

Research Article

Interpreting the Phase Spectrum in Fourier Analysis of Partial Ranking Data

Ramakrishna Kakarala

School of Computer Engineering, Nanyang Technological University, Singapore 637665

Correspondence should be addressed to Ramakrishna Kakarala, ramakrishna@ntu.edu.sg

Received 28 September 2011; Revised 14 February 2012; Accepted 23 February 2012

Academic Editor: Mustapha Ait Rami

Copyright © 2012 Ramakrishna Kakarala. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Whenever ranking data are collected, such as in elections, surveys, and database searches, it is frequently the case that partial rankings are available instead of, or sometimes in addition to, full rankings. Statistical methods for partial rankings have been discussed in the literature. However, there has been relatively little published on their Fourier analysis, perhaps because the abstract nature of the transforms involved impede insight. This paper provides as its novel contributions an analysis of the Fourier transform for partial rankings, with particular attention to the first three ranks, while emphasizing on basic signal processing properties of transform magnitude and phase. It shows that the transform and its magnitude satisfy a projection invariance and analyzes the reconstruction of data from either magnitude or phase alone. The analysis is motivated by appealing to corresponding properties of the familiar DFT and by application to two real-world data sets.

1. Introduction

Ranking data, which arise in scenarios such as elections or database searches, describe how many times a given ordering of objects is chosen. It is frequently the case that when, ranking data are collected, partial ranking data are obtained in addition to, or perhaps instead of, full rankings. A partial or incomplete ranking only specifies the ordering of the top k out of n possibilities and usually indicates that the ranker is either unable to, or indifferent to, the ordering of the remaining $n - k$ items. Full ranking data are obviously a special case of partial ranking data. A classic approach is to treat full ranking data for n items as a function on the symmetric group S_n ; for each permutation $p \in S_n$, the value of $x(p)$ is the number of times the ordering represented by that permutation is chosen [1]. For example, if 3 items are ranked, then $x([2, 1, 3])$ is the number of times the survey respondents chose to rank item 2 first, item 1 second, followed by 3. As discussed in more detail below, partial ranking data

also form functions on S_n that are piecewise constant over cosets of the subgroup fixing the first k items.

The analysis of ranking data, including both full and partial rankings, is well established. Statistical methods exist both for data in the “time domain” (using signal processing terminology), which in this case is the permutation group S_n , and in the “frequency domain” that is obtained through Fourier analysis on the group. Recent papers by Lebanon and Mao [2] and Hall and Miller [3] explore, respectively, the nonparametric modeling and bootstrap analysis of partial ranking data in the time domain. Time domain analysis does not allow such interesting possibilities as using band-limited or “smooth” approximations to the data, on analyzing the strength of various components. Diaconis [1, 4] and Diaconis and Sturmfels [5] use the Fourier transform on S_n to analyze frequency components of both full and partial ranking data. Those papers, while addressing the fundamentals of Fourier analysis in terms of invariant subspaces, do not consider signal processing aspects as considered here. Other papers using the Fourier transform on S_n include Huang et al. [6] for inference on permutations of identities in tracking, and Kondor and Borgwardt [7] to provide labeling-invariant matching of graphs. Kakarala [8] shows that the Fourier transform on S_n may be interpreted in terms of signal processing concepts such as magnitude and phase, but the work is limited to full rankings. In this paper, we take a similar approach to analyze the properties of the Fourier transform on S_n for partial rankings, with particular emphasis on the role of phase in forming the top three ranks, $k \leq 3$.

Underlying our approach is the intuition that, in any frequency-domain approach, whether on the symmetric group S_n or on the more familiar discrete domain $\mathbb{Z}_N = \{0, 1, \dots, N - 1\}$, the Fourier transform values may be separated into magnitudes, which indicate component strengths, and phases, which indicate relative component locations. Such a separation is basic to a signal processing approach, and is well understood in the ordinary discrete Fourier transform (DFT) on \mathbb{Z}_N , and also in two dimensions in the case of images. A familiar demonstration of the importance of phase is to combine the magnitude spectrum of image X with the phase spectrum of image Y and observe that, after inverse transform, the result appears very similar to Y [9]; in other words, phase is more important to our perception of image structure. Therefore, it seems appropriate to ask the following question: what is the role of phase in forming partial ranking data?

The problem of analyzing phase on S_n is not as straightforward as with the DFT on \mathbb{Z}_N , because the Fourier transform on S_n has matrix-valued coefficients, not scalars as with the DFT, making even such elementary concepts as “frequency” nonobvious. Though various papers describe the S_n transform in detail [6] and, code for computing a fast Fourier transform (FFT) on S_n has been published by Kondor [10], the level of abstraction required to understand the S_n transform is high. Therefore, this paper makes a concerted effort to reason from the familiar DFT to explore the relevant concepts on S_n . It shows that the coefficients of the Fourier transform for top k choice partial ranking data are invariant under projections that are determined by the subgroup S_{n-k} . The projection approach provides a relatively simple explanation of the roles of magnitude and phase for partial ranking. The explanation is tested on two real-world data sets.

It should be noted that the concept of partially measured ranking data has interpretations other than the one explored in this paper, which is top k out of n choices data. For example, an “incomplete” ranking specifies a preference among a subset of the choices, not which is most preferred. Among choices A , B , and C , an incomplete ranking might simply say that A is preferred to C , but nothing about A versus B , or B versus C ; mathematically, this may be modelled as a partial order on the choices [2]. Diaconis [4] describes other kinds of

incomplete rankings: “committee selection,” where one chooses the top k out of n choices but does not rank among the choices; “most and least desirable,” where one chooses the most important and least important attributes among n choices but does not specify the order of the middle elements. What is common mathematically to the previous types of data is that they are constant on cosets of a suitably chosen subgroup H of S_n . The mathematical results of this paper concerning magnitude and phase apply to every coset space S_n/H . However, the results provided below on approximation by linear phase or unit-magnitude functions are limited to top k -choice data, whose domain in S_n/S_{n-k} . Though mathematically a special case of partially measured rankings, top k -choice data appears in sufficiently many scenarios to be worth analysis on its own.

