

## MIT Open Access Articles

### *Compression in the Space of Permutations*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Wang, Da et al. "Compression in the Space of Permutations." IEEE Transactions on Information Theory 61, 12 (December 2015): 6417–6431 © 2015 Institute of Electrical and Electronics Engineers (IEEE)

**As Published:** <http://dx.doi.org/10.1109/tit.2015.2485270>

**Publisher:** Institute of Electrical and Electronics Engineers (IEEE)

**Persistent URL:** <http://hdl.handle.net/1721.1/111002>

**Version:** Original manuscript: author's manuscript prior to formal peer review

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike



# Compression in the Space of Permutations

Da Wang, Arya Mazumdar, *Member, IEEE*, and Gregory W. Wornell, *Fellow, IEEE*

**Abstract**—We investigate lossy compression (source coding) of data in the form of permutations. This problem has direct applications in the storage of ordinal data or rankings, and in the analysis of sorting algorithms. We analyze the rate-distortion characteristic for the permutation space under the uniform distribution, and the minimum achievable rate of compression that allows a bounded distortion after recovery. Our analysis is with respect to different practical and useful distortion measures, including Kendall tau distance, Spearman’s footrule, Chebyshev distance and inversion- $\ell_1$  distance. We establish equivalence of source code designs under certain distortions and show simple explicit code designs that incur low encoding/decoding complexities and are asymptotically optimal. Finally, we show that for the Mallows model, a popular nonuniform ranking model on the permutation space, both the entropy and the maximum distortion at zero rate are much lower than the uniform counterparts, which motivates the future design of efficient compression schemes for this model.

**Index Terms**—lossy compressions, mallows model, partial sorting, permutation space

## I. INTRODUCTION

**P**ERMUTATIONS are fundamental mathematical objects and the topic of *codes in permutations* is a well-studied subject in coding theory. A variety of applications that correspond to different metric functions on the symmetric group on  $n$  elements  $\mathcal{S}_n$  have been investigated. For example, some works focus on error-correcting codes in  $\mathcal{S}_n$  with Hamming distance [1], [2], and some others investigate the error correction problem under metrics such as Chebyshev distance [3] and Kendall tau distance [4].

While error correction problems in permutation spaces have been investigated before, the lossy compression problem is largely left unattended. In [5], [6], the authors investigate the lossless compression of a group

This work was supported, in part, by AFOSR under Grant No. FA9550-11-1-0183, and by NSF under Grant No. CCF-1017772 and Grant No. CCF-1318093.

D. Wang was with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139 and now with Two Sigma Investments, New York, NY, 10013 (Email: dawang@alum.mit.edu). A. Mazumdar is with the Department of Electrical and Computer Engineering, University of Minnesota, MN, 55455 (Email: arya@umn.edu). G. W. Wornell is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139 (Email: gww@mit.edu).

The material in this paper was presented in part at International Symposium on Information Theory, Istanbul, Turkey, 2013 and International Symposium on Information Theory, Hawaii, HI, 2014.

of permutations with certain properties, such as efficient rank querying (given an element, get its rank in the permutation) and selection (given a rank, retrieve the corresponding element). By contrast, in this paper we consider the lossy compression (source coding) of permutations, which is motivated by the problems of storing ranking data, and lower bounding the complexity of approximate sorting, which we now describe.

**Storing ranking data:** In applications such as recommendation systems, users rank products and these rankings are analyzed to provide new recommendations. To have personalized recommendation, it may be necessary to store the ranking data for each user in the system, and hence the storage efficiency of ranking data is of interest. Because a ranking of  $n$  items can be represented as a permutation of 1 to  $n$ , storing a ranking is equivalent to storing a permutation. Furthermore, in many cases a rough knowledge of the ranking (e.g., finding one of the top five elements instead of the top element) is sufficient. This poses the question of the number of bits needed for permutation storage when a certain amount of error can be tolerated. In many current applications the cost of lossless storage is usually tolerable and hence lossy compression may not be necessary. However lossy compression is a fundamental topic and it is of theoretical interest to understand the trade-off involved.

**Lower bounding the complexity of approximate sorting:** Given a group of elements of distinct values, comparison-based sorting can be viewed as the process of searching for a true ranking by pairwise comparisons. Since each comparison in sorting provides at most 1 bit of information, the log-size of the permutation set  $\mathcal{S}_n$ ,  $\log_2(n!)$ , provides a lower bound to the required number of comparisons. Similarly, the lossy source coding of permutations provides a lower bound on the number of comparisons to the problem of comparison-based approximate sorting, which can be seen as finding a true permutation up to a certain distortion. Again, the log-size of the code indicates the amount of information (in bits) needed to specify the true permutation, which in turn provides a lower bound on the number of pairwise comparisons needed.

In one line of work, authors of [7] derived both lower and upper bounds for approximate sorting in some range of allowed distortion with respect to the Spearman’s footrule metric [8] (see Definition 1 below). Another line of work concerns an important class of approximate sort-

ing, the problem of *partial sorting*, first proposed in [9] (cf.[10, Chapter 8] for an exposition on the relationships between various sorting problems). Given a set of  $n$  elements  $\mathcal{V}$  and a set of indices  $\mathcal{I} \subset \{1, 2, \dots, n\}$ , a partial sorting algorithm aims to arrange the elements into a list  $[v_1, v_2, \dots, v_n]$  such that for any  $i \in \mathcal{I}$ , all elements with indices  $j < i$  are no greater than  $v_i$ , and all elements with indices  $j' > i$  are no smaller than  $v_i$ . A partial sorting algorithm essentially selects all elements with ranks in the set  $\mathcal{I}$ , and hence is also called *multiple selection*. The information-theoretic lower bound for partial sorting algorithms have been proposed in [11], and the authors of [12] propose a multiple selection algorithms with expected number of comparisons within the information-theoretic lower-bound and an asymptotically negligible additional term.

Comparing with existing work (such as [11]), our analysis framework via rate-distortion theory is more general as we provide an information-theoretic lower bound on the query complexity for *all* approximate sorting algorithms that achieve a certain distortion, and the multiple selection algorithm proposed in [12] turns out to be optimal for the general approximate sorting problem as well. Therefore, our information-theoretic lower bound is tight.

**Remark 1** (Comparison-based sorting implies compression). *It is worth noting that every comparison-based sorting algorithm corresponds to a compression scheme of the permutation space. In particular, the string of bits that represent comparison outcomes in any deterministic (approximate) sorting algorithm corresponds to a (lossy) representation of the permutation.*

*For a more in-depth discussion on the relationship between sorting and compression, see [13] and references therein.*

Beyond the above applications, the rate-distortion theory in permutation spaces is of technical interest on its own because the permutation space does not possess the product structure that a discrete memoryless source induces.

With the above motivations, we consider the problem of lossy compression in permutation spaces in this paper. Following the classical rate-distortion formulation, we aim to determine, given a distortion measure  $d(\cdot, \cdot)$ , the minimum number of bits needed to describe a permutation with distortion at most  $D$ .

The analysis of the lossy compression problem depends on the source distribution and the distortion measure. We are mainly concerned with the permutation spaces with a uniform distribution, and consider different distortion measures based on four distances in the permutation spaces: the Kendall tau distance, Spearman's footrule, Chebyshev distance and inversion- $\ell_1$  distance.

As we shall see in Section II, each of these distortion measures (except inversion- $\ell_1$  distance<sup>1</sup>) has its own operational meaning that may be useful in different applications.

In addition to characterizing the trade-off between rate and distortion, we also show that under the uniform distribution over the permutation space, there are close relationships between some of the distortion measures of interest in this paper. We use these relations to establish the corresponding equivalence of source codes in permutation spaces with different distortion measures. For each distortion measure, we provide simple and constructive achievability schemes, leading to explicit code designs with low complexity.

Finally, we turn our attention to non-uniform distributions over the permutation space. In some applications, we may have prior knowledge about the permutation data, which can be captured in a model of non-uniform distribution. There are a variety of distributional models in different contexts, such as the Bradley-Terry model [15], the Luce-Plackett model [16], [17], and the Mallows model [18]. Among these, we choose the Mallows model due to its richness and applicability in various ranking applications [19], [20], [21]. We analyze the lossless and lossy compression of the permutation space under the Mallows model and with the Kendall tau distance as the distortion measure, and characterize its entropy and end points of its rate-distortion function.

