbf{u}) - d^2(l\mathbf{v}, h\mathbf{v})}{d(g\mathbf{u}, h\mathbf{v}) \cdot d(k\mathbf{u}, l\mathbf{v})} x^2\right] \times dg\, dh\, dk\, dl\,,$$

$$K_p^p(\mathbf{u},\mathbf{v}) = \frac{\int_G \exp\left[-\frac{p}{2}d^2(g\mathbf{u},\mathbf{u})\right] dg \cdot \int_G \exp\left[-\frac{p}{2}d^2(g\mathbf{v},\mathbf{v})\right] dg}{\left(\int_G \exp\left[-\frac{p}{2}d^2(g\mathbf{u},\mathbf{v})\right] dg\right)^2} .$$

If E is a complex vector space (E is a G-Hilbert space), then d is *Euclidean* and eq. ($\mathbb{E}$) has a single solution D_p. Furthermore, D_p fulfills the afore-mentioned consistency requirements (i), (ii) and (iii), except, perhaps, the triangular inequality (iiic) [6].

Notice:

Under such conditions, D_p fullfills at least as many requirements as D'_{p+1} (proposition 2). However, contrary to p in D_p, q has no thermochemical meaning in D'_q.

Proof

The following notation is adopted:

$$\forall(\mathbf{a},\mathbf{b}) \in E^2 \cos(\mathbf{a},\mathbf{b}) = \operatorname{Re}\frac{(\mathbf{a}|\mathbf{b})}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|} .$$

Owing to the isometric character of the action of G on E, the integral over G^4 boils down to an integral over G^3:

$$\Phi_{\mathbf{u},\mathbf{v}}(D_p) = \iiint_{G^3} \exp[p\cos(g\mathbf{u} - h\mathbf{v}, k\mathbf{u} - \mathbf{v})D_p^2(\mathbf{u},\mathbf{v})]\, dg\, dh\, dk .$$

We aim at proving that the map Φ: $R_+ \to [1,+\infty]$, $\Phi(x) = \Phi_{\mathbf{u},\mathbf{v}}(\sqrt{x})$ is biunivocal.

Let us first notice that $\Phi(0) = 1$.

$$\bullet \forall x \in R_+, \quad \Phi'(x) = \iiint_{G^3} p\cos(g\mathbf{u} - h\mathbf{v}, k\mathbf{u} - \mathbf{v}) \times \exp[p\cos(g\mathbf{u} - h\mathbf{v}, k\mathbf{u} - \mathbf{v})\, x]\, dg\, dh\, dk$$

and

$$\Phi'(0) = p\iiint_{G^3} \cos(g\mathbf{u} - h\mathbf{v}, k\mathbf{u} - \mathbf{v})\, dg\, dh\, dk ,$$

$$\Phi'(0) = p\iiint_{G^3} \operatorname{Re}\left[\frac{(g\mathbf{u}|k\mathbf{u}) + (h\mathbf{v}|\mathbf{v}) - (g\mathbf{u}|\mathbf{v}) - (h\mathbf{v}|k\mathbf{u})}{\|g\mathbf{u} - h\mathbf{v}\| \cdot \|k\mathbf{u} - \mathbf{v}\|}\right] dg\, dh\, dk .$$

Successive variable changes $g' = h^{-1}g$, $h' = hg'$, $h'' = h'^{-1}k$ and $h''' = h''^{-1}g'^{-1}$ lead to

$$\Phi'(0) = p\ \mathrm{Re}[\mathbf{I}(\mathbf{J_{uu}} + \mathbf{J_{vv}} + \mathbf{J_{uv}} - \mathbf{J_{vu}})]\,,$$

where

$$\mathbf{I} = \iint_{G^2} \frac{dg'\,dk}{\|g'\mathbf{u} - \mathbf{v}\| \cdot \|k\mathbf{u} - \mathbf{v}\|} = \left(\int_G \frac{dg}{\|g\mathbf{u} - \mathbf{v}\|}\right)^2,$$

$$\mathbf{J_{ab}} = \int_G (g\mathbf{a}|\mathbf{b})\ dg \quad (\mathbf{a}, \mathbf{b} = \mathbf{u}, \mathbf{v})\,.$$

The following proposition holds [1]:

