Letting symmetry guide visualization: multidimensional scaling on groups

Henry Kvinge
Department of Mathematics
Colorado State University
Fort Collins, CO USA
henry.kvinge@colostate.edu

Mark Blumstein
Department of Mathematics
Colorado State University
Fort Collins, CO USA
blumstei@math.colostate.edu

***Abstract*—Multidimensional scaling (MDS) is a fundamental tool for both data visualization and dimensionality reduction. Given a finite collection of points and distances between them, MDS finds a map of these points into Euclidean space which optimally preserves distances. Once in Euclidean space, it is often easier to both visualize and apply standard data analytics and machine learning algorithms. Crucially, MDS automatically provides a measure of how well distances are preserved as a function of the dimension of the target space.**

In this paper we show that when MDS is applied to a set of points on which a group G acts and the metric defining distance is invariant with respect to the action of the group, then MDS can be understood (and its output calculated) in terms of the representation theory of G. In particular when our set of points is G itself, this means that the MDS embedding can be calculated using the Fourier transform on groups. We propose this as an alternative implementation of MDS. We investigate an example in which we apply MDS to permutations from the symmetric group and where distances between permutations are calculated by either the Hamming distance, Cayley distance, or Coxeter distance.

I. Introduction

It is often the case that one wishes to understand the geometry or shape of a given data set, e.g. What symmetries does the data space exhibit, and how can those symmetries be exploited? Groups and their representations give a set of tools for answering these questions.

Given a group we often study its symmetries via a "group representation". A group representation is a vector space on which the group acts. Thus, the idea is to understand the symmetries of a group by considering the simplest possible geometrical setting - vector spaces. As an example, consider the symmetric group of permutations acting on the vector space $\mathbb{R}^n$ by permuting the basis vectors.

The irreducible representations, the smallest invariant subspaces on which the group acts, serve as the building blocks for understanding larger more complicated representations. To understand a group's symmetries, one often begins by finding its irreducible representations.

With these ideas in mind, one begins to sense one way in which group representation theory may be applied to better understand complex data: dimensionality reduction. Given a symmetry group of a data set, the irreducible representations give a natural choice of subspaces on which to project the data.

The idea of using the algebraic and representation-theoretic structure of groups in order to answer questions about data is not novel to this paper. An exemplary text is provided by Diaconis [3], there he demonstrates applications of representation theory to problems in probability and statistics. More recently, ideas from group theory have been applied to problems in machine learning. [6].

Our contribution in this paper is an application of the representation theory of finite groups to a classical dimensionality reduction algorithm - multidimensional scaling (MDS). Like many dimensionality reduction algorithms, the goal of MDS is to find a small number of eigen-directions which capture "most" of the information in a data set.

Given a data set with group of symmetries G, the key observation is that, when viewed as a linear operator, MDS is equivariant with respect to the group action. This is significant as it implies that the output eigenspaces of the MDS algorithm are in fact, representations of the group G, which are a priori known. While the examples of this paper apply only to the case where the data set is exactly a finite group, we see this as the crucial first step in a larger program for studying symmetries in data.

This paper is organized as follows. In Section II-A we review some basic concepts from the theory of finite groups and their representation theory. In Section II-B we review multidimensional scaling. In Section III we investigate how the eigenvalue decomposition of the MDS operator can be understood in terms of group representation theory when there is a group action and the distance metric is invariant with respect to this action. Finally in Section IV we study MDS mappings of permutations from symmetric groups.

II. Background

A. Finite groups and their representations

In this section we will review some of the algebraic prerequisites for this paper. We direct the reader to [4] and [12] for further background. For ease of exposition, we usually choose to work over the field of complex numbers. All the theorems in this section also hold, with modification, when $\mathbb{C}$ is replaced by $\mathbb{R}$.

Recall that a group G is a set with an associated multiplication operation $G \times G \to G$, an identity element 1, and the property that every element g has an inverse g^{-1} such that $gg^{-1} = 1$. Many of the common finite groups arising in nature can be understood as symmetries of a mathematical object.

