Adv. Molec. Struct. Research, *1*, (1993)

Measuring Symmetry in Structural Chemistry[†§]

Hagit Zabrodsky[‡] and David Avnir[*]

Departments of Computer Science[‡] and Organic Chemistry[*]
The Hebrew University of Jerusalem
91904 Jerusalem, Israel

[†]Part 3 of "Continuous Symmetry Measures".[1,2]

[§]Dedicated to Professor Jochanan Blum on the Occasion of his 60th birthday.

Abstract

Symmetry is treated as a continuous property and a Continuous Symmetry Measure (CSM) of structures is defined. The CSM of a structure is defined to be the minimum mean squared distance required to move points of the original structure and change it to a symmetrical structure. This general definition of a symmetry measure enables a comparison of the amount of symmetry of different shapes and the amount of different symmetries of a single shape. We describe a general method of obtaining the minimal distance to the desired shape and apply it to any symmetry element or symmetry group in two or three dimensions. Various examples of the application of the CSM approach to structural chemistry are presented. These include symmetry analysis of distorted tetrahedra and of rotating ethane, symmetry analysis of contours of equi-property such as molecular orbital representations, reconstruction of incomplete structural data and symmetry analysis of structures with uncertain point-location, such as encountered in x-ray data analysis.

1 Introduction

A traditional working-tool in structural chemistry has been symmetry analysis. Symmetry point groups and space groups have been used as reference configurations which either exist or not in the structure under study. We have argued recently[1,2] that this traditional approach fails to capture the richness of shapes and structures, both static and dynamic, which is found in the molecular and supramolecular domains. Most of these are not symmetric. At most they are approximately symmetric, either permanently or if the time-resolution of observation is sufficiently narrow. Consider, for instance, the very weak ($\epsilon_{\max} \approx 200$) forbidden $\pi \rightarrow \pi^*$ transition to the lowest-lying singlet in benzene ($A_g^1 \rightarrow B_{2u}^1$) and compare it with the carbon skeleton of toluene: The D_{6h} symmetry of the benzene hexagon changes to a distinctly different point group, C_{2v}, yet the extinction coefficient increases only to $\epsilon_{\max} = 225$. Current wisdom of accounting for the discrepancy between the major symmetry change and the small effect in the "allowedness" of the transition is to use such arguments that "the methyl perturbs the π system only to a small extent", i.e., the day is saved by resorting to "local" symmetry. Another example is the vibrating water molecule. This is a C_{2v} molecule and its ν_1 and ν_2 vibrational modes preserve this symmetry. But what about ν_3? This vibrational mode distorts the C_{2v} symmetry and again, a legitimate question is by how much does the molecule deviate from C_{2v} after 1% of one cycle, after 10% of it and so forth. Yet another example is the well known phenomenon of removal of the degeneracy of energy levels of a chemical species whenever contained in an environment of symmetry other than its own (a certain arrangement of ligands or a certain packing in the crystal). The degree of removal of degeneracy is directly linked to the "decrease" in the symmetry of the environment, compared to the isolated chemical species. Traditionally, this problem is treated in terms of jumps in the symmetry point group. For instance, the splitting of the degenerate p-orbitals increases from $a_{2u} + e_u$ in a D_{4h} environment to $a_1 + b_1 + b_2$ in a C_{2v} environment[3]. Our next example is the principle of conservation of orbital symmetry which has caused a quantum leap in the understanding of reaction pathways in organic chemistry. It suffices to take one very basic problem to illustrate our point: Consider two ethylenes approaching each other for a [2+2] reaction. The answer to the question of whether that reaction is allowed thermally or photochemically, or whether a suprafacial or antarafacial process will take place, or whether the reaction will take place at all, is very much dependent on the symmetry of alignment of the two reacting molecules or moieties. The extremes are D_{2h} for a parallel approach and C_2 for an orthogonal approach, and it is predicted successfully[4] that the former is needed for a suprafacial photochemical formation of cyclobutane. Most of the time, however, the two ethylenes are not in an ideal D_{2h} arrangement: This may be due to an intramolecular frozen conformation of the two double-bonds, to non-symmetric sterical hindrance caused by substituents on the double bond, and to the dynamical nature of the system (rotations and translations, especially in viscous media).

These are but few examples which illustrate the *need for a continuous scale of symmetry.* A general approach answer this need was layed out in ref.s' 1,2. Here we summarize its main

features, show how the above mentioned examples are approached and extend our theory of continuous symmetry measures (CSM) to three new applications:

1. The CSM of (e.g. molecular orbital) contours.
2. The symmetry of occluded shapes, such that appear in microscopy studies of materials.
3. The CSM of points with uncertain locations, such as can be found in x-ray diffraction analyses in molecular structure determinations.

Several previous studies were led by the need to relax the current strict language of symmetry. Hargittai and Hargittai, emphasize repeatedly in their book[5] the limitations of exact symmetry in the description of many structural problems in chemistry. P.Murray-Rust et. al.[6] and more recently R. Cammi and E. Cavalli[7], have suggested the use of symmetry coordinates to describe nuclear configurations of MX_4 molecules that can be regarded as distorted versions of the T_d symmetrical reference structure. P.G. Mezey and J. Maurani [8,9] extended the point symmetry concept for quasi-symmetric structures by using fuzzy-set theory (terming it "syntopy" and "symmorphy") and provided a detailed demonstration of its application for the case of the water molecule. In another recent study[10], P.G. Mezey used a resolution based similarity method of polycube filling to measure approximate symmetry of molecular distributions. Other relevant contributions are perturbation analyses in spectroscopy[11], and measures on convex-sets[12]. As will be evident below, our approach to the problem of non-ideal symmetry is quite different, being guided by three principles:

1. Non-symmetric shapes should not be treated as a perturbation of an ideal reference. Such shapes, as well as perfectly symmetric ones, should appear on a single continuous scale with no built-in hierarchy of subjective ideality.
2. Assessing symmetry should be detached from referencing to a specific shape; yet the shape of the nearest configuration with the desired symetry, should be obtainable.
3. It should be possible to evaluate the symmetry of a given configuration with respect to any symmetry group, such as the closest one.

These guidelines are implemented as described in the following section.

2 A Continuous Symmetry Measure - Definition

We define the **Continuous Symmetry Measure** (CSM) as a quantifier of the minimum effort required to turn a given shape into a symmetric shape. This effort is measured by the sum of the square distances each point is moved from its location in the original shape

to its location in the symmetric shape. Note that no a priori symmetric reference shape is assumed.

Denote by Ω the space of all shapes of a given dimension, where each shape P is represented by a sequence of n points $\{P_i\}_{i=0}^{n-1}$. We define a metric d on this space as follows:

$$d : \Omega \times \Omega \rightarrow R$$

$$d(P,Q) = d(\{P_i\},\{Q_i\}) = \frac{1}{n}\sum_{i=0}^{n-1} \|P_i - Q_i\|^2$$

This metric defines a distance function between every two shapes in Ω.

We define the **Symmetry Transform** ST as the symmetric shape closest to P in terms of the metric d.

The **Continuous Symmetry Measure (CSM)** of a shape is now defined as the distance to the closest symmetric shape:

$$S = d(P, ST(P))$$

The CSM of a shape $P = \{P_i\}_{i=0}^{n-1}$ is evaluated by finding the symmetry transform $\hat{P}$ of P and computing: $\mathrm{S} = \frac{1}{n}\Sigma_{i=0}^{n-1}\|P_i - \hat{P}_i\|^2$.

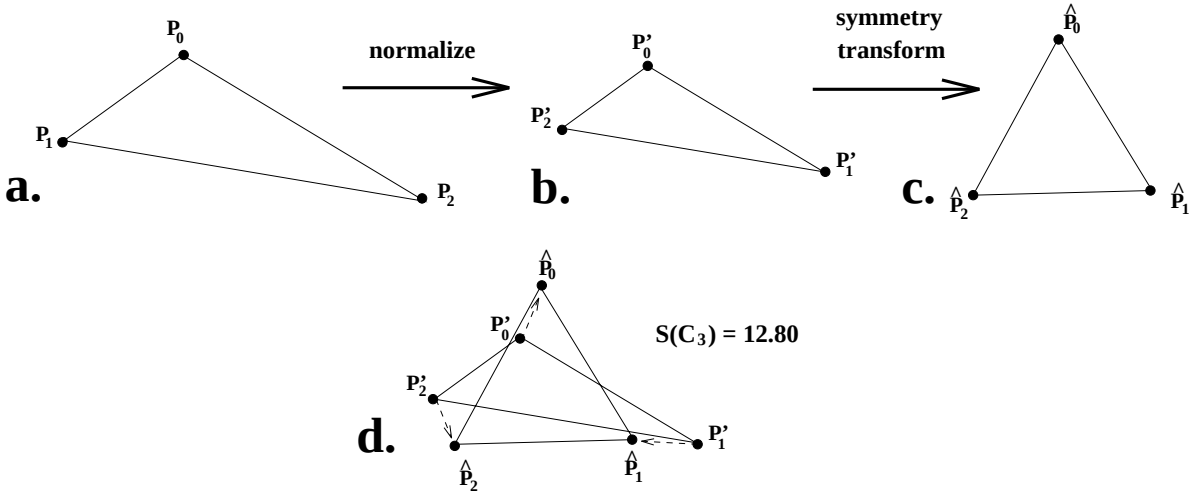

Figure 1: Calculating the CSM of a shape:
a) Original shape $\{P_0, P_1, P_2\}$. b) Normalized shape $\{P'_0, P'_1, P'_2\}$, such that maximum distance to the center of mass is one. c) Applying the symmetry transform to obtain a symmetric shape $\{\hat{P}_0, \hat{P}_1, \hat{P}_2\}$. d) $\mathrm{S}(C_3) = \frac{1}{3}(\|P'_0 - \hat{P}_0\|^2 + \|P'_1 - \hat{P}_1\|^2 + \|P'_2 - \hat{P}_2\|^2)$. CSM values are multiplied by 100 for convenience of handling.

This definition of the CSM implicitly implies invariance to rotation and translation. Normalization of the original shape prior to the transformation additionally allows invariance

to scale (Figure ??). We normalize by scaling the shape so that the maximum distance between points on the contour and the centroid is a given constant (in this paper all examples are given following normalization to 1, however CSM values are multiplied by 100 for convenience of handling). The normalization presents an upper bound of on the mean squared distance moved by points of the shape. Thus the CSM value is limited in range, where CSM=0 for perfectly symmetric shapes (see Appendix ??).

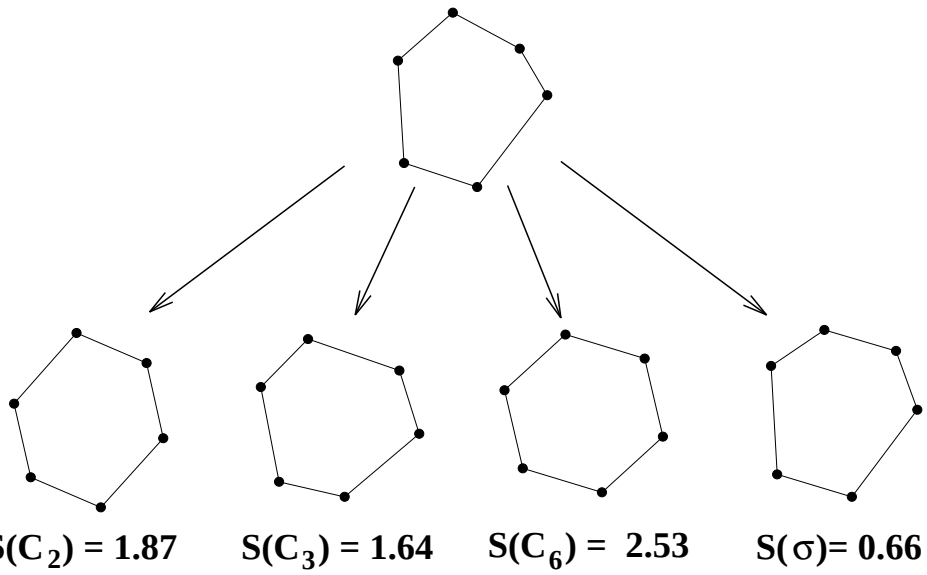

Figure 2: Symmetry Transforms of a 2D polygon and corresponding CSM values.

