

Maximal Length Projections in Group Algebras with Applications to Linear Rank Tests of Uniformity

Anna E. Bargagliotti¹, Michael E. Orrison^{2,*}

¹ *Department of Mathematics, Loyola Marymount University, Los Angeles, CA, USA*

² *Department of Mathematics, Harvey Mudd College, Claremont, CA, USA*

Abstract. Let G be a finite group, let $\mathbb{C}G$ be the complex group algebra of G , and let $p \in \mathbb{C}G$. In this paper, we show how to construct submodules S of $\mathbb{C}G$ of a fixed dimension with the property that the orthogonal projection of p onto S has maximal length. We then provide an example of how such submodules for the symmetric group S_n can be used to create new linear rank tests of uniformity in statistics for survey data that arises when respondents are asked to give a complete ranking of n items.

2000 Mathematics Subject Classifications: 20C05, 15A99, 62G10

Key Words and Phrases: Group algebra, Plancherel formula, singular value decomposition, Wedderburn decomposition, linear rank tests of uniformity

*Corresponding author.

Email addresses: abargagl@lmu.edu (A. Bargagliotti), orrison@hmc.edu (M. Orrison)

1. Introduction

Several nonparametric tests exist to assess whether n items that have been fully ranked m times are significantly different (see Chapter 3 of [5]). The data are the set of m rankings, where each ranking is a permutation of the n items. Testing for differences among the n items amounts to testing whether the m rankings were sampled from a uniform distribution over the symmetric group S_n . If a test rejects the null hypothesis of uniformity, then one could say that the items are significantly different.

To explain in more detail, suppose m respondents in a survey have been asked to rank a collection of n items, say from most preferred to least preferred. Each ranking corresponds naturally to a permutation in the symmetric group S_n . Thus, if we let p denote the resulting distribution on S_n , then we might say that some of the items seem to be somehow different from each other if p appears to be generated by something other than the uniform distribution on S_n .

With that in mind, let S be a subspace of $\mathbb{C}S_n$, the complex group algebra of S_n , that is orthogonal to the one-dimensional subspace spanned by the uniform distribution on S_n . Let p^S denote the orthogonal projection of p onto S . If p had been generated by the uniform distribution on S_n , then we would expect $\|p^S\|$ to be small. We may therefore use the value $\|p^S\|$ to create a statistic for a test for uniformity. In particular, we will use the statistic $mn!\|p^S\|^2$ because it has nice properties (see Theorem 3 in [5]). We call the resulting test the *linear rank test of uniformity associated with S* . If the test statistic is large, then the p-value associated with the test would be small and the test would reject the null hypothesis of uniformity.

Of course, we do not want the value $\|p^S\|$ in our statistic to change under any relabelling of the items. That is why we insist that the subspaces S are also submodules of $\mathbb{C}S_n$, and that the action of S_n on $\mathbb{C}S_n$ is unitary. After all, we do not want the outcome of a test of uniformity to depend on how the collection of items have been labelled.

Fortunately, there are well-known choices for such submodules (see [1, 5]). While these well-known choices are useful, we are interested in constructing new submodules S of a fixed dimension that would maximize the values $\|p^S\|$ and therefore minimize the p-values of the associated tests of uniformity. This would allow us to determine if uniformity should be rejected for a fixed dimension. In other words, if we found that the minimum p-value for a fixed dimension was greater than our predetermined significance level, then there would be no way to reject the null. Knowing the submodules that maximize $\|p^S\|$ and minimize the p-values would therefore provide us with extreme cases.

We can generalize the problem as follows. Let G be a finite group, and let $\mathbb{C}G$ be the complex group algebra of G . View $\mathbb{C}G$ as a left module over itself (i.e., as the left regular $\mathbb{C}G$ -module), and assume that $\mathbb{C}G$ comes equipped with the usual inner product $\langle \cdot, \cdot \rangle$, where if $p, q \in \mathbb{C}G$, then

$$\langle p, q \rangle = \sum_{g \in G} p(g) \overline{q(g)}.$$

The associated norm of p is then $\|p\| = \sqrt{\langle p, p \rangle}$.