PROPOSITION 3

Let G be a compact group, endowed with its Haar measure dg. Let $E = m_1 V_1 \oplus \ldots \oplus m_r V_r$ be the expansion of E into irreducible representations. The character of V_i is noted as χ_i and its degree is denoted as n_i. Let $\mathcal{P}_i$: $V \to m_i V_i$ be the projector of E onto the isotypical representation $m_i V_i$. Then

$$\forall g \in G\,, \quad \forall (\mathbf{a}, \mathbf{b}) \in V^2\,, \quad \int_G (gh\mathbf{a}|h\mathbf{b})dh = \sum_{i=1}^{r} \frac{(\mathcal{P}_i\mathbf{a}|\mathcal{P}_i\mathbf{b})}{n_i} \cdot \chi_i(g)\,.$$

By integration over g: $\mathbf{J_{ab}} = \int_G (g\mathbf{a}|\mathbf{b})\ dg = (\mathcal{P}_1(\mathbf{a})|\mathcal{P}_1(\mathbf{b}))$, wherein $\mathcal{P}_1$ denotes the projector onto the unit representation ($n_1 = 1$). Therefore,

$$\mathbf{J_{uu}} + \mathbf{J_{vv}} + \mathbf{J_{uv}} - \mathbf{J_{vu}} = \|\mathcal{P}_1(\mathbf{u})\|^2 + \|\mathcal{P}_1(\mathbf{v})\|^2 + (\mathcal{P}_1(\mathbf{u})|\mathcal{P}_1(\mathbf{v}))$$
$$+ (\mathcal{P}_1(\mathbf{v})|\mathcal{P}_1(\mathbf{u})) = \|\mathcal{P}_1(\mathbf{u}) - \mathcal{P}_1(\mathbf{v})\|^2\,.$$

Consequently,

$$\Phi'(0) = p\|\mathcal{P}_1(\mathbf{u}) - \mathcal{P}_1(\mathbf{v})\|^2 \left(\int_G \frac{dg}{\|g\mathbf{u} - \mathbf{v}\|}\right)^2 \geqslant 0\,.$$

$$\bullet \forall x \in R_+, \quad \Phi''(x) = \iiint_{G^3} p^2 \cos^2(g\mathbf{u} - h\mathbf{v}, k\mathbf{u} - \mathbf{v})$$
$$\times \exp[p \cos(g\mathbf{u} - h\mathbf{v}, k\mathbf{u} - \mathbf{v})\ x]\ dg\ dh\ dk \geqslant 0\,.$$

Thus, Φ' is an increasing function of x over R_+. Since $\Phi'(0) \geqslant 0$, $\Phi'(x)$ is always positive, and Φ is an increasing function over R_+. But $\Phi(0) = 1$, and Φ is an injection from R_+ into $[1, +\infty]$.

On the other hand, when $x \to +\infty$, $\Phi''(x)$ is equivalent to

$$\mathop{\mathrm{Sup}}_{(g,h,k) \in G^3} p^2 \cos^2(g\mathbf{u} - h\mathbf{v}, k\mathbf{u} - \mathbf{v}) \exp[p \cos(g\mathbf{u} - h\mathbf{v}, k\mathbf{u} - \mathbf{v})\ x] = p^2 e^{px}\,.$$

Therefore, when $x \to +\infty$, $\Phi''(x) \to +\infty$.

It follows that when $x \to +\infty$, $\Phi(x) \to +\infty$, and Φ is surjective from R_+ onto $[1, +\infty]$.

K_p is supposed to be a discriminating pairing product. In particular, $K_p \geqslant 1$. Since Φ is biunivocal from R_+ into $[1, +\infty]$, eq. ($\mathbb{E}$) allows a function D_p to be unequivocally defined. For the same reason, $D_p(\mathbf{u}, \mathbf{v}) = 0$ if and only if $K_p(\mathbf{u}, \mathbf{v}) = 1$, i.e. if and only if $\exists g \in G$, $\mathbf{v} = g\mathbf{u}$ (K_p is discriminating).

The symmetry of eq. ($\mathbb{E}$) is transferred to its solution D_p, and D_p is completely G-invariant.