The general definition of the CSM enables evaluation of a given shape for different types of symmetries (mirror-symmetries, rotational symmetries and any other symmetry-groups - see Section ??). Moreover, this generalization allows comparisons between the different symmetry types, and expressions such as "a shape is more mirror-symmetric than rotationally-symmetric of order two" is valid. An additional feature of the CSM is that we obtain the symmetric shape which is 'closest' to the given one, enabling visual evaluation of the CSM.

An example of a 2D polygon and it's symmetry transforms and CSM values are shown in Figure ??.

In the next Section we describe a geometric algorithm for deriving the Symmetry Transform of a shape. In Section ?? we deal with the initial step of representing a shape by a collection of points.

3 Evaluating the Symmetry Transform

In this Section we describe a geometric algorithm for deriving the Symmetry Transform of a shape represented by a sequence of points $\{P_i\}_{i=0}^{n-1}$. In practice we find the Symmetry Transform of the shape with respect to a given point-symmetry group (see Appendix ?? for a review of algebraic definitions). For simplicity and clarity of explanation, we describe the method by following some examples. Mathematical proofs and derivations are detailed in Section ??.

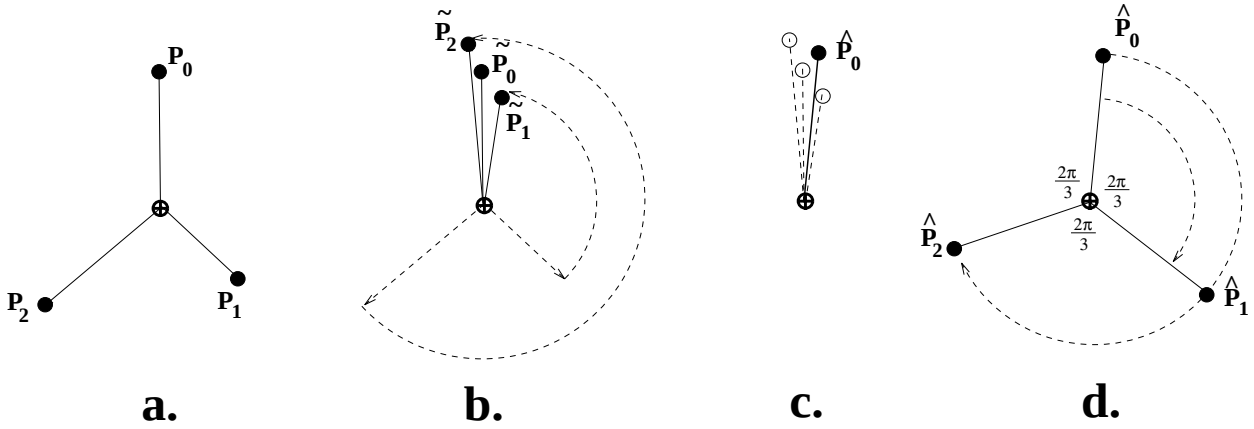

Figure 3: The C_3-symmetry Transform of 3 points: a) original 3 points $\{P_i\}_{i=0}^{2}$. b) Fold $\{P_i\}_{i=0}^{2}$ into $\{\tilde{P}_i\}_{i=0}^{2}$. c) Average $\{\tilde{P}_i\}_{i=0}^{2}$ obtaining $\hat{P}_0 = \frac{1}{3}\sum_{i=0}^{2}\tilde{P}_i$. d) Unfold the average point obtaining $\{\hat{P}_i\}_{i=0}^{2}$.

Following is a geometrical algorithm for deriving the symmetry transform of a shape P having n points with respect to rotational symmetry of order n (C_n-symmetry). This method transforms P into a regular n-gon, keeping the centroid in place.

1. *Fold* the points $\{P_i\}_{i=0}^{n-1}$ (Fig. ??a) by rotating each point P_i counterclockwise about the centroid by $2\pi i/n$ radians (Fig. ??b).
2. Let $\hat{P}_0$ be the average of the points $\{\tilde{P}_i\}_{i=0}^{n-1}$ (Fig. ??c).
3. *Unfold* the points, obtaining the C_n-symmetric points $\{\hat{P}_i\}_{i=0}^{n-1}$ by duplicating $\hat{P}_0$ and rotating clockwise about the centroid by $2\pi i/n$ radians (Fig. ??d).

A 2D shape P having qn points is represented as q sets $\{S_r\}_{r=0}^{q-1}$ of n interlaced points $S_r = \{P_{rn+i}\}_{i=0}^{n-1}$. The C_n-symmetry transform of P is obtained by applying the above algorithm to each set of n points separately, where the folding is performed about the centroid of all the points (Fig. ??).

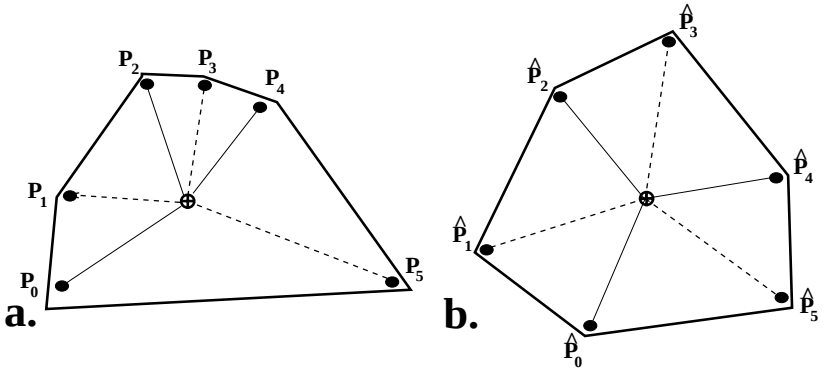

Figure 4: Geometric description of the C_3-symmetry transform for 6 points. The centroid of the points is marked by $\oplus$. a) The original points shown as two sets of 3 points: $S_0 = \{P_0, P_2, P_4\}$ and $S_1 = \{P_1, P_3, P_5\}$. b) The obtained C_3-symmetric configuration.

The procedure for evaluating the symmetry transform for mirror-symmetry is similar: Given a shape represented by $m = 2q$ points and given an initial guess of the symmetry axis, we apply the folding/unfolding method as follows (see Figure **??**):

1. for every pair of points $\{P_0, P_1\}$:
   (a) fold - by reflecting across the mirror symmetry axis obtaining $\{\tilde{P}_0, \tilde{P}_1\}$.
   (b) average - obtaining a single averaged point $\hat{P}_0$.
   (c) unfold - by reflecting back across the mirror symmetry axis obtaining $\{\hat{P}_0, \hat{P}_1\}$.
2. minimize over all possible axis of mirror-symmetry.

The minimization performed in step 2 is, in practice, replaced by an analytic solution (derivation and proof can be found in Appendix **??**).

This method extends to **any** finite point-symmetry group in **any** dimension, where the folding and unfolding are performed by applying the group elements about the center of symmetry (see derivations in Section **??**). The minimization is over all orientations of the symmetry group. In 2D the minimization is performed analytically, in 3D a minimization process is used. A detailed description of the extension to 3D and to any symmetry-group appeared in Ref.2 and two illustrative examples are given in Section **??**.

We briefly mention the case where the number of points m is less than n i.e. less than the number of elements in the symmetry-group G with respect to which we measure symmetry. In this case m should be a factor of n such that there exists a subgroup H of G with n/m elements. In this case, we duplicate each point n/m times and fold/unfold the points with elements of a left coset of G with respect to H. Following the folding/unfolding method, the relocated points will align on symmetry elements of G (on a reflection plane or on a rotation axis for example). Further details of this case and proof can be found in Ref.2.

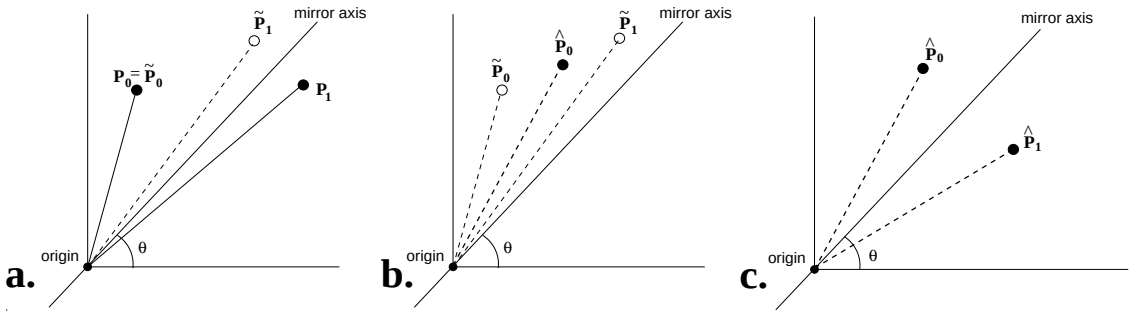

Figure 5: The mirror-symmetry transform of a single mirror 2-association for angle θ: a) mirror-fold the 2-association $\{P_0, P_1\}$ obtaining $\{\tilde{P}_0, \tilde{P}_1\}$. b) The transformed-$P_0$ denoted $\hat{P}_0$ is the average of $\tilde{P}_0$ and $\tilde{P}_1$. c) The transformed-P_1 denoted $\hat{P}_1$ is $\hat{P}_0$ reflected about the symmetry axis. Center of mass of the shape is assumed to be at the origin.

4 Proof of the Folding Method

As described in Section **??**, the CSM of a set of points with respect to a given symmetry group G, is evaluated by first finding the set of points which is G-symmetric and which is closest to the given set in terms of the average distance squared. We must thus prove that the folding method indeed finds the closest symmetric set of points.

The group-theory definitions which are used in this section, are briefly reviewed in Appendix **??**.

Given a finite point-symmetry group G centered at the origin and an ordering of its m elements $\{g_1 = I, \ldots, g_m\}$ and given m general points $P_1, \ldots, P_m$, find m points $\hat{P}_1, \ldots, \hat{P}_m$ and find a rotation matrix R and translation vector w such that $\hat{P}_1, \ldots, \hat{P}_m$ form an ordered orbit under G' (where G' is the symmetry group G rotated by R and translated by w) and bring the following expression to a minimum:

$$\sum_{j=1}^{m} \|P_j - \hat{P}_j\|^2 \tag{1}$$

Since G has a fixed point at the origin and G' has the centroid of orbit $\hat{P}_i$ as a fixed point (see Lemma **??** in Appendix **??**) we have that w is the centroid of orbit $\hat{P}_i$:

$$w = \frac{1}{m} \sum_{i=1}^{m} \hat{P}_i \tag{2}$$

(Note that in the cases where the fixed points of G form an axis or plane, w can be any vector moving the origin to the (rotated) axis or plane passing through the centroid of orbit $\hat{P}_i$. Thus also in these cases w can be considered the centroid of orbit $\hat{P}_i$).