Let \mathcal{S} be the set of all (left) submodules of $\mathbb{C}G$, and let K denote the set $\{\dim S \mid S \in \mathcal{S}\}$ of all possible dimensions of submodules in \mathcal{S} . Also, for each $S \in \mathcal{S}$, let p^S denote the orthogonal projection of p onto S . In this paper, we give a positive answer to the following question:

Given $p \in \mathbb{C}G$ and $k \in K$, is it possible to construct a k -dimensional submodule S of $\mathbb{C}G$ such that $\|p^S\| \geq \|p^{S'}\|$ for all k -dimensional submodules S' of $\mathbb{C}G$?

In what follows, we show how such submodules may be constructed by first using the Wedderburn decomposition of $\mathbb{C}G$ to convert the element $p \in \mathbb{C}G$ into a direct sum of matrices. We then take advantage of the singular value decomposition of each of those matrices to construct maximal length orthogonal projections of p and their associated submodules. We then conclude with a concrete example when the group is the symmetric group S_3 , and p is a data set encoding the results of a survey in which respondents have been asked to rank three items in order of preference.

2. Wedderburn's Theorem and Submodules

Our construction relies on a few well-known facts and theorems about complex group algebras, most of which can be found in [2]. One such theorem is Wedderburn's decomposition theorem:

Theorem 2.1 (Wedderburn). *The group algebra $\mathbb{C}G$ of a finite group G is isomorphic to a direct sum of matrix algebras:*

$$\mathbb{C}G \cong \mathbb{C}^{d_1 \times d_1} \oplus \dots \oplus \mathbb{C}^{d_h \times d_h}.$$

The number h of summands equals the number of conjugacy classes of G , and the d_j are determined by G up to permutation.

Let $D = \bigoplus_{j=1}^h D_j$ be any mapping that realizes the algebra isomorphism in Wedderburn's decomposition theorem, where D_j is the part of D that maps $\mathbb{C}G$ onto the summand corresponding to the complex algebra $\mathbb{C}^{d_j \times d_j}$ of d_j -by- d_j matrices with complex entries. For convenience, we will assume that each $D_j(g)$ is a unitary matrix for all $g \in G$. Such isomorphisms D exist for all G because every matrix representation of G is equivalent to a unitary one.

If S is a submodule of $\mathbb{C}G$, then $D(S) = D_1(S) \oplus \dots \oplus D_h(S)$, where $D_j(S)$ is a submodule of $\mathbb{C}^{d_j \times d_j}$, and thus $\dim S = \sum_{j=1}^h \dim D_j(S)$. By the Wedderburn decomposition theorem, we may manipulate the submodules of $\mathbb{C}G$ by manipulating the submodules of the $\mathbb{C}^{d_j \times d_j}$, viewed as left modules over themselves. Fortunately, the submodules of the $\mathbb{C}^{d_j \times d_j}$ are straightforward to describe.

First, note that if $B \in \mathbb{C}^{d \times d}$, and B is a rank-1 matrix, then there exist nonzero vectors $\mathbf{u}, \mathbf{v} \in \mathbb{C}^d$ such that $B = \mathbf{u}\mathbf{v}^*$ where \mathbf{v}^* is the conjugate transpose of \mathbf{v} . Furthermore, if $A \in \mathbb{C}^{d \times d}$, then $AB = A(\mathbf{u}\mathbf{v}^*) = (A\mathbf{u})\mathbf{v}^*$. It follows that the submodule of $\mathbb{C}^{d \times d}$ that is generated by B is the d -dimensional subspace $\{\mathbf{w}\mathbf{v}^* \mid \mathbf{w} \in \mathbb{C}^d\}$.

In general, if W is a submodule of $\mathbb{C}^{d \times d}$, then $\dim W = cd$ for some c such that $0 \leq c \leq d$. Furthermore, if c is nonzero, then there exist nonzero and linearly independent vectors $\mathbf{v}_1, \dots, \mathbf{v}_c \in \mathbb{C}^d$ such that W is the collection of all matrices of the form

$$\mathbf{w}_1 \mathbf{v}_1^* + \dots + \mathbf{w}_c \mathbf{v}_c^*$$

where $\mathbf{w}_1, \dots, \mathbf{w}_c \in \mathbb{C}^d$. In this case, we will slightly abuse the standard terminology and say that $\mathbf{v}_1, \dots, \mathbf{v}_c$ generate the submodule W .