Furthermore, it is easily checked that if $\mathbf{u}$ belongs to the unit representation of G ($\forall g \in G$, $g\mathbf{u} = \mathbf{u}$), then $\forall \mathbf{v} \in E$, $D_p(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$. Likewise, when $p \to +\infty$, the solution of ($\mathbb{E}$) tends to the standard completely G-invariant distance defined by (proposition 1):

$$\forall (\mathbf{u}, \mathbf{v}) \in E^2, \quad D_\infty(\mathbf{u}, \mathbf{v}) = \operatorname*{Inf}_{g \in G, h \in G} \|g\mathbf{u} - h\mathbf{v}\| = \operatorname*{Inf}_{g \in G} \|g\mathbf{u} - \mathbf{v}\| .$$

At last, expanding $\Phi_{\mathbf{u},\mathbf{v}}(D_0(\mathbf{u}, \mathbf{v}))^{1/p}$ and $[K_p(\mathbf{u}, \mathbf{v})]^p$ in the neighborhood of $p = 0$, the limit equation ($\mathbb{E}$) provides a solution D_0 which coincides with D'_1:

$$D_0(\mathbf{u}, \mathbf{v}) = D'_1(\mathbf{u}, \mathbf{v}) = \frac{1}{\displaystyle\int_G \frac{dg}{d(g\mathbf{u}, \mathbf{v})}} .$$

The map "$D_p(\mathbf{u}, \mathbf{v}) = \sqrt{\ln K_p(\mathbf{u}, \mathbf{v})}$" is a distance on E/G only if $p = +\infty$. For $p < +\infty$, this map does not fulfill the condition (ii.a) and the complete definition equation ($\mathbb{E}$) is needed. Nevertheless, D_p is always greater than $\ln K_p$. More precisely:

PROPOSITION 4

Under the same hypothesis as in theorem 1, the solution D_p of eq. ($\mathbb{E}$) satisfies the following inequalities:

$$D_p^2(\mathbf{u}, \mathbf{v}) \cdot \|\mathcal{P}_1(\mathbf{u} - \mathbf{v})\|^2 \cdot \left(\int_G \frac{dg}{\|g\mathbf{u} - \mathbf{v}\|} \right)^2 \leqslant \ln K_p(\mathbf{u}, \mathbf{v}) \leqslant D_p^2(\mathbf{u}, \mathbf{v}) .$$

Proof

The notations of the proof of theorem 1 are retained.

$$\Phi_{\mathbf{u},\mathbf{v}}(D_p) \leqslant \operatorname*{Sup}_{(g,h,k) \in G^3} \exp[p \cos(g\mathbf{u} - h\mathbf{v}, k\mathbf{u} - \mathbf{v})\, D_p^2(\mathbf{u}, \mathbf{v})] = \exp[p D_p^2(\mathbf{u}, \mathbf{v})] .$$

On the other hand, the exponential function is convex:

$$\Phi_{\mathbf{u},\mathbf{v}}(D_p) \geqslant \exp[p D_p^2(\mathbf{u}, \mathbf{v}) \iiint_{G^3} \cos(g\mathbf{u} - h\mathbf{v}, k\mathbf{u} - \mathbf{v})\, dg\, dh\, dk .$$

Using a result of the proof of theorem 1,

$$\Phi_{\mathbf{u},\mathbf{v}}(D_p) \geqslant \exp[D_p^2(\mathbf{u},\mathbf{v})\Phi'(0)]$$

$$= \exp\left[D_p^2(\mathbf{u},\mathbf{v}) \cdot p \cdot \|\mathcal{P}_1(\mathbf{u}-\mathbf{v})\|^2 \cdot \left(\int_G \frac{dg}{\|g\mathbf{u}-\mathbf{v}\|}\right)^2\right].$$

From eq. ($\mathbb{E}$), $p \ln K_p(\mathbf{u},\mathbf{v}) = \ln \Phi_{\mathbf{u},\mathbf{v}}(D_p(\mathbf{u},\mathbf{v}))$, and the proposition follows. □

In a first approach, the eventual verification of the triangular inequality by the solution D_p necessitates the explicit expression of $D_p(\mathbf{u},\mathbf{v})$. However, the analytical resolution of ($\mathbb{E}$) is not possible in the general case. Nevertheless, in borderline cases, partial resolution can be achieved. The corresponding mathematical study has no direct chemical implication, but it argues in favour of the consistency of the general model.