Example II.1. 1) The set of rotations of the n-gon centered at the origin is a realization of the finite group known as the *cyclic group of order* n, C_n. If we also allow for reflections that preserve the points of the n-gon, then the corresponding group is known as the *dihedral group* D_n and has size $2n$.

2) The set of permutations of n distinct objects, say $\{1, 2, \ldots, n\}$ form a group $\mathcal{S}_n$ known as the *symmetric group*. This group has size $n!$. In this paper we will use cycle notation for permutations. Thus we write a single cycle $\sigma \in \mathcal{S}_n$ with $a_i \in \{1, 2, \ldots, n\}$ as

$$\sigma = (a_1, a_2, \ldots, a_k)$$

where σ sends $a_i \mapsto a_{i+1}$ for $1 \leq i < k$ and $a_k \mapsto a_1$. For permutations with more than one cycle we simply write the corresponding cycles next to each other. For example, the permutation $\sigma \in \mathcal{S}_4$ that sends $1 \mapsto 2$, $2 \mapsto 1$, $3 \mapsto 4$, and $4 \mapsto 3$ is written

$$\sigma = (1, 2)(3, 4). \tag{1}$$

For further information on the rich combinatorics and algebra underlying symmetric groups, see [11].

For a group G, a *representation of* G is a vector space V and a group homomorphism $\rho : G \to Gl_n(V)$, where n is the dimension of V. A G-homomorphism $f : V \to W$ between representations V and W of G is a linear map from V to W which respects the action of G so that for $v \in V$ and $g \in G$

$$f(gv) = gf(v).$$

We write $V \cong W$ if there is a bijective G-homomorphism between V and W.

We call a representation V of G *irreducible* if it does not contain any proper subspace that is stable with respect to the action of G. The following theorem shows that it is often enough to understand irreducible representations of G, since all other representations of G are built from these.

Theorem II.2 (Maschke's Theorem)**.** Suppose that G is a finite group and $\mathbb{F}$ a field whose characteristic does not divide the order of $|G|$. Then if V is an $\mathbb{F}$-vector space and a representation of G, then there are some irreducible G representations W_1, W_2, $\ldots$, W_k such that

$$V \cong W_1 \oplus W_2 \oplus \cdots \oplus W_k. \tag{2}$$

Given a representation V of G, the dual of V, $V^* = \mathrm{Hom}_{\mathbb{C}}(V, \mathbb{C})$, is also a representation of G with the action of $\varphi \in V^*$ being defined such that for all $g \in G$, $v \in V$,

$$(g \cdot \varphi)(v) = \varphi(g^{-1}v). \tag{3}$$

It is often more convenient to work with the *group algebra* over $\mathbb{C}$, $\mathbb{C}[G]$. As a vector space, the elements of G form a basis for $\mathbb{C}[G]$. However, $\mathbb{C}[G]$ also inherits a multiplication strcture from G so that for $a_1g_1 + \cdots + a_ng_n$ and $b_1h_1 + \cdots + b_nh_n$ in $\mathbb{C}[G]$

$$(a_1g_1 + \cdots + a_ng_n)(b_1g'_1 + \cdots + b_ng'_n) \tag{4}$$

$$= \sum_{i,j}^{n} a_ib_jg_ig'_j. \tag{5}$$

This construction can be re-interpreted in terms of dual spaces. G acts on itself by multiplication, so that G^* is a representation of G. We can identify $g \in G$ with $\delta_g \in G^*$ which for $h \in G$ sends $\delta_g(h) = 1$ if $h = g$ and $\delta_g(h) = 0$ otherwise. This identifies G^* with G as a vector space. Furthermore, the convolution operation on elements of G^* agrees with the multiplicative structure defined in (4).

We end this section with a theorem that tells us that understanding all the irreducible representations of G is in many cases equivalent to understanding G.

Theorem II.3 (Artin-Wedderburn)**.** Let G be a finite group and let $V_1, V_2, \ldots, V_r$ be a complete set of isomorphism classes of irreducible representations of G. Then

$$\mathbb{C}[G] \cong \bigoplus_{i}^{r} \mathrm{Mat}_{\mathbb{C}}(V_i). \tag{6}$$

That is, $\mathbb{C}[G]$ is actually isomorphic to a direct product of matrix algebras, and these matrix algebras are invariant to multiplication by elements of the group.