3. Frobenius Inner Products and Plancherel's Formula

We now turn our attention to relating the norm of a vector $p \in \mathbb{C}G$ to the matrices $D_1(p), \dots, D_h(p)$. Let $A, B \in \mathbb{C}^{d \times d}$. The *Frobenius inner product* of A and B is given by

$$\langle A, B \rangle_F = \text{Tr}(AB^*) = \sum_{i,j} a_{ij} \overline{b_{ij}}$$

where a_{ij} and b_{ij} are the ij th entries of A and B , respectively. The *Frobenius norm* $\|A\|_F$ of $A \in \mathbb{C}^{d \times d}$ is then defined to be $\sqrt{\langle A, A \rangle_F}$.

Let $p, q \in \mathbb{C}G$. Because the $D_j(g)$ are unitary matrices for all $g \in G$, the Plancherel formula (see, for example, Theorem 6.8 in [2]) gives us that

$$\langle p, q \rangle = \frac{1}{|G|} \sum_{j=1}^h d_j \langle D_j(p), D_j(q) \rangle_F.$$

Thus, for all $p \in \mathbb{C}G$,

$$\|p\|^2 = \frac{1}{|G|} \sum_{j=1}^h d_j \|D_j(p)\|_F^2.$$

To work more easily with each of the $\|D_j(p)\|_F$, we next focus on the singular value decompositions of the $D_j(p)$. Doing so will lead to the submodules of $\mathbb{C}G$ for which we are searching.

4. Singular Value Decompositions and Our Construction

Let $A \in \mathbb{C}^{d \times d}$. A singular value decomposition of A is a factorization $A = U\Sigma V^*$ where U and V are d -by- d unitary matrices, and Σ is a diagonal matrix whose entries are nonnegative real numbers, and where, by convention, if the i th entry on the diagonal of Σ is denoted by σ_i , then $\sigma_1 \geq \dots \geq \sigma_d$. A singular value decomposition of A always exists, and the resulting *singular values* $\sigma_1, \dots, \sigma_d$ of A are unique.

Let $U\Sigma V^*$ be a singular value decomposition of A , and denote the i th columns of U and V by \mathbf{u}_i and \mathbf{v}_i , respectively. We may then use the outer product form of the singular value decomposition to express A as

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^* + \dots + \sigma_n \mathbf{u}_d \mathbf{v}_d^*.$$

Doing so is useful in practice because, if c is such that $1 \leq c \leq d$, then a rank- c approximation A_c of A that minimizes $\|A - A_c\|_F$ is given by

$$A_c = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^* + \cdots + \sigma_c \mathbf{u}_c \mathbf{v}_c^*.$$

See [6] for a discussion of this and other facts about the singular value decomposition. It follows that if we want a cd -dimensional submodule W of $\mathbb{C}^{d \times d}$ such that the orthogonal projection (using the Frobenius inner product) of A onto W has maximal length (with respect to the Frobenius norm), then a possible choice for W is the submodule of $\mathbb{C}^{d \times d}$ generated by $\mathbf{v}_1, \dots, \mathbf{v}_c$. In this case, the orthogonal projection of A onto W is precisely A_c . Furthermore, because the \mathbf{v}_i form an orthonormal basis of \mathbb{C}^d , we have that

$$\|A_c\|_F^2 = \sigma_1^2 + \cdots + \sigma_c^2.$$

Let $p \in \mathbb{C}G$, and let $k \in K$. We are now ready to construct a k -dimensional submodule S of $\mathbb{C}G$ such that $\|p^S\| \geq \|p^{S'}\|$ for all k -dimensional submodules S' of $\mathbb{C}G$. First, for each j , let

$$\sigma_1^{(j)} \mathbf{u}_1^{(j)} (\mathbf{v}_1^{(j)})^* + \cdots + \sigma_{d_j}^{(j)} \mathbf{u}_{d_j}^{(j)} (\mathbf{v}_{d_j}^{(j)})^*$$

be a singular value decomposition for $D_j(p)$.