The rest of the paper is organized as follows. We first present the problem formulation in Section II. We then analyze the geometry of the permutation spaces and show that there exist close relationships between some distortion measures of interest in this paper in Section III. In Section IV, we derive the rate-distortion functions for different permutation spaces. In Section V, we provide achievability schemes for different permutation spaces under different regimes. After that, we turn our attention to non-uniform distributional model over the permutation space and analyze the lossless and lossy compression for Mallows model in Section VI. We conclude with a few remarks in Section VII.

## II. PROBLEM FORMULATION

In this section we discuss aspects of the formulation of the rate-distortion problem for permutation spaces. We first introduce the distortion measures of interest in Section II-B, and then provide a mathematical formulation of the rate-distortion problem in Section II-C.

<sup>1</sup>We are interested in inversion- $\ell_1$  distance due to its extremal property shown in Equation (7), which is useful when we derive results for other permutation spaces. Further use of this metric in the context of smooth representation of permutations can be found in [14].

### A. Notation and facts

Let  $\mathcal{S}_n$  denote the symmetric group of  $n$  elements. We write an element of  $\mathcal{S}_n$  as an array of natural numbers with values ranging from  $1, \dots, n$  and every value occurring only once in the array. For example,  $\sigma = [3, 4, 1, 2, 5] \in \mathcal{S}_5$ . This is also known as the *vector notation* for permutations. The identity of the symmetric group  $\mathcal{S}_n$  (identity permutation) is denoted by  $\text{Id} = [1, 2, \dots, n]$ . For a permutation  $\sigma$ , we denote its permutation inverse by  $\sigma^{-1}$ , where  $\sigma^{-1}(x) = i$  when  $\sigma(i) = x$ , and  $\sigma(i)$  is the  $i$ -th element in array  $\sigma$ . For example, the permutation inverse of  $\sigma = [2, 5, 4, 3, 1]$  is  $\sigma^{-1} = [5, 1, 4, 3, 2]$ . Given a metric  $d : \mathcal{S}_n \times \mathcal{S}_n \rightarrow \mathbb{R}^+ \cup \{0\}$ , we define a *permutation space*  $\mathcal{X}(\mathcal{S}_n, d)$ .

Throughout the paper, we let  $[a : b] \triangleq \{a, a + 1, \dots, b-1, b\}$  for any two integers  $a$  and  $b$ , and use  $\sigma[a : b]$  as a shorthand for the vector  $[\sigma(a), \sigma(a+1), \dots, \sigma(b)]$ .

We make use of the following version of *Stirling's approximation*:

$$\left(\frac{m}{e}\right)^m e^{\frac{1}{12m+1}} < \frac{m!}{\sqrt{2\pi m}} < \left(\frac{m}{e}\right)^m e^{\frac{1}{12m}}, m \geq 1. \quad (1)$$

### B. Distortion measures

There exists many natural distortion measures on the permutation group  $\mathcal{S}_n$  [22]. In this paper we choose a few distortion measures of interest in a variety of application settings, including Spearman's footrule ( $\ell_1$  distance between two permutation vectors), Chebyshev distance ( $\ell_\infty$  distance between two permutation vectors), Kendall tau distance and the inversion- $\ell_1$  distance (see Definition 5).

Before introducing definitions for these distortion measures, we define the concept of *ranking*. Given a list of items with values  $v_1, v_2, \dots, v_n$  such that  $v_{\sigma^{-1}(1)} \succ v_{\sigma^{-1}(2)} \succ \dots \succ v_{\sigma^{-1}(n)}$ , where  $a \succ b$  indicates  $a$  is preferred to  $b$ , we say the permutation  $\sigma$  is the *ranking* of this list of items, where  $\sigma(i)$  provides the rank of item  $i$ , and  $\sigma^{-1}(r)$  provides the index of the item with rank  $r$ . Note that sorting via pairwise comparisons is simply the procedure of rearranging  $v_1, v_2, \dots, v_n$  to  $v_{\sigma^{-1}(1)}, v_{\sigma^{-1}(2)}, \dots, v_{\sigma^{-1}(n)}$  based on preferences obtained from pairwise comparisons.

Given two rankings  $\sigma_1$  and  $\sigma_2$ , we measure the total deviation of ranking and maximum deviation of ranking by the Spearman's footrule and the Chebyshev distance respectively.

**Definition 1** (Spearman's footrule [8]). *Given two permutations  $\sigma_1, \sigma_2 \in \mathcal{S}_n$ , the Spearman's footrule between  $\sigma_1$  and  $\sigma_2$  is*

$$d_{\ell_1}(\sigma_1, \sigma_2) \triangleq \|\sigma_1 - \sigma_2\|_1 = \sum_{i=1}^n |\sigma_1(i) - \sigma_2(i)|.$$

**Definition 2** (Chebyshev distance). *Given two permutations  $\sigma_1, \sigma_2 \in \mathcal{S}_n$ , the Chebyshev distance between  $\sigma_1$  and  $\sigma_2$  is*

$$d_{\ell_\infty}(\sigma_1, \sigma_2) \triangleq \|\sigma_1 - \sigma_2\|_\infty = \max_{1 \leq i \leq n} |\sigma_1(i) - \sigma_2(i)|.$$

The Spearman's footrule in  $\mathcal{S}_n$  is upper bounded by  $\lfloor n^2/2 \rfloor$  (cf. Table I) and the Chebyshev distance in  $\mathcal{S}_n$  is upper bounded by  $n - 1$ .

Given two lists of items with ranking  $\sigma_1$  and  $\sigma_2$ , let  $\pi_1 \triangleq \sigma_1^{-1}$  and  $\pi_2 \triangleq \sigma_2^{-1}$ , then we define the number of pairwise adjacent swaps on  $\pi_1$  that changes the ranking of  $\pi_1$  to the ranking of  $\pi_2$  as the Kendall tau distance.

**Definition 3** (Kendall tau distance [23]). *The Kendall tau distance  $d_\tau(\sigma_1, \sigma_2)$  from one permutation  $\sigma_1$  to another permutation  $\sigma_2$  is defined as the minimum number of transpositions of pairwise adjacent elements required to change  $\sigma_1$  into  $\sigma_2$ .*

The Kendall tau distance is upper bounded by  $\binom{n}{2}$ .

**Example 1** (Kendall tau distance). *The Kendall tau distance for  $\sigma_1 = [1, 5, 4, 2, 3]$  and  $\sigma_2 = [3, 4, 5, 1, 2]$  is  $d_\tau(\sigma_1, \sigma_2) = 7$ , as one needs at least 7 transpositions of pairwise adjacent elements to change  $\sigma_1$  to  $\sigma_2$ . For example,*

$$\begin{aligned} \sigma_1 &= [1, 5, 4, \mathbf{2}, \mathbf{3}] \\ &\rightarrow [1, 5, \mathbf{4}, \mathbf{3}, 2] \rightarrow [1, \mathbf{5}, \mathbf{3}, 4, 2] \rightarrow [1, \mathbf{3}, 5, 4, 2] \\ &\rightarrow [3, \mathbf{1}, \mathbf{5}, 4, 2] \rightarrow [3, 5, \mathbf{1}, 4, 2] \rightarrow [3, \mathbf{5}, 4, 1, 2] \\ &\rightarrow [3, 4, 5, 1, 2] = \sigma_2. \end{aligned}$$

Being a popular global measure of disarray in statistics, Kendall tau distance also has a natural connection to sorting algorithms. In particular, given a list of items with values  $v_1, v_2, \dots, v_n$  such that  $v_{\sigma^{-1}(1)} \succ v_{\sigma^{-1}(2)} \succ \dots \succ v_{\sigma^{-1}(n)}$ ,  $d_\tau(\sigma^{-1}, \text{Id})$  is the number of swaps needed to sort this list of items in a bubble-sort algorithm [24].

Finally, we introduce a distortion measure based on the concept of inversion vector, another measure of the order-ness of a permutation.

**Definition 4** (inversion, inversion vector [25]). *An inversion in a permutation  $\sigma \in \mathcal{S}_n$  is a pair  $(\sigma(i), \sigma(j))$  such that  $i < j$  and  $\sigma(i) > \sigma(j)$ .*

We use  $I_n(\sigma)$  to denote the total number of inversions in  $\sigma \in \mathcal{S}_n$ , and

$$K_n(k) \triangleq |\{\sigma \in \mathcal{S}_n : I_n(\sigma) = k\}| \quad (2)$$

to denote the number of permutations with  $k$  inversions.

