BEST NON-SPHERICAL SYMMETRIC LOW RANK APPROXIMATION [*]

MILI I. SHAH [†] AND DANNY C. SORENSEN [‡]

Abstract. The symmetry preserving singular value decomposition (SPSVD) produces the best symmetric (low rank) approximation to a set of data. These symmetric approximations are characterized via an invariance under the action of a symmetry group on the set of data. The symmetry groups of interest consist of all the non-spherical symmetry groups in three dimensions. This set includes the rotational, reflectional, dihedral, and inversion symmetry groups. In order to calculate the best symmetric (low rank) approximation, the symmetry of the data set must be determined. Therefore, matrix representations for each of the non-spherical symmetry groups have been formulated. These new matrix representations lead directly to a novel reweighting iterative method to determine the symmetry of a given data set by solving a series of minimization problems. Once the symmetry of the data set is found, the best symmetric (low rank) approximation in the Frobenius norm and matrix 2-norm can be established by using the SPSVD.

Key words. singular value decomposition, symmetry, symmetry operation, symmetry constraints, rotation, reflection, dihedral, inversion, large scale, protein dynamics

AMS subject classifications. 15A18, 65F15

1. Introduction. This paper is concerned with the approximation of a set of points in the three-dimensional real space $\Re^3$ that is known to have spatial symmetries, perhaps slightly perturbed by noise. To address this structured approximation problem, we developed a symmetry preserving singular value decomposition (SPSVD) in [16]. This SPSVD was shown to provide the best symmetric approximation (in the 2- and Frobenius- norms) to a set of spatial data and low rank symmetric SVD approximations were obtained by truncation of this SPSVD. Here, this work is extended in two ways. First, we provide a new proof of optimality that rigorously establishes the low rank approximation obtained via SPSVD truncation is in fact the best low rank symmetric approximation (of specified rank k) to the given data set. Second, the symmetries of interest are generalized from considering only reflections or rotations to considering all the non-spherical symmetry groups in three dimensions (see Figure 1.1). Here we provide a systematic characterization of all the non-spherical symmetry groups in terms of matrix representations of their generators. These matrix representations are determined through the calculation of a few (at most three) vectors that determine the major and minor axes of symmetry. These new characterizations are essential for extending the methodology we developed in [16] to all the non-spherical symmetry groups. This generalization is necessary for many applications. Specifically, in the area of protein dynamics many proteins exhibit symmetries that are more complex than just reflective or rotational symmetry. In this paper, we demonstrate that the SPSVD applied to such proteins is more accurate in decreasing noise that may occur during molecular dynamic simulations when compared to the conventional singular value decomposition (SVD). Moreover, taking advantage of symmetry reduces the computational costs and storage requirements of the SPSVD as compared to the SVD.

[†]Loyola College, Department of Mathematical Sciences, 4501 N. Charles St., Baltimore, MD, 21210 (mishah@loyola.edu).

[‡]Rice University, Department of Computational and Applied Mathematics, 6100 Main St. MS 134, Houston, Texas, 77005-1892 (sorensen@rice.edu).

$$C_8 \qquad \overline{C_8} \qquad C_{16}C_8$$

$$D_8 \qquad \overline{D_8} \qquad D_8C_8 \qquad D_{16}D_8$$
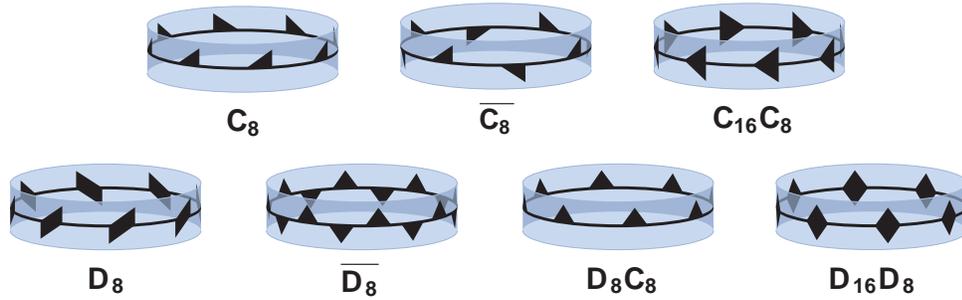
FIG. 1.1. *Example of each of the seven infinite series that define all the non-spherical symmetry groups in three dimensions for $k = 8$. C_8 consists of rotations about one axis through the center c by angles $\pi/4$, $\overline{C_8}$ consists of rotations of $\pi/4$ about one axis through the center c immediately followed by a reflection across the plane perpendicular to the axis, $C_{16}C_8$ is C_8 combined with reflections about the plane perpendicular to the axis, D_8 is C_8 combined with rotations by π about 8 horizontal axes through c that form equal angles $\pi/8$ with each other, $\overline{D_8}$ is D_8 combined with 8 reflection planes containing the main axis of symmetry, D_8C_8 is C_8 along with 8 reflection planes containing the main axis of symmetry, and $D_{16}D_8$ is D_8 along with one plane perpendicular to the main axis of symmetry (Adapted from http://en.wikipedia.org/wiki/Image:Uniaxial.png 06/10/07 with permission from Andrew Kepert).*

Calculating the SPSVD is a two-step process. In the first step, a matrix representation for the symmetry of a given data set is determined. This process is presented as a novel iterative reweighting method: a scheme which is rapidly convergent in practice and seems to be extremely effective in ignoring outliers of the data. In the second step, the best approximation that maintains the symmetry calculated from the first step is computed. This approximation is designated the SPSVD of the data set.

There has been considerable research related to the first step of the SPSVD, symmetry detection and formulation [1,4,9,12–14,18,22]. However, this previous research is not applicable to the work presented here since certain information necessary to characterize the symmetry of the data in matrix form is not available. For example, both the angle and axis of rotation are necessary in order to compute the standard Rodrigues matrix. However, in many situations, only the angle of rotation is known. Therefore, it is necessary to create a new formulation of rotation in the case where the axis of symmetry is not given. An exception is the work done by Pinsky et al. [14]. In this paper, Pinsky et al. describe methods to determine inversions, rotations, and reflections. It should be noted that their work for rotations and reflections is equivalent to our earlier work shown in [16]. In the case of inversions, the method shown in this paper is equivalent to the Pinsky et al. formulation. Another point concerning earlier research is that no matrix formulation has been given to characterize many of the symmetry groups in three dimensions. The solution to this problem is presented here by formulating a concise matrix representation for each of the seven infinite series, which defines all the non-spherical symmetry groups in three dimensions. The cyclic, inversion, and dihedral symmetry groups are included in this series.

In addition, there has been some work that is related to the second step of the SPSVD, symmetric approximations [2, 10, 11, 22, 23]. Specifically, Zabrodsky et al. [23] define the *Folding Method*, which is equivalent to the symmetric approximation discussed in this paper. However, their proof does not reveal the best symmetric low rank approximation to a set, nor can it be efficiently calculated for large scale matrices as is possible with the SPSVD.

This paper is organized as follows. Section 2 defines symmetry and characterizes the generators for each of the symmetry groups of the seven infinite series in readily computed matrix form. Section 3 describes an algorithm to identify the symmetry group for a given set of correctly matched data. Section 4 extends this identification of symmetry groups by creating an iterative method that effectively ignores outliers that are inherent in a noisy data set during symmetry detection. Section 5 constructs the SPSVD which produces the best symmetric low rank approximation to a set of data. Finally, Section 6 presents applications of the SPSVD to protein dynamics.

Throughout the discussion, $\|\cdot\|$ shall denote the 2-norm and $\|\cdot\|_F$ shall represent the Frobenius norm. The term *smallest eigenvalue* will refer to the algebraically smallest eigenvalue of a symmetric matrix. The n-dimensional identity matrix will be denoted as $\mathbf{I}_n$, while the three-dimensional identity matrix will be denoted as $\mathbf{I}$. All vectors are column vectors.

2. Defining Symmetry. Symmetry may be classified as a set of invertible linear transformations from $\Re^3 \to \Re^3$ that satisfy the group properties:
- The inverse of a transformation belonging to the set also belongs to the set.
- The product of two transformations belonging to the set also belongs to the set.

As a result, the linear transformations are isomorphic to a group of nonsingular matrices [17]. Moreover, this group of nonsingular matrices must be scale preserving [20]. In other words, if a symmetry group contains more than one element, then the matrices must be orthogonal.

In order to characterize specific symmetry groups, certain definitions are now offered. The number of elements in the group is called the *order* of the group. Note that groups may be either finite or infinite. For example, the set of invertible real $n \times n$ matrices forms an infinite group, while the set, $\{\mathbf{I}, -\mathbf{I}\}$, forms a finite group of order 2. A group may be defined by its generator. Here, a subset of a group is a *generator* if every element of the group can be written as a (finite) product of elements of the subset and their inverses [5]. Conventionally, generators are represented by $\langle \cdot \rangle$. For example, $\{\mathbf{I}, -\mathbf{I}\}$ is generated by $\langle -\mathbf{I} \rangle$. When a group, $\mathbb{G}$, acts on a set, $\mathcal{S}$, it permutes the elements of $\mathcal{S}$. For a specific element $\mathbf{s} \in \mathcal{S}$, the movement of $\mathbf{s}$ is defined as the *orbit* of $\mathbf{s}$; i.e.,

$$\mathcal{O}_{\mathbb{G}}(\mathbf{s}) = \{\mathbf{Gs} : \mathbf{G} \in \mathbb{G}\}.$$

Therefore, if $\mathbb{G} = \{\mathbf{I}_2, -\mathbf{I}_2\}$, then

$$\mathcal{O}_{\mathbb{G}}\left(\begin{array}{c} 1 \\ 1 \end{array} \right) = \left\{ \left(\begin{array}{c} 1 \\ 1 \end{array} \right), \left(\begin{array}{c} -1 \\ -1 \end{array} \right) \right\}.$$

As stated above, a group may be characterized by its generator. In the case of symmetry, this characterization implies that there exists a finite set of orthogonal matrices that can generate the full symmetry group of interest. It can be shown that this generator is a composition of reflections and rotations. Therefore, once matrix representations for reflection and rotation are formulated, then all symmetry groups may be defined in terms of these two representations [16].