The points $\hat{P}_1, \ldots, \hat{P}_m$ form an orbit of G', thus the following must be satisfied:

$$\hat{P}_i = g'_i \hat{P}_1 = R^t g_i R(\hat{P}_1 - w) + w \quad i = 1 \ldots m \tag{3}$$

where g'_i is the matrix representation of the ith symmetry element of G' and is equal to the ith symmetry element g_i of G rotated by R and translated by w.

Using Lagrange multipliers with Eq.s **??**-**??** we must minimize the following:

$$\sum_{j=1}^{m} \|P_j - \hat{P}_j\|^2 \; + \sum_{j=1}^{m} \lambda_j^t (\hat{P}_j - R^t g_j R(\hat{P}_1 - w) + w) \; + \varepsilon (w - \frac{1}{m} \sum_{j=1}^{m} \hat{P}_j)$$

where ε, λ_j for $j = 1 \ldots m$ are the Lagrange multipliers.

Equating the derivatives to zero we obtain:

$$\sum_{j=1}^{m} (P_j - \hat{P}_j) = 0 \tag{4}$$

and using the last constraint (Eq. **??**) we obtain:

$$w = \frac{1}{m} \sum_{j=1}^{m} P_j \tag{5}$$

i.e. the centroid of $P_1, \ldots, P_m$ coincides with the centroid of $\hat{P}_1, \ldots, \hat{P}_m$ (in terms of the symmetry measure defined in Section **??**, the centroid of a configuration and the centroid of the closest symmetric configuration is the same for any point symmetry group G).

Noting that g'_j are isometries and distance preserving, we have from the derivatives:

$$\sum_{j=1}^{m} g'^{t}_j (P_j - \hat{P}_j) = \sum_{j=1}^{m} R^t g_j^t R (P_j - \hat{P}_j) = 0$$

Expanding using the constraints we obtain:

$$m\hat{P}_0 = \sum_{j=1}^{m} R^t g_j^t R P_j - \sum_{j=1}^{m} R^t g_j^t R w$$

or

$$\hat{P}_1 - w = \frac{1}{m} \sum_{j=1}^{m} R^t g_j^t R (P_j - w) \tag{6}$$

The geometric interpretation of Eq. **??** is the folding method as described in Section **??**, thus proving that the folding method results in the G-symmetric set of points closest to the given set.

Given $n = qm$ points (i.e. q sets of m points) $\{P_1^i, \ldots, P_m^i\}$ $i = 1 \ldots q$ we obtain the result given in Eq. **??** for each set of m points separately, i.e. for $i = 1 \ldots q$:

$$\hat{P}_1^i - w = \frac{1}{m}\sum_{j=1}^{m} R^t g_j^t R(P_j^i - w) \qquad (7)$$

where $w = \frac{1}{n}\sum_{i=1}^{q}\sum_{j=1}^{m} P_j^i$ is the centroid of all n points. The geometric interpretation of Eq. ?? is the folding method as described in Section ?? for $m = qn$ points.

5 Examples of Measuring Symmetry in 3D: Tetrahedricity and Rotating Ethane

5.1 Tetrahedricity of Phosphates

We return now to the general question: given any number of vertices in space, what is its symmetry measure with respect to any symmetry group, subgroup or class. As explained in the previous section, the generalized approach is to divide the given points into sets and to apply the folding/unfolding method separately on each set, while evaluating the CSM value over all the given points. For example, let us analyze the tetrahedricity of a tetrahedron with a branched connected set of 5 points $P_1 \ldots P_5$ as shown in Figure ??a, which models a tetrahedron with a central atom and apply the CSM folding/unfolding method to evaluate its T_d symmetry. The connectivity constrains the division of points into sets (Part 4; in preparation) and restricts the center point (P_5) to be a one-point set. We thus divide the points into 2 sets: $\{P_1 \ldots P_4\}$ and $\{P_5\}$. The closest symmetric configuration will have point P_5 relocated to a position where all 24 of the T_d-symmetry group elements leave it in place. The only such position is at the origin (centroid of the configuration marked as an open circle in Figure ??) where all symmetry planes and axes intersect. Points $P_1 \ldots P_4$ will be relocated to form a perfect T_d-symmetric configuration of four points, i.e. each point will lie on a C_3-rotation axis (see Fig. ??b). $S(T_d)$ (or any $S(G)$) is then calculated by considering the full set of $\hat{P}_1 \ldots \hat{P}_5$.

To illustrate it, we now analyze the distorted phosphate tetrahedron $Cd_2P_2O_7^{6}$ using our method. We first recall that our method evaluates the distance from tetrahedricity and not from a specific tetrahedron; and that rather than reporting the deviation in terms of a table of many coordinates (as done in reference 6), we provide a single ($S(T_d)$) value. To obtain it, the 3D position coordinates of the four oxygens and phosphorus were taken from ref.13 (also used by Bürgi et. al., p. 1790 in their paper) as:

$$\begin{array}{lcccc} P_1 & = (& 0.0 & 0.0 & 1.645) \\ P_2 & = (& 0.0 & 1.518860 & -0.347028) \\ P_3 & = (& -1.286385 & -0.700083 & -0.391603) \\ P_4 & = (& 1.179085 & -0.755461 & -0.372341) \end{array}$$

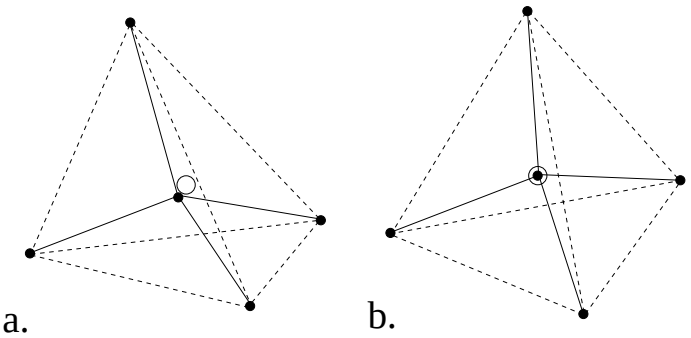

Figure 6: A distorted tetrahedron with a central atom, analyzed as a connected configuration of 5 points. The open circle marks the centroid of the configuration; b) the closest T_d-symmetric configuration.

with an additional center point 0.0. By applying the folding method as described above, the symmetry measure obtained in this example is $S(T_d) = 0.17$ and the closest symmetric shape is a regular tetrahedron with arm length 1.537Å. (By comparison, a set of 10 symmetry displacement coordinates is used in ref. 6 to report the deviation of this tetrahedron from ideality). In a further example Murray-Rust et. al. used the symmetry coordinates to evaluate the threefold axes of 1-methyl-1-silabicycloheptane (Section 5 in ref. 6). They found that the distorted S_iC_4 structure (Fig. 5 in ref. 6) is better described with the three-fold axis passing through one vertex (point $\mathbf{C}_1$ in their notation) rather than through another ($\mathbf{C}_2$ in their notation). Using the CSM method with respect to C_{3v}-symmetry we easily support their conclusion as follows:
Given the coordinates:

$$\begin{array}{lllllll} P_1 & = & \mathbf{C}_1 & = (& 0.0 & 0.0 & 1.645) \\ P_2 & = & \mathbf{C}_2 & = (& 0.0 & 0.87461971 & -0.48480962) \\ P_3 & = & \mathbf{C}_3 & = (& 0.75128605 & -0.39338892 & -0.52991926) \\ P_4 & = & \mathbf{C}_4 & = (& -0.75128605 & -0.39338892 & -0.52991926) \end{array}$$

with S_i at the origin, the $S(C_{3v})$ of the configuration was calculated by the method described in Section **??** and found to be $S(C_{3v}) = 0.02$ when the threefold axis passes through point $\mathbf{C}_1$, compared to $S(C_{3v}) = 1.16$ when the threefold axis is constrained to pass through point $\mathbf{C}_2$. Using the folding method we can also measure the C_{3v}-symmetry of the configuration with the constraint that three of the configuration points are equatorial. In this case the S value increases to 5.26, with the three-fold axes passing through point $\mathbf{C}_2$.

5.2 The Rotating Tetrahedra of Ethane[1]

[1]This section is also Erratum to Section 4.4 in ref. 2. We thank A. Cotton and Y. Pinto for drawing our attention to an error.

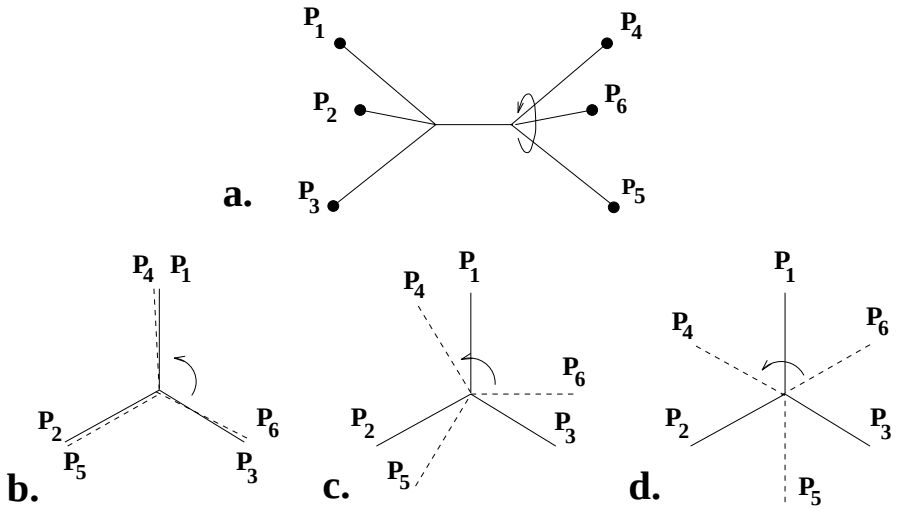

Figure 7: Modeling the C-C rotation in ethane.
a) Only the right hand tetrahedron moves; b) the cycle starts with the eclipsed D_{3h} rotamer; c) one of the six chiramers (see text); d) the D_{3d} staggered rotamer.

Another mechanism which strongly affects molecular symmetry is intramolecular rotation. Consider, for instance, one of the most basic examples, namely, the rotation of the two ethane tetrahedra around the C-C bond (Fig. **??**a). Current wisdom allows an exceedingly poor description of that process from the symmetry point of view: Ethane is D_{3d} when staggered (Fig. **??**d), D_{3h} when eclipsed (Fig. **??**b) and D_3 anywhere in between, including the rotamer which is only 1° away from any of the extremes. Doesn't physical intuition dictate that it is more natural to ask about that 1° rotamer, how much D_{3h} or D_{3d} it contains? Or for that matter, how much D_{3h} and D_{3d} exists at any point in a full 360° cycle? As has already become evident throughout this paper, the CSM method allows one to select any symmetry group and follow its gradual changes along such a full 360° cycle of rotation. We demonstrate it on two perfect tetrahedral structures connected along one of the tetrahedral arms and rotating with respect to each other around the connecting arm. We model this by stabilizing one of the tetrahedra and rotating the other, beginning the cycle with the two tetrahedra perfectly aligned (eclipsed) and rotating the second tetrahedron anti-clockwise. For simplicity in evaluating the S value, we considered only the tetrahedral arms not invloved in the C-C bond. Fig. **??** displays the result where the S value is given as a function of the cycle. (Fig. **??**a shows a full 360° cycle and Fig. **??**b shows a detail). The following observations are made:

- The D_{3h} profile of the rotation and the D_{3d} profile are similar, but shifted from each other by 60°. That is, there is as much D_{3d}-ness in the eclipsed structure as there is D_{3h}-ness in the staggered structure. This is intuitive since the distance (rotation or projection) from an ϵ° position to the 0° position (eclipsed) which determines the

$S(D_{3h})$ value, is equivalent to the distance (rotation or projection) from a $60° + \epsilon°$ position to the 60° position (staggered) and which determines the $S(D_{3d})$ value.