Theorem II.3 is closely related to the Fourier transform on G. More specifically the Fourier transform on G and its inverse allow one to move from one side of 6 to the other. Note that the Fourier transform is mostly commonly stated in terms of functions on G (i.e. in terms of G^*). For further details see [3, Chapter 2].

Example II.4. Consider the cyclic group of order n, C_n. Theorem II.3, shows that there are two distinguished bases for elements of $\mathbb{C}[C_n]$ (or alternatively C_n^*). The first is given by elements of C_n. That is, if C_n is generated by g with $g^n = 1$, then one basis of $\mathbb{C}[C_n]$ is

$$\{1, g, g^2, \ldots, g^{n-1}\} \tag{7}$$

(alternatively $\{\delta_1, \delta_g, \ldots, \delta_{g^{n-1}}\}$). There is another basis for $\mathbb{C}[C_n]$ coming from the decomposition of $\mathbb{C}[C_n]$ into irreducible representations via Theorem II.3 (which are all 1-dimensional since C_n is commutative). When written in terms of functions on C_n, this is exactly the familiar discrete Fourier basis.

Example II.5. The isomorphism classes of irreducible representations of the symmetric group $\mathcal{S}_n$ are indexed by partitions of n. Thus $\mathcal{S}_3$ has three different irreducible representations, one corresponding to (3), one corresponding to $(2, 1)$, and one corresponding to $(1, 1, 1)$. Because $\mathcal{S}_n$ is not a commutative group for $n > 2$, the right side of (6) consists of a product of non-trivial matrix algebras. The product structure allows one to handle each of these matrix algebras individually. That is, one can choose to compute with some but not others (a broad generalization of thresholding in signal processing). This approach was used in the context of multi-object tracking in [7].

B. Multidimensional scaling

This section begins with a review of the classical MDS algorithm, our main reference is [1]. We conclude the section by reframing the classical algorithm using notation and language that is better suited for our representation theory applications later in the paper.

The classical MDS algorithm is a dimensionality reduction algorithm for metric spaces. The classical algorithm takes in a Euclidean distance matrix D for points in a high dimensional space, and outputs the coordinates of a set of points (typically in a low dimensional Euclidean space) whose inter-point distances form a Euclidean matrix which best approximates D. The optimality properties of MDS are well-known so we do not discuss them here, see section 14.4 of [1] for details.

To begin, let (X, d) be a finite metric space, so that X is a finite set and $d : X \times X \to \mathbb{R}_{\geq 0}$ is the metric capturing a notion of distance between points in X. The MDS algorithm is as follows:

1) Let $X = \{x_1, x_2, \ldots, x_n\}$. Create a matrix A such that $A_{ij} = -\frac{1}{2}d(x_i, x_j)^2$.
2) Let H be the centering matrix, that is $H = I_n - 11^T$ where I_n is the $n \times n$ identity matrix and 11^T is the $n \times n$ matrix of all 1's.
3) Define $B = HAH$, and compute its eigendecomposition. Because B is symmetric, its eigenvalues will necessarily be real. By the spectral theorem for symmetric matrices, we may decompose B as $B = V\Lambda V^T$, where V is orthogonal consisting of eigenvectors of B, and Λ is a diagonal matrix of eigenvalues, with the standard convention that Λ is put in descending order.
4) If d is not a Euclidean distance metric, it will be necessary to work with a new matrix $\hat{B}$, which is formed by throwing away all eigenvector/eigenvalue pairs corresponding to negative eigenvalues in B. The point is that $\hat{B}$ is positive semi-definite and therefore serves as the inner product matrix for some collection of points in Euclidean space. This collection of points is the desired output of the algorithm.
5) The embedding of X into $\mathbb{R}^k$ for $1 \leq k \leq n$ is obtained as follows: If $\hat{B} = \hat{V}\hat{\Lambda}\hat{V}^T$, the algorithm returns n points in $\mathbb{R}^k$ where the coordinates of the ith point is given by taking the first k-values of row i of the matrix $\hat{V}\hat{\Lambda}^{1/2}$.