DEFINITION 2.1. *A set of points* $\mathcal{S} \subset \Re^3 \bigcap \mathbf{w}^\perp$ *is reflectively symmetric with respect to the hyperplane* $\mathcal{H}$ *if for every point* $\mathbf{s} \in \mathcal{S}$, *there exists a point* $\hat{\mathbf{s}} \in \mathcal{S}$ *such that* $\hat{\mathbf{s}} = \mathbf{s} + \tau \mathbf{w}$ *for some scalar* τ *with* $\mathbf{s} + \frac{\tau}{2}\mathbf{w} \in \mathcal{H}$. *Here, a hyperplane* $\mathcal{H}$ *is specified by a constant* γ *and a vector* $\mathbf{w}$ *via* $\mathcal{H} := \{\mathbf{x} : \gamma + \mathbf{w}^T \mathbf{x} = 0\}$. *In this case, the vector* $\mathbf{w}$ *is called the* normal to the plane. *Note that the center* $\mathbf{c} \equiv \frac{1}{m} \sum_{\mathbf{s} \in \mathcal{S}} \mathbf{s}$ *of*

the point set lies in the plane of symmetry, where m is the number of elements in $\mathcal{S}$. Moreover, since the data is assumed to be mean-adjusted, the center is at the origin, $\mathbf{c} = 0$ which implies $\gamma = 0$.

The following lemma is an immediate consequence of the fact that for each $\mathbf{s} \in \mathcal{S}$ there is a reflected point $\hat{\mathbf{s}} = \mathbf{s} + \tau\mathbf{w} \in \mathcal{S}$.

LEMMA 2.2. *A set $\mathcal{S}$ is reflectively symmetric with respect to a hyperplane $\mathcal{H}$ with unit normal $\mathbf{w}$ if and only if*

$$\mathcal{S} = \mathbf{W}\mathcal{S} = (\mathbf{I} - 2\mathbf{w}\mathbf{w}^T)\mathcal{S},$$

where $\mathbf{W} = \mathbf{I} - 2\mathbf{w}\mathbf{w}^T$ is known as the reflection matrix.

DEFINITION 2.3. *A set of points $\mathcal{S} \subset \Re^3 \bigcap \mathbf{q}^\perp$ is k-fold rotationally symmetric about an axis $\mathbf{q} \in \Re^3$ if there exist an 3×3 orthogonal matrix $\mathbf{C}_k$ such that for every point $\mathbf{s} \in \mathcal{S}$, there are exactly $k - 1$ distinct points $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{k-1} \in \mathcal{S}$ with $\mathbf{C}_k^i\mathbf{s} = \mathbf{s}_i$ for $i = 1, 2, \ldots, k - 1$.* The unit vector $\mathbf{q}$ is called the *axis of symmetry*, while $\mathbf{C}_k$ is known as the *rotation matrix*. Lemma 2.4 gives an expression for the rotation matrix $\mathbf{C}_k$.

LEMMA 2.4. *A set $\mathcal{S}$ is k-fold rotationally symmetric with respect to an axis of symmetry $\mathbf{q}$ if and only if there exists some $\mathbf{Q} \in \Re^{3 \times 2}$ and $\mathbf{G}_k \in \Re^{2 \times 2}$ such that for $i = 1, 2, \ldots, k - 1$,*

$$\mathcal{S} = \mathbf{C}_k^i\mathcal{S} = (\mathbf{I} - \mathbf{Q}\mathbf{G}_k\mathbf{Q}^T)^i\mathcal{S},$$

where $[\mathbf{q}, \mathbf{Q}]$ forms an orthogonal matrix and $\mathbf{I}_2 - \mathbf{G}_k$ is a 2×2 orthogonal matrix that describes a plane rotation through an angle of $\theta = 2\pi/k$ degrees.

Using these formulations for reflection and rotation that are discussed in greater detail in [15], the seven infinite series that define all the non-spherical symmetry groups in three dimensions may now be generated (see Figure 1.1). The classification is split into two sets, as adapted from Weyl [20]: (orientation preserving) proper rotations and (non-orientation preserving) improper rotations. Observe that the groups of proper rotations for the seven infinite series are given by $\mathbb{C}_k$ and $\mathbb{D}_k$. In these cases, the cyclic group $\mathbb{C}_k$ represents rotations about one axis through the center $\mathbf{c}$ by angles $2\pi/k$, and the dihedral group $\mathbb{D}_k$ consists of these rotations combined with rotations by π about k horizontal axes through $\mathbf{c}$ that form equal angles π/k with each other. Therefore, the cyclic group is generated by $\langle \mathbf{C}_k \rangle$ while the dihedral group is generated by $\langle \mathbf{C}_k, \mathbf{C}_2 \rangle$. Notice that in the case of dihedral symmetry, the axis of symmetry for $\mathbf{C}_k$ is perpendicular to the axis of symmetry for $\mathbf{C}_2$.

The improper rotations may be added to the classification of the seven infinite series in only two ways as outlined in Weyl [20]:

1. Adding the reflection $\mathbf{Z}$ about the center $\mathbf{c}$ (also known as inversion). In other words, $\mathbf{Z}$ carries any point $\mathbf{p}$ to its symmetric counterpart $\mathbf{p}'$ found by lengthening the straight line $\mathbf{pc}$ to $\mathbf{cp}'$. Therefore, for a group $\mathbb{G}$ of proper rotations $\mathbf{S}$, a new group $\overline{\mathbb{G}} = \mathbb{G} + \mathbf{Z}\mathbb{G}$ is formed such that $\mathbf{Z}\mathbb{G}$ contains the improper rotations $\mathbf{Z}\mathbf{S}$.

2. Substituting some proper rotations $\mathbf{S}$ by improper rotations $\mathbf{Z}\mathbf{S}$ as stated above. Hence, if all proper rotations $\mathbb{P}'$ in the difference $\mathbb{G}/\mathbb{P}$, where $\mathbb{P}$ is a subgroup of a proper rotation group $\mathbb{G}$ of index 2, is replaced with $\mathbf{Z}\mathbb{P}'$, a new group of improper rotations $\mathbb{G}\mathbb{P}$ is formed. Note that half of this new group consists of the proper rotations $\mathbb{P}$, while the other half is improper.

Starting with the first method, the set of improper rotations are constructed. Adjoining the inversion, $\mathbf{Z}$, to the cyclic group, $\mathbb{C}_k$, results in the group of k-fold

inversions, $\overline{\mathbb{C}}_k$. A body is said to be $\overline{\mathbb{C}}_k$ if it is invariant under the combined transformations of rotation of $2\pi/k$ degrees about an axis and reflection in the plane perpendicular to that axis. Since this symmetry group is a composition of rotation and reflection, it is generated by $\langle \mathbf{C}_k \mathbf{W}_h \rangle$, where $\mathbf{W}_h$ represents reflection along the plane perpendicular (horizontal) to the axis of symmetry for $\mathbf{C}_k$. It should be noted that the order of transformations does not matter for this case since $\mathbf{C}_k \mathbf{W}_h = \mathbf{W}_h \mathbf{C}_k$. The antiprismatic symmetric group, $\overline{\mathbb{D}}_k$, is formed when $\mathbf{Z}\mathbb{D}_k$ is appended to the dihedral group, $\mathbb{D}_k$. This attachment results in the addition of k reflection planes (between the binary axes) containing the main axis of symmetry to the elements of the dihedral group. Therefore, the antiprismatic group may be generated by $\langle \mathbf{C}_k, \mathbf{C}_2, \mathbf{W}_v \rangle$. Here, the axis of symmetry for $\mathbf{C}_2$ is perpendicular to the axis of $\mathbf{C}_k$ (as is the case for dihedral symmetry), and the plane of reflection $\mathbf{W}_v$ runs along (vertical) the axis of symmetry of $\mathbf{C}_k$.

Using the second procedure on $\mathbb{C}_k$ and $\mathbb{D}_k$ results in the following three groups. Beginning with the group $\mathbb{C}_{2k}$, the group $\mathbb{C}_k$ is indeed a subgroup of order 2. Thus, if the substitution, as outlined in the second method, is performed, then the prismatic group $\mathbb{C}_{2k}\mathbb{C}_k$ is constructed. This group contains rotations of $2\pi/k$ degrees about an axis along with reflections about the perpendicular plane. Thus, the group is generated by $\langle \mathbf{C}_k, \mathbf{W}_h \rangle$, where $\mathbf{W}_h$ is reflection along the plane perpendicular (horizontal) to the axis of symmetry of $\mathbf{C}_k$. Next, consider the dihedral groups. The only subgroups of $\mathbb{D}_k$ of order 2 are $\mathbb{C}_k$ and $\mathbb{D}_{k/2}$ (for k even). The pyramidal group, $\mathbb{D}_k\mathbb{C}_k$, constructs k reflective planes running through the main axis of symmetry along with the base rotational group $\mathbb{C}_k$, while the bipyramidal group, $\mathbb{D}_{2k}\mathbb{D}_k$, appends a perpendicular plane of symmetry to the dihedral group. Therefore, the pyramidal group is generated by $\langle \mathbf{C}_k, \mathbf{W}_v \rangle$, and the bipyramidal group is generated by $\langle \mathbf{C}_k, \mathbf{C}_2, \mathbf{W}_h \rangle$. Again, note that $\mathbf{W}_v$ is the plane that runs along (vertical) the axis of symmetry of $\mathbf{C}_k$, while $\mathbf{W}_h$ is the plane that runs perpendicular (horizontal) to the axis of symmetry of $\mathbf{C}_k$. For both cases, $\mathbf{W}_v$ and $\mathbf{W}_h$, the reflection matrix takes the form $\mathbf{I} - 2\mathbf{w}\mathbf{w}^T$. Also, the axis of symmetry for $\mathbf{C}_2$ is perpendicular to the axis of symmetry for $\mathbf{C}_k$.