Simplification of the problem can be achieved by restriction:

- to simple groups ($G = \{e, \sigma\}$) and simple representations (of degree 1);
- to subsets F of E;
- to a local differential analysis ($\mathbf{v} = \mathbf{u} + d\mathbf{u}$) [7].

4. Representations of the group $\mathcal{S}_2 = \{e, \sigma\}$

It has been proved elsewhere that pairing products on all $\mathcal{S}_2$-Hilbert spaces E are discriminating [2,4]. $\mathcal{S}_2$ has only two irreducible representations (of degree one): the unit representation V_1 with the character χ_1 ($\chi_1(e) = \chi_1(\sigma) = 1$), and the representation V_2 with the character χ_2 ($\chi_2(e) = 1$, $\chi_2(\sigma) = -1$): E is split as a direct sum: $E = m_1 V_1 \oplus m_2 V_2$. A vector $\mathbf{u}$ in E is written as $\mathbf{u} = \mathbf{u}_1 \oplus \mathbf{u}_2$, with $\mathbf{u}_1 \in m_1 V_1$ and $\mathbf{u}_2 \in m_2 V_2$ ($eu_1 = u_1 = \sigma u_1$; $eu_2 = u_2 = -\sigma u_2$). It is easily verified that

$$K_p(\mathbf{u},\mathbf{v}) = K_p(\mathbf{u}_1,\mathbf{v}_1) \cdot K_p(\mathbf{u}_2,\mathbf{v}_2)$$

$$= \exp[-\tfrac{1}{2}\|\mathbf{u}_1 - \mathbf{v}_1\|^2]\left(\frac{\cosh[p\|\mathbf{u}_2\|^2]\cosh[p\|\mathbf{v}_2\|^2]}{\cosh^2[p(\mathbf{u}_2|\mathbf{v}_2)^2]}\right)^{1/p}.$$

On the other hand,

$$\Phi_{\mathbf{u},\mathbf{v}}(D_p) = \tfrac{1}{8} \sum_{g,h,k=e,\sigma} \exp[p\cos(g\mathbf{u} - h\mathbf{v}, k\mathbf{u} - \mathbf{v})\, D_p^2(\mathbf{u},\mathbf{v})].$$

Φ is calculated by using the relations $e\mathbf{u}_1 = \mathbf{u}_1$, $e\mathbf{u}_2 = \mathbf{u}_2$, $\sigma\mathbf{u}_1 = \mathbf{u}_1$, $\sigma\mathbf{u}_2 = -\mathbf{u}_2$:

$$\Phi_{\mathbf{u},\mathbf{v}}(D_p(\mathbf{u},\mathbf{v})) = \tfrac{1}{8} \sum_{g,h,k=\pm 1}$$

$$\times \exp\left[p \frac{\|\mathbf{u}_1 - \mathbf{v}_1\|^2 + \mathrm{Re}(g\mathbf{u}_2 - h\mathbf{v}_2 | k\mathbf{u}_2 - \mathbf{v}_2)}{\sqrt{\|\mathbf{u}_1 - \mathbf{v}_1\|^2 + \|g\mathbf{u}_2 - h\mathbf{v}_2\|^2}\sqrt{\|\mathbf{u}_1 - \mathbf{v}_1\|^2 + \|k\mathbf{u}_2 - \mathbf{v}_2\|^2}} D_p^2(\mathbf{u},\mathbf{v}) \right].$$

No analytical solution of the equation "$\Phi_{\mathbf{u},\mathbf{v}}(D_p(\mathbf{u},\mathbf{v})) = K_p(\mathbf{u},\mathbf{v})$" is available in the general case.