2. Background Material

Fourier analysis on the symmetric group S_n is normally described in abstract terms involving group representation theory, which makes the subject difficult to understand for non-specialists. As mentioned in the Introduction, we use analogy to the better known DFT on \mathbb{Z}_N . The DFT is defined for data x by the familiar pair of equations for transform and inverse:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N}, \quad x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j2\pi kn/N}. \quad (2.1)$$

Each complex-valued DFT coefficient is expressed in terms of magnitude and phase by writing $X[k] = |X[k]|e^{j\phi(k)}$, where the absolute value determines the magnitude, and the angle $\phi(k)$ measures the starting value at $n = 0$ in the period of the constituent sinusoid $e^{j2\pi kn/N}$. The translation property of the DFT shows that the transform of the circularly shifted function $y[n] = x[n+t]$ has coefficients $Y[k] = X[k]e^{j2\pi kt/N}$, which shows that the magnitude does not change but the phase changes linearly, that is, $\phi(k) \mapsto \phi(k) + 2\pi kt/N$. Hence, phase is closely connected with location.

Suppose now that the data x has the additional symmetry of having a subperiod, that is, $x[n+M] = x[n]$ where M divides N . Then, it is well known that the DFT coefficients $X[k]$ are zero unless k is a multiple of N/M . For example, if $N = 128$ and $M = 4$, then, of the 128 possible DFT coefficients, only four are nonzero: $X[0]$, $X[32]$, $X[64]$, and $X[96]$. It is helpful to see the previous example in a different way to better understand the discussion of the symmetric group below. Suppose that we define \tilde{x} as the data within one period, that is, $\tilde{x}[n] = x[n]$ for $n = 0, \dots, 3$, and $\tilde{x}[n] = 0$ otherwise. Let τ denote the periodic pulse train of Kronecker δ functions defined as follows:

$$\tau[n] = \sum_{m=0}^{31} \delta[n - 4m]. \quad (2.2)$$

Then, $x = \tilde{x} * \tau$, where $*$ denotes circular convolution over 128 points. We have, therefore, by the convolution property of the DFT that $X[k] = \tilde{X}[k]T[k]$, where both \tilde{X} and T are the respective DFTs on 128 points of \tilde{x} and τ . It is easy to see that $T[k] = 32$ for $k = 0, 32, 64, 96$ but $T[k] = 0$ otherwise. We might consider the function $P[k] = T[k]/32$ a *projection* of the DFT coefficients; the term projection is appropriate because P takes values of either 0 or 1, and therefore $P[k]P[k] = P[k]$ for all k . With the projection so defined, we have that $X[k] = P[k]X[k]$, which shows that the data are invariant to the projection and therefore lie

in its image. The projection approach helps considerably below in formulating the transform for partial rankings on the symmetric group.

The symmetric group S_n is the collection of all $n!$ possible permutations of the set $\{1, 2, \dots, n\}$. If p and q represent two permutations in S_n , then the product pq denotes q applied first followed by p . For example, if $n = 4$ and $p = [2, 3, 4, 1]$, which indicates that $p(1) = 2, p(2) = 3, p(3) = 4, p(4) = 1$, and similarly $q = [3, 4, 1, 2]$, then $pq = [4, 1, 3, 2]$. With that product, S_n forms a group, with identity-denoted e and inverse p^{-1} being the unique permutation that exactly undoes the action of p , that is, $p^{-1}p = pp^{-1} = [1, 2, \dots, n]$. For example, the inverse of $[2, 3, 4, 1]$ is $[4, 1, 3, 2]$.

Data consisting of full rankings form functions on S_n in the manner described in the Introduction. The same domain also serves for partial ranking data. If we have data where only the first k of the n items is ranked, then, for each $p \in S_n$, let us define the value of $x(p)$ to be the number of times the first k elements of p is chosen. The definition leads to piecewise constant functions on S_n . An example illustrates the approach. Suppose $n = 3$ items are to be ranked in an election given to 600 voters, but the respondents give only their top choices as follows: item 1 gets 100 votes, item 2 gets 200 votes, and item 3 gets 300 votes. Then, we construct x on S_3 by extending the votes to all permutations p based on first item, so that $x([1, 2, 3]) = x([1, 3, 2]) = 100$, and similarly for the other 4 choices of p . If we were to view the previous construction in group-theoretic terms, the function x is such that it is constant on left cosets of the subgroup S_2 fixing the first element, that is, $x(p) = x(ps)$ for all $s \in S_2$ where $s(1) = 1$ by definition, and $p(1) = t$ for the item t being chosen. Though the constant vote given to each coset is mathematically convenient, it does not capture certain effects that may be interesting; for example, if I choose oranges as my favorite fruit, I may be more likely to choose apples than durians as my next favorite, even if I am not required to state my next favorite. Nevertheless, due to its convenience, we use the constant on cosets approach in the remainder of this paper.

A detailed example helps to illustrate the model. In the famous American Psychological Association (APA) election data [1], which is available online (<http://www.stat.ucla.edu/data/hand-daly-lunn-mconway-ostrowski/ELECTION.DAT>), 5,738 voters provided full rankings of each of 5 candidates for president. The full rankings form a function on S_5 and are shown plotted in Figure 1(a) against the $5! = 120$ elements of the group arranged in lexicographic order. In the same election, many voters chose not to submit full rankings but provided instead partial rankings. Specifically, 5,141 voters submitted only their top choice, 2,462 voters submitted only their first and second choices in order, and 2,108 voters submitted only their top three choices in order. Consequently, there were a total of 9,711 voters giving only partial rankings, more than the 5,738 that gave full rankings. After forming piecewise constant functions as described above, the partial ranking data are displayed in Figures 1(b)–1(d).