- The maximal $S(D_{3h})$ value is at the 60° staggered rotamer which is the farthest away from the perfect D_{3h} eclipsed rotamer (0°, $S(D_{3h}) = 0$). Three such maxima are observed in a full cycle, corresponding to the three staggered rotamers. Similarly, the maximal $S(D_{3d})$ value is for the eclipsed rotamer at 0°, and again, there are three maxima in a full cycle corresponding to the three eclipsed rotamers.

- Fig. **??** points to a rotamer (Fig. **??**c) which is neither eclipsed nor staggered, but in between, at $30° + n \cdot 60°$. We term these special chiral (!) rotamers at $30° + n \cdot 60°$ - **chiramers** (Fig. **??**c). There are six of these in a full cycle, compared to three eclipsed and three staggered (which are, of course, achiral). These six chiramers, which are at the crossing of the continuous D_{3h} and D_{3d} profile lines, are also the six maxima of the contuinuous C_{3v} profile line. Note that the C_{3v} line (which is equivalent to the σ-line) coincides with either the D_{3h} line (0°-30°, 90°-120°, ...) or with the D_{3d} line (30°-90°, 150°-210°, ...), whichever gives the lower S value. To understand this, we note that C_{3v} is a sub-group of both D_{3h} and D_{3d} and that in ethane, the nearest C_{3v} object at any point of the full cycle must be either of the two achiral rotamers. Thus, the chiramer, is also the most chiral rotamer of ethane.

Finally, we wish to make a brief preliminary comment on what seems to us an important application of our approach: Many thermodynamic and kinetic quantities vary cyclically with internal rotations. A commonly presented quantity is the (repulsion) potential. It is then interesting to see, how this property varies with the symmetry rather than with the traditional torsion angle. The results for a model sinusoidal potential (Fig. **??**a) are shown in Fig. **??**b. Let us first detail how the potential follows this new process coordinate: The D_{3h} potential line varies smoothly with S, starting at the eclipsed $S = 0$ value and dropping to zero potential at the staggered $S = 22.22$ value, then it reverses and climbs back up to the maximum potential completing 120° of the cycle. This drop and rise in potential along the symmetry coordinate is repeated continuously, completing a full 360° cycle. The behaviour of D_{3d} is a mirror image, starting with the maximal potential at the eclipsed $S = 22.22$ value and dropping to zero potential at the staggered $S = 0$ value. The behaviour of C_{3v} is interesting: up to $S = 5.95$ it follows the D_{3h} line, but then it continues to drop along the D_{3d} line back towards the $S = 0$ value (the line then climbs back up). Perhaps most notable is that the lines of the two symmetry groups bifurcate at the 30°-chiramers. What does such symmetry/potential bifurcation mean? In general, it may mean that for symmetry governed processes, such a crossing point is where the process may select to proceed one way or the other depending on which symmetry is prefered.

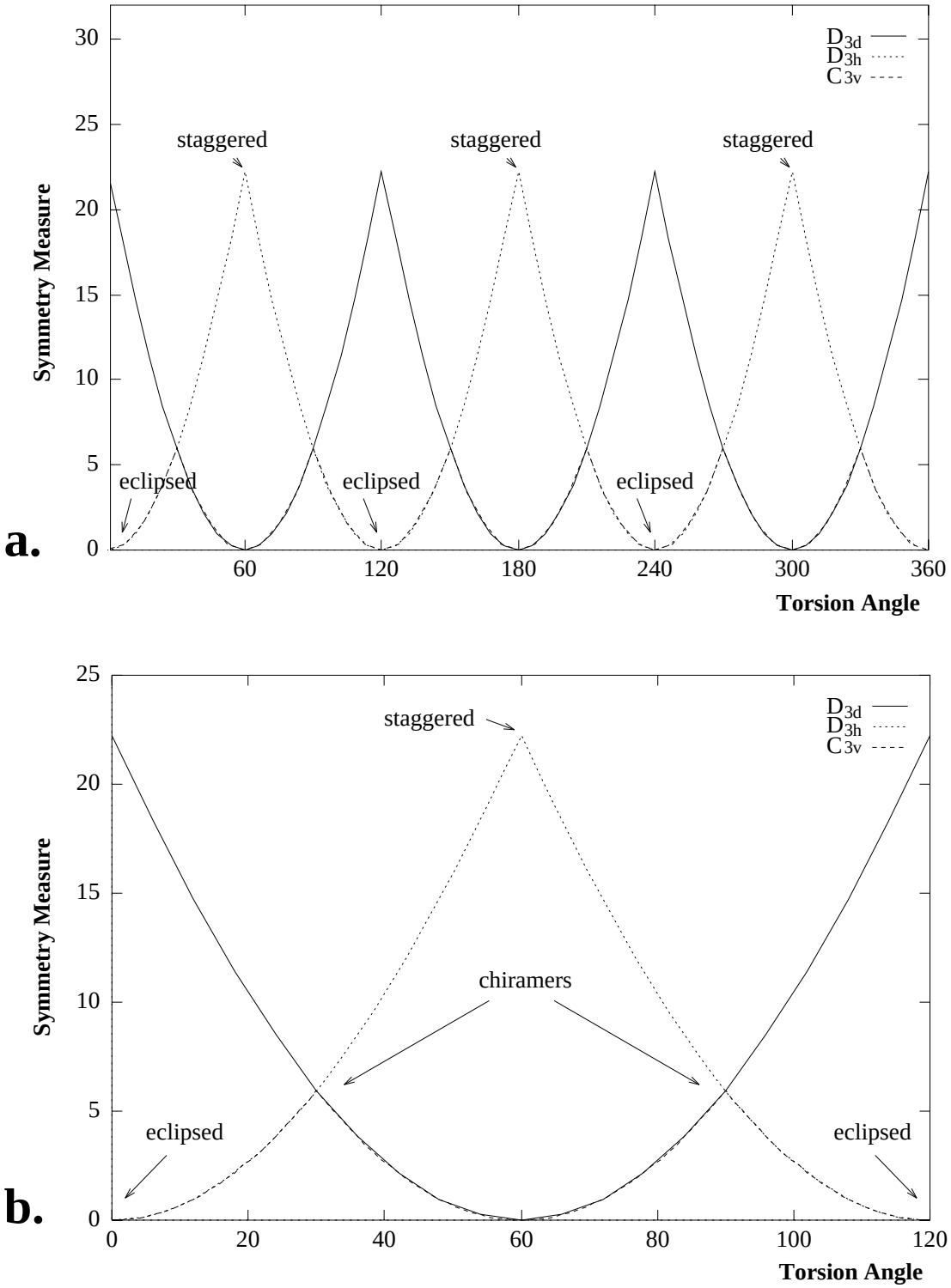

Figure 8: D_{3h} (——), D_{3d} (- - - -) and C_{3v} (- - - -) for rotating ethane. a) a full cycle b) a detail of one third of the cycle.

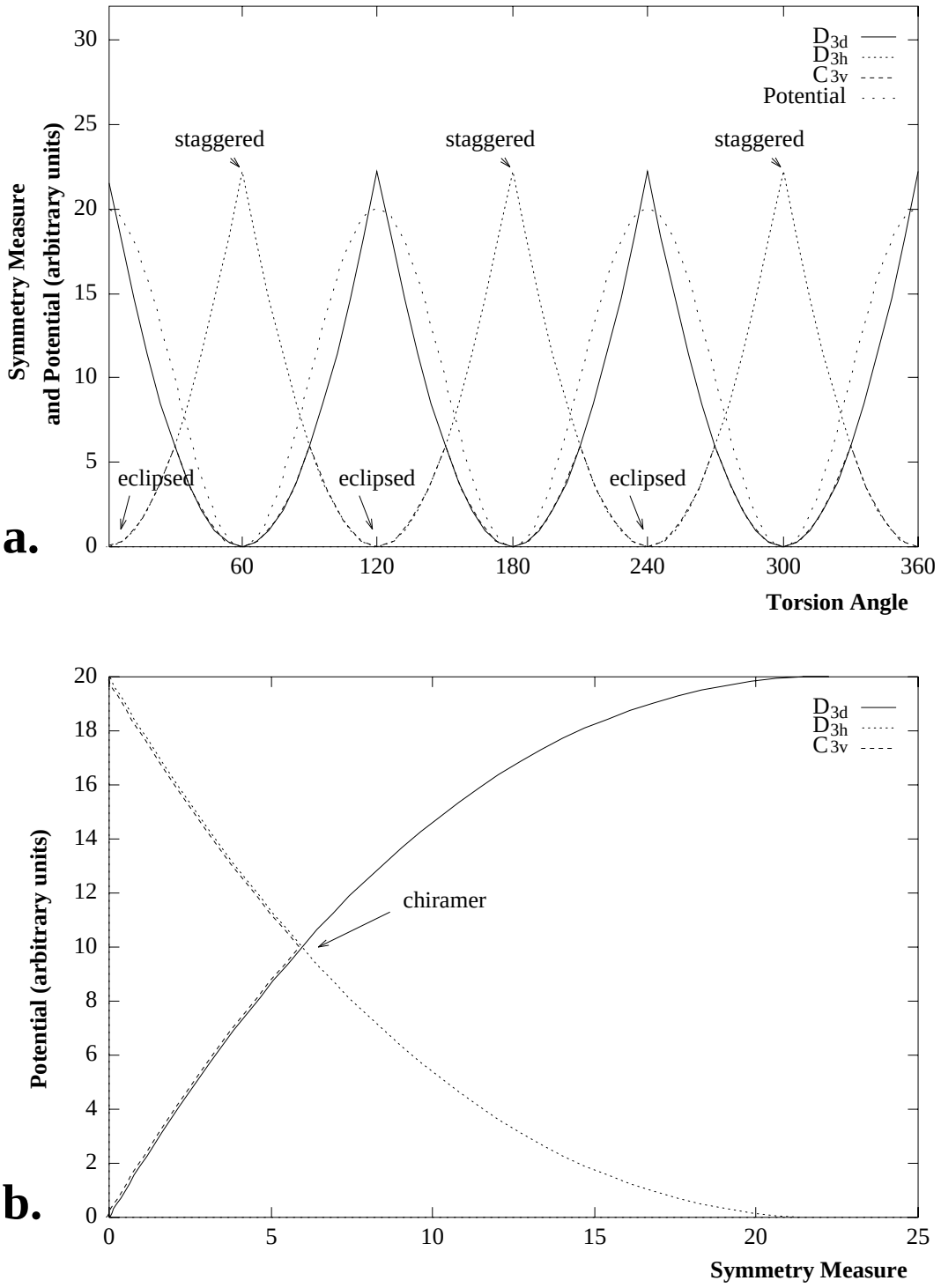

Figure 9: a) A sinusoidal potential is superimposed on Fig. 8a. b) The C-C rotation in ethane presented in the plane of potential vs symmetry measure of D_{3d}, D_{3h} and C_{3v}.