(1) $m_1 = 0$: $E = m_2 V_2$:

Calculations lead to:

$$K_p^p(\mathbf{u},\mathbf{v}) = \frac{\cosh[p\|\mathbf{u}\|^2]\cosh[p\|\mathbf{v}\|^2]}{\cosh^2[p\,\mathrm{Re}(\mathbf{u}|\mathbf{v})]},$$

$$\Phi_{\mathbf{u},\mathbf{v}}(D_p(\mathbf{u},\mathbf{v})) = \tfrac{1}{2}\left\{ \cosh[pD_p^2(\mathbf{u},\mathbf{v})] + \cosh\left[p\frac{\|\mathbf{u}\|^2 - \|\mathbf{y}\|^2}{\|\mathbf{u}-\mathbf{v}\| \cdot \|\mathbf{u}+\mathbf{v}\|} D_p^2(\mathbf{u},\mathbf{v}) \right] \right\}.$$

• By restriction to the sphere F_r of radius r and centered at 0, the equation is resolved:

$$D_p(\mathbf{u},\mathbf{v}) = \sqrt{\frac{1}{p}\,\mathrm{invcosh}\left[\frac{2\cosh^2 pr^2}{\cosh^2(pr^2\cos(\mathbf{u},\mathbf{v}))} - 1 \right]}.$$

The triangular inequality has been numerically verified for $m_2 = 1$, $E = \mathbb{C}$ and for $F_r = F_1$ (half-circle of unit radius in the complex plane wherein curvilinear coordinates of $\mathbf{u}$ boil down to a single number $\theta_{\mathbf{u}} \in [0, \pi[$, and: $\cos(\mathbf{u},\mathbf{v}) = \cos(\theta_{\mathbf{u}} - \theta_{\mathbf{v}})$).

• If $m_2 = 1$, then $E \cong \mathbb{C}$ is an irreducible representation of $\mathcal{S}_2$. By restriction to the Euclidean (real) subspace $F = \mathbb{R}$, the action of $\mathcal{S}_2$ is defined by: $\forall x \in \mathbb{R}$, $ex = x$, $\sigma x = -x$. Again, the equation can be resolved explicitly:

$$\forall (x,y) \in \mathbb{R}^2, \quad D_p(x,y) = \sqrt{\frac{1}{p}\,\mathrm{invcosh}\left[\frac{\cosh px^2 \cosh py^2}{\cosh^2 pxy} \right]}.$$

The triangular inequality has not been proved analytically, but was always satisfied in an iterative computation for $0 \leqslant x \leqslant 10$, $0 \leqslant y \leqslant 10$, $0 \leqslant z \leqslant 10$, with increments $\Delta x = \Delta y = \Delta z = 0.05$.

Thus, it is very likely that D_p is a distance on $R/\mathcal{S}_2 = R_+$ or R_-. D_p does not depend explicitly on the Euclidean distance "$|a - b|$", unlike distances δ defined by: $\delta(x,y) = |f(x) - f(y)|$, where f: $R \to R$, is an injective map, or by: $\delta(x,y) = \theta(|x-y|)$, with θ: $R_+ \to R_+$, $\theta(0) = 0$, $\forall (\mathbf{u},\mathbf{v}) \in (R^+)^2$, $\theta(\mathbf{u}+\mathbf{v}) \leqslant \theta(\mathbf{u}) + \theta(\mathbf{v})$ [8].

(2) $m_1 = m_2 = 1$. $E \cong \mathbb{C}\oplus\mathbb{C}$ *(regular representation of* $\mathcal{S}_2$*)*

By restriction to the Euclidean (real) plane $F = \mathbb{R}\oplus\mathbb{R}$, $u_1 = y$, $u_2 = x$: $\mathcal{S}_2$ acts by reflexion with respect to the y-axis ($ex = x$, $\sigma x = -x$, $ey = \sigma y = y$). Equation ($\mathbb{E}$) takes the form ($u = (x, y)$, $v = (x', y')$):

$$2+2\exp[p\mathbf{a}D_p^2] + 2\exp[p\mathbf{b}D_p^2] + \exp[p\mathbf{c}D_p^2] + \exp[p\mathbf{d}D_p^2]$$

$$= 8\exp[p(y-y')^2]\,\frac{\cosh px^2 \cosh px'^2}{\cosh^2 pxx'},$$

where

$$\mathbf{a} = \frac{(y-y')^2 - (x+x')(x-x')}{\sqrt{(y-y')^2 + (x+x')^2}\cdot\sqrt{(y-y')^2 + (x-x')^2}},$$

$$\mathbf{b} = \frac{(y-y')^2 + (x+x')(x-x')}{\sqrt{(y-y')^2 + (x+x')^2}\cdot\sqrt{(y-y')^2 + (x-x')^2}},$$