An advantage of placing both full and partial rankings on the same domain is that we may apply the same Fourier transform in both cases. The Fourier transform on S_n , which is formally obtained from the theory of group representations, has important differences to the DFT. We review some basic facts from the literature [4]. First, the Fourier coefficients on S_n are matrix valued, unlike the scalar values of the DFT. Second, they are indexed by arithmetic partitions of n with nonincreasing elements, which are roughly analogous to the frequency index k of the DFT. For example, for $n = 5$, the seven such partitions are $(5), (4, 1), (3, 2), (3, 1^2) := (3, 1, 1), (2^2, 1), (2, 1^3),$ and (1^5) . For every partition λ of n , the Fourier basis elements belonging to it are collected into a square-matrix-denoted D_λ whose dimensions n_λ are calculated using standard formulas [4]. For S_5 , the seven partitions

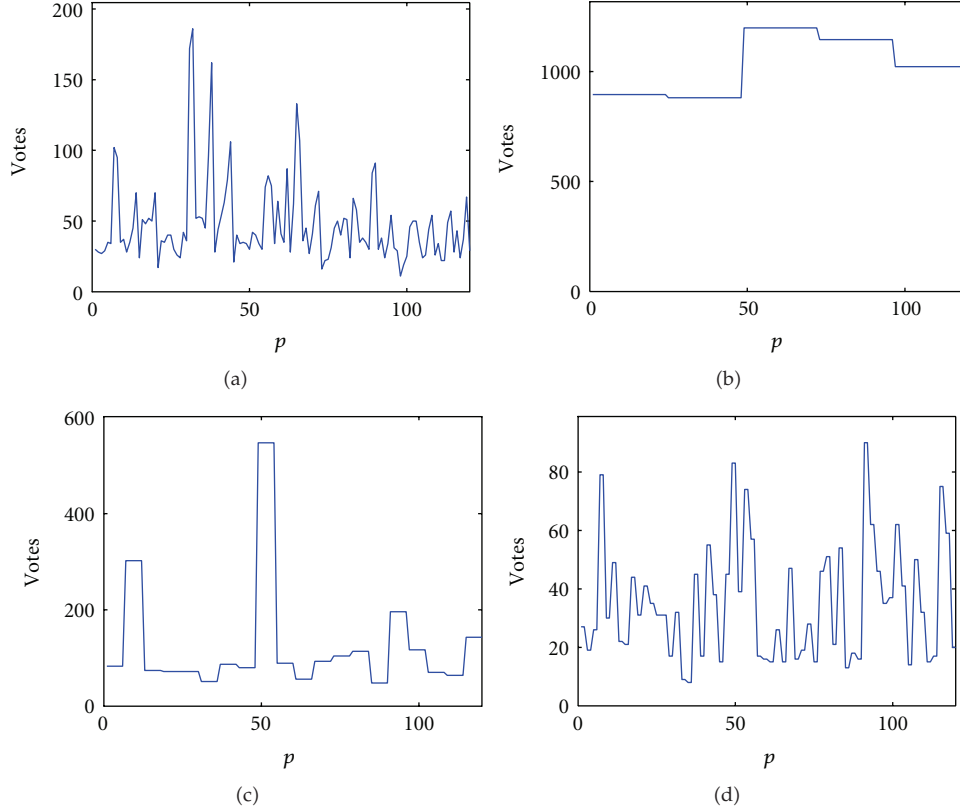


Figure 1: APA election data with both full and partial rankings are shown plotted in lexicographic order on S_5 . The four subplots indicate as follows: (a) full ranking data; (b) votes where only the top candidate was given; (c) votes where the top two candidates in order were given; (d) votes where the top three were given. Note that, in each case, the votes peak at different locations: (a) the maximum votes (186) go to the ordering 23154, that is, candidate 2 has top preference, 3 second, followed by 1, 5, and 4 in decreasing preference; (b) the maximum votes (1,198) go to candidate 3; (c) the maximum votes (547) go to the ordering 31; (d) the maximum votes (90) go to the ordering 451. Note that, in (d), the second most popular ordering, getting 83 votes, is 312, which is more consistent with the result of (b) and (c).

described previously have square basis matrices D with respective dimensions 1, 4, 5, 6, 5, 4, 1, giving a total of 120 basis functions on S_5 , where the number 120 is obtained by summing squares of dimensions. The basis may be constructed using real-valued functions, using the Young orthogonal representation (YOR). The Fourier transform and its inverse are, respectively, written

$$X(\lambda) = \sum_{p \in S_n} x(p) D_\lambda(p)^\top, \quad x(p) = \sum_{\lambda \vdash n} \frac{n_\lambda}{n!} \text{Trace}[X(\lambda) D_\lambda(p)]. \quad (2.3)$$

The symbol $\lambda \vdash n$ on the right hand sum indicates a sum over all partitions for which D_λ is defined. Algorithms for constructing the D matrices are given in Huang et al. [6, Algs 3,4] and are used in obtaining the experimental results of this paper. In particular, we have $D_n(p) = 1$, so that $X(n)$ is a scalar containing the “d.c.” value of the signal, and $D_{1^n}(p)$ is also

scalar alternating between +1 and -1 in the manner similar to the Nyquist frequency $k = N/2$ in the DFT.

Two important properties of the Fourier transform are relevant to this paper: the Fourier basis matrices D_λ that are obtained from the YOR are orthogonal, $D_\lambda(p)D_\lambda(p)^\top = I$, which mimics the exponential unitarity $e^{j\omega}e^{-j\omega} = 1$ in the DFT; under a left translation of the data on S_n obtained by $x(p) \mapsto x(sp)$, the coefficients undergo the transformation $X(\lambda) \mapsto X(\lambda)D_\lambda(s)$, and, under a right translation $x(p) \mapsto x(sp)$, the coefficients transform as $X(\lambda) \mapsto D_\lambda(s)X(\lambda)$. Those two properties suggest an interpretation of the matrix-valued Fourier coefficients in terms of magnitude and phase [8]. The Fourier coefficient may be written using the matrix polar decomposition as $X(\lambda) = \hat{X}(\lambda)O(\lambda)$, where \hat{X} representing magnitude is the positive semidefinite matrix obtained as the square root of XX^\top , and O is an orthogonal matrix representing phase. A standard result in matrix theory [11, page 190] shows that the magnitude \hat{X} is unique, though the phase O needs not be unless X is nonsingular. Under left translation by s , the magnitude remains invariant while the phase changes by $O(\lambda) \mapsto D_\lambda(s)O(\lambda)$, which is analogous to the phase shift $\phi(k) \mapsto 2\pi kt/N + \phi(k)$ for the DFT. Note that both magnitude and phase may be computed using the singular value decomposition (SVD), $X = USV^\top$, by setting $\hat{X} = USU^\top$ and $O = UV^\top$. Below, we use the polar decomposition of magnitude and phase and analyze its properties for partial ranking data.