In conclusion, the seven infinite series in three dimensions are

$$\mathbb{C}_k, \ \overline{\mathbb{C}}_k, \ \mathbb{C}_{2k}\mathbb{C}_k \qquad\qquad \text{for } k = 1, 2, \ldots$$
$$\mathbb{D}_k, \ \overline{\mathbb{D}}_k, \ \mathbb{D}_k\mathbb{C}_k, \ \mathbb{D}_{2k}\mathbb{D}_k \qquad\qquad \text{for } k = 1, 2, \ldots$$

The generator for each symmetry group along with the order of the group can be seen in Table 2.1. A note should be made with regards to the order of $\overline{\mathbb{C}}_k = \langle \mathbf{C}_k \mathbf{W}_h \rangle$. Here, k is assumed to be even, since

$$\mathbf{C}_k \mathbf{W}_h = \mathbf{I} - 2\mathbf{q}\mathbf{q}^T - \mathbf{Q}\mathbf{G}_k\mathbf{Q}^T,$$

where reflection about the normal $\mathbf{q}$ is represented by $\mathbf{W}_h = \mathbf{I} - 2\mathbf{q}\mathbf{q}^T$ and k-fold rotation about the axis $\mathbf{q}$ is denoted as $\mathbf{C}_k = \mathbf{I} - \mathbf{Q}\mathbf{G}_k\mathbf{Q}^T$. Therefore,
- For j odd, $\mathbf{W}_h^j = \mathbf{W}_h$ and $(\mathbf{C}_k\mathbf{W}_h)^j = \mathbf{C}_k^j\mathbf{W}_h$.
- For j even, $\mathbf{W}_h^j = \mathbf{I}$ and $(\mathbf{C}_k\mathbf{W}_h)^j = \mathbf{C}_k^j$.

Hence, there is a difference in operation generated by $\mathbf{C}_k\mathbf{W}_h$ depending on whether k is even or odd. If k is odd, reflection ($\mathbf{W}_h$) and k-fold rotation ($\mathbf{C}_k$) must exist independently as the following demonstrates:

$$\mathbf{C}_k^k \mathbf{W}_h^k = \mathbf{W}_h \Rightarrow \mathbf{W}_h \in \langle \mathbf{C}_k \mathbf{W}_h \rangle,$$

Notation	Generator	Order	Description
$\mathbb{C}_k$	$\langle \mathbf{C}_k \rangle$	k	Rotations of $2\pi/k$ about one axis
$\overline{\mathbb{C}_k}$	$\langle \mathbf{C}_k \mathbf{W}_h \rangle$	k (even)	C_k followed by perpendicular reflection
$\mathbb{C}_{2k}\mathbb{C}_k$	$\langle \mathbf{C}_k, \mathbf{W}_h \rangle$	k	C_k along with perpendicular reflection
$\mathbb{D}_k$	$\langle \mathbf{C}_k, \mathbf{C}_2 \rangle$	$2k$	C_k along with π rotations about k axes
$\mathbb{D}_k\mathbb{C}_k$	$\langle \mathbf{C}_k, \mathbf{W}_v \rangle$	$2k$	C_k along with k reflection planes
$\overline{\mathbb{D}_k}$	$\langle \mathbf{C}_k, \mathbf{C}_2, \mathbf{W}_v \rangle$	$4k$	D_k along with k reflection planes
$\mathbb{D}_{2k}\mathbb{D}_k$	$\langle \mathbf{C}_k, \mathbf{C}_2, \mathbf{W}_h \rangle$	$4k$	D_k along with perpendicular reflection

TABLE 2.1
The seven infinite series

and

$$\left. \begin{array}{l} \mathbf{C}_k^{k-1}\mathbf{W}^{k-1} = \mathbf{C}_k^{k-1} \\ \mathbf{C}_k^{k+1}\mathbf{W}^{k+1} = \mathbf{C}_k^{1} \end{array} \right\} \Rightarrow \mathbf{C}_k \in \langle \mathbf{C}_k \mathbf{W}_h \rangle.$$

Thus, $\overline{\mathbb{C}_k} = \mathbb{C}_{2k}\mathbb{C}_k$ if k is odd. This is not necessarily the case when k is even [8]. Therefore, when dealing with $\overline{\mathbb{C}_k}$, k is assumed to be even.

Now that the seven infinite series have been formed, methods to calculate the symmetry group for a given set of correctly matched data may now be constructed.

3. Calculating Symmetry. In the previous section, the generators for each of the symmetry groups of the seven infinite series were formulated. Here, this research is extended to the computation of the generator for a given set of correctly matched data. Methods to match the set of data given are shown in [1, 6, 23].

As discussed in the introduction, there has been considerable research in the area of symmetry detection. However, this work generally does not take advantage of information that is inherent in the data set, such as knowledge of the generator. Instead, the research assumes the information a priori. This section formulates methods to calculate the generator of symmetry for each of the seven infinite series by taking advantage of the correct matching of the data.

Classification of the generators of symmetry may be split into three sections: those that can be formulated with one axis, two axes, and three axes. To begin, the series that can be constructed with just one axis of symmetry will be considered. This discussion will be followed by methods to calculate matrix formulations for the series composed of double and triple axes.

3.1. Single axis computation. The cyclic $\mathbb{C}_k$, inversion $\overline{\mathbb{C}_k}$, and prismatic $\mathbb{C}_{2k}\mathbb{C}_k$ groups may all be constructed using just one axis of symmetry. This fact is obvious for the cyclic $\mathbb{C}_k$ and inversion $\overline{\mathbb{C}_k}$ groups, but may not be so apparent for the prismatic $\mathbb{C}_{2k}\mathbb{C}_k = \langle \mathbf{C}_{2k}, \mathbf{W}_h \rangle$ group since the generator contains two elements. However, both elements need the same axis $\mathbf{q}$ to formulate its content because

$$\mathbf{C}_k = \mathbf{I} - \mathbf{Q}\mathbf{G}_k\mathbf{Q}^T$$
$$\mathbf{W}_h = \mathbf{I} - 2\mathbf{q}\mathbf{q}^T,$$

where the columns of $\mathbf{Q}$ span the space orthogonal to $\mathbf{q}$ and $\mathbf{W}_h$ represents reflection along the plane perpendicular (horizontal) to the axis of symmetry $\mathbf{q}$.

In order to develop a means to calculate the axis of symmetry, one must recall that the data set of m points is assumed to be correctly matched. Therefore, for the

k-fold cyclic group, the data is split into k matrices $\mathbf{X}_j \in \Re^{(3 \times m/k)}$ for $j = 0, \ldots, k-1$ such that

$$\mathbf{X}_j = \mathbf{C}_k^j \mathbf{X}_0 = (\mathbf{I} - \mathbf{Q}\mathbf{G}_k\mathbf{Q}^T)^j \mathbf{X}_0, \tag{3.1}$$

where $\mathbf{C}_k$ is the k-fold rotation matrix. For the case of k-fold inversion, the data is again split into k matrices. However, here k is assumed to be even as discussed in the previous section. Therefore, for $j = 0, \ldots, k-1$

$$\mathbf{X}_j = (\mathbf{C}_k \mathbf{W}_h)^j \mathbf{X}_0,$$

where $\mathbf{C}_k \mathbf{W}_h$ is the k-fold inversion matrix. In other words,

$$\mathbf{X}_j = \left[\mathbf{q}\mathbf{q}^T + \mathbf{Q}(\mathbf{I} - \mathbf{G}_k)^j \mathbf{Q}^T \right] \mathbf{X}_0$$

for j even, and

$$\mathbf{X}_j = \left[-\mathbf{q}\mathbf{q}^T + \mathbf{Q}(\mathbf{I} - \mathbf{G}_k)^j \mathbf{Q}^T \right] \mathbf{X}_0$$

for j odd. Finally, for the k-fold prismatic group, the data is split into $2k$ matrices such that for $j = 0, 1, \ldots, k-1$,

$$\mathbf{X}_{2j} = \left[\mathbf{q}\mathbf{q}^T + \mathbf{Q}(\mathbf{I} - \mathbf{G}_k)^{2j} \mathbf{Q}^T \right] \mathbf{X}_0,$$
$$\mathbf{X}_{2j+1} = \left[-\mathbf{q}\mathbf{q}^T + \mathbf{Q}(\mathbf{I} - \mathbf{G}_k)^{2j+1} \mathbf{Q}^T \right] \mathbf{X}_0.$$

Using these relationships, a characteristic for the axis of symmetry for each of the symmetry groups may be formed. This formulation is an extension of our characterization of the cyclic group in [16].

LEMMA 3.1. *Suppose $\mathbf{X}_0$ has full rank and that $\mathbf{G}_k$ is nonsingular. Then $\mathbf{q}$ is a major axis of symmetry if and only if*

$$\mathbf{q}^T \mathbf{M} = 0,$$

where

$$\mathbf{M} = (k-1)\mathbf{X}_0 - \sum_{j=1}^{k-1} \mathbf{X}_j$$

for the cyclic group,

$$\mathbf{M} = (k-1)\mathbf{X}_0 - \sum_{j=1}^{k-1} (-1)^j \mathbf{X}_j$$

for the inversion group, and

$$\mathbf{M} = (2k-1)\mathbf{X}_0 - \sum_{j=1}^{2k-1} (-1)^j \mathbf{X}_j$$

for the prismatic group.