The size of the embedding dimension is determined by the values of the eigenvalues. In practice, one often takes just the coordinates determined by the first 2 or 3 eigenvalues, which allows for visualization of the data.

Also, step 4 in the algorithm deserves a comment. Aside from dimensionality reduction, one of the interesting aspects of MDS is that its eigenvalues give a rough measure of how far off from Euclidean the data space X is. It turns out (theorem 14.2.1 [1]) that the matrix B is positive semi-definite if and only if the distance matrix D formed by the points of X is Euclidean.

For this paper it will be helpful to use the following operator notation for MDS (following [5]).

Define the function $K_A : X \times X \to \mathbb{C}$ so that for $(x_1, x_2) \in X \times X$,

$$K_A(x_1, x_2) = -\frac{1}{2}d(x_1, x_2)^2. \tag{8}$$

In other words, if we formed an $n \times n$ matrix out of the function K_A by taking the ijth spot of the matrix to be $K_A(x_i, x_j)$ we would simply get back the A matrix as defined above. In the same way, we define the function K_B corresponding to matrix B.

Equation 14.2.9 of [1] gives the following formula to compute $K_B(x_i, x_j)$ for any i and j:

$$\begin{aligned} K_B(x_1, x_2) = K_A(x_1, x_2) - \frac{1}{n}\sum_{y \in X} K_A(y, x_2) \\ -\frac{1}{n}\sum_{y \in X} K_A(x_1, y) + \frac{1}{n^2}\sum_{(y_1, y_2) \in X \times X} K_A(y_1, y_2). \end{aligned} \tag{9}$$

Definition II.6. Given a finite dimensional metric space (X, d) the *MDS operator* is a $\mathbb{C}$-linear map

$$T : X^* \to X^*$$

defined for $\phi \in X^*$, and $x \in X$ by

$$T(\phi)(x) \doteq \sum_{y \in X} \phi(y) K_B(y, x).$$

Since X is assumed to contain a finite number of points, T is a linear operator on a finite dimensional $\mathbb{C}$-vector space. One readily sees that this is the same definition of the MDS operator as before: the function ϕ determines an $n \times 1$ vector with ith spot given by $\phi(x_i)$, and the formula for $T(\phi)$ is the formula for matrix multiplication of matrix B by vector ϕ.

Note also that we are allowing for coefficients in $\mathbb{C}$. This change is only in order to simplify the representation theory later. This does not change the algorithm however: any complex eigenvector of the real symmetric matrix B gives a pair of real eigenvectors by taking the real and imaginary parts.

III. MDS AND GROUP ACTIONS

In this section we discuss how the natural algebraic structure of a group can be used to both better understand the MDS algorithm and also to produce alternative ways of calculating an MDS embedding. We say that G acts on the left of a set X if there is a map $G \times X \to X$ which is compatible with the algebraic structure of the group i.e. for every $g, h \in G$, and $x \in X$: $g \cdot (h \cdot x) = (gh) \cdot x$, and $e \cdot x = x$, where e is the identity of the group. A similar definition may be made for right-actions. Note that any group acts on itself by left multiplication, right multiplication, and conjugation.

Let (X, d) be a finite metric space (so that X is a finite set). Let G be a finite group that acts on the left of X. The metric d is said to be left G-invariant if $d(gx, gy) = d(x, y)$ for all $x, y \in X$.

Given a left G-action on X, there is an induced left action on X^* defined for $\phi \in X^*$, $g \in G$, and $x \in X$ by,

$$(g\phi)(x) := \phi(g^{-1}x). \tag{10}$$

Also recall that the MDS operator is a $\mathbb{C}$-linear map, $T : X^* \to X^*$.

The following theorem and algorithm are the main results of this paper.