3. Fourier Analysis of Partial Rankings

In the previous section, we saw that translational symmetry in the DFT domain results in a projection invariance $X[k] = P[k]X[k]$ for the DFT coefficients. Inspired by that result and noting that our method of placing partial ranking data on S_n results in a kind of translational symmetry, we look for the relevant projection characteristics of the Fourier coefficient matrices on S_n . Finding the projection characteristics provides significant reduction in computational complexity and also shows the role of phase for partial ranking data as discussed below. For that purpose, define for each subgroup H of S_n and each $\lambda \vdash n$ the matrix

$$P_H(\lambda) = \frac{1}{|H|} \sum_{h \in H} D_\lambda(h). \quad (3.1)$$

Then, it is known [12, page 111] that $P_H = P_H^\top$ and $P_H P_H = P_H$, so that P_H is an orthogonal projection. The main result of this paper is now stated.

Theorem 3.1. *Let x denote a function on S_n that is piecewise constant with respect to a subgroup, that is, $x(p) = x(ph)$ for every $p \in S_n$ and h in the subgroup H with $|H|$ elements. Then, each Fourier coefficient of x is invariant under the corresponding projection: $X(\lambda) = P_H(\lambda)X(\lambda)$, and that is true of its magnitude as well $\hat{X}(\lambda) = P_H(\lambda)\hat{X}(\lambda)$.*

Proof. The projection invariance of X follows from the translational property of the S_n Fourier transform, from which $x(p) = x(ph)$ results in $X = D(h)X$, when averaged over all elements of H result in $X = P_H X$. (This fact has been shown in the literature; see [12] and Kondor [13, Section 5]). To prove that \hat{X} is invariant, note that P_H being a projection means that there exists an orthogonal matrix U such that $P_H = UI_r U^\top$, where I_r is the identity matrix up to

the first $r = \text{Rank}[P_H]$ entries. Then, $\widehat{X}' = U^\top \widehat{X} U$ is the unique positive semidefinite square root of $X' = U^\top X U$. Since $X' = I_r X'$, $\widehat{X}' \widehat{X}' = X' (X')^\top$ implies that $\widehat{X}' = I_r \widehat{X}' I_r$, so that \widehat{X}' is zero outside the upper left $r \times r$ subblock. Consequently, $\widehat{X}' = I_r \widehat{X}'$, and, therefore, $\widehat{X} = P_H \widehat{X}$. \square

The theorem may be applied to partial ranking data consisting of k out of n elements ranked by using the subgroup $H = S_{n-k}$ that fixes the first k elements and varies the remaining ones. Table 1 shows the ranks of the projections for the first three values of k . The reader may note that Diaconis [1] provides essentially the same numbers as in Table 1, though not obtained through projections. For $k = 1$, only two frequencies λ are involved, each with rank 1. The dimension of the representation $D_{(n-1,1)}$ is $n - 1$, and consequently the projection $P_H X$ has only $n - 1$ degrees of freedom. Therefore, the n degrees of freedom for first-choice-only data ($k = 1$) are divided between the one-dimensional ‘‘d.c.’’ value obtained for frequency $\lambda = (n)$ and the $n - 1$ degrees of freedom for $\lambda = (n - 1, 1)$.

The theorem and table are illustrated with examples in the next section.

We examine the roles that magnitude and phase play in partial ranking data by appealing to the more familiar DFT for intuition. If X is the DFT of real-valued data x , with magnitude-phase decomposition $X = |X|e^{j\phi}$, then the inverse DFT of the magnitude $|X|$ alone is the zero-phase signal

$$x_{zp}[n] = \frac{1}{N} \sum_{k=0}^{N-1} |X[k]| e^{j2\pi kn/N}. \quad (3.2)$$

The zero-phase signal has certain properties: its peak value occurs at the origin since $x_{zp}[0] \geq |x_{zp}[n]|$; it is symmetric with respect to sign inversion, since $x_{zp}[-n] = x_{zp}[n]$. We may shift the peak of x_{zp} from 0 to any desired location q by applying the linear phase shift $\phi \mapsto \phi - 2\pi kq/N$. The resulting linear phase signal is

$$x_{lp}[n] = \frac{1}{N} \sum_{k=0}^{N-1} |X[k]| e^{-j2\pi kq/N} e^{j2\pi kn/N}. \quad (3.3)$$

The properties of the linear phase signal x_{lp} are now as follows: its peak value occurs at $n = q$; it is symmetric about q since $x_{lp}[q - n] = x_{lp}[q + n]$. In other words, we see that, in the absence of phase, the basic components add directly to peak at the starting point, and by shifting the starting point to any given location produces a linear phase version of the signal. Analogous to the zero-phase signal, we may define the unit-magnitude signal by applying the inverse DFT to only the phase:

$$x_{um}[n] = \frac{1}{N} \sum_{k=0}^{N-1} e^{j\phi(k)} e^{j2\pi kn/N}. \quad (3.4)$$

For the DFT, magnitude, and phase, each contains half the degrees of freedom of the original signal, and therefore both are equally important to exact numerical reconstruction. The concepts discussed also apply for the symmetric group S_n as we now show.

Table 1: Rank of projection matrices $P_{n-k}(\lambda)$ for various k and λ . All of the required λ for $k \leq 3$ are shown [1].