Proof. The case for the inversion group will only be shown here since the case for the cyclic group is shown in [16] and the case for the prismatic group follows this proof closely. First, note that if $\mathbf{q}$ is an axis of symmetry, then $\mathbf{q}^T\mathbf{Q} = 0$ must be true and thus, for j even,

$$\mathbf{q}^T\mathbf{X}_j = \mathbf{q}^T\Big[\mathbf{q}\mathbf{q}^T + \mathbf{Q}(\mathbf{I} - \mathbf{G}_k)^j\mathbf{Q}^T\Big]\mathbf{X}_0 = \mathbf{q}^T\mathbf{X}_0,$$

while for j odd,

$$\mathbf{q}^T\mathbf{X}_j = \mathbf{q}^T\Big[-\mathbf{q}\mathbf{q}^T + \mathbf{Q}(\mathbf{I} - \mathbf{G}_k)^j\mathbf{Q}^T\Big]\mathbf{X}_0 = -\mathbf{q}^T\mathbf{X}_0,$$

which implies

$$\mathbf{q}^T\mathbf{M} = \mathbf{q}^T\Big[(k-1)\mathbf{X}_0 - \sum_{j=1}^{k-1}(-1)^j\mathbf{X}_j\Big]$$

$$= (k-1)\mathbf{q}^T\mathbf{X}_0 - \sum_{j=1}^{k-1}(-1)^j\mathbf{q}^T\mathbf{X}_j$$

$$= (k-1)\mathbf{q}^T\mathbf{X}_0 - \sum_{j=1}^{k-1}\mathbf{q}^T\mathbf{X}_0$$

$$= 0.$$

Now, suppose $\hat{\mathbf{q}}$ is any unit vector that satisfies $\hat{\mathbf{q}}^T\mathbf{M} = 0$ (in place of $\mathbf{q}$). Note that

$$\sum_{j=1}^{k-1}(-1)^j\mathbf{X}_j = \sum_{j=1}^{k-1}\Big[\mathbf{q}\mathbf{q}^T + (-1)^j\mathbf{Q}(\mathbf{I} - \mathbf{G}_k)^j\mathbf{Q}^T\Big]\mathbf{X}_0$$

$$= \Big[(k-1)\mathbf{q}\mathbf{q}^T + \mathbf{Q}\Big(\sum_{j=1}^{k-1}(-1)^j(\mathbf{I} - \mathbf{G}_k)^j\Big)\mathbf{Q}^T\Big]\mathbf{X}_0$$

$$= \Big[(k-1)\mathbf{q}\mathbf{q}^T - \mathbf{Q}\mathbf{Q}^T\Big]\mathbf{X}_0$$

$$= k\mathbf{q}\mathbf{q}^T\mathbf{X}_0 - \mathbf{X}_0$$

and $(\mathbf{I} - \mathbf{G}_k)^k = \mathbf{I}$ implies $\sum_{j=1}^{k-1}(-1)^j(\mathbf{I} - \mathbf{G}_k)^j = -\mathbf{I}$ when $\mathbf{G}_k$ is nonsingular. From this, it follows that

$$\mathbf{M} = (k-1)\mathbf{X}_0 - \sum_{j=1}^{k-1}(-1)^j\mathbf{X}_j = k\Big(\mathbf{I} - \mathbf{q}\mathbf{q}^T\Big)\mathbf{X}_0.$$

Therefore, since $\mathbf{X}_0$ is full rank,

$$0 = \hat{\mathbf{q}}^T\mathbf{M} = k\hat{\mathbf{q}}^T\Big(\mathbf{I} - \mathbf{q}\mathbf{q}^T\Big)\mathbf{X}_0$$

implies that $\hat{\mathbf{q}} = \mathbf{q}(\hat{\mathbf{q}}^T\mathbf{q})$. Since both $\mathbf{q}$ and $\hat{\mathbf{q}}$ are unit length, it follows from Cauchy–Schwarz that $\hat{\mathbf{q}} = \pm\mathbf{q}$. $\square$

Therefore, the solution of the optimization problem

$$\min_{\|\mathbf{q}\|=1} \|\mathbf{q}^T\mathbf{M}\|_F \tag{3.2}$$

specifies the approximate axis of symmetry $\mathbf{q}$. Thus, $\mathbf{q}$ can be computed as follows:

LEMMA 3.2. *The solution $\mathbf{q}$ to the minimization problem (3.2) is the unit eigenvector (up to sign) corresponding to the smallest eigenvalue of $\mathbf{M}\mathbf{M}^T$.*

3.2. Double axes computation. For the dihedral $\mathbb{D}_k$ and pyramidal $\mathbb{D}_k\mathbb{C}_k$ groups, two axes – the major $\mathbf{q}$ axis and minor $\mathbf{p}$ axis – are necessary in order to compute the generator of each group. Here, the major axis of symmetry $\mathbf{q}$ denotes the axis of rotation for the $\mathbf{C}_k$ rotation, whereas the minor axis $\mathbf{p}$ represents the axis of symmetry for the $\mathbf{C}_2$ rotation for the dihedral group or the normal of $\mathbf{W}_v$ for the pyramidal group.

To calculate the generator for the dihedral and pyramidal groups, one takes advantage of the correctly matched data set. In the case of the dihedral $\mathbb{D}_k$ group, the data is matched within $2k$ matrices

$$\mathbf{X}_{2j} = \mathbf{C}_k^j \mathbf{X}_0$$
$$\mathbf{X}_{2j+1} = \mathbf{C}_2 \mathbf{X}_{2j} = \mathbf{C}_2 \mathbf{C}_k^j \mathbf{X}_0$$

for $j = 0, 1, \ldots, k-1$. Similarly, the pyramidal $\mathbb{D}_k\mathbb{C}_k$ group is matched for $j = 0, 1, \ldots, k-1$

$$\mathbf{X}_{2j} = \mathbf{C}_k^j \mathbf{X}_0$$
$$\mathbf{X}_{2j+1} = \mathbf{W}_v \mathbf{X}_{2j} = \mathbf{W}_v \mathbf{C}_k^j \mathbf{X}_0,$$

where $\mathbf{W}_v$ is the plane that runs along (vertical) the axis of symmetry.

Solving the following minimization problem:

$$\min_{\|\mathbf{q}\|=1} \|\mathbf{q}^T \mathbf{M}\|_F \tag{3.3}$$

gives the major axis of symmetry $\mathbf{q}$ for the dihedral and pyramidal groups where

$$\mathbf{M} = (2k-1)\mathbf{X}_0 - \sum_{j=1}^{2k-1} (-1)^j \mathbf{X}_j \tag{3.4}$$

for dihedral symmetry and

$$\mathbf{M} = (2k-1)\mathbf{X}_0 - \sum_{j=1}^{2k-1} \mathbf{X}_j \tag{3.5}$$

for pyramidal symmetry. This property is a consequence of the following lemma.

LEMMA 3.3. *Suppose $\mathbf{X}_0$ has full rank and that $\mathbf{G}_k$ is nonsingular. Then $\mathbf{q}$ is a major axis of symmetry if and only if*

$$\mathbf{q}^T \mathbf{M} = 0.$$

Proof. The proof follows the proof technique of Lemma 3.1 □

LEMMA 3.4. *The solution $\mathbf{q}$ to the minimization problem (3.3) is the unit eigenvector corresponding to the smallest eigenvalue of $\mathbf{M}\mathbf{M}^T$, where $\mathbf{M}$ is defined in Equation (3.4) for the dihedral group and in Equation (3.5) for the pyramidal group.*

The procedure for calculating the minor axis follows a similar method as the major axis. For both the dihedral and pyramidal cases, the data is configured into two matrices

$$\hat{\mathbf{X}}_0 = [\mathbf{X}_0, \mathbf{X}_2, \ldots, \mathbf{X}_{2k-2}]$$
$$\hat{\mathbf{X}}_1 = [\mathbf{X}_1, \mathbf{X}_3, \ldots, \mathbf{X}_{2k-1}].$$

Therefore,

$$\hat{\mathbf{X}}_1 = \mathbf{C}_2\hat{\mathbf{X}}_0$$

for the dihedral group, and

$$\hat{\mathbf{X}}_1 = \mathbf{W}_v\hat{\mathbf{X}}_0$$

for the pyramidal group. Thus, calculating the minor axis for the dihedral group follows the form of the 2-fold cyclic group. In other words, the minor axis is calculated as the axis of symmetry of Lemma 3.2, namely the minor axis is the eigenvector associated with the smallest eigenvalue of $\mathbf{N}\mathbf{N}^T$, where

$$\mathbf{N} = \hat{\mathbf{X}}_0 - \hat{\mathbf{X}}_1.$$

The case of the pyramidal group follows a different form. Since $\hat{\mathbf{X}}_1$ is reflectively symmetric to $\hat{\mathbf{X}}_0$, calculating the minor axis reduces to calculating the normal to the plane of symmetry. In our earlier work [16], we showed that the normal to the plane of symmetry can be calculated by solving

$$\min_{\|\mathbf{w}\|=1}\{\|\hat{\mathbf{X}}_0 - \mathbf{W}_v\hat{\mathbf{X}}_1\|_F\}, \tag{3.6}$$

where $\mathbf{W}_v = \mathbf{I} - 2\mathbf{w}\mathbf{w}^T$.

LEMMA 3.5. *The solution* $\mathbf{w}$ *to the minimization problem (3.6) is the unit eigenvector corresponding to the smallest eigenvalue of the symmetric indefinite matrix*

$$\mathbf{N} = \hat{\mathbf{X}}_0\hat{\mathbf{X}}_1^T + \hat{\mathbf{X}}_1\hat{\mathbf{X}}_0^T.$$

In conclusion, calculating the generator for the double axes symmetry groups results in computing a major axis $\mathbf{q}$ and a minor axis $\mathbf{p}$. Once these axes are known, the full symmetry group can be formed as

$$\langle\mathbf{C}_2, \mathbf{C}_k\rangle = \langle\mathbf{I} - 2\mathbf{P}\mathbf{P}^T, \mathbf{I} - \mathbf{Q}\mathbf{G}_k\mathbf{Q}^T\rangle,$$

for the case of the dihedral group and

$$\langle\mathbf{W}_v, \mathbf{C}_k\rangle = \langle\mathbf{I} - 2\mathbf{p}\mathbf{p}^T, \mathbf{I} - \mathbf{Q}\mathbf{G}_k\mathbf{Q}^T\rangle$$

for the case of the pyramidal group. Here, the columns of $\mathbf{P}$ and $\mathbf{Q}$ span the space perpendicular to $\mathbf{p}$ and $\mathbf{q}$, respectively.