6 Point Selection for Representation of Contours

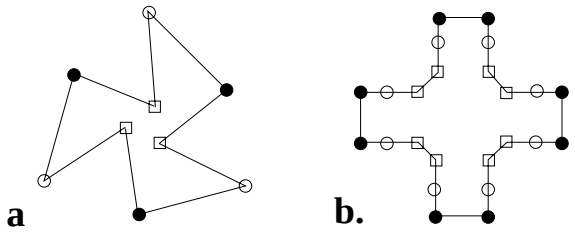

Figure 10: Dividing m selected points into interlaced sets: a) C_n-symmetry - one possibility. b) D_n-symmetry - one of the $m/2n$ possibilities

As symmetry has been defined on a sequence of points, representing a given shape by points must precede the application of the symmetry transform. Thus, for the case of equi-property contours, such as electronic orbital contours, one represents it as a string of equally spaced points (as dense as one wishes) and then perform the CSM folding-unfolding procedure as usual. As described in Section **??**, when a multiple of n points are given (where n is the number of elements in the symmetry group), the points must be divided into sets of n points. In general, this problem is exponential. However when the points are cyclically connected or ordered such as along a contour, the ordering of the points restricts the possible divisions into sets. For example in 2D, points along the contour of a C_n-symmetric shape form orbits which are interlaced as shown in Figure **??**a for C_3-symmetry. Thus, given a set of $m = nq$ ordered points there is only one possible division of the points into q sets of n points - the q sets must be interlaced (as was shown in Fig. **??**). In the case of D_n-symmetry the $m = 2nq$ ordered points, form q orbits which are interlaced and partially inverted as shown in Figure **??**b for D_4-symmetry. Thus, given a set of $m = 2nq$ ordered points there are $m/2n = q$ possible division of the points.

In Figure **??** we demonstrate the application of the contour CSM analysis on the lone-pair orbital of a distorted water molecule (perhaps a frozen moment of a vibration, or a water molecule in a matrix of amorphous ice, or a water molecule trapped in a micropore). The ratio of length of the two O-H bonds is 0.9 (instead of 1.0) and the H-O-H angle is 104°.[14] Two contours are shown, each of which has been represented as a string of about 200 points. The CSM with respect to mirror symmetry was evaluated by the above mentioned $m = nq$ pairing. It is seen quantitatively (Fig. **??**) that the distortive effects of the unequal bonds, fades away from the inner to the outer contours.

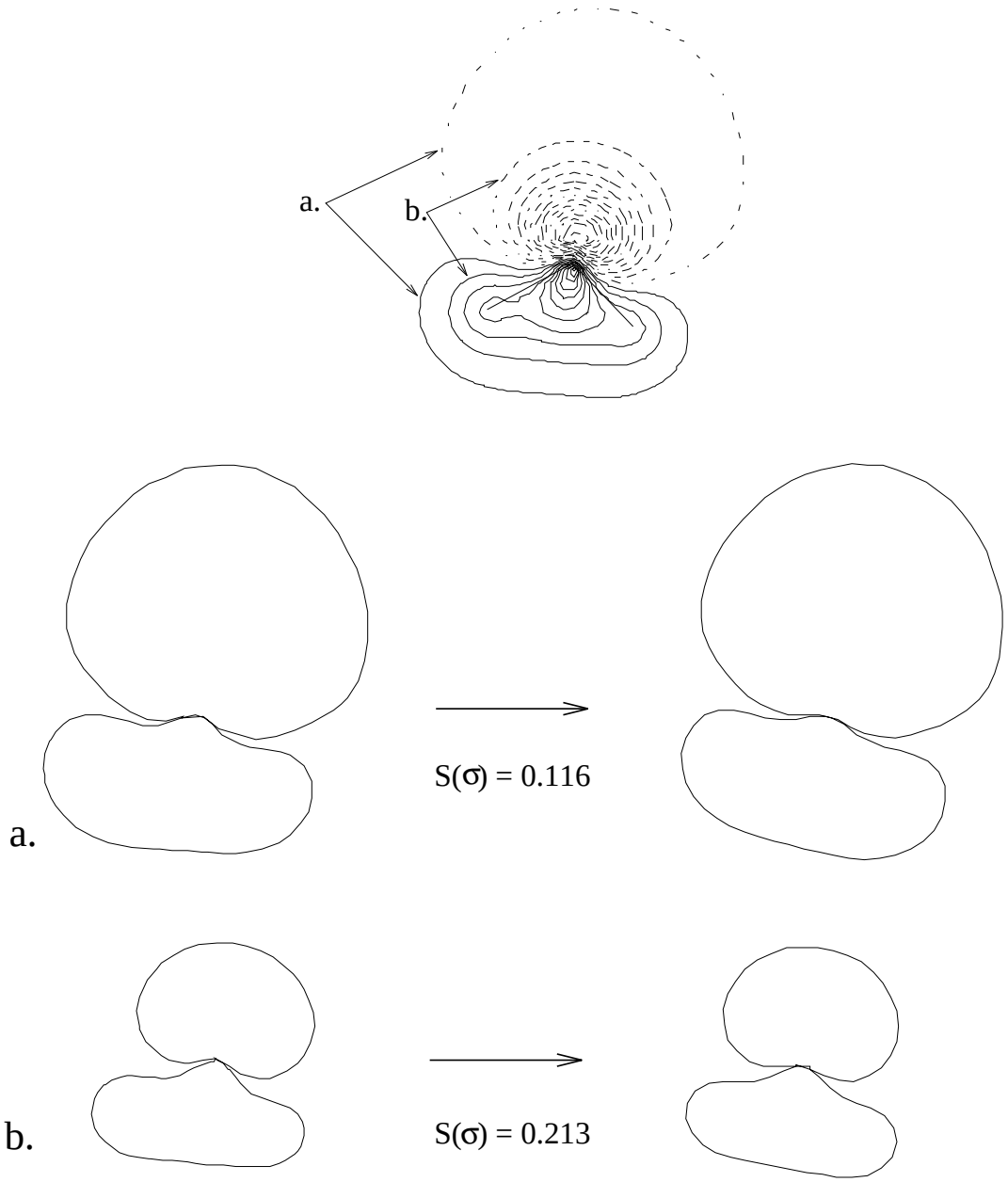

Figure 11: Two equi-amplitude contours of the wave function of the lone-pair orbital of distorted water molecule are shown[14,15]. The two contours are spaced by 0.05 Bohr$^{-3/2}$, and the value of the outer one is 0.576 Bohr$^{-3/2}$. S(σ) values are indicated in the figure.

7 Symmetry of Occluded Shapes

Here we address a problem which is commonly encountered in microscopy studies of particulate materials: Given a collection of similar objects, such that the individual shapes occlude each other, how can the symmetry of these objects be extracted using methods of automated image analysis. The method we apply to solve the problem is to evaluate the symmetry of the occluded shapes, locate the center of symmetry and reconstruct the symmetric shape most similar to the unoccluded original.

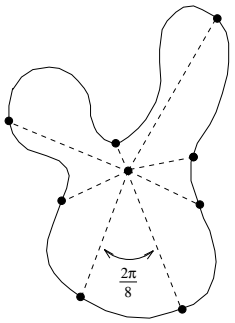

Figure 12: Selection at equal angles: points are distributed along the contour at regular angular intervals around the centroid.

As described in Section **??**, a shape can be represented by points selected at equal distances along the contour. Another method of representing a shape is by selecting points at regular angular intervals (selection-by-angle) about a point (Fig. **??**). Angular selection is usually about the centroid of the shape. However, angular selection of points about a point other than the centroid will give a different symmetry measure value (Fig. **??**). We define the **center of symmetry** of a shape as that point about which angular selection gives the minimum symmetry measure value. When a symmetric shape is not occluded the center of symmetry aligns with the centroid of the shape. However, the center of symmetry of truncated or occluded objects does not align with its centroid but aligns with the (unknown) centroid of the unoccluded shape. Thus the center of symmetry of a shape is robust under truncation and occlusion.

To locate the center of symmetry, we use an iterative procedure of gradient descent that converges from the centroid of an occluded shape to the center of symmetry. Denote by center of selection, that point about which points are selected using angular selection. We initialize the iterative process by setting the centroid as the center of selection. At each step we compare the symmetry value of points angularly selected about the center of selection and about points in its immediate neighborhood. That point about which angular selection gives minimum symmetry value, is set to be the new center of selection. If center of selection does not change the neighborhood size is decreased. The process is terminated when neighborhood size reaches a predefined minimum size. The center of selection at the end of the process is taken as the center of symmetry.

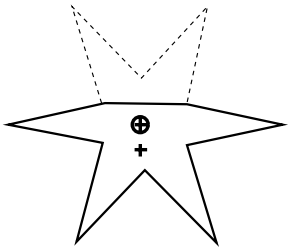

Figure 13: The symmetry value obtained by angular selection about the center of mass (marked by +) is greater than the symmetry value obtained by angular selection about the center of symmetry (marked by $\oplus$).

The closest symmetric shape obtained by angular selection about the center of symmetry (Fig. **??**c) is visually more similar to the original (Fig. **??**a) than that obtained by angular selection about the centroid of the occluded shape (Fig. **??**b). We consider the former a reconstruction of the unoccluded shape.

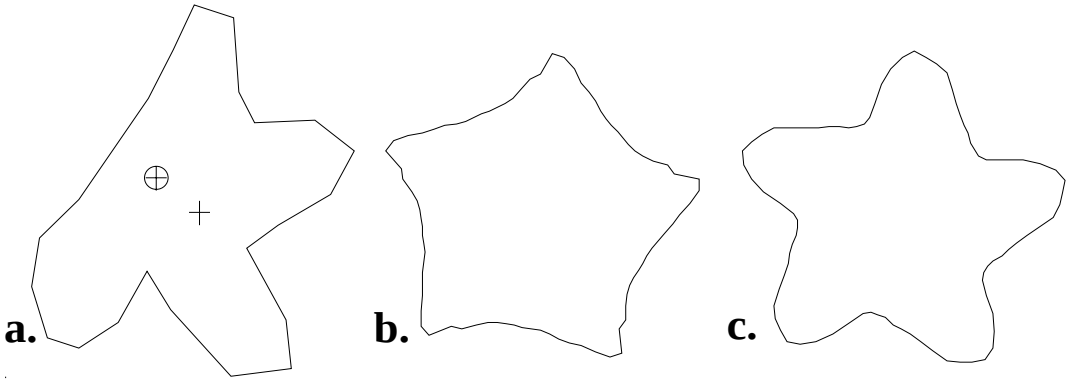

Figure 14: a) Original occluded shape, its centroid (+) and its center of symmetry($\oplus$). b,c) The closest C_5-symmetric shapes following angular selection about the centroid (b) and about the center of symmetry (c).

The process of reconstructing the occluded shape can be improved by altering the method of evaluating the symmetry of a set of points. As described in Section **??** the symmetry of a set of points is evaluated by folding, averaging and unfolding about the centroid of the points. We alter the method as follows:

1. The folding and unfolding (steps 1 and 3) will be performed about the center of selection rather than about the centroid of the points.

2. Rather than averaging the folded points (step 2), we apply a more robust clustering method: we average over the folded points, drop the points furthest from the average (this is justified by noticing that such points are due to occlusion) and then reaverage (see Figure **??**).