For each vector $\mathbf{c} = (c_1, \dots, c_h) \in \mathbb{N}^h$ of nonnegative integers such that $0 \leq c_j \leq d_j$, let $\sigma_{\mathbf{c}}^2 = |G|^{-1} \sum_{j=1}^h d_j \left(\sum_{i=1}^{c_j} (\sigma_i^{(j)})^2 \right)$ and let $d_{\mathbf{c}} = \sum_{j=1}^h c_j d_j$. Let V_{c_j} be the invariant subspace of $\mathbb{C}^{d_j \times d_j}$ that is generated by the vectors $\mathbf{v}_1^{(j)}, \dots, \mathbf{v}_{c_j}^{(j)}$, and let $V_{\mathbf{c}} = V_{c_1} \oplus \cdots \oplus V_{c_h}$. If we define $\widetilde{V}_{\mathbf{c}} = D^{-1}(V_{\mathbf{c}})$, then note that $\dim \widetilde{V}_{\mathbf{c}} = \dim V_{\mathbf{c}} = d_{\mathbf{c}}$ and $\|p^{\widetilde{V}_{\mathbf{c}}}\|^2 = \sigma_{\mathbf{c}}^2$.

Let \mathbf{C}_k denote the set of vectors $\mathbf{c} = (c_1, \dots, c_h) \in \mathbb{N}^h$ of nonnegative integers such that $0 \leq c_j \leq d_j$ and $d_{\mathbf{c}} = k$. By the above discussion, we have the following theorem:

Theorem 4.1. *Let $p \in \mathbb{C}G$, and let $k \in K$. If $\mathbf{c} \in \mathbf{C}_k$ has the property that $\sigma_{\mathbf{c}}^2 \geq \sigma_{\mathbf{c}'}^2$ for all $\mathbf{c}' \in \mathbf{C}_k$ and $S = \widetilde{V}_{\mathbf{c}}$, then $\|p^S\| \geq \|p^{S'}\|$ for all k -dimensional submodules S' of $\mathbb{C}G$.*

Thus, given $p \in \mathbb{C}G$ and $k \in K$, we are able to use singular value decompositions of the images of p under each of the D_j to construct a k -dimensional submodule S such that $\|p^S\|$ is maximal.

Note that in the process, we may also construct analogous spaces for the right regular module of $\mathbb{C}G$ by using the associated $\mathbf{u}_i^{(j)}$ to generate right submodules instead. Furthermore, because we are using only the singular values of the $D_j(p)$ to determine the lengths of the resulting projections, we immediately have the following corollary, which is perhaps somewhat surprising:

Corollary 4.2. *Let $p \in \mathbb{C}G$, and let $k \in K$. If S is a k -dimensional left submodule of $\mathbb{C}G$ such that $\|p^S\|$ is maximal, and R is a k -dimensional right submodule of $\mathbb{C}G$ such that $\|p^R\|$ maximal, then $\|p^S\| = \|p^R\|$.*

Finally, as noted in the introduction, we are sometimes interested in maximizing the value $\|p^S\|$ when S is a submodule of $\mathbb{C}G$ that is orthogonal to the one-dimensional submodule spanned by the uniform distribution on G . With that in mind and using the

notation above, note that if D_1 denotes the trivial representation of G , then $S = \widetilde{V}_{\mathbf{c}}$ will be orthogonal to the one-dimensional submodule spanned by the uniform distribution if and only if $\mathbf{c} = (c_1, \dots, c_h)$ has the property that $c_1 = 0$. Thus, if we let \mathbf{C}_k^0 denote the set of such vectors in \mathbf{C}_k , and we let \mathcal{O} denote the set of submodules in $\mathbb{C}G$ that are orthogonal to the submodule spanned by the uniform distribution on G , then we arrive at the following theorem:

Theorem 4.3. *Let $p \in \mathbb{C}G$, and let $k \in K$. If $\mathbf{c} \in \mathbf{C}_k^0$ has the property that $\sigma_{\mathbf{c}}^2 \geq \sigma_{\mathbf{c}'}^2$ for all $\mathbf{c}' \in \mathbf{C}_k^0$ and $S = \widetilde{V}_{\mathbf{c}}$, then $\|p^S\| \geq \|p^{S'}\|$ for all k -dimensional submodules S' in \mathcal{O} .*

5. Example

In this section, we present an example of the construction described above when applied to a function defined on the symmetric group S_3 . We then use our construction to create a new linear rank test of uniformity for S_3 .