3.3. Triple axes computation. The remaining two groups of the seven infinite series need three axes – the major $\mathbf{q}$, minor $\mathbf{p}$, and semi-minor $\mathbf{w}$ – in order to calculate their generator. The generator for both the antiprismatic $\overline{\mathbb{D}_k}$ group and the bipyramidal $\mathbb{D}_{2k}\mathbb{D}_k$ group consist of the dihedral group $\mathbb{D}_k = \langle\mathbf{C}_k, \mathbf{C}_2\rangle$ plus a reflection operator. Here, the reflection is vertical $\mathbf{W}_v$ for the case of the antiprismatic group, and horizontal $\mathbf{W}_h$ for the case of the bipyramidal group. Since each generator contains the dihedral group, the major and minor axes are computed as stated in the previous section. This section will concentrate on determining the third axis, the semi-minor axis $\mathbf{w}$.

Here, the data is split into 2 matrices, $\hat{\mathbf{X}}_0$ and $\hat{\mathbf{X}}_1$, where

$$\hat{\mathbf{X}}_0 = [\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_{2k-1}]$$

contains the dihedral group and

$$\hat{\mathbf{X}}_1 = [\mathbf{W}\mathbf{X}_0, \mathbf{W}\mathbf{X}_1, \ldots, \mathbf{W}\mathbf{X}_{2k-1}]$$

contains the reflection of the dihedral group. Again, $\mathbf{W} = \mathbf{W}_v$ for the antiprismatic group and $\mathbf{W} = \mathbf{W}_h$ for the bipyramidal group. Then the semi-minor axis $\mathbf{w}$ can be calculated as discussed in Lemma 3.5. In other words, the semi-minor axis $\mathbf{w}$ is just the eigenvector associated with the smallest eigenvalue of

$$\mathbf{N} = \hat{\mathbf{X}}_0 \hat{\mathbf{X}}_1^T + \hat{\mathbf{X}}_1 \hat{\mathbf{X}}_0^T.$$

4. Noisy Symmetry. Up to this point, we have only considered data that is perfectly symmetric. However, most applications will involve imperfectly measured observations; the given data is generally noisy. Therefore, during symmetry detection, there may be a need to weight certain elements in the data set higher than others. For instance, when calculating the generator of a protein dynamics trajectory, one may wish to place more emphasis on the docking site since this site determines the function of the protein and is where most of the dynamics occur. On the other hand, the side chains, generally, have more noise and less influence on the overall dynamics of the trajectory. Thus, less weight should be placed on those regions. This section begins by calculating the generator of symmetry for a known weighting. This is followed by an introduction to a novel iterative method that automatically chooses weightings to effectively ignore outliers of the data set.

4.1. General Weighting. It has been demonstrated that for each symmetry group, calculating the optimal axis (axes) of symmetry reduces to an optimization problem: To calculate an axis of symmetry

$$\mathbf{q} = \underset{\|\mathbf{q}\|=1}{\operatorname{argmin}} \|\mathbf{q}^T \mathbf{M}\|_F,$$

where $\mathbf{M}$ is described in the previous section, and to calculate the normal

$$\mathbf{w} = \underset{\|\mathbf{w}\|=1}{\operatorname{argmin}} \|\mathbf{X}_0 - \mathbf{W}\mathbf{X}_1\|_F,$$

where $\mathbf{W} = \mathbf{I} - 2\mathbf{w}\mathbf{w}^T$. A weighting may be inserted into these minimization problems to de-emphasize anomalies in the supposed symmetry relation. In each case, a diagonal weighting matrix $\mathbf{D} = \operatorname{diag}\{\delta_i\}$ is introduced, where the jth diagonal weights the jth column of the matrix of the objective function. Thus, the optimization problems become: To calculate an axis of symmetry

$$\mathbf{q} = \underset{\|\mathbf{q}\|=1}{\operatorname{argmin}} \|[\mathbf{q}^T \mathbf{M}]\mathbf{D}\|_F, \tag{4.1}$$

and to calculate the normal

$$\mathbf{w} = \underset{\|\mathbf{w}\|=1}{\operatorname{argmin}} \|[\mathbf{X}_0 - \mathbf{W}\mathbf{X}_1]\mathbf{D}\|_F, \tag{4.2}$$

where $\mathbf{W} = \mathbf{I} - 2\mathbf{w}\mathbf{w}^T$.

LEMMA 4.1. *The solution $\mathbf{q}$ to the minimization problem (4.1) is the unit eigenvector corresponding to the smallest eigenvalue of*

$$\mathbf{M}\mathbf{D}^2\mathbf{M}^T.$$

where $\mathbf{M}$ *may be any of the forms defined in Lemma 3.1, Equation (3.4), or Equation (3.5).*

LEMMA 4.2. *The solution* $\mathbf{w}$ *to the minimization problem (4.2) is the unit eigenvector corresponding to the smallest eigenvalue of the symmetric indefinite matrix*

$$\mathbf{N} = \mathbf{X}_0 \mathbf{D}^2 \mathbf{X}_1^T + \mathbf{X}_1 \mathbf{D}^2 \mathbf{X}_0^T.$$

Since the minimizations are with respect to the Frobenius norm, both the above optimization problems (4.1) and (4.2) can be expanded column-wise into

$$\min_{\|\mathbf{w}\|=1} \sum_{j=1}^{m} \delta_j \mathbf{w}^T \mathbf{M}_i \mathbf{w}, \tag{4.3}$$

where

$$\mathbf{M}_i = (\mathbf{M}\mathbf{e}_i)(\mathbf{M}\mathbf{e}_i)^T$$

for calculation of the axis of symmetry, and

$$\mathbf{M}_i = \left\| \mathbf{x}_i^{(0)} - \mathbf{x}_i^{(1)} \right\|^2 \mathbf{I} + 2\left(\mathbf{x}_i^{(0)} \mathbf{x}_i^{(1)T} + \mathbf{x}_i^{(1)} \mathbf{x}_i^{(0)T} \right)$$

for calculation of the normal to the plane of reflective symmetry.

Note that the calculation of the minor axis and semi-minor axis is similar to the methods given above. Once the major axis has been calculated, the data is split into two matrices $\hat{\mathbf{X}}_0$ and $\hat{\mathbf{X}}_1$ as described in Sections 3.2 and 3.3. The orthogonality properties between the axes are accomplished by projecting a guess for the (semi-) minor axis onto $\mathcal{Q}$ by the projection matrix $\mathbf{Q}\mathbf{Q}^T$, where the columns of $\mathbf{Q}$ span $\mathcal{Q}$, the space perpendicular to the axis $\mathbf{q}$.

4.2. Discrepancy Weighting. An iterative reweighting scheme is now developed to construct a $\mathbf{D}$ that diminishes the influence of outliers in the SPSVD. This weighting is adapted from our previous work [16], but here it is generalized for the generator of each of the seven infinite series.

Given a guess $\mathbf{z}$ to the normal/axis of symmetry, the weight δ_i of the minimization (4.3) is set as

$$\delta_i = (\mathbf{z}^T \mathbf{M}_i \mathbf{z})^{-1}.$$

Therefore, if $\mathbf{z}$ is a good approximation to the normal/axis, then $\mathbf{z}^T \mathbf{M}_i \mathbf{z}$ will be small; thus δ_i will be a large weight.

Define

$$F(\mathbf{z}, \mathbf{w}) = \sum_{i=1}^{m} \delta_i \mathbf{w}^T \mathbf{M}_i \mathbf{w} = \sum_{i=1}^{m} \frac{f_i(\mathbf{w})}{f_i(\mathbf{z})}$$

where $f_i(\mathbf{z}) = \mathbf{z}^T \mathbf{M}_i \mathbf{z}$. The best normal/axis with respect to this weighting, may be found as the $\mathbf{w}$ that solves the respective minimization problem described in Lemma 4.1 or Lemma 4.2. Note that the approximate $\mathbf{w}$ associated with this weighting solves

$$\min_{\|\mathbf{w}\|=1} F(\mathbf{z}, \mathbf{w}), \tag{4.4}$$
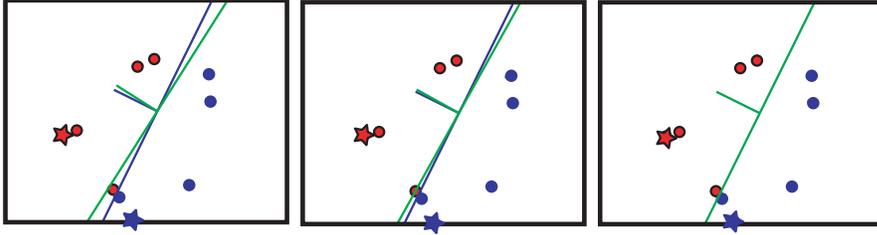
FIG. 4.1. *Convergence of discrepancy weighting. Notice how as the iterates progress, less emphasis is placed on the outliers (stars). Adapted from [16]*

which suggests an iterative reweighting scheme that adjusts the vector $\mathbf{z}$ to optimally diminish the effect of outliers. Beginning with an initial guess $\mathbf{z}_0$, iterate

$$\mathbf{z}_{p+1} = \underset{\|\mathbf{w}\|=1}{\operatorname{argmin}} F(\mathbf{z}_p, \mathbf{w}), \quad p = 0, 1, 2, \ldots \tag{4.5}$$

until $\|\mathbf{z}_{p+1} - \mathbf{z}_p\|$ is sufficiently small. Notice that the fixed point to this iteration will solve the following max-min problem

$$\max_{\|\mathbf{z}\|=1} \left\{ \min_{\|\mathbf{v}\|=1} F(\mathbf{z}, \mathbf{v}) \right\} \tag{4.6}$$

as the following lemma indicates.