Denote  $i' = \sigma(i)$  and  $j' = \sigma(j)$ , then  $i = \sigma^{-1}(i')$  and  $j = \sigma^{-1}(j')$ , and thus  $i < j$  and  $\sigma(i) > \sigma(j)$  is equivalent to  $\sigma^{-1}(i') < \sigma^{-1}(j')$  and  $i' > j'$ .

A permutation  $\sigma \in \mathcal{S}_n$  is associated with an inversion vector  $\mathbf{x}_\sigma \in \mathcal{G}_n \triangleq [0 : 1] \times [0 : 2] \times \dots \times [0 : n - 1]$ ,

where  $\mathbf{x}_\sigma(i')$ ,  $1 \leq i' \leq n-1$  is the number of inversions in  $\sigma$  in which  $i' + 1$  is the first element. Formally, for  $i' = 2, \dots, n$ ,

$$\mathbf{x}_\sigma(i'-1) = |\{j' \in [1 : n] : j' < i', \sigma^{-1}(j') > \sigma^{-1}(i')\}|.$$

Let  $\pi \triangleq \sigma^{-1}$ , then the *inversion vector* of  $\pi$ ,  $\mathbf{x}_\pi$ , measures the deviation of ranking  $\sigma$  from Id. In particular, note that

$$\begin{aligned} \mathbf{x}_\pi(k) &= |\{j' \in [1 : n] : j' < k, \pi^{-1}(j') > \pi^{-1}(k)\}| \\ &= |\{j' \in [1 : n] : j' < k, \sigma(j') > \sigma(k)\}| \end{aligned}$$

indicates the number of elements that have *larger ranks and smaller item indices* than that of the element with index  $k$ . In particular, the rank of the element with index  $n$  is  $n - \mathbf{x}_\pi(n-1)$ .

**Example 2.** Given 5 items such that  $v_4 \succ v_1 \succ v_2 \succ v_5 \succ v_3$ , then the inverse of the ranking permutation is  $\pi = [4, 1, 2, 5, 3]$ , with inversion vector  $\mathbf{x}_\pi = [0, 0, 3, 1]$ . Therefore, the rank of the  $v_5$  is  $n - \mathbf{x}_\pi(n-1) = 5 - 1 = 4$ .

The mapping from  $\mathcal{S}_n$  to  $\mathcal{G}_n$  is one-to-one as the inversion vectors exactly describes the execution of the algorithm insertion sort [24].

With these, we define the inversion- $\ell_1$  distance.

**Definition 5** (inversion- $\ell_1$  distance). Given two permutations  $\sigma_1, \sigma_2 \in \mathcal{S}_n$ , we define the inversion- $\ell_1$  distance,  $\ell_1$  distance of two inversion vectors, as

$$d_{\mathbf{x}, \ell_1}(\sigma_1, \sigma_2) \triangleq \sum_{i=1}^{n-1} |\mathbf{x}_{\sigma_1}(i) - \mathbf{x}_{\sigma_2}(i)|. \quad (3)$$

**Example 3** (inversion- $\ell_1$  distance). The inversion vector for permutation  $\sigma_1 = [1, 5, 4, 2, 3]$  is  $\mathbf{x}_{\sigma_1} = [0, 0, 2, 3]$ , as the inversions are  $(4, 2), (4, 3), (5, 4), (5, 2), (5, 3)$ . The inversion vector for permutation  $\sigma_2 = [3, 4, 5, 1, 2]$  is  $\mathbf{x}_{\sigma_2} = [0, 2, 2, 2]$ , as the inversions are  $(3, 1), (3, 2), (4, 1), (4, 2), (5, 1), (5, 2)$ . Therefore,

$$d_{\mathbf{x}, \ell_1}(\sigma_1, \sigma_2) = d_{\ell_1}([0, 0, 2, 3], [0, 2, 2, 2]) = 3.$$

As we shall see in Section III, all these distortion measures are related. While the operational significance of the inversion- $\ell_1$  distance may not be as clear as other distortion measures, some of its properties provide useful insights in the analysis of other distortion measures.

**Remark 2.** While Spearman's footrule and Chebyshev distance operate on the ranking domain, inversion vector and Kendall tau distance can be viewed as operating on the inverse of the ranking domain.

### C. Rate-distortion problems

With the distortions defined in Section II-B, in this section we define rate-distortion problems under both average-case and worst-case distortions.

**Definition 6** (Codebook for average-case distortion). An  $(n, D_n)$  source code  $\bar{\mathcal{C}}_n \subseteq \mathcal{S}_n$  for  $\mathcal{X}(\mathcal{S}_n, d)$  under the average-case distortion is a set of permutations such that for a  $\sigma$  that is drawn from  $\mathcal{S}_n$  according to a distribution  $P$  on  $\mathcal{S}_n$ , there exists an encoding mapping  $f_n : \mathcal{S}_n \rightarrow \bar{\mathcal{C}}_n$  that

$$\mathbb{E}_P[d(f_n(\sigma), \sigma)] \leq D_n. \quad (4)$$

The mapping  $f_n : \mathcal{S}_n \rightarrow \bar{\mathcal{C}}_n$  can be assumed to satisfy

$$f_n(\sigma) = \arg \min_{\sigma' \in \bar{\mathcal{C}}_n} d(\sigma', \sigma)$$

for any  $\sigma \in \mathcal{S}_n$ .

In most parts of this paper we focus on the case  $P$  is uniformly distributed over the symmetric group  $\mathcal{S}_n$ , except in Section VI, where a distribution arising from the Mallows model is used. In both cases the source distribution has support  $\mathcal{S}_n$ , and we define the worst-case distortion as follows.

**Definition 7** (Codebook for worst-case distortion). An  $(n; D_n)$  source code  $\hat{\mathcal{C}}_n \subseteq \mathcal{S}_n$  for  $\mathcal{X}(\mathcal{S}_n, d)$  under the worst-case distortion is a set of permutations such that for any  $\sigma \in \mathcal{S}_n$ , there exists an encoding mapping  $f_n : \mathcal{S}_n \rightarrow \hat{\mathcal{C}}_n$  that

$$\max_{\sigma \in \mathcal{S}_n} d(f_n(\sigma), \sigma) \leq D_n. \quad (5)$$

The mapping  $f_n : \mathcal{S}_n \rightarrow \hat{\mathcal{C}}_n$  can be assumed to satisfy

$$f_n(\sigma) = \arg \min_{\sigma' \in \hat{\mathcal{C}}_n} d(\sigma', \sigma)$$

for any  $\sigma \in \mathcal{S}_n$ .

**Definition 8** (Rate function). For a class of source codes  $\{\mathcal{C}_n\}$  that achieve a distortion  $D_n$ , let  $A(n, D_n)$  be the minimum size of such codes, and we define the minimal rate for distortions  $D_n$  as

$$R(D_n) \triangleq \frac{\log A(n, D_n)}{\log n!}.$$

In particular, we denote the minimum rate of the codebook under average-case distortion with uniform source distribution and worst-case distortions by  $\bar{R}(D_n)$  and  $\hat{R}(D_n)$  respectively.

Similar to the classical rate-distortion setup, we are interested in deriving the trade-off between distortion level  $D_n$  and the rate  $R(D_n)$  as  $n \rightarrow \infty$ . In this work we show that for the distortions  $d(\cdot, \cdot)$  and the sequences of distortions  $\{D_n, n \in \mathbb{Z}^+\}$  of interest,  $\lim_{n \rightarrow \infty} R(D_n)$  exists.

For Kendall tau distance and inversion- $\ell_1$  distance, a close observation shows that in regimes such as  $D_n = O(n)$  and  $D_n = \Theta(n^2)$ ,  $\lim_{n \rightarrow \infty} R(D_n) = 1$  and  $\lim_{n \rightarrow \infty} R(D_n) = 0$  respectively. In these two regimes, the trade-off between rate and distortion is really shown in the higher order terms in  $\log A(n, D_n)$ , i.e.,

$$r(D_n) \triangleq \log A(n, D_n) - \log n! \lim_{n \rightarrow \infty} R(D_n). \quad (6)$$

For convenience, we categorize the distortion  $D_n$  under Kendall tau distance or inversion- $\ell_1$  distance into three regimes. We say  $D$  is small when  $D = O(n)$ , moderate when  $D = \Theta(n^{1+\delta})$ ,  $0 < \delta < 1$ , and large when  $D = \Theta(n^2)^2$ .

We choose to omit the higher order term analysis for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$  because its analysis is essentially the same as  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ , and the analysis for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$  is still open.