In Chapter 8 of [3], Diaconis provides the results of a survey in which respondents were asked to rank where they want to live: in the city, in the suburbs, or in the country. We will use the same setting for our example, but we will adjust the data slightly to better highlight the usefulness of our construction. In particular, suppose we have 68 respondents and that their rankings are

π	city	suburbs	country	$p(\pi)$
id	1	2	3	12
(12)	2	1	3	18
(23)	1	3	2	8
(13)	3	2	1	10
(123)	2	3	1	7
(132)	3	1	2	13

where $p(\pi)$ is the number of people who have chosen π . For example, 18 people chose the transposition (12), and 13 people chose the three-cycle (132).

We will use the Wedderburn decomposition of $\mathbb{C}S_3$ to convert p into a direct sum of matrices. First, we will construct an algebra isomorphism from $\mathbb{C}S_3$ to $\mathbb{C}^{1 \times 1} \oplus \mathbb{C}^{2 \times 2} \oplus \mathbb{C}^{1 \times 1}$.

Let D_1 be the one-dimensional trivial representation of S_3 . Thus $D_1(\pi)$ is the one-by-one identity matrix for all $\pi \in S_3$.

Next, let D_2 be the two-dimensional unitary representation of S_3 defined by setting

$$D_2((12)) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad D_2((23)) = \frac{1}{2} \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{bmatrix}.$$

This suffices to define $D_2(\pi)$ for all $\pi \in S_3$ because S_3 is generated by the transpositions (12) and (23). Note that D_2 is also used in Chapter 8 of [3].

Finally, let D_3 be the one-dimensional sign representation of S_3 where

$$D_3(\text{id}) = D_3((123)) = D_3((132)) = [1]$$

and

$$D_3((12)) = D_3((23)) = D_3((13)) = [-1].$$

The representations D_1 , D_2 , and D_3 form a complete set of irreducible unitary representations of S_3 , and thus $D_1 \oplus D_2 \oplus D_3$ is an algebra isomorphism from $\mathbb{C}S_3$ to $\mathbb{C}^{1 \times 1} \oplus \mathbb{C}^{2 \times 2} \oplus \mathbb{C}^{1 \times 1}$ (see, for example, Theorem 2.15 in [2]). Applying this isomorphism to the function p yields

$$[68] \oplus \begin{bmatrix} -7.0000 & 3.4641 \\ -6.9282 & 11.0000 \end{bmatrix} \oplus [-4].$$

Given this output, we may now easily construct the k -dimensional submodules of $\mathbb{C}S_3$ described in Theorem 2 for all k such that $1 \leq k \leq 6$.

For example, suppose we want a 3-dimensional submodule S of $\mathbb{C}S_3$ such that $\|p^S\| \geq \|p^{S'}\|$ for all 3-dimensional submodules S' of $\mathbb{C}S_3$. In this case, because $\mathbb{C}S_3$ contains only two 1-dimensional irreducible submodules, we know that S must be the direct sum of an irreducible 1-dimensional submodule and an irreducible 2-dimensional submodule.

For the 1-dimensional submodule, we will use the trivial submodule of $\mathbb{C}S_3$. This is because the projection of p onto the trivial submodule is larger than the projection of p onto the sign submodule. This is easy to see in this case because $|68| > |-4|$. Note also that we do not need to compute the singular value decomposition of one-by-one matrices because doing so would just return the matrices themselves.