Theorem III.1. Let (X, d) be a finite metric space and assume that d is invariant with respect to the action of a finite group G. Then, the MDS operator T is equivariant with respect to the action of G on X^*. That is, for $g \in G$ and $\phi \in X^*$,

$$T(g\phi) = gT(\phi). \tag{11}$$

Proof. First, we prove that the function $K_B(x, y)$ is G-invariant, and use this invariance to show equivariance of T.

Recall that $K_A(x, y) \doteq -\frac{1}{2}d^2(x, y)$. Given a left-invariant metric d, it is immediate that the function K_A is G-invariant: $K_A(gx, gy) = K_A(x, y)$ for all $g \in G$ and $x, y \in X$. Using equation 9, we see that $K_B(x, y)$ is also G-invariant.

For $g \in G$, $\phi \in X^*$, and $x \in X$ we compute:

$$\begin{aligned}
T(g \cdot \phi)(x) &:= \sum_{y \in X} \phi(g^{-1}y)K_B(y, x) &(12)\\
&= \sum_{y \in X} \phi(y)K_B(gy, x) &(13)\\
&= \sum_{y \in X} \phi(y)K_B(y, g^{-1}x) &(14)\\
&:= T(\phi)(g^{-1}x) &(15)\\
&:= g \cdot T(\phi)(x). &(16)
\end{aligned}$$

Line 13 is a re-ordering of the summation. Line 14 follows from the G-invariance of K_B, specifically, $K_B(gy, x) = K_B(gy, gg^{-1}x) = K_B(y, g^{-1}x)$. Therefore, T is G-equivariant. □

With the same hypotheses of the previous theorem we have:

Corollary III.2. Given a G-invariant metric on X, the eigenspaces of the MDS operator T are representations of the group G.

Proof. The previous theorem implies that T is G-equivariant, i.e. $T(g \cdot \phi) = g \cdot T(\phi)$ for every $\phi \in X^*$. Suppose that λ is an eigenvalue of T, and that ϕ is a corresponding eigenvector. Then,

$$\begin{aligned}
T(g \cdot \phi) &= g \cdot T(\phi)\\
&= g\lambda\phi\\
&= \lambda T(g \cdot \phi).
\end{aligned}$$

Therefore, $g\phi$ is also an eigenvector with eigenvalue λ. If E_λ is the eigenspace of T corresponding to eigenvalue λ, then E_λ has the structure of a vector space. The computation above shows that E_λ is preserved by the action of G, which is of course, the definition of a group representation. □

It follows from the results above that given a group of symmetries of a data set, computation of the MDS algorithm can be advanced if the irreducible representations of G are already known. Fortunately, this is the case for most groups which arise in applications. In this case, we a priori know the eigenvectors of T without having to do any computation.

Finally, given the irreducible representations $\{V_i\}$, we rank them by their corresponding eigenvalues $\{\lambda_i\}$ computed by letting T act on each V_i. This ranking essentially tells us which irreducible representation stores the most important information for embedding X into Euclidean space.

It may be the case that by using the geometry of G and other known parameters of the data space X, certain representations are naturally relevant while others are not. In this case, one can reduce computational expense by computing the eigenvalues of the MDS operator in a lower dimensional space by excluding the irrelevant representations from computation.

Thus, we have the following alternative algorithm for calculating the MDS embedding.

Algorithm 1 Representation-theoretic MDS

1: **Inputs:** A finite metric space (X, d), where a group G acts on X, and the metric is G-invariant. Also, a target dimension k for the embedding dimension.
2: Decompose X^* into irreducible G-representations $X^* \cong V_1 \oplus \cdots \oplus V_r$
3: Calculate the eigenvalue λ_i for T on each V_i. Order $V_1, \ldots, V_r$ so that $\lambda_1, \ldots, \lambda_r$ is decreasing.
4: Given the eigenvectors and eigenvalues, follow the steps of the classical MDS algorithm to compute the embedding coordinates.