Note that the higher order terms  $r(D_n)$  may behave differently under average and worst-case distortions, and in this paper we restrict our attention to the worst-case distortion.

### III. RELATIONSHIPS BETWEEN DISTORTION MEASURES

In this section we show how the four distortion measures defined in Section II-B are related to each other, which is summarized in (7) and (8). These relationships imply equivalence in some lossy compression schemes, which we exploit to derive the rate-distortion functions in Section IV.

For any  $\sigma_1 \in \mathcal{S}_n$  and  $\sigma_2$  randomly uniformly chosen from  $\mathcal{S}_n$ , the following relations hold:

$$\begin{aligned} nd_{\ell_\infty}(\sigma_1, \sigma_2) &\geq d_{\ell_1}(\sigma_1, \sigma_2) \\ &\geq d_\tau(\sigma_1^{-1}, \sigma_2^{-1}) \\ &\geq d_{\mathbf{x}, \ell_1}(\sigma_1^{-1}, \sigma_2^{-1}), \end{aligned} \quad (7)$$

$$\begin{aligned} nd_{\ell_\infty}(\sigma_1, \sigma_2) &\stackrel{w.h.p.}{\lesssim} d_{\ell_1}(\sigma_1, \sigma_2) \\ &\lesssim d_\tau(\sigma_1^{-1}, \sigma_2^{-1}) \\ &\stackrel{w.h.p.}{\lesssim} d_{\mathbf{x}, \ell_1}(\sigma_1^{-1}, \sigma_2^{-1}), \end{aligned} \quad (8)$$

where  $x \lesssim y$  indicates  $x < c \cdot y$  for some constant  $c > 0$ ,

and  $\stackrel{w.h.p.}{\lesssim}$  indicates  $\lesssim$  with high probability.

Next, we provide detailed arguments for (7) and (8) by analyzing the relationship between different pairs of distortion measures.

<sup>2</sup>In the small distortion region with  $R(D_n) = 1$ ,  $r(D_n)$  is negative while in the large distortion region where  $R(D_n) = 0$ ,  $r(D_n)$  is positive.

1) *Spearman's footrule and Chebyshev distance*: Let  $\sigma_1$  and  $\sigma_2$  be any permutations in  $\mathcal{S}_n$ , then by definition,

$$d_{\ell_1}(\sigma_1, \sigma_2) \leq n \cdot d_{\ell_\infty}(\sigma_1, \sigma_2), \quad (9)$$

and additionally, a scaled Chebyshev distance lower bounds the Spearman's footrule with high probability. More specifically, for any  $\pi \in \mathcal{S}_n$ , let  $\sigma$  be a permutation chosen uniformly from  $\mathcal{S}_n$ , then

$$\mathbb{P}[c_1 \cdot n \cdot d_{\ell_\infty}(\pi, \sigma) \leq d_{\ell_1}(\pi, \sigma)] \geq 1 - O(1/n) \quad (10)$$

for any positive constant  $c_1 < 1/3$  (See Appendix A-A for proof).

2) *Spearman's footrule and Kendall tau distance*: The following theorem is a well-known result on the relationship between the Kendall tau distance and the  $\ell_1$  distance of permutation vectors.

**Theorem 1** ([8]). *Let  $\sigma_1$  and  $\sigma_2$  be any permutations in  $\mathcal{S}_n$ , then*

$$d_{\ell_1}(\sigma_1, \sigma_2)/2 \leq d_\tau(\sigma_1^{-1}, \sigma_2^{-1}) \leq d_{\ell_1}(\sigma_1, \sigma_2). \quad (11)$$

3) *inversion- $\ell_1$  distance and Kendall tau distance*: We show that the inversion- $\ell_1$  distance and the Kendall tau distance are related via Theorem 2.

**Theorem 2.** *Let  $\sigma_1$  and  $\sigma_2$  be any permutations in  $\mathcal{S}_n$ , then for  $n \geq 2$ ,*

$$\frac{1}{n-1} d_\tau(\sigma_1, \sigma_2) \leq d_{\mathbf{x}, \ell_1}(\mathbf{x}_{\sigma_1}, \mathbf{x}_{\sigma_2}) \leq d_\tau(\sigma_1, \sigma_2). \quad (12)$$

*Proof:* See Appendix A-B. ■

**Remark 3.** *The lower and upper bounds in Theorem 2 are tight in the sense that there exist permutations  $\sigma_1$  and  $\sigma_2$  that satisfy the equality in either lower or upper bound. For equality in lower bound, when  $n = 2m$ , let  $\sigma_1 = [1, 3, 5, \dots, 2m-3, 2m-1, 2m, 2m-2, \dots, 6, 4, 2]$ ,  $\sigma_2 = [2, 4, 6, \dots, 2m-2, 2m, 2m-1, 2m-3, \dots, 5, 3, 1]$ , then  $d_\tau(\sigma_1, \sigma_2) = n(n-1)/2$  and  $d_{\mathbf{x}, \ell_1}(\sigma_1, \sigma_2) = n/2$ , as  $\mathbf{x}_{\sigma_1} = [0, 0, 1, 1, 2, 2, \dots, m-2, m-2, m-1, m-1]$ ,  $\mathbf{x}_{\sigma_2} = [0, 1, 1, 2, 2, 3, \dots, m-2, m-1, m-1, m]$ . For equality in upper bound, note that  $d_\tau(\text{Id}, \sigma) = d_{\mathbf{x}, \ell_1}(\text{Id}, \sigma)$ .*

Theorem 2 shows that in general  $d_\tau(\sigma_1, \sigma_2)$  is not a good approximation to  $d_{\mathbf{x}, \ell_1}(\sigma_1, \sigma_2)$  due to the  $1/(n-1)$  factor. However, (13) shows that Kendall tau distance scaled by a constant actually provides a lower bound to the inversion- $\ell_1$  distance with high probability. In particular, for any  $\pi \in \mathcal{S}_n$ , let  $\sigma$  be a permutation chosen uniformly from  $\mathcal{S}_n$ , then

$$\mathbb{P}[c_2 \cdot d_\tau(\pi, \sigma) \leq d_{\mathbf{x}, \ell_1}(\pi, \sigma)] \geq 1 - O(1/n) \quad (13)$$

for any positive constant  $c_2 < 1/2$  (See Appendix A-C for proof).

TABLE I  
CHARACTERIZATION OF MAXIMUM, MEAN AND VARIANCE OF  
VARIOUS DISTANCES.

	Max	Mean	Variance
$n \cdot \ell_\infty$	$n(n-1)$	$< n^2$	$\Theta(n^3)$
$\ell_1$	$\lfloor n^2/2 \rfloor$	$n^2/3 + o(n^2)$	$2n^3/45 + o(n^3)$
Kendall-tau	$n(n-1)/2$	$n^2/4 + o(n^2)$	$n^3/36 + o(n^3)$
inversion- $\ell_1$	$n(n-1)/2$	$> n^2/8$	$< n^3/6$

Results in both (10) and (13) are concentration results in the sense that the mean for distances are  $\Theta(n^2)$  and the standard deviation for the distances are  $\Theta(n^{3/2})$ . Related quantities are summarized in Table I, where results on  $\ell_1$  distance and Kendall tau distance are from [8, Table 1], and results on  $\ell_\infty$  distance and inversion- $\ell_1$  distance are derived in Appendix A-A and Appendix A-C. Therefore, these distance are concentrated around mean and separated probabilistically.

**Remark 4.** *The constants in (10) and (13) may be improved if both of the permutations in question are chosen randomly, instead of one being fixed. However as the techniques are exactly same, we refrain from providing those expressions.*

#### IV. TRADE-OFFS BETWEEN RATE AND DISTORTION

In this section we present some of the main results of this paper—the trade-offs between rate and distortion in permutation spaces. Throughout this section we assume the permutations are uniformly distributed over  $\mathcal{S}_n$ .

We first present Theorem 3, which shows how a lossy source code under one distortion measure implies a lossy source code under another distortion measure. Building on these relationships, Theorem 4 shows that all distortion measures in this paper essentially share the same rate-distortion function. Last, in Section IV-B, we present results on the trade-off between rate and distortion for  $\mathcal{X}(\mathcal{S}_n, d_\tau)$  and  $\mathcal{X}(\mathcal{S}_n, d_{x, \ell_1})$  when the distortion leads to degenerate rates  $R(D_n) = 0$  and  $R(D_n) = 1$ .