$k \setminus \lambda$	(n)	$(n-1, 1)$	$(n-2, 2)$	$(n-2, 1^2)$	$(n-3, 3)$	$(n-3, 2, 1)$	$(n-3, 1^3)$
1	1	1	0	0	0	0	0
2	1	2	1	1	0	0	0
3	1	3	3	3	1	2	1

Using the inverse transform (2.3), we define the zero-phase signal x_{zp} on S_n corresponding to the data x as

$$x_{zp}(p) = \sum_{\lambda \vdash n} \frac{n_\lambda}{n!} \text{Trace} \left[\widehat{X}(\lambda) D_\lambda(p) \right]. \quad (3.5)$$

Noting that $D_\lambda(e) = I$ for the identity permutation, we see that the positive semidefiniteness of \widehat{X} implies that $\text{Trace}[\widehat{X}U] \leq \text{Trace}[\widehat{X}]$ for every orthogonal matrix U , as easily seen by using the eigen-decomposition $\widehat{X} = V\Lambda V^\top$ and applying the circular invariance of trace. Consequently, $x_{zp}(e) \geq x_{zp}(p)$ for all p . Furthermore, there is inversion symmetry since $x_{zp}(p^{-1}) = x_{zp}(p)$ due to the trace property $\text{Trace}[\widehat{X}D^\top] = \text{Trace}[D\widehat{X}^\top] = \text{Trace}[\widehat{X}D]$. The properties of a zero-phase signal are formally similar to those of an ‘‘autocorrelation,’’ which we define on S_n as follows:

$$a_x(s) = \sum_{p \in S_n} x(p)x(ps). \quad (3.6)$$

The connection between zero-phase signals and autocorrelations is made clear in a theorem stated below.

Reasoning as above, we see that we may shift the peak of the zero-phase signal to any given permutation q by the linear phase transformation $\widehat{X} \mapsto \widehat{X}D_\lambda(q^{-1})$, resulting in the linear-phase signal

$$x_{lp}(p) = \sum_{\lambda \vdash n} \frac{n_\lambda}{n!} \text{Trace} \left[\widehat{X}(\lambda) D_\lambda(q)^\top D_\lambda(p) \right]. \quad (3.7)$$

Properties of the linear-phase signal are established in the following theorem, the proof of which is given in an earlier paper [8].

Theorem 3.2. *For every real-valued function x on S_n with Fourier transform X , we have the following.*

- (a) *The transform X is symmetric with respect to matrix transpose if and only if x is symmetric with respect to inversion:*

$$\forall \lambda \vdash n \quad X(\lambda) = X(\lambda)^\top \iff x(p) = x(p^{-1}) \quad \forall p \in S_n. \quad (3.8)$$

- (b) *$X(\lambda)$ is positive semidefinite for all λ if and only if there exists a function y such that x is the autocorrelation of y , that is, $x = a_y$ using the notation of (3.6).*

- (c) *Symmetric functions are precisely those with linear-phase transforms: there exists $q \in S_n$ such that $X(\lambda) = S(\lambda)D_\alpha(q^{-1})$ with $S(\lambda) = S(\lambda)^\top$ if and only if $x(pq) = x(p^{-1}q)$ for all $p \in S_n$.*

The theorem shows that each linear-phase signal is inversion symmetric about its peak location q , that is, $x_{lp}(qp) = x_{lp}(qp^{-1})$. As above, we may define the unit-magnitude signal by using only the phase O in the polar decomposition $X = \widehat{X}O$ in the inverse DFT on S_n as follows:

$$x_{um}(p) = \sum_{\lambda \vdash n} \frac{n_\lambda}{n!} \text{Trace}[O(\lambda)D_\lambda(p)]. \quad (3.9)$$

Noting that the polar decomposition $X = \widehat{X}O$ of an $m \times m$ matrix places $m(m+1)/2$ degrees of freedom in the positive definite matrix \widehat{X} and $m(m-1)/2$ in the orthogonal matrix O , we see that magnitude is slightly more important (by m) to numerically reconstructing full ranking data. However, the situation is much different when partial rank data is involved. By examining Table 1 and using Theorem 3.1, we show that the unit-magnitude signal is nearly complete in the case of first rank data.

Theorem 3.3. *If x is top choice only data ($k = 1$) on S_n , then there exist constants α and β such that $x = \alpha x_{um} + \beta$.*

The proof follows after noting that, by Theorem 3.1 and Table 1, the magnitude $\widehat{X}(n-1, 1)$ is a scalar, so that $\alpha = \widehat{X}(n-1, 1)$ and $\beta = (n!)^{-1}X(n)(1 - \alpha/|X(n)|)$.

4. Examples

Consider the group S_5 used for the APA data shown in Figure 1. For the top two choice data ($k = 2$), the ranks of the projections in Table 1 show that the $5 \times 4 = 20$ degrees of freedom are allocated as follows: 1 in the d.c. term $X(5)$; 8 in the term $X(4, 1)$; the remaining 11 degrees of freedom allocated as 5 and 6, respectively, in each of the Fourier coefficients for $(3, 2)$ and $(3, 1^2)$. By choosing a basis in which $P(4, 1) = I_2$, we obtain the following for the nonzero entries of the Fourier coefficient and its magnitude (rounded to integers):

$$X(4, 1) = \begin{pmatrix} -729 & 1452 & 986 & 505 \\ -2808 & 237 & -1885 & -59 \end{pmatrix}, \quad \widehat{X}(4, 1) = \begin{pmatrix} 1964 & 94 \\ 94 & 3389 \end{pmatrix}. \quad (4.1)$$

Each matrix is actually 4×4 , and the zero entries are not shown.