We do not expect the above algorithm to be efficient for arbitrary X. When $X = G$ then the second step above becomes the Fourier transform on G. In many common groups fast transforms and software packages with implementations of these transforms do exist. For example, Clausen's fast Fourier transform on $\mathcal{S}_n$, brings the complexity of transform from $n!^2$ (for a naive approach) to $n!(n+1)n(n-1)/3$ [2]. Implementations of this transform can be found in [9]. Fast transforms also exist for all commutative groups, wreath groups for the symmetric group, and supersolvable groups. We expect that this would compare favorably to the complexity of the MDS algorithm on a finite group of size. See [10] and [8] for more information.

We finally note that the trivial representation never contributes any information used for an MDS embedding.

Proposition III.3. The MDS eigenvalue corresponding to the trivial representation of a group is always zero.

IV. Examples

In this section we take our metric space X to be equal to a permutation group G. There are many metrics on permutation groups, the choice of which depends on the application. For this example we consider three popular choices: the Hamming distance, the Cayley distance, and the Coxeter distance.

- *Hamming distance:* For permutations $\sigma, \pi \in \mathcal{S}_n$, the Hamming distance $H : \mathcal{S}_n \times \mathcal{S}_n \to \mathbb{Z}_{\geq 0}$ is defined as
$$H(\sigma, \pi) = \#\{1 \leq k \leq n \mid \sigma(k) \neq \pi(k)\ \}.$$
- *Cayley distance:* For permutations $\sigma, \pi \in \mathcal{S}_n$, the Cayley distance $T : \mathcal{S}_n \times \mathcal{S}_n \to$ between σ and π is defined to be the minimum number of transpositions that we have to multiply σ by to get π.
- *Coxeter distance:* The definition of Coxeter distance is identical to the definition of the Cayley distance except that now we can only multiply by adjacent transpositions (that is, transpositions which send $k \mapsto k+1$ and $k+1 \mapsto k$ for $1 \leq k \leq n$, the keeps every other element of $\{1, 2, \ldots, n\}$ fixed) from a fixed side. We chose to multiply by adjacent transpositions on the left.

We applied the MDS operator to the permutations of $\mathcal{S}_3$ and studied the correspondence of MDS eigenvalues and irreducible representations of $\mathcal{S}_3$. Recall that $\mathbb{C}[G]$ decomposes as
$$\mathbb{C}[\mathcal{S}_3] \cong V^{(3)} \oplus V^{(2,1)} \oplus V^{(2,1)} \oplus V^{(1,1,1)}. \qquad (17)$$
The representation $V^{(3)}$ is the trivial representation and by Proposition III.3 can be ignored. $V^{(2,1)}$ is called the *sign representation* and is 1-dimensional. Finally $V^{(2,1)}$ is frequently referred to as the *standard representation*.

The MDS eigenvalues for these representations are summarized in Table 1. We see here for example that when one takes Hamming distance and Coxeter distance, the most important irreducible representations are the standard representations. In the case of the Coxeter distance, we see in particular that one of the isomorphic copies of the standard representation captures much more information about the embedding into Euclidean space (with an MDS eigenvalue of $\sqrt{6}$) than the other (with an MDS eigenvalue of $\sqrt{2}$). Thus, in analogy to principal component analysis we can say that much of the "energy" of the embedding is captured by the first standard representation. Comparing the projection $\mathcal{S}_3$ into $\mathbb{R}^2$ with respect to the Coxeter distance in Figure 2 with the matrix of all pairwise distances in Figure 4 we can verify that our embedding based on one of the standard representations is indeed reasonable.

On the other hand by consulting the pairwise distance matrix for the Hamming distance on $\mathcal{S}_3$ in Figure 5 we see that the corresponding MDS map of $\mathcal{S}_3$ into $\mathbb{R}^2$ is less good. This agrees with Table 1 where we see that the "energy" measured by the MDS eigenvalue size is more evenly spread among irreducible representations.