$$\mathbf{c} = \frac{(y-y')^2 - (x+x')^2}{(y-y')^2 + (x+x')^2},$$

$$\mathbf{d} = \frac{(y-y')^2 - (x-x')^2}{(y-y')^2 + (x-x')^2}.$$

On the x-axis (and parallel lines) we find again the distance

$$D_p(x, x') = \sqrt{\frac{1}{p}\,\mathrm{invcosh}\left[\frac{\cosh px^2 \cosh px'^2}{\cosh^2 pxx'}\right]}.$$

Given $x \geqslant 0$ and $x' \geqslant 0$, when $p \to +\infty$, $D_p(x, x')$ tends to the Euclidean distance $|x - x'|$ and is equivalent to

$$D_p(x, x') \underset{p\to+\infty}{\sim} \sqrt{(x-x')^2 + \frac{\ln 2}{p}}.$$

Finally, the solution of ($\mathbb{E}$) on a line parallel to the y-axis ($x = x_0$) is defined by

$$2 + \exp[pD_p^2] + 4\exp\left[\frac{p|y-y'|}{\sqrt{(y-y')^2 + 4x_0^2}}D_p^2\right]$$

$$+ \exp\left[p\frac{(y-y')^2 - 4x_0^2}{(y-y')^2 + 4x_0^2}D_p^2\right] = 8\exp[p(y-y')^2].$$

When $x_0 \to +\infty$, the solution of ($\mathbb{E}$) on the line to infinity is explicitly given by

$$D_p(y, y') = \sqrt{\frac{1}{p}\,\text{invcosh}[4\exp(p(y-y')^2) - 3]}\,.$$

The location of the various expressions found for the solution D_p of ($\mathbb{E}$) is summarized in fig. 1 [9].

Let us come back to the chemical interpretation. Upon a simple change of the axis in R^2, the regular representation of S_2 is also a permutation of coordinates X and Y such that: $X = \frac{1}{\sqrt{2}}(y + x)$, $Y = \frac{1}{\sqrt{2}}(y - x)$. If X and Y are regarded as ligand parameters at equivalent sites of a symmetrized skeleton (the two substituents of a ketone [3], of a carbene [3], etc.), a pair of points $u(X, Y)$ and $\sigma u(Y, X)$ represents a molecule. The "gap" D_p between two points belonging to different pairs is independent of the points considered. If the triangular inequality is satisfied in a whole plane, D_p is a distance for the set of the skeletal analogs with two equivalent skeletal sites and real ligand parameters (see fig. 1).

5. Concluding remarks

Most of the preceding definitions might be extended to regular metric spaces without an Euclidean or Hermitian structure [10]. A non-Euclidean version of the theorem [6a] could be useful in the following way. Suppose that D_G is a completely