LEMMA 4.3. *If $\mathbf{v} = \mathbf{z}$ is a fixed point of the minimization problem (4.4), then $\mathbf{z}$ is a solution to the max-min problem (4.6), and $F(\mathbf{z}, \mathbf{v}) = m$.*

The above lemma explains that a fixed point of iteration (4.5) solves the max-min problem (4.6). The existence of a fixed point to the iteration (4.5) is shown in Theorem 4.4.

THEOREM 4.4. *There is a point $\mathbf{z}_*$ of unit norm such that*

$$\mathbf{z}_* = \underset{\|\mathbf{w}\|=1}{\operatorname{argmin}} F(\mathbf{z}_*, \mathbf{w}). \tag{4.7}$$

The proofs of these results are essentially the same as the proofs given for the corresponding results (Lemma 3.3 and Theorem 3.4) in [16], and hence are not repeated here. Together, these results show that there is at least one fixed point that solves (4.7) and that any such point solves the max-min problem (4.6). Iteration (4.5) is designed to produce such a fixed point and hence solve the max-min problem (4.6).

Remark: Theorem 4.4 assumes $\Phi(\mathbf{z}) \neq 0$. This is a reasonable assumption since the only way $\Phi(\mathbf{z}) = 0$ is if $\|\mathbf{x}_j^{(0)}\| = \|\mathbf{x}_j^{(1)}\| = \ldots = \|\mathbf{x}_j^{(k-1)}\|$ for some n-tuplet $(\mathbf{x}_j^{(0)}, \mathbf{x}_j^{(1)}, \ldots, \mathbf{x}_j^{(k-1)})$, where k is dependent on the order of the symmetry group in consideration. Since the sets are assumed to be noisy, it is unlikely that these norms are precisely equal in practice. We have created another formulation that solves this problem which involves taking the inner-product between the current and previous iterate. Though the analysis of this inner-product weighting is not as complete as the weighting presented in this paper. Details can be found in the technical report [15].

The convergence history depicted in Figure 4.1 is typical, and iteration (4.5) seems to be convergent in practice, though no analytic proof showing the convergence of the iterates $\mathbf{z}_p$ has been given. However, the sequence of function values does converge. This fact is established with the following theorem, Theorem 4.5.

THEOREM 4.5. *The sequence*

$$F(\mathbf{z}_p, \mathbf{z}_{p+1}) \to m,$$

as $p \to \infty$.

This discrepancy weighting has been very effective in ignoring anomalies in real-life applications. Specifically, we have tested this method in the area of molecular dynamics. Results can be seen in Section 6.

A note should be made in the case of formulating an iterative search for the double and triple axes formulation. The algorithm begins by searching for an optimal major and (semi-) minor axes of symmetry with no weights. To preserve the orthogonality conditions, the (semi-) minor axis is projected onto the space perpendicular to the major axis with a projection matrix, as described before. Then, the iteration calculates the weights by alternatively projecting the major/(semi-) minor axis onto the space perpendicular to the (semi-) minor/major axis. This projection is required to preserve the orthogonality conditions. The iteration continues until the current and previous major/(semi-) minor axis is within a specified tolerance. We have observed that the iteration may begin to oscillate between the best major and (semi-) minor axes if the orthogonality conditions between the axes are skewed due to noise. However, unless the noise is extreme, these oscillations will occur close to the pre-defined tolerance.

5. Optimal Symmetry Preserving SVD. In the previous sections, the seven infinite series were generated by the composition of two transformations: reflection and rotation. Once constructed, these orthogonal transformations will build the best symmetric approximation to a set of data that preserves the specified symmetry. The construction is specified as the statement of Theorem 5.1. This new result fully generalizes [16]. Moreover, this new proof also rigorously establishes the leading rank k truncation of the SPSVD as the best symmetric low rank approximation of rank k to the original given data.

Once the symmetry of the data is known and represented as $\{\mathbf{R}_i\}$, for $i = 0, 1, \ldots, k - 1$, where k is the order of the group, the best symmetric approximation may be constructed with a procedure based upon the following theorem. Here, the data set is assumed to be correctly matched. Recall, methodology for calculating such a matching is presented in $[1, 6, 23]$.

THEOREM 5.1. *Suppose a given data set $\mathbf{X}$ is split into correctly matched subsets $\mathbf{X}_i$ where*

$$\mathbf{R}_i^T \mathbf{X}_i = \mathbf{X}_0 + \mathbf{E}_i.$$

and $\mathbf{E}_i$ is the error resulting from a noisy symmetric data set. Then the best symmetric approximation

$$\hat{\mathbf{X}} = \begin{pmatrix} \hat{\mathbf{X}}_0 \\ \hat{\mathbf{X}}_1 \\ \vdots \\ \hat{\mathbf{X}}_{k-1} \end{pmatrix}$$

with regards to both the 2-norm and Frobenius norm can be found by minimizing

$$\min_{\hat{\mathbf{X}}_j = \mathbf{R}_j \hat{\mathbf{X}}_0} \left\| \begin{pmatrix} \mathbf{X}_0 \\ \vdots \\ \mathbf{X}_{k-1} \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{X}}_0 \\ \vdots \\ \hat{\mathbf{X}}_{k-1} \end{pmatrix} \right\|^2.$$

The solution to this minimization problem can be calculated with the SVD of

$$\mathbf{USV}^T = \begin{pmatrix} \hat{\mathbf{X}}_0 \\ \vdots \\ \hat{\mathbf{X}}_{k-1} \end{pmatrix}$$

where

$$\mathbf{U} = \frac{1}{\sqrt{k}} \begin{pmatrix} \mathbf{U}_0 \\ \vdots \\ \mathbf{U}_{k-1} \end{pmatrix}, \quad \mathbf{S} = \sqrt{k}\mathbf{S}_0, \quad \mathbf{V} = \mathbf{V}_0,$$

and

$$\mathbf{U}_j = \mathbf{R}_j\mathbf{U}_0, \quad \text{for} \quad j = 0, 1, 2, \ldots, k-1,$$

with

$$\mathbf{U}_0\mathbf{S}_0\mathbf{V}_0^T = \frac{1}{k}(\mathbf{X}_0 + \mathbf{R}_1^T\mathbf{X}_1 + \mathbf{R}_2^T\mathbf{X}_2 + \cdots + \mathbf{R}_{k-1}^T\mathbf{X}_{k-1}).$$

Moreover, the best rank-ℓ symmetric approximation to the original data set is

$$\sum_{j=1}^{\ell} \sigma_j \mathbf{u}_j \mathbf{v}_j^T$$

where $\mathbf{u}_j$ and $\mathbf{v}_j$ are the jth column of $\mathbf{U}$ and $\mathbf{V}$, respectively, and σ_j is the jth singular value of $\mathbf{S}$ ordered decreasingly.

  Proof. Consider

$$\left\| \begin{pmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_{k-2} \\ \mathbf{X}_{k-1} \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{X}}_0 \\ \hat{\mathbf{X}}_1 \\ \vdots \\ \hat{\mathbf{X}}_{k-2} \\ \hat{\mathbf{X}}_{k-1} \end{pmatrix} \right\|^2 = \left\| \mathbf{B} \left[\begin{pmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_{k-2} \\ \mathbf{X}_{k-1} \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{X}}_0 \\ \hat{\mathbf{X}}_1 \\ \vdots \\ \hat{\mathbf{X}}_{k-2} \\ \hat{\mathbf{X}}_{k-1} \end{pmatrix} \right] \right\|^2$$

$$= \left\| \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_{k-1} \\ \mathbf{X}_0 \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{X}}_1 \\ \hat{\mathbf{X}}_2 \\ \vdots \\ \hat{\mathbf{X}}_{k-1} \\ \hat{\mathbf{X}}_0 \end{pmatrix} \right\|^2$$

where the orthogonal matrix $\mathbf{B}$ is given by

$$\mathbf{B} = \begin{pmatrix} 0 & \mathbf{I} & & \\ & \ddots & \ddots & \\ & & 0 & \mathbf{I} \\ \mathbf{I} & & & 0 \end{pmatrix}.$$

For $j = 1, 2, \ldots, k-1$, define the orthogonal matrices

$$\mathbf{B}_j = \frac{1}{\sqrt{j+1}} \begin{pmatrix} \mathbf{R}_{k-j}^T & \sqrt{j}\mathbf{I} \\ -\sqrt{j}\mathbf{I} & \mathbf{R}_{k-j} \end{pmatrix} \text{ and } \hat{\mathbf{B}}_j = \begin{pmatrix} \mathbf{I}_{3(k-(j+1))} & & \\ & \mathbf{B}_j & \\ & & \mathbf{I}_{3(j-1)} \end{pmatrix}.$$

Let

$$\mathbf{Z}_0 = \mathbf{X}_0$$
$$\mathbf{Z}_j = \mathbf{R}_{k-j}^T \mathbf{X}_{k-j} + \mathbf{R}_{k-(j-1)}^T \mathbf{X}_{k-(j-1)} + \ldots$$
$$+ \mathbf{R}_{k-1}^T \mathbf{X}_{k-1} + \mathbf{X}_0$$
$$\mathbf{N}_j = \frac{-1}{\sqrt{j(j+1)}} \left(j\mathbf{X}_{k-j} - \mathbf{R}_{k-j}\mathbf{Z}_{j-1} \right).$$