Finally, one can ask why groups of points seem clustered in the MDS projection of $\mathcal{S}_3$ into $\mathbb{R}^2$ with respect to Cayley distance in Figure 6. The answer comes from the fact that the irreducible representation that contains the most energy in this case is the sign representation. This is a 1-dimensional representation with all permutations receiving a value of 1 or

Metric	standard rep.	sign rep.
Hamming distance	4.5, 4.5	3
Cayley distance	2, 2	2.5
Coxeter distance	6, 2	1.5

Fig. 1. This table summarizes the eigenvalues of the MDS operator on irreducible representations of $\mathbb{C}[\mathcal{S}_3]$ (see (17)). Note that when we use Coxeter distance, the two isomorphic copies of the representation corresponding to partition $(2,1)$ have different MDS eigenvalues. Thus in this case, when projecting permutations from $\mathcal{S}_3$ to Euclidean space, it is important to pick not only the correct isomorphism class of irreducible $\mathcal{S}_3$-representation but also the correct copy from the decomposition (17) of $\mathcal{S}_3$.

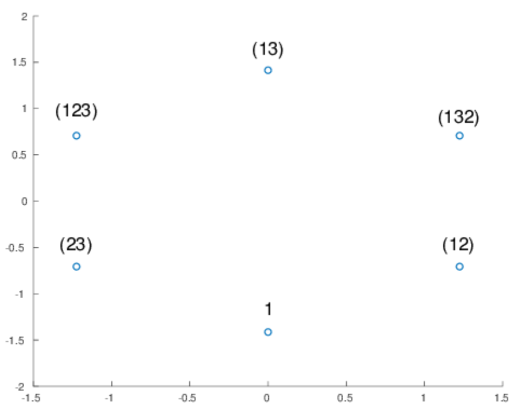

Fig. 2. Projection of the permutations in $\mathcal{S}_3$ into $\mathbb{R}^2$ with respect to the Coxeter distance using MDS. Since one copy of the 2-dimensional irreducible representation $V^{(2,1)}$ in the decomposition (17) of $\mathcal{S}_3$ has by far the largest associated MDS eigenvalue, the best projection into $\mathbb{R}^2$ projects directly onto this representations.

-1. The clustering of points arises from the fact that points are first sent to a fixed distance on both sides of 0 on the x-axis in Figure 6. Their vertical position is determined by one coordinate of a standard representation.

V. Conclusion

In this paper we showed that representation theory can play a vital role in understanding how to visualize and perform dimensionality reduction on data from a space where the notion of distance between points is invariant with respect to the action of a group.

We close by suggesting a number of avenues for further research.

- For the algorithm described in Section III to be useful in practice, it seems likely that one needs an efficient way to calculating the MDS eigenvalues on simple representations of G with respect to a given metric. We hope that in particularly important cases, such as symmetric groups, one will be able to find a combinatorial/algebraic method of discovering what these eigenvalues should be.
- While in this paper we focused on metric spaces which are invariant with respect to the action of a group, the general observations that when the metric is invariant with respect to some action then the MDS operator is

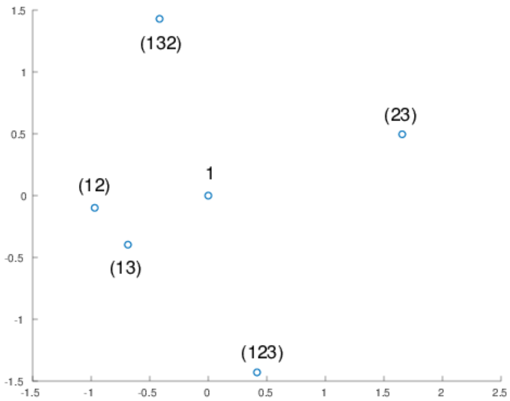

Fig. 3. Projection of the permutations in $\mathcal{S}_3$ into $\mathbb{R}^2$ with respect to the Hamming metric using MDS. Since the 2-dimensional irreducible representations of $\mathcal{S}_3$ corresponding partition $(2,1)$ have the largest associated MDS eigenvalue, the best projection of this data into $\mathbb{R}^2$ projects directly onto these representations.