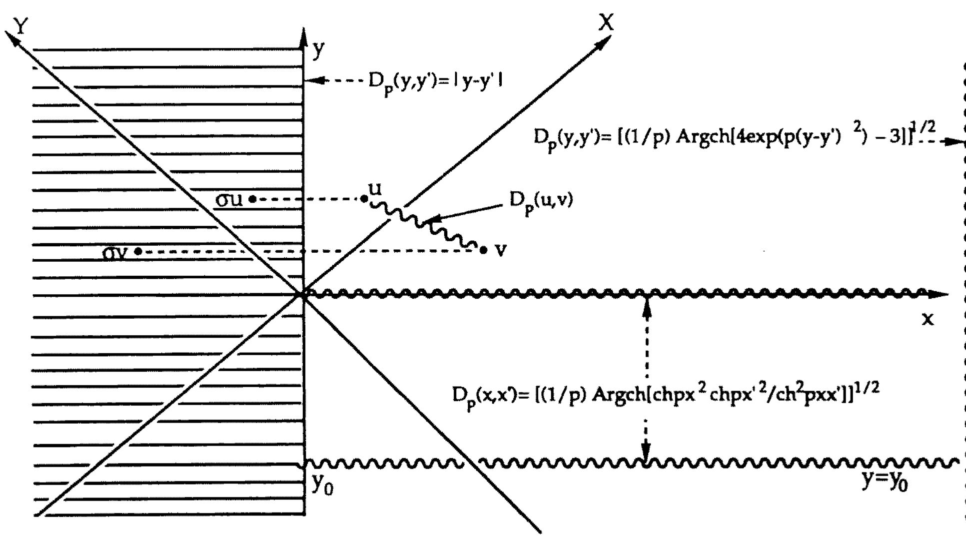

Fig. 1. Lines admitting explicit solutions of eq. ($\mathbb{E}$) for the representation of S_2 in the Euclidean real plane $F = \mathbb{R}\oplus\mathbb{R}$ by the reflexion with respect to the y-axis ($ex = x$, $\sigma x = -x$, $ey = \sigma y = y$). The coordinates X and Y are permutated by σ and may be regarded as ligands parameters of molecules with two equivalent skeletal sites.

G-invariant distance corresponding to a pairing product K_p on an Euclidean vector space E. $(E/G, D_G)$ is a metric space. G is realizable as a subgroup of the orthogonal group $O_n(E)$. The representation of G induces representations of bigger groups H, $O_n(E) \supset H \supset G$. Now, neither the complete G-invariance $(\forall(g,h) \in (H-G)^2$, $D_G(g\mathbf{u}, \mathbf{v}) = D_G(\mathbf{u}, \mathbf{v}))$ nor the regular G-invariance (isometry: $\forall g \in H D_G(g\mathbf{u}, g\mathbf{v}) = D_G(\mathbf{u}, \mathbf{v})$) is valid any longer. Therefore, the previous process leading from (E, d) to $(E/G, D_G)$ cannot be iterated to go from $(E/G, D_G)$ to $(E/H, D_H)$, where D_H would be the completely H-invariant distance on E/G (and on E itself).

Nevertheless, such an iteration could be done if G is a normal subgroup of H (then, $\forall g \in H$, $D_G(g\mathbf{u}, g\mathbf{v}) = D_G(\mathbf{u}, \mathbf{v})$): when H is the direct product of two subgroups G and K, two ways would be possible to get a completely H-invariant distance extension. The first way directly gives the distance D_H on E/H. The second way would proceed in two steps: first, a distance on E/G is defined. Since G is normal in H, the process would be valid again from the metric space $(E/G, D_G)$ instead of (E, d): another completely H-invariant distance D'_H on $H = G \cdot K$ would be obtained. Comparison of D_H with D'_H is challenging, but even in the most non-trivial case $H = (\mathbb{S}_2)^2$, $E = \mathbb{R} \oplus \mathbb{R}$ the calculation of D_H and D'_H is not an evident task.

Efforts are in progress to exhibit new discriminating pairing products, and then explore the question of triangular inequality of the solution of eq. $(\mathbb{E})$. Due to the failure to prove so far that D_p is a distance [11], the differential resolution of $(\mathbb{E})$ will be addressed in order to endow spaces E/G with a metric related to the kinetics of transformations between skeletal analogs [7].

References and notes

[1] R. Chauvin, Paper I of this series, J. Math. Chem. 16 (1994) 245.
[2] R. Chauvin, J. Phys. Chem. 96 (1992) 4701.
[3] R. Chauvin, J. Phys. Chem. 96 (1992) 4706.
[4] R. Chauvin, Paper II of this series, J. Math. Chem. 16 (1994) 257.
[5] M. Pavel, *Fundamentals of Pattern Recognition* (M. Decker Ed., 1989).
[6] (a) The possibility of a more general formulation for non-Euclidean metric spaces is outlined here. At first sight, the Euclidean expression $\Phi_{\mathbf{u},\mathbf{v}}(x) = \iiint_{G^3} \exp[p \cos(g\mathbf{u} - h\mathbf{v}, k\mathbf{u} - \mathbf{v}) D_p^2(\mathbf{u}, \mathbf{v})]$ $dg\, dh\, dk$ can be tentatively extended by

$$\Phi_{\mathbf{u},\mathbf{v}}(x) = \iiiint_{G^4} \exp\left[\frac{p}{2} \frac{d^2(g\mathbf{u}, l\mathbf{v}) + d^2(k\mathbf{u}, h\mathbf{v}) - d^2(g\mathbf{u}, k\mathbf{u}) - d^2(l\mathbf{v}, h\mathbf{v})}{d(g\mathbf{u}, h\mathbf{v} \cdot d(k\mathbf{u}, l\mathbf{v})} x^2\right] dg\, dh\, dk\, dl\,.$$