##### A. Rate-distortion functions

**Theorem 3** (Relationships of lossy source codes). *For both worst-case distortion and average-case distortion with uniform distribution, a following source code on the left hand side implies a source code on the right hand side:*

- 1)  $(n, D_n/n)$  source code for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty}) \Rightarrow (n, D_n)$  source code for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$ ,
- 2)  $(n, D_n)$  source code for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1}) \Rightarrow (n, D_n)$  source code for  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ ,
- 3)  $(n, D_n)$  source code for  $\mathcal{X}(\mathcal{S}_n, d_\tau) \Rightarrow (n, 2D_n)$  source code for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$ ,

- 4)  $(n, D_n)$  source code for  $\mathcal{X}(\mathcal{S}_n, d_\tau) \Rightarrow (n, D_n)$  source code for  $\mathcal{X}(\mathcal{S}_n, d_{x, \ell_1})$ .

The relationship between source codes is summarized in Fig. 1.

**Remark 5** (Non-equivalence of lossy source codes for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$  and  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$ ). *It is worth noting that in general, an  $(n, D_n)$  source code for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$  does not imply an  $(n, D_n/(nc_1) + O(1))$  source code for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$  in spite of the relationship shown in (8), even under the average-case distortion. This is exemplified in Example 4 below.*

In [26], it was shown incorrectly that lossy source codes for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$  and  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$  are equivalent, leading to an over-generalized version of Theorem 3.

**Example 4.** *When  $n = km$  and  $m = n^\delta$ , we define the following  $k$  sets with size  $m$*

$$\begin{aligned} \mathcal{I}_1 &= \{2, 3, \dots, m, m+1\}, \\ \mathcal{I}_2 &= \{m+2, m+3, \dots, 2m, 2m+1\}, \\ &\dots \\ \mathcal{I}_j &= \{(j-1)m+2, (j-1)m+3, \dots, jm, jm+1\}, \\ &\dots \\ \mathcal{I}_k &= \{(k-1)m+2, (k-1)m+3, \dots, n, 1\} \end{aligned}$$

and construct the following  $k$  subsequences for any permutation  $\sigma \in \mathcal{S}_n$ :

$$\mathbf{s}_j(\sigma) = [\sigma(j_1), \sigma(j_2), \dots, \sigma(j_m)], 1 \leq j \leq k$$

where for each  $j$ ,  $j_p \in \mathcal{I}_j$  for any  $1 \leq p \leq m$  and  $j_1 < j_2 < \dots < j_m$ .

Given any permutation  $\sigma$ , we can encode it as  $\hat{\sigma}$  by sorting each of its subsequences  $\mathbf{s}_j(\sigma)$ ,  $1 \leq j \leq k$ . Then the overall distortion  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$  satisfies

$$\begin{aligned} D_{\ell_1} &\leq (k-1)m^2/2 + [(m-1)^2 + 2n] = O(km^2/2) \\ &= O(n^{1+\delta}). \end{aligned}$$

Therefore, this source code is an  $(n, D_n)$  source code for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$ . However, for any  $\sigma$  in  $\mathcal{S}_n$ , if  $\sigma(1) \neq 1$ ,

$$d_{\ell_\infty}(\sigma, \hat{\sigma}) \geq (k-1)m + 2 - 1 \geq (k-1)m = \Theta(n).$$

Hence this encoding achieves average distortion  $\Theta(n)$  in  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$ . Therefore, while this code is  $D_n$  for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$ , it is not  $D_n/n$  for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$ .

Similarly, one can find a code that achieves distortion  $O(n^{1+\delta})$  for  $\mathcal{X}(\mathcal{S}_n, d_{x, \ell_1})$  but not  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ .

The proof of Theorem 3 is based on the relationships between various distortion measures investigated in Section III and we defer the proof details in Appendix B-A.

Below shows that, for the uniform distribution on  $\mathcal{S}_n$ , the rate-distortion function is the same for both

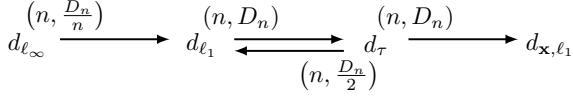


Fig. 1. Relationship between source codes. An arrow indicates a source code in one space implies a source in another space.

average- and worst-case, apart from the terms that are asymptotically negligible.

**Theorem 4** (Rate-distortion functions). *For permutation spaces  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ ,  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ , and  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$ ,*

$$\begin{aligned} \bar{R}(D_n) &= \hat{R}(D_n) \\ &= \begin{cases} 1 & \text{if } D_n = O(n), \\ 1 - \delta & \text{if } D_n = \Theta(n^{1+\delta}), \quad 0 < \delta \leq 1. \end{cases} \end{aligned}$$

For the permutation space  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$ ,

$$\begin{aligned} \bar{R}(D_n) &= \hat{R}(D_n) \\ &= \begin{cases} 1 & \text{if } D_n = O(1), \\ 1 - \delta & \text{if } D_n = \Theta(n^\delta), \quad 0 < \delta \leq 1. \end{cases} \end{aligned} \quad (14)$$

The rate-distortion functions for all these spaces are summarized in Fig. 2.

*Proof sketch:* The achievability comes from the compression schemes<sup>3</sup> proposed in Section V. The average-case converse for  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$  can be shown via the geometry of permutation spaces in Appendix A. Then because a  $D$ -ball in  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$  has the largest volume (cf. (7)), a converse for other permutation spaces can be inferred.

The rest of the proof follows from the simple fact that an achievability scheme for the worst-case distortion is also an achievability scheme for the average-case distortion, and a converse for the average-case distortion is also a converse for the worst-case distortion.

We present the detailed proof in Appendix B-B. ■

Because the rate-distortion functions under average-case and worst-case distortion coincide, if we require

$$\lim_{n \rightarrow \infty} \mathbb{P}[d(f_n(\sigma), \sigma) > D_n] = 0 \quad (15)$$

instead of  $\mathbb{E}[d(f_n(\sigma), \sigma)] \leq D_n$  in Definition 6, then the asymptotic rate-distortion trade-off remains the same.

Given the number of elements  $n$  and a distortion level  $D$ , we can compute the number of bits needed by first computing  $\delta$  via the asymptotic relationship  $\log D / \log n - 1$  (for permutation spaces  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ ,  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ , and  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$ ) or  $\log D / \log n$  (for permutation space  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$ ), then obtain the number of bits needed via  $(1 - \delta)n \log_2 n$ .

<sup>3</sup>Achievability results can also follow from simple random choice construction of covering codes, which are quite standard [27]. Instead we provide explicit constructions.

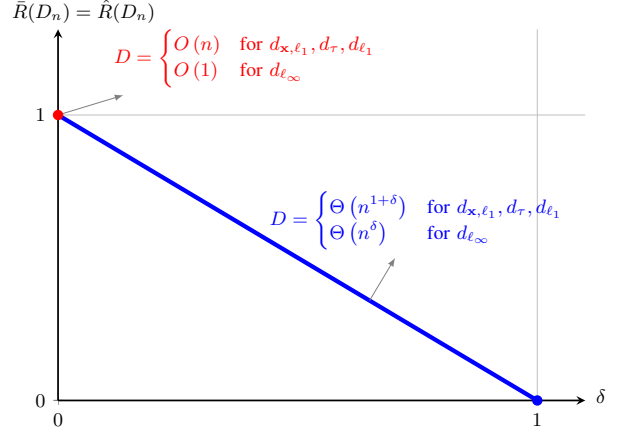


Fig. 2. Rate-distortion function for permutation spaces  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ ,  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ ,  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$ , and  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$ .

### B. Higher order term analysis

As mentioned in Section II, for small- and large-distortion regimes it is of interest to understand the trade-off between rate and distortion via the higher order term defined in (6). In this section we present the analysis for both regimes in permutation spaces  $\mathcal{X}(\mathcal{S}_n, d_\tau)$  and  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ .

**Theorem 5.** *In the permutation space  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ , when  $D_n = an^\delta, 0 < \delta \leq 1$ , for the worst-case distortion,  $\underline{r}_\tau^s(D_n) \leq r(D_n) \leq \overline{r}_\tau^s(D_n)$ , where*

$$\overline{r}_\tau^s(D_n) = \begin{cases} -a(1 - \delta)n^\delta \log n + O(n^\delta), & 0 < \delta < 1 \\ -n \left[ \log \frac{(1+a)^{1+a}}{a^a} \right] + o(n), & \delta = 1 \end{cases}, \quad (16)$$

$$\underline{r}_\tau^s(D_n) = \begin{cases} -n^\delta \frac{a \log 2}{2} + O(1), & 0 < a < 1 \\ -n^\delta \frac{\log [2a]!}{[2a]} + O(1), & a \geq 1 \end{cases}. \quad (17)$$

When  $D_n = bn^2, 0 < b \leq 1/2$ ,  $\underline{r}_\tau^1(D_n) \leq r(D_n) \leq \overline{r}_\tau^1(D_n)$ , where

$$\overline{r}_\tau^1(D_n) = \max \{0, n \log 1/(2be^2)\}, \quad (18)$$

$$\underline{r}_\tau^1(D_n) = n \log [1/(2b)] + O(\log n). \quad (19)$$

**Remark 6.** *Some of the results above for  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ , since their first appearances in the conference version [28], have been improved subsequently by [29]. More specifically, for the small distortion regime, [29, Lemma 7, Lemma 10] provides an improved upper bound and show that  $\underline{r}_\tau^s(D_n) = \overline{r}_\tau^s(D_n)$  in (16). For the large distortion regime, [29, Lemma 11] shows a lower bound that is tighter than (18).*

**Theorem 6.** *In the permutation space  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ , when  $D_n = an^\delta, 0 < \delta \leq 1$ ,*

$$\underline{r}_{\mathbf{x}, \ell_1}^s(D_n) \leq r(D_n) \leq \overline{r}_{\mathbf{x}, \ell_1}^s(D_n),$$



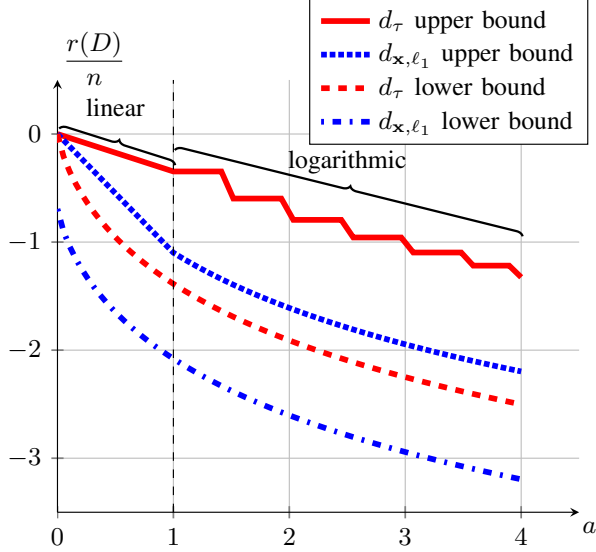


Fig. 3. Higher-order trade-off between rate and distortion in the small distortion regime with  $D = an$ . The slope discontinuities of the  $d_\tau$  upper bound in the range of  $a \geq 1$  is due to the flooring in (17).

where  $\overline{r_{\mathbf{x}, \ell_1}^s}(D_n) = \overline{r_\tau^s}(D_n) - n^\delta \log 2$  (cf. (16)) and

$$\overline{r_{\mathbf{x}, \ell_1}(D_n)} = \begin{cases} -\lfloor n^\delta \rfloor \log(2a - 1) & a > 1 \\ -\lfloor an^\delta \rfloor \log 3 & 0 < a \leq 1 \end{cases}.$$

When  $D_n = bn^2$ ,  $0 < b \leq 1/2$ ,

$$\overline{r_{\mathbf{x}, \ell_1}^1}(D_n) \leq r(D_n) \leq \overline{r_{\mathbf{x}, \ell_1}^1}(D_n),$$

where  $\overline{r_{\mathbf{x}, \ell_1}^1}(D_n) = \overline{r_\tau^1}(D_n)$  (cf. (18)) and  $\overline{r_{\mathbf{x}, \ell_1}^1}(D_n) = n \log \lceil 1/(4b) \rceil + O(1)$ .

*Proof for Theorem 5 and Theorem 6:* The achievability is presented in Section V-D and Section V-E. For converse, note that for a distortion measure  $d$ ,

$$|\mathcal{C}_n| N_d(D_n) \geq n!,$$

where  $N_d(D_n)$  is the maximum size of balls with radius  $D_n$  in the corresponding permutation space  $\mathcal{X}(\mathcal{S}_n, d)$  (cf. Appendix A for definitions), then a lower bound on  $|\mathcal{C}_n|$  follows from the upper bound on  $N_d(D_n)$  in Lemma 15 and Lemma 17. We omit the details as it is analogous to the proof of Theorem 4. ■

The bounds to  $r(D_n)$  of both Kendall tau distance and inversion- $\ell_1$  distance in both small and large distortion regimes are shown in Fig. 3 and Fig. 4.

## V. COMPRESSION SCHEMES

Though the permutation space has a complicated structure, in this section we show two rather straightforward compression schemes, *sorting subsequences* and *component-wise scalar quantization*, which are optimal

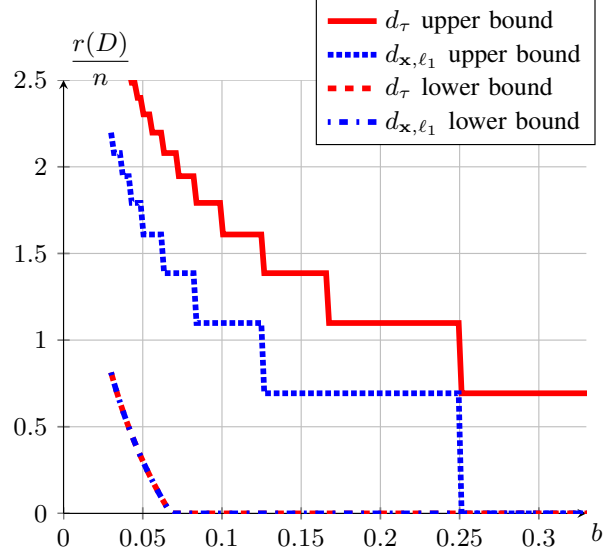


Fig. 4. Higher-order trade-off between rate and distortion in the large distortion regime with  $D = bn^2$ . The lower bounds for  $d_\tau$  and  $d_{\mathbf{x}, \ell_1}$  are identical.

as they achieve the rate-distortion functions in Theorem 4. We first describe these two key compression schemes in Section V-A and Section V-B respectively. Then in Sections V-C to V-E, we show that by simply applying these schemes with proper parameters, we can achieve the corresponding trade-offs between rate and distortion shown in Section IV.

The equivalence relationships in Theorem 3 suggest these two compression schemes achieve the same asymptotic performance. In addition, it is not hard to see that in general sorting subsequences has higher time complexity (e.g.,  $O(n \log n)$  for moderate distortion regime) than the time complexity of component-wise scalar quantization (e.g.,  $O(n)$  for moderate distortion regime). However, these two compression schemes operate on the permutation domain and the inversion vector of permutation domain respectively, and the time complexity to convert a permutation from its vector representation to its inversion vector representation is  $\Theta(n \log n)$  [24, Exercise 6 in Section 5.1.1]. Therefore, the cost of transforming a permutation between different representations should be taken into account when selecting the compression scheme.

### A. Quantization by sorting subsequences

In this section we describe the basic building block for lossy source coding in permutation space  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$ ,  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$  and  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ : sorting subsequences, either of the given permutation  $\sigma$  or of its inverse  $\sigma^{-1}$ . This operation reduces the number of possible permutations and thus the code rate, but introduces distortion.

By choosing the proper number of subsequences with proper lengths, we can achieve the corresponding rate-distortion function.

More specifically, we consider a code obtained by the sorting the first  $k$  subsequences with length  $m$ ,  $2 \leq m \leq n$ ,  $km \leq n$ :

$$\mathcal{C}(k, m, n) \triangleq \{f_{k,m}(\sigma) : \sigma \in \mathcal{S}_n\}$$

where  $\sigma' = f_{k,m}(\sigma)$  satisfies

$$\begin{aligned} \sigma'[im + 1 : (i+1)m] \\ &= \text{sort}(\sigma[im + 1 : (i+1)m]), \quad 0 \leq i \leq k, \\ \sigma'(j) &= \sigma(j), \quad j > km. \end{aligned}$$

This procedure is illustrated in Fig. 5.