	1	(12)	(23)	(13)	(123)	(132)
1	0	1	1	3	2	2
(12)	1	0	2	2	3	3
(23)	1	2	0	2	1	3
(13)	3	2	2	0	1	1
(123)	2	3	1	1	0	2
(132)	2	1	3	1	2	0

Fig. 4. The matrix of pairwise distances for Coxeter distance on $\mathcal{S}_3$.

equivariant with respect to that action still holds. Thus if look at metrics which are invariant with respect to the action of an algebra A, then the MDS operator will still be stable on irreducible representations of A.

- The ideas in this paper can be easily extended to actions of a compact Lie group. Riemannian homogeneous spaces such as spheres, orthogonal groups, and Grassmann manifolds play an important role in applications, and we hope to study these examples further. For a discussion of an analogue of MDS to infinite metric spaces see [5].

References

[1] J. Bibby, J. Kent, and K. Mardia. *Multivariate analysis*. Academic Press, London, 1979.

	1	(12)	(23)	(13)	(123)	(132)
1	0	2	2	2	3	3
(12)	2	0	3	3	2	2
(23)	2	3	0	3	2	2
(13)	2	3	3	0	2	2
(123)	3	2	2	2	0	3
(132)	3	2	2	2	3	0

Fig. 5. The matrix of pairwise distances for Hamming distance on $\mathcal{S}_3$.

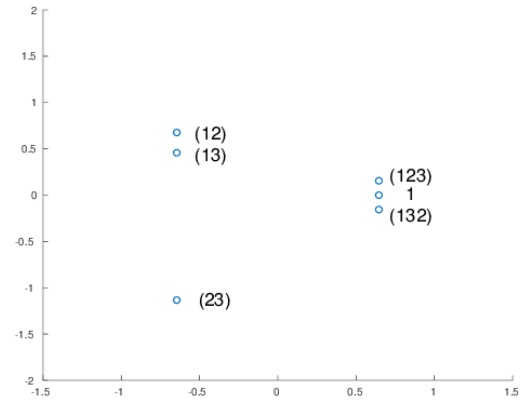

Fig. 6. Projection of the permutations in $\mathcal{S}_3$ into $\mathbb{R}^2$ with respect to Cayley distance using MDS. Since the 2-dimensional irreducible representations of $\mathcal{S}_3$ corresponding partition $(2,1)$ have the largest associated MDS eigenvalue, the best projection of this data into $\mathbb{R}^2$ projects directly onto these representations.

[2] M. Clausen. Fast generalized fourier transforms. *Theoretical Computer Science*, 67(1):55–63, 1989.

[3] P. Diaconis. *Group representations in probability and statistics*, volume 11 of *Institute of Mathematical Statistics Lecture Notes—Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1988.

[4] W. Fulton and J. Harris. *Representation theory*, volume 129 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1991. A first course, Readings in Mathematics.

[5] L. Kassab. Multidimensional scaling: Infinite metric measure spaces, 2018. Master's thesis, Colorado State University. In preparation.

[6] R. Kondor. *Group theoretical methods in machine learning*. PhD thesis, Columbia University, 2008.

[7] R. Kondor, A. Howard, and T. Jebara. Multi-object tracking with representations of the symmetric group. In *Artificial Intelligence and Statistics*, pages 211–218, 2007.

[8] D. K. Maslen and D. N. Rockmore. Generalized FFTs—a survey of some recent results. In *Groups and computation, II (New Brunswick, NJ, 1995)*, volume 28 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 183–237. Amer. Math. Soc., Providence, RI, 1997.

[9] G. Plumb, D. Pachauri, R. Kondor, and V. Singh. S n fft: a julia toolkit for fourier analysis of functions over permutations. *The Journal of Machine Learning Research*, 16(1):3469–3473, 2015.

[10] D. N. Rockmore. Some applications of generalized FFTs. In *Groups and computation, II (New Brunswick, NJ, 1995)*, volume 28 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 329–369. Amer. Math. Soc., Providence, RI, 1997.

[11] B. E. Sagan. *The symmetric group*, volume 203 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2001. Representations, combinatorial algorithms, and symmetric functions.

[12] J.-P. Serre. *Linear representations of finite groups*. Springer-Verlag, New York-Heidelberg, 1977. Translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42.