To illustrate the properties of phase for partial ranking data on S_n , we reconstruct each of the partial rank signals in Figure 1 using only zero and linear phase and show the results in Figure 2. In Figure 2(d), we see a strikingly good fit between the partial rank data with two preferences and its linear phase approximation: numerically, we have $\|x - x_{lp}\| \cdot \|x + x_{lp}\|^{-1} = 0.08$, where $\|\cdot\|$ is the l_2 norm. This suggests that the phase structure of the two-preference data is relatively simple, and the inversion symmetry property indicates that voters are equally content with transposing the order of the two top preferences moving away from the peak. The result is made more interesting by noting that, of the 20 degrees of freedom in top-two preference data, only 6 are constrained by the magnitude spectrum given by the \widehat{X} matrices;

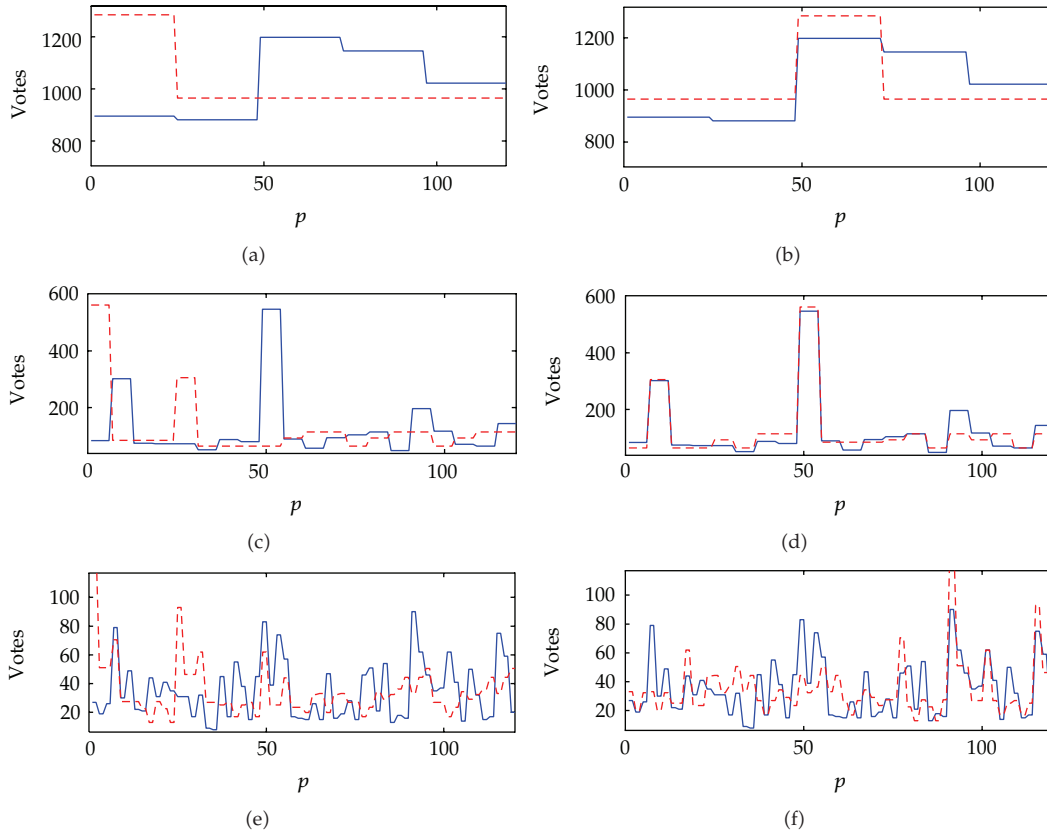


Figure 2: Partial ranking data from the APA election shown plotted in lexicographic order on S_5 , with comparisons to the zero-phase (magnitude only) approximations in the left column, and the linear-phase approximations in the right. The linear-phase approximations are in each case adjusted so that the location of their peak value q matches that of the original data. The six subplots indicate as follows: (a) first preference only (blue) compared to zero phase (red dashed line); (b) first preference only with linear-phase approximation; (c) and (d) the same for the top two preferences; (e) and (f) the same for the top three preferences. Note that the two-preference data in (d) is well fit by the linear-phase approximation.

hence, adding only the linear phase term necessary to shift the peak should not be sufficient to reconstruct 92% of the signal, but it is.

The different levels of fit between the partial rank data and its linear-phase approximations may be understood also by considering the degrees of freedom involved. On the domain S_5 for the APA data, first preference data has 5 degrees of freedom. From Table 1, we see that are two frequencies involved, both with rank 1. As discussed above, $\hat{X}(n-1, 1)$ is a scalar. Consequently, the magnitude spectrum constrains 2 out of the 5 degrees of freedom. The case $k = 2$ is discussed above, and, for $k = 3$, we have that 24 out of the 60 degrees of freedom are constrained by magnitudes. However, as n increases, the degrees of freedom for the magnitude spectrum do not increase, because the ranks of the projection matrices are independent of n . For example, for $n = 50$, the magnitude spectrum for top three choices data ($k = 3$) constrains only 24 out of the 117,600 degrees of freedom. Consequently, for three choices data with large n , the phase spectrum by far exceeds the magnitude component in constraining data.

To illustrate the role of phase for top-choice data for large n , we examine the college rankings from 2009 by US News and World Report that is available online (http://supportingadvancement.com/potpourri/us_news_and_world_report/us_news_rankings.htm). In this data, $n = 65$ American universities are ranked on 17 numerical categories, including acceptance rate, percentage of classes with fewer than 20 students, and alumni giving rate. We consider each category as a voter giving a vote to only the university having the top category value. In the event of ties, which happens only in one category—the percentage of need met for full time students, where 23 universities met 100% of the need—all of the universities having the top value were given a vote. Figure 3 shows the data is poorly fit by with zero phase, as expected, but the shape of the data is well fit up to a scale factor by the unit-magnitude signal as expected from Theorem 3.3.

5. Discussion

We have seen in the previous section that the fit between partial ranking data and its linear phase approximation can be surprisingly good, especially in the case of the APA data for $k = 2$. The quality of linear phase fit is not limited to partial rank data. Full ranking data, which are discussed in [8], may also show a good linear phase approximation. Consider the German survey data, which consists of full rankings of four items by 2,262 voters [14]. Figure 4(a) shows that the data is well reconstructed by a linear-phase approximation; in fact, the linear-phase approximation reproduces 93% of the original signal as measured by $\|x - x_{lp}\| \cdot \|x + x_{lp}\|^{-1}$. Similarly, Figure 4(b) shows that the full ranking data for the APA election is well approximated (78%) by its linear-phase version. However, with full ranking data, the magnitude spectrum dominates: on S_4 , as with the German survey data, we obtain 17 out of the 24 degrees of freedom in the full-ranking data from the magnitude spectrum, while, on S_5 , we obtain 73 of the 120 or 62.5% of the d.o.f. from the magnitude spectrum of full rankings. Therefore, with full ranking data, we should not be as surprised by the quality of fit by linear-phase approximation as we might be with partial rank data.