Then  $|\mathcal{C}(k, m, n)| = n!/(m!^k)$ , and we define the (log) size reduction as

$$\begin{aligned} \Delta(k, m) &\triangleq \log \frac{n!}{|\mathcal{C}(k, m, n)|} = k \log m! \\ &\stackrel{(a)}{=} k \left[ m \log(m/e) + \frac{1}{2} \log m + O\left(\frac{1}{m}\right) \right], \end{aligned}$$

where (a) follows from Stirling's approximation in (1). Therefore,

$$\Delta(k, m) = \begin{cases} km \log m + o(km \log m) & m = \Omega(1) \\ k \log m! & m = \Theta(1) \end{cases}.$$

We first calculate the worst-case and average-case distortions for permutation space  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ :

$$\hat{D}_{d_\tau}(k, m) = k \frac{m(m-1)}{2} \leq km^2/2 \quad (20)$$

$$\bar{D}_{d_\tau}(k, m) = k \frac{m(m-1)}{4} \leq km^2/4 \quad (21)$$

where (20) is from (38).

**Remark 7.** Due to the close relationship between the Kendall tau distance and the Spearman's footrule shown in (11), the following codebook via the inverse permutations  $\{\sigma^{-1}\}$  is an equivalent construction to the codebook for Kendall tau distance above.

- 1) Construct a vector  $a(\sigma)$  such that for  $1 \leq i \leq k$ ,

$$a(i) = j \text{ if } \sigma^{-1}(i) \in [(j-1)m+1, jm], 1 \leq j \leq k.$$

Then  $a$  contains exactly  $m$  values of integers  $j$ .

- 2) Form a permutation  $\pi'$  by replacing the length- $m$  subsequence of  $a$  that corresponds to value  $j$  by vector  $[(j-1)m+1, (j-1)m+2, \dots, jm]$ .

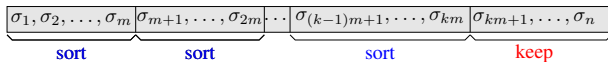


Fig. 5. Quantization by sorting subsequences.

It is not hard to see that the set of  $\{\pi'^{-1}\}$  forms a codebook with the same size with distortion in Kendall tau distance upper bounded by  $km^2/2$ .

Similarly, for permutation space  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$  and  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$ , we consider sorting subsequences in the inverse permutation domain, where

$$\mathcal{C}'(k, m, n) \triangleq \{\pi^{-1} : \pi = f_{k,m}(\sigma^{-1}), \sigma \in \mathcal{S}_n\}.$$

It is straightforward that  $\mathcal{C}'(k, m, n)$  has the same cardinality as  $\mathcal{C}(k, m, n)$  and hence code rate reduction  $\Delta(k, m)$ . And the worst-case and average-case distortions satisfy

$$\hat{D}_{\ell_\infty}(k, m) = m - 1 \quad (22)$$

$$\bar{D}_{\ell_\infty}(k, m) \leq m - 1 \quad (23)$$

$$\hat{D}_{\ell_1}(k, m) = k \lfloor m^2 \rfloor / 2 \leq km^2/2 \quad (24)$$

$$\bar{D}_{\ell_1}(k, m) = k(m^2 - 1)/3, \quad (25)$$

where (24) comes from Table I and (25) comes from (37).

### B. Component-wise scalar quantization

To compress in the space of  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ , component-wise scalar quantization suffices, due to the product structure of the inversion vector space  $\mathcal{G}_n$ .

More specifically, to quantize the  $k$  points in  $[0 : k-1]$ , where  $k = 2, \dots, n$ , with  $m$  uniformly spaced points, the maximal distortion is

$$\hat{D}_{\mathbf{x}, \ell_1}(k, m) = \lceil (k/m - 1) / 2 \rceil, \quad (26)$$

Conversely, to achieve distortion  $\hat{D}_{\mathbf{x}, \ell_1}$  on  $[0 : k-1]$ , we need

$$m = \left\lceil k / \left( 2\hat{D}_{\mathbf{x}, \ell_1} + 1 \right) \right\rceil \quad (27)$$

points.

### C. Compression in the moderate distortion regime

In this section we provide compression schemes in the moderate distortion regime, where for any  $0 < \delta < 1$ ,  $D_n = \Theta(n^\delta)$  for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$  and  $D_n = \Theta(n^{1+\delta})$  for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$ ,  $\mathcal{X}(\mathcal{S}_n, d_\tau)$  and  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ . While Theorem 3 indicates a source code for  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$  can be transformed into source codes for other spaces under both average-case and worst-case distortions, we develop explicit compression schemes for each permutation spaces as the transformation of permutation representations incur additional computational complexity and hence may not be desirable.

1) *Permutation space*  $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$ : Given distortion  $D_n = \Theta(n^\delta)$ , we apply the sorting subsequences scheme in Section V-A and choose  $m = D_n + 1$ , which ensures the maximal distortion is no more than  $D_n$ , and  $k = \lfloor n/m \rfloor$ , which indicates

$$\begin{aligned} km &= \lfloor n/m \rfloor m = n + O(n^\delta) \\ \log m &= \delta \log n + o(1) \\ \Delta(k, m) &= km \log m + o(km \log m) \\ &= \delta n \log n + O(n). \end{aligned}$$

2) *Permutation spaces*  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$  and  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ : Given distortion  $D_n = \Theta(n^{1+\delta})$ , we apply the sorting subsequences scheme in Section V-A and choose

$$\begin{aligned} m &= (1/\alpha) \lfloor D_n/n \rfloor \leq D_n/(n\alpha) \\ k &= \lfloor n/m \rfloor, \end{aligned}$$

then

$$\begin{aligned} km &= n - |O(n^\delta)| \\ D &\leq \alpha km^2 \leq D_n \\ \Delta(k, m) &= \delta n \log n - n \log(\alpha e) + o(n), \end{aligned}$$

where the constant  $\alpha$  depends on the distortion measure and whether we are considering worst-case or average-case distortion, as shown in (20), (21), (24) and (25), and is summarized in Table II.

3) *Permutation space*  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ : Given distortion  $D_n = \Theta(n^{1+\delta})$ , we apply the component-wise scalar quantization scheme in Section V-B and choose the quantization error of the coordinate with range  $[0 : k-1]$   $D^{(k)}$  to be

$$D^{(k)} = \frac{kD}{(n+1)^2},$$

then

$$\begin{aligned} m_k &= \left\lceil k / \left( 2 + D^{(k)} + 1 \right) \right\rceil = \left\lceil \frac{k(n+1)^2}{2kD_n + (n+1)^2} \right\rceil \\ &\leq \left\lceil \frac{(n+1)^2}{2D_n} \right\rceil, \end{aligned}$$

and the overall distortion and the codebook size satisfy

$$\begin{aligned} D &= \sum_{k=2}^n \frac{(n-1)(n+2)}{(n+1)^2} D_n \leq D_n, \\ \log |\mathcal{C}_n| &= \sum_{k=2}^n \log m_k \leq n \log \left\lceil \frac{(n+2)^2}{2D_n} \right\rceil \\ &= (1-\delta)n \log n + O(n). \end{aligned}$$

TABLE II  
VALUES OF  $\alpha$  FOR DIFFERENT COMPRESSION SCENARIOS.

$\alpha$	average-case	worst-case
$\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$	1/3	1/2
$\mathcal{X}(\mathcal{S}_n, d_\tau)$	1/4	1/2

D. *Compression in the small distortion regime*

In this section we provide compression schemes in the small distortion regime for  $\mathcal{X}(\mathcal{S}_n, d_\tau)$  and  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ , where for any  $a > 0, 0 < \delta < 1, D_n = an^\delta$ .

1) *Permutation space*  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ : When  $a \geq 1$ , let  $m = \lfloor 2a \rfloor$  and  $k = \lfloor n^\delta/m \rfloor$ , then

$$\begin{aligned} \Delta(k, m) &= k \log m! \\ &\geq (n^\delta/m - 1) \log m! = \frac{\log \lfloor 2a \rfloor!}{\lfloor 2a \rfloor} n^\delta + O(1). \end{aligned}$$

And the worst-case distortion is upper bounded by

$$km^2/2 \leq \frac{n^\delta m}{2} \leq an^\delta = D_n.$$

When  $0 < a < 1$ , let  $m = 2$  and  $k = \lfloor D_n/2 \rfloor$ , then

$$\Delta(k, m) = k \log m! = \left\lfloor \frac{D_n}{2} \right\rfloor \log 2 = \frac{a \log 2}{2} n^\delta + O(1).$$

And the worst-case distortion is no more than  $km^2/2 \leq D_n$ .