Then

$$\left\| \begin{pmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_{k-2} \\ \mathbf{X}_{k-1} \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{X}}_0 \\ \hat{\mathbf{X}}_1 \\ \vdots \\ \hat{\mathbf{X}}_{k-2} \\ \hat{\mathbf{X}}_{k-1} \end{pmatrix} \right\|^2 = \left\| \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_{k-1} \\ \mathbf{X}_0 \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{X}}_1 \\ \hat{\mathbf{X}}_2 \\ \vdots \\ \hat{\mathbf{X}}_{k-1} \\ \hat{\mathbf{X}}_0 \end{pmatrix} \right\|^2$$

$$= \left\| \hat{\mathbf{B}}_j \hat{\mathbf{B}}_{j-1} \ldots \hat{\mathbf{B}}_1 \left[\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_{k-1} \\ \mathbf{X}_0 \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{X}}_1 \\ \hat{\mathbf{X}}_2 \\ \vdots \\ \hat{\mathbf{X}}_{k-1} \\ \hat{\mathbf{X}}_0 \end{pmatrix} \right] \right\|^2$$

$$= \left\| \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_{k-(j+1)} \\ \frac{1}{\sqrt{j+1}}\mathbf{Z}_j \\ \mathbf{N}_j \\ \vdots \\ \mathbf{N}_1 \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{X}}_1 \\ \hat{\mathbf{X}}_2 \\ \vdots \\ \hat{\mathbf{X}}_{k-(j+1)} \\ \sqrt{j+1}\hat{\mathbf{X}}_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right\|^2$$

since

$$\mathbf{B}_j \begin{pmatrix} \mathbf{X}_{k-j} \\ \frac{1}{\sqrt{j}}\mathbf{Z}_{j-1} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{j+1}}\mathbf{Z}_j \\ \mathbf{N}_j \end{pmatrix} \text{ and } \mathbf{B}_j \begin{pmatrix} \hat{\mathbf{X}}_{k-j} \\ \sqrt{j}\hat{\mathbf{X}}_0 \end{pmatrix} = \begin{pmatrix} \sqrt{j+1}\hat{\mathbf{X}}_0 \\ 0 \end{pmatrix}.$$

If this process continues until $j = k-1$, then

$$\left\| \begin{pmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_{k-1} \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{X}}_0 \\ \hat{\mathbf{X}}_1 \\ \vdots \\ \hat{\mathbf{X}}_{k-1} \end{pmatrix} \right\|^2 = \left\| \begin{pmatrix} \frac{1}{\sqrt{k}}\mathbf{Z}_{k-1} \\ \mathbf{N}_{k-1} \\ \vdots \\ \mathbf{N}_1 \end{pmatrix} - \begin{pmatrix} \sqrt{k}\hat{\mathbf{X}}_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right\|^2$$

$$= \left\| \frac{1}{\sqrt{k}}\mathbf{Z}_{k-1} - \sqrt{k}\hat{\mathbf{X}}_0 \right\|^2 + \sum_{i=1}^{k-1} \|\mathbf{N}_j\|^2.$$

Hence, the best symmetric rank-ℓ approximation to the original data set is determined by the best rank-ℓ approximation, $\hat{\mathbf{X}}_0$, to $\frac{1}{k}\mathbf{Z}_{k-1}$ for both the Frobenius norm and 2-norm [7]. □

Intuitively, Theorem 5.1 states that the best symmetric approximation, $\hat{\mathbf{X}}$, to a data set $\mathbf{X}$ is given by first finding the best symmetric approximation, $\hat{\mathbf{X}}_0$, to $\mathbf{X}_0$ by calculating the average of $\mathbf{R}^{k-i}\mathbf{X}_i$, for $i = 1, 2, \ldots, k-1$, where $\mathbf{R}^{k-i}\mathbf{X}_i$ is the transformation of $\mathbf{X}_i$ onto $\mathbf{X}_0$. Notice if $\mathbf{X}$ is a perfectly symmetric set, then $\mathbf{R}^{k-i}\mathbf{X}_i = \mathbf{X}_0$, so the average of $\mathbf{R}^{k-i}\mathbf{X}_i$ for $i = 0, 1, \ldots, k-1$ will equal $\mathbf{X}_0$. Next, to determine the symmetric approximation, $\hat{\mathbf{X}}$, multiply $\hat{\mathbf{X}}_0$ by $\mathbf{R}^i$ to get $\hat{\mathbf{X}}_i$ and concatenate the $\hat{\mathbf{X}}_i$ to form $\hat{\mathbf{X}}$. This step forces $\hat{\mathbf{X}}$ to be a perfectly symmetric set, and Theorem 5.1 proves that this is, in fact, the best symmetric approximation to $\mathbf{X}$ with respect to the Frobenius norm and matrix 2-norm. Note that this result is identical to the conclusions of Zabrodsky et al. in [23]. However, Theorem 5.1 also presents a way to efficiently calculate the best symmetric low rank approximation to the data set by taking an SVD.

It is observed in Theorem 5.1 that the best symmetric ℓ-rank approximation to a data set $\mathbf{X}$ is given by $\mathbf{U}_\ell \mathbf{S}_\ell \mathbf{V}_\ell^T$, where the SVD of the symmetric approximation $\hat{\mathbf{X}} = \mathbf{U}\mathbf{S}\mathbf{V}^T$. Here, $\mathbf{U}_\ell$ and $\mathbf{V}_\ell$ represent the leading ℓ columns of $\mathbf{U}$ and $\mathbf{V}$, and $\mathbf{S}_\ell$ denotes the leading $\ell \times \ell$ principal submatrix of $\mathbf{S}$. One may construct $\mathbf{U}_\ell, \mathbf{S}_\ell$, and $\mathbf{V}_\ell$ in a straightforward manner using the ARPACK software on a serial computer or P_ARPACK on a parallel platform. This is useful for the large data sets that often appear in applications such as molecular dynamics where the matrices are on the order of tens of thousands [16, 21]. It may seem counterintuitive to use ARPACK on such dense systems. However, for large data sets, it is computationally more efficient to calculate only the leading ℓ terms (singular values and vectors) using ARPACK instead of computing all of the singular values and then discarding $n - \ell$ of them. One may either specify ℓ or utilize a restarting scheme to adjust ℓ until $\sigma_\ell \geq \text{tol} * \sigma_1 > \sigma_{\ell+1}$. The important computational point is that only matrix-vector products are needed to calculate

$$\mathbf{u} = \frac{1}{k}(\mathbf{X}_0 + \mathbf{R}_1^T \mathbf{X}_1 + \mathbf{R}_2^T \mathbf{X}_2 + \cdots + \mathbf{R}_{k-1}^T \mathbf{X}_{k-1})\mathbf{v},$$

and this is essentially the same work per iteration one would require to compute the corresponding standard SVD of $\mathbf{X}$ without the symmetry constraint.

6. Experimental Results. A number of experiments were performed on a Mac 2.16 GHz Intel Core 2 Duo machine. These experiments are an extension of the SPSVD approximations shown in our previous paper [16], which only presents data sets with 2-fold rotational or reflectional symmetries of size approximately $10,000 \times 10,000$. However, in [16] there is also a discussion of symmetrizing the motions (modes) of the molecule. In addition, low rank symmetric approximations are made by looking at just the first few modes of the molecule. Here, the concentration is placed on the structure of the molecule. However, if a trajectory for each molecule is presented, then the symmetric motions of the molecule can be calculated as shown in our previous work [16]. In this paper, the generator for each molecule represents more complex symmetries such as 5-fold rotational symmetry to represent single axis computation (Figure 6.1) and dihedral symmetry to represent double axes computation (Figure 6.2). In addition, the best symmetric approximation (SPSVD) to the original data set is calculated. Since molecules are chiral, reflective symmetry is not possible with these structures. Thus, no example is given to illustrate triple axes computations. However, simulations have been performed on contrived data sets with triple

(a) 1SAC   (b) 1SAC (red) with symmetric approxima- tion (blue)   (c) 1SAC (red) with noise added (blue)   (d) 1SAC (red) with full symmetric approx- imation of the noisy molecule (blue)
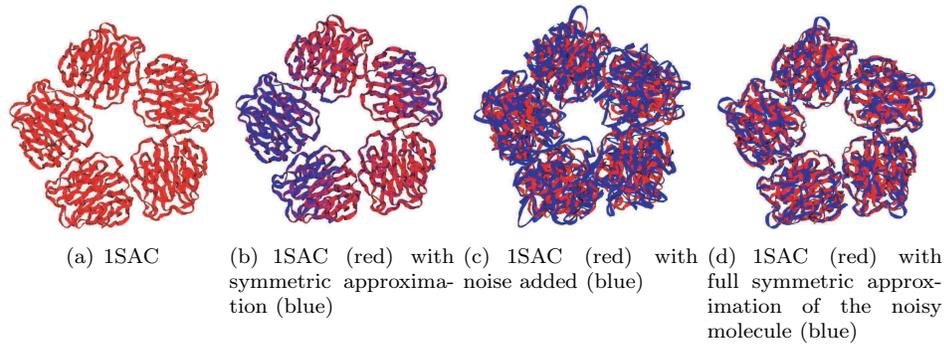
FIG. 6.1. *Different approximations of ISAC.*

axes generators with results that are similar to the single and double axes generators presented here.

The data sets consist of molecules acquired from the Protein Data Bank (PDB) (http://www.rcsb.org/pdb). The first molecule, serum amyloid P-component (1SAC), which exhibits 5-fold rotational symmetry, has been linked to Alzheimer's disease [19], while the second molecule, superoxide dismutase (1IDS), which exhibits 2-fold dihedral symmetry, has been shown to reduce radiofibrosis in breast cancer patients [3].

1SAC is a 5-fold rotationally symmetric molecule that consists of 8245 atoms (Figure 6.1(a)). Thus, the matrix of coordinate points is of size 3×8245. Calculating the generator for this molecule took less than a second to compute. The tolerance for each iterations of the fixed point iteration (4.5) is shown in Figure 6.3(a). Notice that in Figure 6.1(b) the best symmetric approximation (SPSVD) (blue) is superimposed on top of the original (red) data set. For purposes of illustration, artificial noise is introduced into the molecule (Figure 6.1(c)) and the SPSVD is applied to the noisy molecule (Figure 6.1(d)). The SPSVD averages out the noise introduced into the molecule and the resulting SPSVD approximation has a better fit to the original molecule. This conclusion is analytically supported by noting that the relative error in the 2-norm between the noisy and original molecule is approximately 0.142, while the relative error between the symmetric (noisy) molecule and the original molecule is approximately 0.066. In other words, the SPSVD approximation cuts the noise level by more than a half.