It is reasonable to wonder what we gain by approximating data that we already have in exact form. Diaconis [1] states a general principle in analyzing data: “if you’ve found some structure, take it out, and look at what’s left.” The results in this and the previous section show cases where linear-phase structure exists in full rank and, more surprisingly, given the degrees of freedom argument, in partial-rank data. The high level of fit in the cases we have analyzed suggest that, once we remove the linear phase structure, there is little left. it would be interesting to apply linear phase approximation to a larger variety of data sets to see whether such symmetry is common. Also, a potential application of the linear-phase formulation is that it provides a way of reasoning about ranking data with reduced complexity, where phase is essentially eliminated except for a single component. It would be interesting to apply the linear phase approximation as a simplifying means to compare graphs up to relabeling of data [7].

5.1. Complexity

One of the limitations of ranking data is that the size of the domain S_n increases as $n!$, making it impractical to capture a complete set of fully ranked data for n much larger than 10. Furthermore, the complexity of the group theoretic FFT for S_n is $\mathcal{O}(n! \log^2 n!)$, as shown in Maslen [15, Theorem 1.1]. This is very difficult to compute for $n > 10$. However, partial ranking data and their spectral analysis allow data for much larger n to be analyzed. For

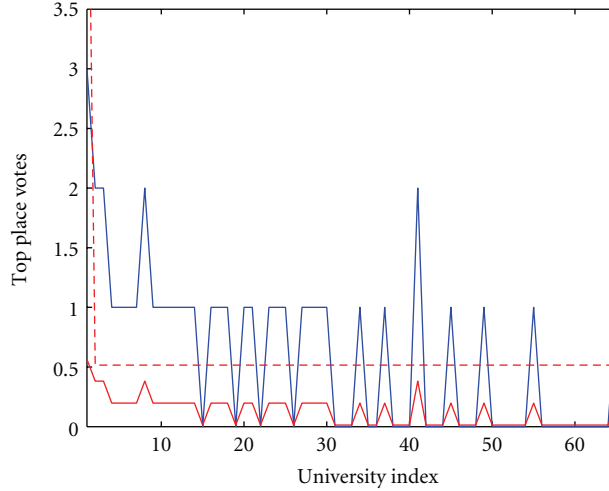


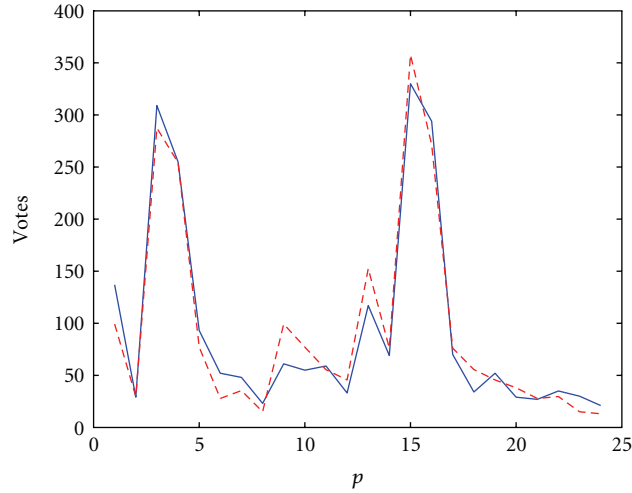
Figure 3: Top place votes for 65 American universities are shown plotted in the solid blue line. The dashed red line shows the reconstruction using the zero-phase version x_{zp} , and the solid red line shows the result using x_{um} , the unit-magnitude signal.

example, the number of data points for the top 3 out of n choices is $n!/(n-3)! \approx n^3$, which remains tractable for n up to 100. Maslen [15] showed that the group-theoretic FFT on S_n when adapted for $k = 3$ has $\mathcal{O}(n^4)$ complexity; in comparison, the ordinary FFT on \mathbb{Z}_m for $m = (100)^4$ can be completed in 3 seconds on a 2.6 GHz quad-core Xeon processor. Therefore, we see that processing only partial rank data allows a capability of roughly an order-of-magnitude increase in n over fully-ranked data. If we restrict to only top choice data ($k = 1$), then there is a linear-time algorithm for computing the Fourier transform [16].

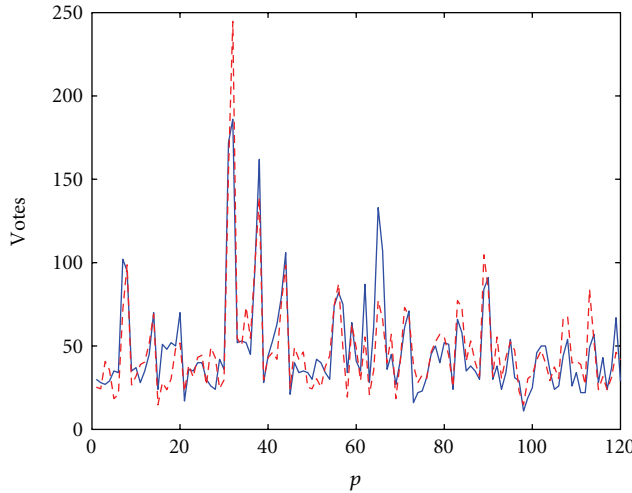
Knowing the complexity of the transform helps to determine the complexity of either the zero-phase (3.5) or the unit-magnitude (3.9) approximations. Each of those approximations requires the following three steps: computing the forward transform, separating each coefficient matrix into magnitude and phase components, and computing the inverse transform. The inverse transform has the same complexity as the forward transform. The magnitude-phase separation requires performing an SVD of each matrix coefficient, followed by two matrix multiplications for the magnitude, or one for the phase. The cost of each SVD is $\mathcal{O}(n_\lambda^3)$, where n_λ is the size of each representation λ . Unfortunately, there are no simple, closed-form, expressions for n_λ . However, when using partial rank data, the number of coefficients involved is relatively small due to the projection property. From Table 1, we see that there are only 7 coefficient matrices for top-three choices data ($k = 3$), the largest of which has rank 3. Note that the ranks listed in the Table are independent of n . We may use reduced SVDs for these 7 coefficients, resulting in efficient calculation of the magnitude-phase separation due to their low ranks. Consequently, for large n , the cost of either the zero-phase or the unit-magnitude approximation is dominated by the cost of the forward and inverse transforms, which are each $\mathcal{O}(n^4)$ for top-three choice data.