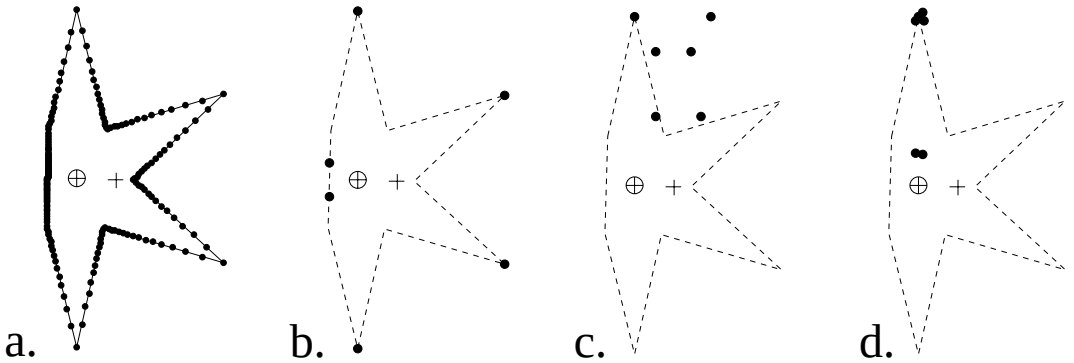

Figure 15: Improving the averaging of folded points:
a) An occluded shape with points selected using angular selection about the center of symmetry (marked as $\oplus$). b) A single set (orbit) of the selected points of a) is shown. c) folding the points about the centroid of the shape (marked as $+$), points are clustered sparsely. d) folding the points about the center of symmetry of the shape, points are clustered tightly. Eliminating the extremes (the two furthest points) and averaging will result in smaller averaging error and better reconstruction.

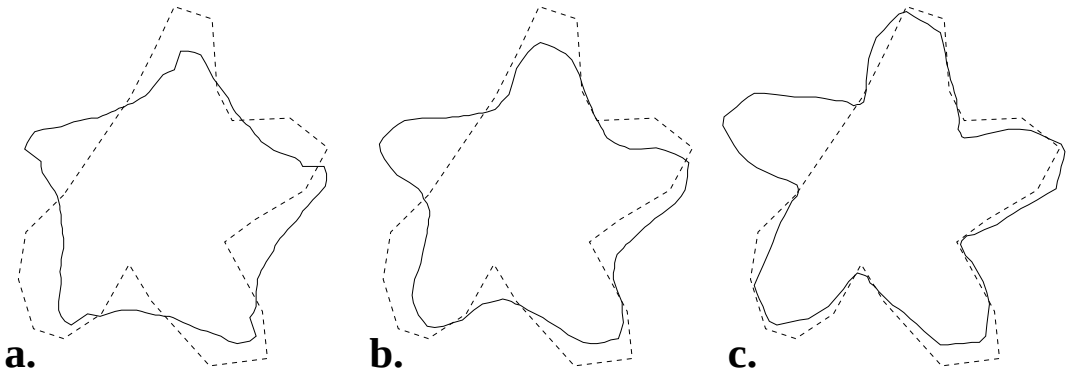

Figure 16: Reconstruction of an occluded almost symmetric shape. The original shape is shown as a dashed line. The reconstructed shape is shown as a solid line.
a) The closest symmetric shape following angular selection about the centroid. b) The closest symmetric shape following angular selection about the center of symmetry. c) The closest symmetric shape following angular selection about the centroid and altered symmetry evaluation (see text).

The improvement in reconstruction of an occluded shape is shown in Figure **??**. This method improves both the shape and the localization of the reconstruction.

8 Symmetry of Points with Uncertain Locations

Information obtained from any analytical instrument has a certain degree of uncertainty. In structural chemistry, the uncertainty may be in the location of the atoms, as obtained by, e.g. diffraction methods, due to all known causes (crystal imperfections, thermal motion, etc.)[16]. We address ourselves now, to this problem, again focusing on symmetry issues.

Quite often the data is given as a collection of probability distribution functions of point locations. Given points with such uncertain locations, the following questions are of interest:

- What is the most probable symmetric shape represented by the data.
- What is the probability distribution of symmetry measure values for the given data.

8.1 The Most Probable Symmetric Shape

Figure **??**a shows a configuration of points whose locations are given by a normal distribution function (marked as rectangles having width and length proportional to the standard deviation). In this section we show a method of evaluating the most probable symmetric shape closest to the data. For simplicity we derive the method with respect to rotational symmetry of order n (C_n-symmetry). The solution for mirror symmetry is similar (see Appendix **??**).

Given n points in 2D whose positions are given as normal probability distributions: $Q_i \sim \mathcal{N}(P_i, \Lambda_i) \quad i = 0 \ldots n-1$, we find the C_n-symmetric configuration of points $\{\hat{P}_i\}_0^{n-1}$ which is most probable.

Denote by ω the center of mass of $\hat{P}_i$: $\quad \omega = \frac{1}{n}\sum_{i=0}^{n-1} \hat{P}_i$.

Having that $\{\hat{P}_i\}_0^{n-1}$ are C_n-symmetric, the following must be satisfied:

$$\hat{P}_i = R_i(\hat{P}_0 - \omega) + \omega \tag{8}$$

for $i = 0 \ldots \quad n-1$ where R_i is a matrix representing a rotation of $2\pi i/n$ radians.

Thus, given the measurements $Q_0, \ldots, Q_{n-1}$ we need to find the most probable $\hat{P}_0$ and ω. We find $\hat{P}_0$ and ω that maximize $\mathrm{Prob}(\{\mathrm{P_i}\}_{\mathrm{i=0}}^{\mathrm{n-1}} \mid \omega, \hat{\mathrm{P}}_0)$ under the symmetry constraints of Eq. **??**.

Considering the normal distribution we have:

$$\prod_{i=0}^{n-1} k_i \exp(-\frac{1}{2}(\hat{P}_i - P_i)^t \Lambda_i^{-1}(\hat{P}_i - P_i)$$

where $k_i = \frac{1}{2}\pi \mid \Lambda_i \mid^{1/2}$. Having log being a monotonic function, we maximize:

$$\log \prod_{i=0}^{n-1} k_i \exp(-\frac{1}{2}(\hat{P}_i - P_i)^t \Lambda_i^{-1}(\hat{P}_i - P_i)$$

Thus we need to find those parameters which maximize:

$$-\frac{1}{2}\sum_{i=0}^{n-1}(\hat{P}_i - P_i)^t \Lambda_i^{-1}(\hat{P}_i - P_i)$$

under the symmetry constraint of Eq. **??**.

Substituting Eq. **??**, taking the derivative with respect to $\hat{P}_0$ and equating to zero we obtain:

$$\underbrace{(\sum_{i=0}^{n-1} R_i^t \Lambda_i^{-1} R_i)}_{A}\hat{P}_0 + \underbrace{\sum_{i=0}^{n-1} R_i^t \Lambda_i^{-1}(I - R_i)}_{B}\omega = \underbrace{\sum_{i=0}^{n-1} R_i^t \Lambda_i^{-1} P_i}_{E} \tag{9}$$

Note that $R_0 = I$ where I is the identity matrix.
When the derivative with respect to ω is zero:

$$\underbrace{(\sum_{i=0}^{n-1}(I - R_i)^t \Lambda_i R_i)}_{C}\hat{P}_0 + \underbrace{\sum_{i=0}^{n-1}(I - R_i)^t \Lambda_i^{-1}(I - R_i)}_{D}\omega = \underbrace{\sum_{i=0}^{n-1}(I - R_i)^t \Lambda_i^{-1} P_i}_{F} \tag{10}$$

Notice that when all Λ_i are equal (i.e. all points have the same uncertainty, which is equivalent to the cases in the previous sections where point location is known with no uncertainty), Eqs. **??**-**??** reduce to Eqs. **??**-**??** in Section **??**.

Reformulating Eqs. **??** and **??** in matrix formation we obtain:

$$\underbrace{\begin{pmatrix} A & B \\ C & D \end{pmatrix}}_{U}\underbrace{\begin{pmatrix} \hat{P}_0 \\ \omega \end{pmatrix}}_{V} = \underbrace{\begin{pmatrix} E \\ F \end{pmatrix}}_{Z}$$

Noting that U is symmetric we solve by inversion $V = U^{-1}Z$ and obtain the parameters ω and $\hat{P}_0$, and obtain the most probable C_n-symmetric configuration, given the measurements $\{Q_i\}_{i=0}^{n-1}$.

Similar to the representation in Section **??**, given $m = qn$ measurements $\{Q_i\}_{i=0}^{m-1}$, we consider them as q sets of n interlaced measurements: $\{Q_{iq+j}\}_{i=0}^{n-1}$ for $j = 0 \ldots q-1$ (see Figure **??**). The derivations given above are applied to each set of n measurements separately, inorder to obtain the most probable C_n-symmetric set of points $\{\hat{P}_i\}_{i=0}^{m-1}$.

Thus the symmetry constraints that must be satisfied are:

$$\hat{P}_{iq+j} = R_i(\hat{P}_j - \omega) + \omega$$

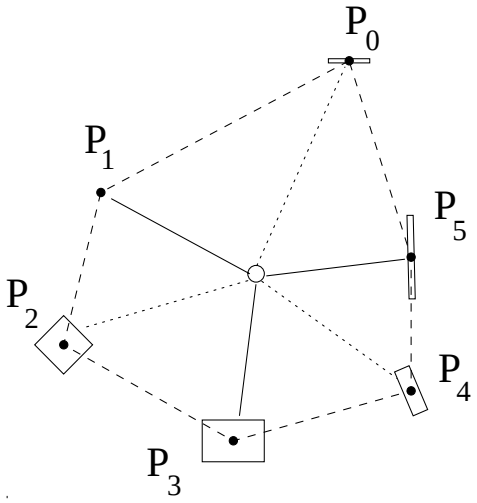

Figure 17: A configuration of 6 measurements represented as two interlaced sets of three measurements. One set is marked by the solid lines and the other set is marked by the dotted lines.

for $j = 0 \ldots \ q-1$ and $i = 0 \ldots \ n-1$ where, again, R_i is a matrix representing a rotation of $2\pi i/n$ radians and ω is the centeroid of all points $\{\hat{P}_i\}_{i=0}^{m-1}$.
As derived in Eq. **??**, we obtain for $j = 0 \ldots q-1$:

$$\underbrace{(\sum_{i=0}^{n-1} R_i^t \Lambda_{iq+j}^{-1} R_i)}_{A_j} \hat{P}_j + \underbrace{\sum_{i=0}^{n-1} R_i^t \Lambda_{iq+j}^{-1} (I - R_i)}_{B_j} \omega \ = \underbrace{\sum_{i=0}^{n-1} R_i^t \Lambda_{iq+j}^{-1} P_{iq+j}}_{E_j} \tag{11}$$

and equating to zero, the derivative with respect to ω, we obtain, similar to Eq. **??**:

$$\sum_{j=0}^{q-1} \underbrace{(\sum_{i=0}^{n-1} (I - R_i)^t \Lambda_{iq+j} R_i)}_{C_j} \hat{P}_j + \underbrace{\sum_{j=0}^{q-1} \sum_{i=0}^{n-1} (I - R_i)^t \Lambda_{iq+j}^{-1} (I - R_i)}_{D} \omega \ = \underbrace{\sum_{j=0}^{q-1} \sum_{i=0}^{n-1} (I - R_i)^t \Lambda_{iq+j}^{-1} P_{iq+j}}_{F} \tag{12}$$

Reformulating Eqs. **??** and **??** in matrix formation we obtain:

$$\underbrace{\begin{pmatrix} A_0 & & & B_0 \\ & A_1 & & B_1 \\ & & \ddots & \vdots \\ & & A_{q-1} & B_{q-1} \\ C_0 & C_1 & \cdots C_{q-1} & D \end{pmatrix}}_{U} \underbrace{\begin{pmatrix} \hat{P}_0 \\ \hat{P}_1 \\ \vdots \\ \hat{P}_{q-1} \\ \omega \end{pmatrix}}_{V} = \underbrace{\begin{pmatrix} E_0 \\ E_1 \\ \vdots \\ E_{q-1} \\ F \end{pmatrix}}_{Z}$$

Noting that U is symmetric we solve by inversion $V = U^{-1}Z$ and obtain the parameters ω and $\{\hat{P}_j\}_{j=0}^{q-1}$, and obtain the most probable C_n-symmetric configuration, $\{\hat{P}_j\}_{j=0}^{m-1}$ given the measurements $\{Q_i\}_{i=0}^{m-1}$.