However, it is not obvious that the requirement (iib) follows from such an expression, unless we were able to prove that the inequality $|d^2(g\mathbf{u}, l\mathbf{v}) + d^2(k\mathbf{u}, h\mathbf{v}) - d^2(g\mathbf{u}, k\mathbf{u}) - d^2(l\mathbf{v}, h\mathbf{v})| \leqslant 2d(g\mathbf{u}, h\mathbf{v}) \cdot d(k\mathbf{u}, l\mathbf{v})$ is everywhere satisfied. This problem will be at least formally circumvented by dividing the unknown x^2 by a factor $C_m(\mathbf{u}, \mathbf{v})^{f(p)}$, where $f(p) = p/|p - p_0|$, $p_0 \neq 0$ and:

$$C_m(\mathbf{u},\mathbf{v}) = \mathrm{Max}\left\{\frac{d^2(g\mathbf{u}, l\mathbf{v} + d^2(k\mathbf{u}, h\mathbf{v}) - d^2(g\mathbf{u}, k\mathbf{u}) - d^2(l\mathbf{v}, h\mathbf{v})}{d(g\mathbf{u}, h\mathbf{v}) \cdot d(k\mathbf{u}, l\mathbf{v})}; (g,h,k) \in G^3\right\} (C_m(\mathbf{u},\mathbf{v}) \geqslant 1).$$

See: R. Chauvin, Entropy in dissimilarity and chiralty measures, submitted for publication (1994).

(b) The equation can be formulated in an even more general context than the context of metric spaces E. Let $\underline{A}$ and $\underline{B}$ be two fuzzy subgroups of G (which may be realized as fuzzy symmetry subgroups of molecular models **u** and **v**), and let $\underline{C}$ be a conjugacy link between them [1]. A definition of $D_p(\underline{A},\underline{B})$ can be written by the equation: $\Phi_{\underline{A},\underline{B}}(D_p(\underline{A},\underline{B}))D_p(\underline{A},\underline{B}) = 1/X(\underline{A},\underline{B})$, where X is the conjugacy index, and

$$\Phi_{\underline{A},\underline{B}}(x) = \iiint_{G^3} \exp\left[\frac{p}{2}\frac{\ln^2 \mu_{\underline{A}}(g) + \ln^2 \mu_{\underline{B}}(h) - \ln^2 \mu_{\underline{C}}(kg) - \ln^2 \mu_{\underline{C}}(hk)}{\ln \mu_{\underline{C}}(hkg) \ln \mu_{\underline{C}}(k)} x^2\right] dg\, dh\, dk.$$

[7] R. Chauvin, Paper IV of this series, J. Math. Chem. 16 (1994) 285.

[8] For example, the function $\delta(x,y) = |x^2 - y^2|^{1/2}$ is a distance on R/S_2. Notice that δ corresponds to D'_0, the limit of the series D'_q for $q = 0$ (see proposition 2): $D'_0 = [\prod_{g\in G} d(g\mathbf{u},\mathbf{v})]^{1/|G|}$.

[9] The real projection of the regular representation of S_2 is studied in ref. [7].

[10] Not only the Euclidean structure of E, but also the isometric character of the action of G on E, can be ommitted in the very first definitions of K_p and $\Phi_{\mathbf{u},\mathbf{v}}$: averaging over *all* the relative orientations of **u** and **v** must then be added under the integral symbol.

[11] The topology defined by a distance D_p over a set of G-representative vectors of E would be equivalent to the starting Euclidean topology. A sphere of radius R centered at 0 for a distance D_p ($D_p(\mathbf{0},\mathbf{u}) = R$) would be a set of G-representative vectors of the sphere radius R centered at 0 for the Euclidean metric. However, D_p would not be invariant to translations, and spheres centered elsewhere would be different from Euclidean spheres. In connection with such questions, the differential geometry of E/G endowed with a distance D_p could be tackled from a differential resolution of ($\mathbb{E}$), see ref. [7].