5.2. Approximation and Compression

It is reasonable to wonder whether we may obtain signal compression by approximating partial rank data by either (3.5) or (3.9). Clearly, for large n and small k , the zero-phase



(a) Full ranking German survey data (blue) and its linear phase approximation (dashed red)



(b) Full ranking APA data (blue) and its linear phase approximation (dashed red)

Figure 4: Two examples illustrating linear-phase approximation for full ranking data. Lexicographic ordering is used for permutations on the horizontal axes of both graphs. Note that the data in (a) has domain S_4 , while (b) has domain S_5 .

approximation (3.5) is poor because magnitude constrains only a small number of degrees of freedom, as described in the previous section. Conversely, the phase spectrum constrains much of the data; as discussed previously, phase constrains all but 24 of the 117,600 degrees of freedom for $n = 50$, meaning that it is really a very minor compression. To summarize using (3.5) to replace the signal is too much compression, while using (3.9) is too little compression.

The error in approximating data with either its zero-phase version (3.5) or unit-magnitude (3.9) may be determined as follows. Considering the inverse transform on S_n is determined by Fourier coefficients $X(\lambda)$, we see that the error in zero-phase approximation is governed by $\|X(\lambda) - \hat{X}(\lambda)\|^2$, where the norm means the sum of squared entries. Consequently,

due to the submultiplicative property of the matrix norm, we estimate a relative error at each λ of

$$\frac{\|X(\lambda) - \hat{X}(\lambda)\|^2}{\|X(\lambda)\|^2} \leq \|O(\lambda) - I\|^2 \leq 2n_\lambda. \quad (5.1)$$

Here, n_λ is the dimension of $O(\lambda)$, and we used the identity $\|\hat{X}\|^2 = \|X\|^2$. A similar calculation for unit-magnitude approximation shows that the error at each λ is

$$\frac{\|X(\lambda) - O(\lambda)\|^2}{\|X(\lambda)\|^2} \leq 2n_\lambda. \quad (5.2)$$

These are weak upper bounds, and it would be desirable to improve on them in future work.

6. Summary

This paper analyzes the properties of the Fourier spectrum for partial ranking data and shows that the transform coefficients satisfy a projection invariance. The coefficients may be converted to magnitude and phase components, with the magnitude also showing projection invariance. We show that first rank data is essentially determined by its phase spectrum, but that as n increases, the phase dominates magnitude in forming partial rank data.

Acknowledgment

The author thanks the anonymous reviewers for their comments, which greatly improved the paper.

References

- [1] P. Diaconis, "A generalization of spectral analysis with application to ranked data," *The Annals of Statistics*, vol. 17, no. 3, pp. 949–979, 1989.
- [2] G. Lebanon and Y. Mao, "Non-parametric modeling of partially ranked data," *Journal of Machine Learning Research*, vol. 9, pp. 2401–2429, 2008.
- [3] P. Hall and H. Miller, "Modeling the variability of rankings," *The Annals of Statistics*, vol. 38, no. 5, pp. 2652–2677, 2010.
- [4] P. Diaconis, *Group Representations in Probability and Statistics*, Institute of Mathematical Statistics, Hayward, Calif, USA, 1988.
- [5] P. Diaconis and B. Sturmfels, "Algebraic algorithms for sampling from conditional distributions," *The Annals of Statistics*, vol. 26, no. 1, pp. 363–397, 1998.
- [6] J. Huang, C. Guestrin, and L. Guibas, "Fourier theoretic probabilistic inference over permutations," *Journal of Machine Learning Research*, vol. 10, pp. 997–1070, 2009.
- [7] R. Kondor and K. Borgwardt, "The skew spectrum of graphs," in *Proceedings of the International Conference on Machine Learning (ICML)*, A. McCallum and S. Roweis, Eds., pp. 496–503, Omnipress, 2008.
- [8] R. Kakarala, "A signal processing approach to Fourier analysis of ranking data: the importance of phase," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1518–1527, 2011.
- [9] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and Systems*, Prentice-Hall, 2nd edition, 1996.
- [10] R. Kondor, " S_n ob: a C++ library for fast Fourier transforms on the symmetric group," 2006, <http://www.its.caltech.edu/~risi/index.html>.

- [11] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, Computer Science and Applied Mathematics, Academic Press, Orlando, Fla, USA, 2nd edition, 1985.
- [12] E. Hewitt and K. A. Ross, *Abstract Harmonic Analysis. Vol. II: Structure and Analysis for Compact Groups. Analysis on Locally Compact Abelian Groups*, Springer, New York, NY, USA, 1970.
- [13] R. Kondor, "The skew spectrum of functions on finite groups and their homogeneous spaces," *Representation Theory*. In press, <http://arxiv.org/abs/0712.4259>.
- [14] M. Croon, "Latent class models for the analysis of rankings," in *New Developments in Psychological Choice Modeling*, G. D. Solte, H. Feger, and K. C. Klauer, Eds., pp. 99–121, North-Holland, 1989.
- [15] D. K. Maslen, "The efficient computation of Fourier transforms on the symmetric group," *Mathematics of Computation*, vol. 67, no. 223, pp. 1121–1147, 1998.
- [16] M. Clausen and R. Kakarala, "Computing Fourier transforms and convolutions of S_n - 1-invariant signals on S_n in time linear in n ," *Applied Mathematics Letters*, vol. 23, no. 2, pp. 183–187, 2010.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