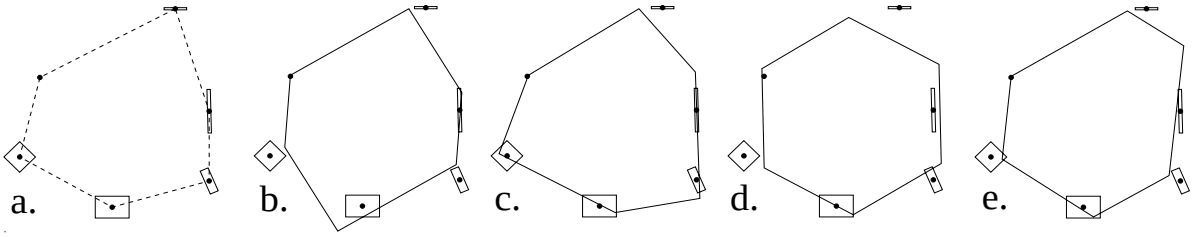

Figure 18: a) A configuration of 6 measurements given by a normal distribution function (marked as rectangles having width and length proportional to the standard deviation). The most probable symmetric shapes with respect to: b) C_2-symmetry. c) C_3-symmetry. d) C_6-symmetry. e) mirror-symmetry.

Several examples are shown in Figure ?? where for a given set of measurements (Figure ??a), the most probable symmetric shapes are shown (Figure ??b-e for C_2, C_3, C_6 and mirror symmetry respectively)

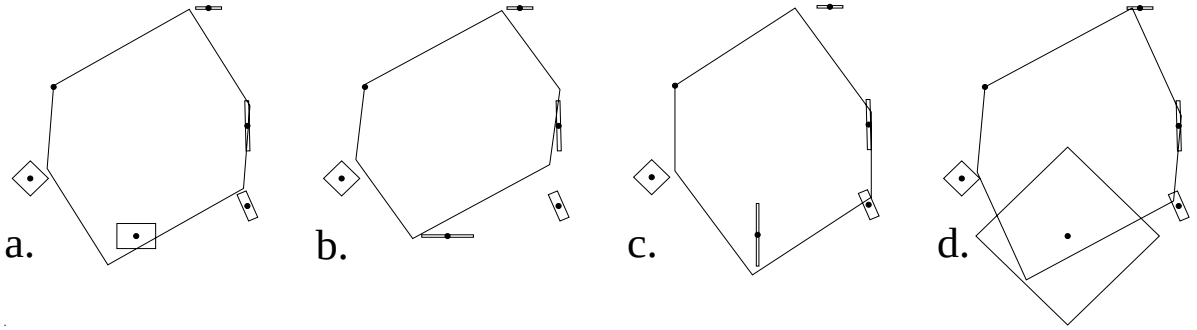

Figure 19: The most probable C_2-symmetric shape for a set of measurements after varying the probability distribution of the bottom measurement. Distributions are normal distributions marked as rectangles having width and length proportional to the standard deviation.

Figure ?? shows an example of varying the probability distribution of a measurement on the resulting symmetric shape. Figure ??a shows the most probable C_2-symmetric shape for the set of measurements of Figure ??a. Figures ??b-d show the most probable C_2-symmetric shape after varying the distribution of the bottom measurement.

8.2 The Probability Distribution of Symmetry Values

Here the question of interest is not the closest symmetric configuration, but rather the symmetry measure or the probability distribution of the symmetry measure values given the probability distributions of the point locations.

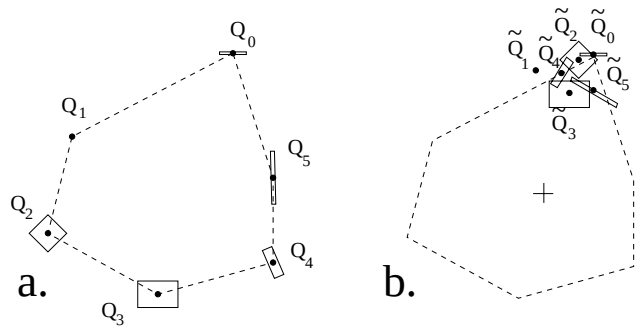

Figure 20: a) A configuration of 6 measurements Q_i. b) Each measurement Q_i was rotated $2\pi i/6$ radians about the centroid (marked as '+') obtaining measurement $\tilde{Q}_i$.

Consider the configuration of measurements in 2D given in Figure **??**a. Each measurement Q_i is a normal probability distribution $Q_i \sim \mathcal{N}(P_i, \Lambda_i)$. W.l.g. we assume the centroid of the configuration is at the origin. In order to evaluate the C_n-symmetry distribution (in our case $n = 6$) we rotate each measurement Q_i by $2\pi i/n$ radians about the origin obtaining the configuration of measurements $\tilde{Q}_i$ as in Figure **??**b.

Denote by X_i the 2-dimensional random variable having a normal distribution equal to that of measurement $\tilde{Q}_i$ i.e.

$$\begin{aligned} E(X_i) &= R_i P_i \\ \mathrm{Cov}(X_i) &= R_i \Lambda_i R_i^t \end{aligned}$$

where R_i denotes (as in Section **??**) the rotation matrix of $2\pi i/n$ radians.

Denote by Y_i the 2-dimensional random variable:

$$Y_i = X_i - \frac{1}{n} \sum_{j=0}^{n-1} X_j$$

in matrix notation:

$$\underbrace{\begin{pmatrix} Y_0 \\ \vdots \\ Y_{n-1} \end{pmatrix}}_{\mathbf{Y}} = A \underbrace{\begin{pmatrix} X_0 \\ \vdots \\ X_{n-1} \end{pmatrix}}_{\mathbf{X}}$$

or $\mathbf{Y} = A\mathbf{X}$ where $\mathbf{Y}$ and $\mathbf{X}$ are of dimension $2n$ and A is the $2n \times 2n$ matrix:

$$A = \frac{1}{n}\begin{pmatrix} n-1 & 0 & -1 & 0 & -1 & \cdots \\ 0 & n-1 & 0 & -1 & 0 & \cdots \\ -1 & 0 & \ddots & 0 & -1 & \cdots \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & \cdots & & n-1 \end{pmatrix}$$

And we have

$$E(\mathbf{X}) = \begin{pmatrix} E(X_0) \\ \vdots \\ E(X_{n-1}) \end{pmatrix} \quad \mathrm{Cov}(\mathbf{X}) = \begin{pmatrix} \mathrm{Cov}(X_0) & & \\ & \ddots & \\ & & \mathrm{Cov}(X_{n-1}) \end{pmatrix}$$

$$E(\mathbf{Y}) = AE(\mathbf{X}) \qquad \mathrm{Cov}(\mathbf{Y}) = A\mathrm{Cov}(\mathbf{X})A^t$$

The matrix $A\mathrm{Cov}(\mathbf{X})A^t$, being symmetric and positive definite, we find the $2n \times 2n$ matrix S diagonalizing $\mathrm{Cov}(\mathbf{Y})$ i.e.

$$SA\mathrm{Cov}(\mathbf{X})A^tS^t = D$$

where D is a diagonal matrix (of rank $2(n-1)$).

Denote by $\mathbf{Z} = (Z_0, \ldots, Z_{n-1})^t$ the $2n$-dimensional random variable $SA\mathbf{X}$.

$$\begin{aligned} E(\mathbf{Z}) &= SAE(\mathbf{X}) \\ \mathrm{Cov}(\mathbf{Z}) &= SA\mathrm{Cov}(\mathbf{X})A^tS^t = D \end{aligned}$$

Thus the random variables Z_i that compose $\mathbf{Z}$ are independent and, being linear combinations of X_i, they are of normal distribution.

The symmetry measure, as defined in Section **??**, is equivalent, in the current notations, to $s = \mathbf{Y}^t\mathbf{Y}$. Having S orthonormal we have

$$s = (SA\mathbf{X})^tSA\mathbf{X} = \mathbf{Z}^t\mathbf{Z}$$

If $\mathbf{Z}$ were a random variable of standard normal distribution, we would have s being of a χ^2 distribution of order $2(n-1)$. In the general case Z_i are normally distributed but not standard and $\mathbf{Z}$ cannot be standardized globally. We approximate the distribution of s as a normal distribution with

$$\begin{aligned} E(s) &= E(\mathbf{Z})^tE(\mathbf{Z}) + \mathrm{trace}D^tD \\ \mathrm{Cov}(s) &= 2\mathrm{trace}(D^tD)(D^tD) + 4E(\mathbf{Z})^tD^tDE(\mathbf{Z}) \end{aligned}$$

In Figure **??**e we display distributions of the symmetry measure as obtained for the examples in Figure **??**a-d.

Application of the method described here to thermal ellipsoids in x-ray analysis, is in progress.

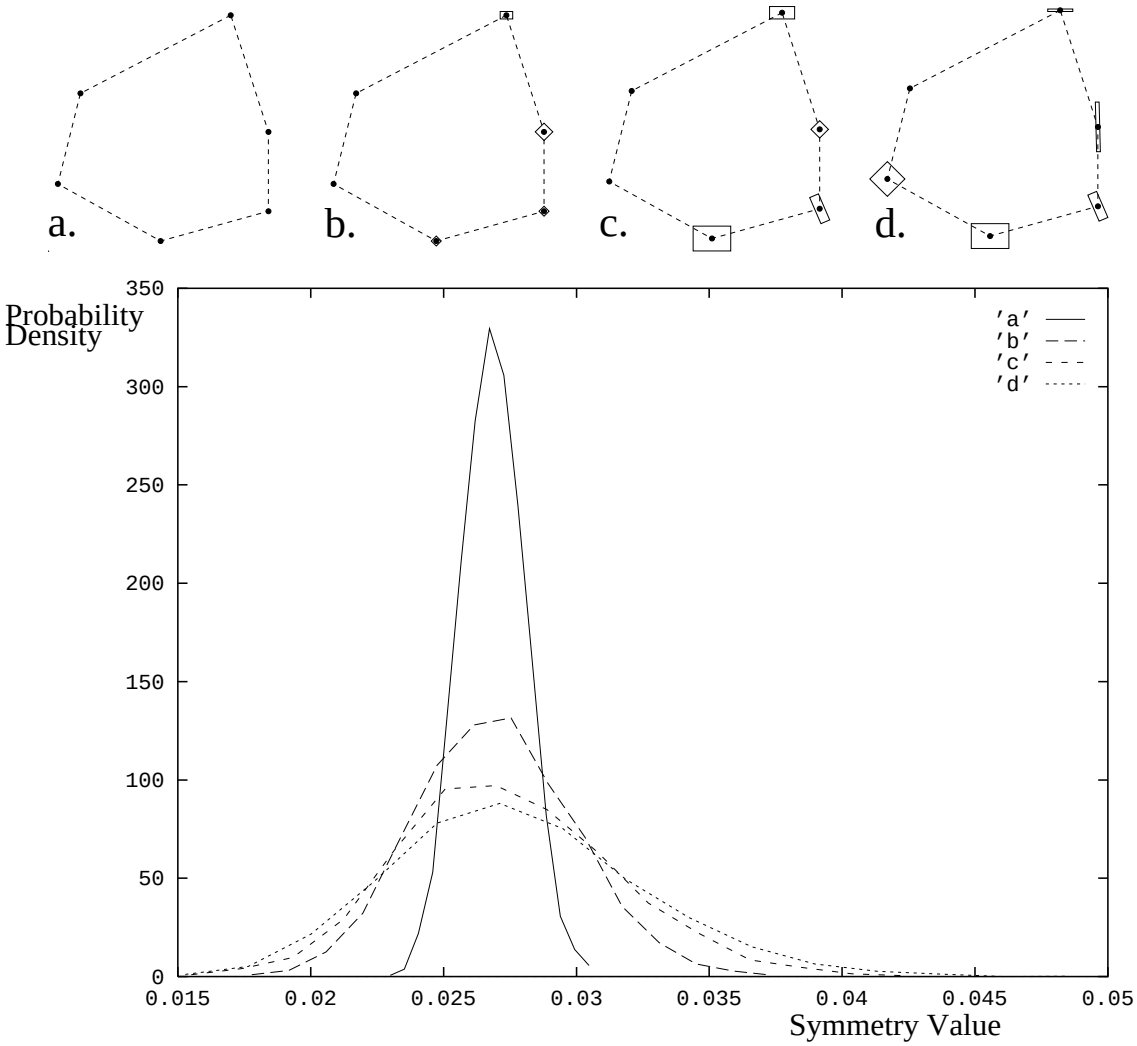

Figure 21: a-d) Some examples of configurations of measurements. e) Probability distribution of symmetry values (with respect to C_6-symmetry) for the configurations a-d.