(a) 1IDS with genera- tor   (b) 1IDS (red) with SPSVD approxima- tion (blue)   (c) 1IDS (red) with a low rank approxi- mation of the noisy molecule (blue)   (d) 1IDS (red) with a low rank symmetric approximation of the noisy molecule (blue)
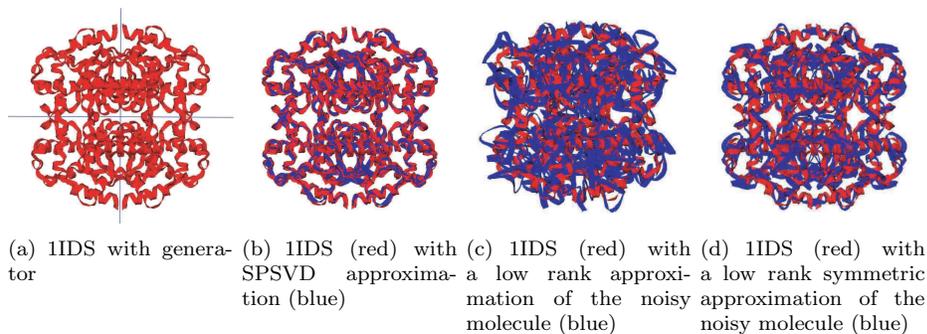
FIG. 6.2. *Different approximations of 1IDS.*

(a) Fixed point iteration for 1SAC

(b) Fixed point iteration for 1IDS
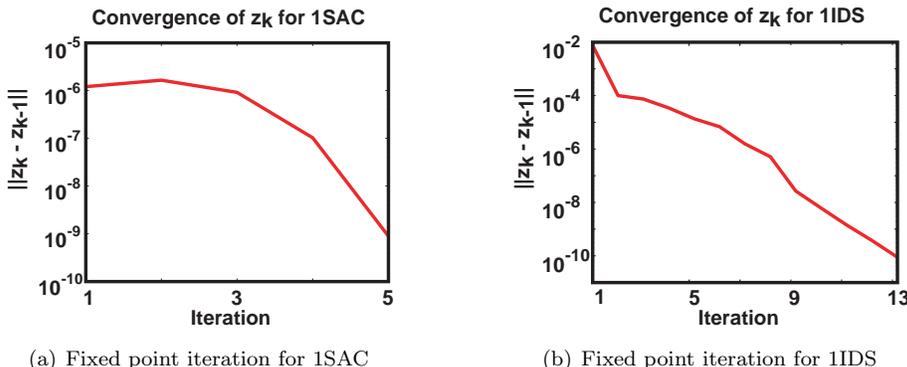
FIG. 6.3. *Fixed point iteration.*

In the case of the 6272 molecule 1IDS, calculation of the generator took approximately 5 seconds to compute. The increased computational time is a result of needing two axes, the major and minor, in order to formulate the generator for 2-fold dihedral symmetry. The tolerance for each step of the fixed point iteration being applied to 1IDS is shown in Figure 6.3(b). The final iteration's major and minor axes is shown on top of 1IDS in Figure 6.2(a), whereas the best symmetric approximation (SPSVD) (blue) is shown on top of the original (red) data set in Figure 6.2(b).

A simulated trajectory of matrix size $18,816 \times 100$ is constructed in order to compare low rank approximations obtained from the SVD and SPSVD with results appearing in Figure 6.2(c) and 6.2(d), respectively. Details of the formulation of this simulated trajectory can be found in [16, 21]. As with the full SPSVD approximation, the rank-6 SPSVD approximation better fits the original data set when compared to the rank-6 SVD approximation. This conclusion is analytically supported by noting that the relative error between a rank-6 approximation of the noisy molecule and the original molecule is approximately 0.195, while the relative error between the a rank-6 symmetric (noisy) molecule and the original molecule is approximately 0.098. In other words, a low rank SPSVD approximation also cuts the noise level by more than a half when compared to a standard low rank SVD approximation to the molecule.

A note should be made with regards to the computational cost and storage requirements of calculating an SPSVD (low rank) approximation compared to an SVD (low rank) approximation. First, the computational cost for the SPSVD is far less than the cost of an SVD approximation since the only the base set of the SVD, which is $1/k$ the size of the original data, has to be calculated. Then the full SPSVD may be formed by computing the orbit of the base set. Second, with regards to storage requirements, only the base set and generator have to be stored for an SPSVD approximation, which equals a storage reduction of approximately $1/k$. The full approximation can later be obtained by calculating the orbit of the base set. In conclusion, the SPSVD approximation not only reduces noise but the storage and computational requirements are also decreased when compared to conventional SVD methods.

7. Conclusion. This paper focuses on the formulation and application of a symmetry preserving singular value decomposition (SPSVD). The SPSVD extends the singular value decomposition by constructing the best low rank approximation to a data set that also preserves the data set's inherent symmetry. Among others, reflective, rotational, inversion, and dihedral symmetry groups are considered here.

In order to calculate an SPSVD, a matrix representation of the symmetry group of interest needs to be obtained. This step is established by an iterative reweighting process that effectively ignores anomalies in the data set. Once the symmetry is known, then the SPSVD may be built using only matrix-vector products and is no more expensive than conventional SVD methods. Additionally, the SPSVD may reduce noise that has been introduced into the data set. In conclusion, the SPSVD is an efficient method for calculating the best symmetric (low rank) approximation to a set of data in both the Frobenius norm and the matrix 2-norm.

REFERENCES

[1] M. ATALLAH, *On symmetry detection*, IEEE Trans. Computers, C–34 (1985), pp. 663–666.
[2] N. AUBRY, W.-Y. LIAN, AND E. S. TITI, *Preserving symmetries in the proper orthogonal decomposition*, SIAM J. Sci. Comput., 14 (1993), pp. 483–505.
[3] F. CAMPANA, S. ZERVOUDIS, B. PERDEREAU, E. G. E, A. FOURQUET, C. BADIU, G. TSAKIRIS, AND S. KOULALOGLOU, *Topical superoxide dismutase reduces post-irradiation breast cancer fibrosis*, Journal of Cellular and Molecular Medicine, 8 (2004), pp. 109–116.
[4] O. COLLIOT, A. V. TUZIKOV, R. M. CESAR, AND I. BLOCH, *Approximate reflectional symmetries of fuzzy objects with an application in model-based object recognition*, Fuzzy Sets and Systems, 147 (2004), pp. 141–163.
[5] D. S. DUMMIT AND R. M. FOOTE, *Abstract Algebra*, John Wiley and Sons, Inc., New York, 1999.
[6] E. EADES, *Symmetry finding algorithms*, Computational Morphology (G. T. Toussaint, Ed.), Amsterdam: North-Holland (1988), pp. 41–51.
[7] G. H. GOLUB AND C. F. V. LOAN, *Matrix Computations (3rd ed.)*, Johns Hopkins University Press, Baltimore, MD, USA, 1996.
[8] L. H. HALL, *Group Theory and Symmetry in Chemistry*, McGraw-Hill, New York, 1969.
[9] M. KAZHDAN, B. CHAZELLE, D. DOBKIN, T. FUNKHOUSER, AND S. RUSINKIEWICS, *A reflective symmetry descriptor for 3D models*, Algorithmica, 38 (2003), pp. 201–225.
[10] M. KIRBY AND L. SIROVICH, *Low-dimensional procedure for the characterization of human faces*, J. Opt. Soc. Am., 4 (1987), pp. 519–524.
[11] ———, *Application of the Karhunen-Loeve procedure for the characterization of human faces*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12 (1990), pp. 103–108.
[12] G. MAROLA, *On the detection of the axes of symmetry of symmetric and almost symmetric planar images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (1989), pp. 104–108.
[13] D. O'MARA AND R. OWENS, *Measuring bilateral symmetry in digital images*, in TENCON Digital Signal Processing Applications, 1996, pp. 151–156.
[14] M. PINSKY, D. CASANOVA, P. ALEMANY, S. ALVAREZ, D. AVNIR, C. DRYZUN, Z. KIZNER, AND A. STERKIN, *Symmetry operation measures*, Journal of Computational Chemistry, (2007).
[15] M. I. SHAH, *A symmetry preserving singular value decomposition*, PhD Thesis, Tech Report TR07-06, Department of Computational and Applied Mathematics, Rice University, 2007.
[16] M. I. SHAH AND D. C. SORENSEN, *A symmetry preserving singular value decomposition*, SIAM Journal on Matrix Analysis and Applications, 28 (2006), pp. 749–769.
[17] V. I. SMIRNOV, *Linear Algebra and Group Theory*, McGraw–Hill Book Company, Inc., New York, 1961.
[18] C. SUN AND J. SHERRAH, *3D symmetry detection using the Extended Gaussian Image*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19 (1997), pp. 164–168.
[19] G. A. TENNENT, L. B. LOVAT, AND M. B. PEPYS, *Serum amyloid p component prevents proteolysis of the amyloid fibrils of alzheimers disease and systemic amyloidosis*, in Proceedings of the National Academy of Sciences of the United States of America, vol. 92(10), 1995, pp. 4299–4303.
[20] H. WEYL, *Symmetry*, Princeton University Press, Princeton, NJ, 1952.
[21] W. WRIGGERS, Z. ZHANG, M. SHAH, AND D. C. SORENSEN, *Simulating nanoscale functional motions of biomolecules*, Molecular Simulation, 32 (2006), pp. 803–815.
[22] H. ZABRODSKY, S. PELEG, AND D. AVNIR, *Continuous symmetry measures*, J. American Chem. Soc, 114 (1992), pp. 7843–7851.
[23] ———, *Symmetry as a continuous feature*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19 (1997), pp. 246–247.