On the other hand, for the 2-dimensional submodule, we will use the singular value decomposition of the matrix

$$\begin{bmatrix} -7.0000 & 3.4641 \\ -6.9282 & 11.0000 \end{bmatrix}$$

which is

$$\begin{bmatrix} -4.5876 & 5.4827 \\ -8.2669 & 9.8798 \end{bmatrix} + \begin{bmatrix} -2.4124 & -2.0186 \\ 1.3387 & 1.1202 \end{bmatrix}.$$

The matrix on the left corresponds to a 2-dimensional submodule with the property that the projection of p onto it has maximal length when compared with the projections of p onto all irreducible 2-dimensional submodules of $\mathbb{C}S_3$. By Theorem 2, it follows that the inverse image under $D_1 \oplus D_2 \oplus D_3$ of

$$[68] \oplus \begin{bmatrix} -4.5876 & 5.4827 \\ -8.2669 & 9.8798 \end{bmatrix} \oplus [0]$$

will generate the 3-dimensional submodule S for which we were looking, and that this inverse image will in fact be the projection p^S .

Finally, as noted in the introduction, we were drawn to these kinds of constructions because of linear rank tests of uniformity [1]. For example, we could ask what would happen if we were to use the 2-dimensional submodule created above, which is orthogonal to the trivial submodule, in a linear rank test of uniformity applied to the function p .

	means	marginals	pairs	probabilities	ours
statistic	1.0882	6.7647	1.3235	7.0000	6.3841
p-value	0.5804	0.1489	0.7236	0.2206	0.0411

Table 1: The outcome of several linear rank tests of uniformity when applied to our data set p .

Furthermore, we could ask how the outcome of this test compares to those of other well-known linear rank tests of uniformity. The results of such a comparison are presented in Table I.

In Table I, we see that standard linear rank tests of uniformity such as the means test, marginals test, pairs test, and probabilities test (see Chapter 3 of [5]) fail to reject the null hypothesis of uniformity because they lead to large p-values. On the other hand, the test associated with the 2-dimensional submodule constructed above leads to a rejection of the null hypothesis of uniformity with a p-value of .0411, which is the smallest p-value for any such test associated with a 2-dimensional submodule by Theorem 4. This is because our new test corresponds to a submodule that contains a large projection of the data, while the other tests correspond to submodules that contain only small projections of the data.

6. Future Directions

This paper arose because we wanted to better understand linear rank tests of uniformity. In that setting, distributions defined on symmetric groups are the appropriate objects of study. It is common, however, to encounter *partially ranked data* where respondents do not fully rank a set of items. For example, they might be asked to list only their top k choices from a list of n items where $k < n$. In this case, the associated distributions can be viewed as distributions on cosets of a subgroup of the symmetric group, and it would be interesting to see what the analogues of Theorem 2 and Theorem 4 would be in this case.

Furthermore, it would be interesting to have analogues of Theorem 2 and Theorem 4 for data defined on homogeneous spaces of finite groups in general (i.e., for data defined on cosets of a subgroup of a finite group). Section 2.3 of [4] seems like a good place for interested readers to start working with such data, especially when it comes to finding matrices whose singular value decompositions are as useful as the ones we have described in this paper. See also Chapter 5 of [3] for more information about data defined on homogeneous spaces of finite groups.

ACKNOWLEDGEMENTS Special thanks to Francis Su for helpful conversations and suggestions.

References

- [1] Anna E. Bargagliotti and Michael E. Orrison. Linear rank tests of uniformity: understanding inconsistent outcomes and the construction of new tests. *J. Nonparametr. Stat.*, 24(2):481–495, 2012.
- [2] Michael Clausen and Ulrich Baum. *Fast Fourier transforms*. Bibliographisches Institut, Mannheim, 1993.
- [3] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.
- [4] A. Fässler and E. Stiefel. *Group theoretical methods and their applications*. Birkhäuser Boston, Inc., Boston, MA, 1992. Translated from the German by Baoswan Dzung Wong.
- [5] John I. Marden. *Analyzing and modeling rank data*, volume 64 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1995.
- [6] G. W. Stewart. On the early history of the singular value decomposition. *SIAM Rev.*, 35(4):551–566, 1993.