Acknowledgments: We thank Prof. S. Peleg for his guidance and Prof. S. Shaik for his continuous interest in this project[14]. D.A. is a member of the F. Haber Research Center for Molecular Dynamics and of the Farkas Center for Light-Energy Conversion.

Appendices

A The Bounds of S values

Following the definition of the CSM in Section **??**, the CSM values are limited to the range $0 \ldots 1$ (where 1 is the normalization scale). The lower bound of the CSM is obvious from the fact that the average of the square of the distances moved by the object points, is necessarily non-negative. The upper bound of the average is limited to 1 since the object is previously normalized to maximum distance of 1 and by translation of all vertex points to the center of mass, a symmetric shape is obtained.

The upper bound on the CSM can be tightened for specific cases. For instance in 2D one can show that the maximum S' value for a triangle, with respect to C_3 is 1/3: Consider the 3 vertices of a normalized triangle P_1, P_2, P_3 in 2D (the centroid is at the origin). W.l.g. assume $P_1 = (0, 1)$ and that P_2 has a positive x-coordinate and denote by (x, y) the coordinates of P_2. Given the constraint that the centroid is at the origin, one has $P_3 = (-x, -1 - y)$. In fact P_2 is limited to a circle sector due to the centroid constraint and the normalization constraint (limiting all P_i's to be in the unit circle). Given these notations, we have that the S' value of the triangle with respect to C_3-symmetry, is given by:

$$\frac{1}{3}(1 + y^2 + y - \sqrt{3}x + x^2)$$

Considering the limited range of the P_2 coordinates, the maximum value is obtained when $P_2 = (0, 0)$ or $P_2 = (0, -1)$ (which are equivalent cases) and the maximum CSM value is 1/3.

The maximum CSM value is actually obtained for extreme cases such as a polygon of m vertices ($m = qn$) whose contour outlines a regular q-gon (i.e. every q-th vertex of the m-gon coincides with a vertex of a regular q-gon). For details, see Appendix in Part I[1].

B Orbits

We first review some basic definitions required for our proofs and derivations.

The **orbit** of x under a group G is the set $\{gx \mid g \in G\}$.
x and y belong to the same orbit if $y = gx$ for some $g \in G$.

Given a finite group G and given an ordering of its elements: $g_1, g_2, \ldots, g_m$, the **orbit** under G of a point x in Euclidean space is $x_1, \ldots, x_m$ such that $x_i = g_i x$ for $i = 1 \ldots m$. If

$g_1 = e$ (the identity element of G) then $x_i = g_i x_1$ $i = 1 \ldots m$.

Lemma 1 *The centroid of an orbit of finite point-symmetry group G is invariant under G.*

A point $x \in X$ is a **general point** (or is in general position) with respect to G if for all $g \in G, g \neq e$ (where e is the identity in G) we have $gx \neq x$.

Lemma 2 *If x is a general point with respect to G then all points in the orbit of x are general points. Furthermore for $g_1, g_2 \in G$ $g_1 \neq g_2 \Rightarrow g_1 x \neq g_2 x$.*

Thus if x is a general point its orbit contains $N(G)$ different points ($N(G)$ is the number of elements in group G).

Lemma 3 *If the orbit of x has a point in common with the orbit of y under G, then the two orbits are equal.*

For any $x \in X$ the group $G^x = \{g \in G \mid gx = x\}$ is called the **isotropy subgroup** of G at x and it contains all elements of group G that leave x invariant. If x is a general point, its isotropy subgroup contains a single element of G - the identity i.e. $G^x = \{e\}$.

Lemma 4 *If G is finite, the number of different points in the orbit containing x is $N(G)/N(G^x)$.*

(proof is immediate from the 1-1 relationship between points in the orbit of x and the left cosets of G^x . Each left coset of G^x consists of all elements of G that map x to a specific point y).

C Finding the Optimal Orientation in 2D

Following the derivation in Section **??** we derive, here, an analytic solution to finding the orientation (rotation matrix R) which minimizes Eq. **??** under the constraints given in Eq.s **??**-**??**. In Part I[1] (Appendix A.2) we gave the derivation for the specific case of the D_1 group having the two elements: $\{E, \sigma\}$.

In 2D there are 2 classes of point symmetry groups: the class C_n having rotational symmetry of order n and the class D_n having rotational symmetry of order n and n reflection axes. The problem of finding the minimizing orientation is irrelevant for the C_n symmetry groups and R is usually taken as I (the identity matrix). We derive here a solution for the orientation in the case where G is a D_n symmetry group.

The 2n elements of the D_n-symmetry group $(g_1, \ldots, g_{2n})$ are the n elements $E, C_n^1, C_n^2, \ldots, C_n^{n-1}$ $(g_1, \ldots, g_n$ respectively) and the n elements obtained by applying a reflection σ on each of these elements: $\sigma, \sigma C_n^1, \sigma C_n^2, \ldots, \sigma C_n^{n-1}$ $(g_{n+1}, \ldots, g_{2n}$ respectively). We denote the orientation of the symmetry group as the angle θ between the reflection axis and the y axis. Thus

$$R = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

W.l.g. we assume the centroid (w) is at the origin. The matrix representation of the rotational elements of D_n is then $g_i' = R^t g_i R = g_i$ for $i = 1 \ldots n$. The matrix representation of the operation σ is given by:

$$R_f = R^t g_{n+1} R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} -\cos 2\theta & -\sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{pmatrix}$$

and $g_i' = R_f g_{i-n}$ for $i = n+1 \ldots 2n$. Thus from Section **??** we must minimize the following over θ:

$$\begin{aligned} \sum_{j=1}^{2n} \|P_j - \hat{P}_1\|^2 &= \sum_{j=1}^{2n} \|g_j'^t P_j - \hat{P}_1\|^2 \\ &= \sum_{j=1}^{2n} \|g_j'^t P_j - \frac{1}{2n} \sum_{i=1}^{2n} g_j'^t P_j\|^2 \\ &= \sum_{j=1}^{2n} \| \sum_{i=1}^{n} g_i^t P_i + \sum_{i=n+1}^{2n} R_f g_i^t P_i - 2n g_j'^t P_j\|^2 \end{aligned} \tag{13}$$

Denoting by x_i, y_i the coordinates of the point $g_i'^t P_i$ and taking the derivative of Eq. **??** with respect to θ we obtain:

$$\boxed{\tan 2\theta = \frac{\sum_{i=1}^{n} \sum_{j=n+1}^{2n} (x_i y_j + x_j y_i)}{\sum_{i=1}^{n} \sum_{j=n+1}^{2n} (x_i x_j - y_i y_j)}} \tag{14}$$

which is an analytic solution for the 2D case of orientation. However in higher dimensions a minimization procedure is used.

D The Most Probable Mirror Symmetric Shape

In Section ?? we described a method for finding the most probable rotationaly symmetric shape given measurements of point location. The solution for mirror symmetry is similar. In this case, given m measurements (where $m = 2q$), the unknown parameters are $\{\hat{P}_j\}_{j=0}^{q-1}, \omega$ and θ where θ is the angle of the reflection axis. However these parameters are redundant and we reduce the dimensionality of the problem by replacing the 2 dimensional ω with the one dimensional x_0 representing the x-coordinate at which the reflection axis intersects the x-axis. Additionally we replace R_i the rotation matrix with $R = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$ the reflection about an axis at an angle θ to the x-axis. The angle θ is found analyticaly (see ref. 1) thus the dimensionality of the problem is $2q + 1$ (rather than $2q + 2$) and elimination of the last row and column of matrix U (see Section ??) allows an inverse solution as in the rotational symmetry case.

REFERENCES

1. Zabrodsky, H.; Peleg, S.; Avnir, D. *J. Am. Chem. Soc.* 1992, *114*, 7843.
2. Zabrodsky, H.; Peleg, S.; Avnir, D. *J. Am. Chem. Soc.* 1993, *115*, 8278.
3. Table 6-12 in ref.5.
4. Fleming, I. *Frontier Orbitals and Organic Chemical Reactions*; Wiley: Chichester, 1987.
5. Hargittai, I.; Hargittai, M.; *Symmetry Through the Eyes of a Chemist*; VCH: Weinheim, 1986.
6. Murray-Rust, P.; Bürgi, H.B.; Dunitz , J. D. *Acta Cryst.* 1978, *B34*, 1787.
7. Cammi, R.; Cavalli, E., *Acta Cryst.* 1992, *B48*, 245.
8. Mezey, P.G., in "New Theoretical Concepts for Understanding Organic Reactions", Bertrán, J.; Csizmadia, I.G., eds., Kluwer: Dordrecht, 1989, p. 55 and p. 77.
9. Maruani, J.; Mezey. P.G. *Compt. Rend. hebd. Séanc. Acad. Sci. Paris*, II, 1987, *305*, 1051 (Erratum: ibid, 1988, *306*, 1141); Mezey, P.G.; Maruani, *J. Molec. Phys.* 1990 *69*, 97; idem, Int. J. Quant. Chem. 1993, *45*, 177.
10. Mezey. P.G., *J. Math. Chem.* 1992, *11*, 27.
11. E.g. Bunker, P.R. "Molecular Symmetry and Spectroscopy", Academic Press: New York, 1979, chapter 11.
12. Grünbaum, B., *Proc. Pure Math: Am. Math. Soc.* 1963, *7*, 233.
13. Calvo C., *Can. J. Chem.* 1969, *47*, 3409-3416.
14. The contours were computes as Slater-type orbitals, represented by three Gaussian functions (STO-3G) using GAMESS (General Atomic and Molecular Electronic Structure System) program.[15] We thank Dr. David Danovich and Prof. Sasson Shaik for suggesting this molecule and for performing the calculations.
15. Schmidt, M.W.; Baldridge, K.K.; Boatz, J.A.; Jensen, J.H.; Koseki, S.; Gordon, M.S.; Nguyen, K.A.; Windus, T.L.; Elbert, S.T., *Quant. Chem. Program Exchange Bulletin* 1990, *10*, 52.
16. Stout, G.H.; Jensen, L.H., "X-Ray Structure Determination", 2nd ed., Wiley, Newyork, 1989.