2) *Permutation space*  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ : When  $a > 1$ , let

$$m_k = \begin{cases} k & k \leq n - \lfloor n^\delta \rfloor \\ \lceil k/(2a-1) \rceil & k > n - \lfloor n^\delta \rfloor \end{cases}, \quad k = 2, \dots, n$$

then the distortion  $D^{(k)}$  for each coordinate  $k$  satisfies

$$D^{(k)} \leq \begin{cases} a & k \leq \lfloor n^\delta \rfloor \\ 0 & k > \lfloor n^\delta \rfloor \end{cases}, \quad k = 2, 3, \dots, n,$$

and hence overall distortion is  $\sum_{k=2}^n D^{(k)} = (\lfloor n^\delta \rfloor)a \leq D_n$ . In addition, the codebook size

$$|\hat{\mathcal{C}}_n| = \prod_{k=2}^n m_k \leq (1/(2a-1))^{\lfloor n^\delta \rfloor} \prod_{k=2}^n k.$$

Therefore,  $\log |\hat{\mathcal{C}}_n| \leq \log n! - \lfloor n^\delta \rfloor \log(2a-1) + O(\log n)$ .

When  $a \leq 1$ , let

$$m_k = \begin{cases} \lceil k/3 \rceil & k < \lfloor D_n \rfloor \\ k & k \geq \lfloor D_n \rfloor \end{cases}, \quad k = 2, \dots, n$$

and apply uniform quantization on the coordinate  $k$  of the inversion vector with  $m_k$  points, Then the distortion  $D^{(k)}$  for each coordinate  $k$  satisfies

$$D^{(k)} \leq \begin{cases} 1 & k < \lfloor D_n \rfloor \\ 0 & k \geq \lfloor D_n \rfloor \end{cases}, \quad k = 2, 3, \dots, n,$$

and hence overall distortion is  $\sum_{k=2}^n D^{(k)} = \lfloor D_n \rfloor - 1 \leq D_n$ . In addition, the codebook size

$$\begin{aligned} |\hat{\mathcal{C}}_n| &= \prod_{k=2}^n m_k \leq \prod_{k=2}^{\lfloor D_n \rfloor - 1} (k+3)/3 \prod_{k=\lfloor D_n \rfloor}^n k \\ &= \frac{1}{3^{\lfloor D_n \rfloor - 1}} \lfloor D_n \rfloor (\lfloor D_n \rfloor + 1) (\lfloor D_n \rfloor + 2) \prod_{k=5}^{n-1} k. \end{aligned}$$

Therefore,  $\log |\hat{C}_n| \leq \log n! - \lceil an^\delta \rceil \log 3 + O(\log n)$ .

### E. Compression in the large distortion regime

In this section we provide compression schemes in the large distortion regime for  $\mathcal{X}(\mathcal{S}_n, d_\tau)$  and  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ , where for any  $0 < b < 1/2$ ,  $D_n = bn^2$ .

1) *Permutation space*  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ : Let  $k = \lceil 1/(2b) \rceil$  and  $m = \lfloor n/k \rfloor$ , then

$$\begin{aligned} \Delta(k, m) &= k \log m! \geq k \log(n/k - 1)! \\ &\geq k \lfloor n/k \rfloor \log(n/k) - n/k \log e + O(\log n) \\ &= n \log(n/e) - n \log \lceil 1/(2b) \rceil + O(\log n). \end{aligned}$$

Hence  $\hat{r}(D_n) = \log n! - \Delta(k, m) \leq \log \lceil 1/(2b) \rceil + O(\log n)$ . And the worst-case distortion is upper bounded by

$$km^2/2 \leq n^2/(2k) \leq n^2/(1/b) = bn^2.$$

2) *Permutation space*  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ : Let  $m_k = \lfloor k/(4b(k-1) + 1) \rfloor$ ,  $k = 2, \dots, n$ . The distortion  $D^{(k)}$  for each coordinate  $k$  satisfies

$$D^{(k)} = \left\lfloor \frac{1}{2} \left( \frac{k}{m} - 1 \right) \right\rfloor \leq \lceil 2b(k-1) \rceil, k = 2, 3, \dots, n,$$

and hence overall distortion  $\sum_{k=2}^n D^{(k)} \leq \sum_{k=2}^n 2b(k-1) + 1 \leq (b+1/n)n(n-1)$ . In addition, the codebook size

$$|\hat{C}_n| = \prod_{k=2}^n m_k \leq \prod_{k=2}^n \left\lfloor \frac{k-1}{4b(k-1)} \right\rfloor \leq \left\lfloor \frac{1}{4b} \right\rfloor^{n-1}.$$

Therefore,  $\log |\hat{C}_n| \leq n \log \lceil 1/(4b) \rceil + O(1)$ .

## VI. COMPRESSION OF PERMUTATION SPACE WITH MALLOW'S MODEL

In this section we depart from the uniform distribution assumption and investigate the compression of a permutation space with a non-uniform model—Mallows model [18], a model with a wide range of applications such as ranking, partial ranking, and even algorithm analysis (see [30, Section 2e] and the references therein). In the context of storing user ranking data, the Mallows model (or more generally, the mixture of Mallows model) captures the phenomenon that user rankings are often similar to each other. In the application of approximate sorting, the Mallows model may be used to model our prior knowledge that permutations that are similar to the reference permutation are more likely.

**Definition 9** (Mallows model). *We denote a Mallows model with reference permutation (mode)  $\pi$  and parameter  $q$  as  $\mathcal{M}(\pi, q)$ , where for each permutation  $\sigma \in \mathcal{S}_n$ ,*

$$\mathbb{P}[\sigma; \mathcal{M}(\pi, q)] = \frac{q^{d_\tau(\sigma, \pi)}}{Z_{q, \pi}},$$

where normalization  $Z_{q, \pi} = \sum_{\sigma \in \mathcal{S}_n} p^{d_\tau(\sigma, \pi)}$ . In particular, when the mode  $\pi = \text{Id}$ ,  $Z_q \triangleq Z_{q, \text{Id}} = [n]_q!$  [30, (2.9)], where  $[n]_q!$  is the  $q$ -factorial  $[n]_q! = [n]_q [n-1]_q \dots [1]_q$  and  $[n]_q$  is the  $q$ -number

$$[n]_q \triangleq \begin{cases} \frac{1-q^n}{1-q} & q \neq 1 \\ n & q = 1 \end{cases}.$$

As we shall see, the entropy of the permutation space with a Mallows model is in general  $\Theta(n)$ , implying lower storage space requirement and *potentially* lower query complexity for sorting. Since the Mallows model is specified via the Kendall tau distance, we use Kendall tau distance as the distortion measure, and focus our attention on the average-case distortion.

Noting the Kendall tau distance is right-invariant [22], for the purpose of compression, we can assume the mode  $\pi = \text{Id}$  without loss of generality, and denote the Mallows model by  $\mathcal{M}(q) \triangleq \mathcal{M}(\text{Id}, q)$ .

### A. Repeated insertion model

The Mallows model can be generated through a process named *repeated insertion model* (RIM), which is introduced in [31] and later applied in [21].

**Definition 10** (Repeated insertion model). *Given a reference permutation  $\pi \in \mathcal{S}_n$  and a set of insertion probabilities  $\{p_{i,j}, 1 \leq i \leq n, 1 \leq j \leq i\}$ , RIM generates a new output  $\sigma$  by repeated inserting  $\pi(i)$  before the  $j$ -th element in  $\sigma$  with probability  $p_{i,j}$  (when  $j = i$ , we append  $\pi(i)$  at the end of  $\sigma$ ).*

**Remark 8.** *Note that the insertion probabilities at step  $i$  is independent of the realizations of earlier insertions.*

The  $i$ -th step in the RIM process involves sampling from a multinomial distribution with parameter  $p_{i,j}, 1 \leq j \leq i$ . If we denote the sampling outcome at the  $i$ -th step of the RIM process by  $a_i, 1 \leq i \leq n$ , then  $a_i$  indicates the location of insertion. By Definition 10, a vector  $\mathbf{a} = [a_1, a_2, \dots, a_n]$  has an one-one correspondence to a permutation, and we called this vector  $\mathbf{a}$  an *insertion vector*.

**Lemma 7.** *Given a RIM with reference permutation  $\pi = \text{Id}$  and insertion vector  $\mathbf{a}_\sigma$ , then the corresponding permutation  $\sigma$  satisfies*

$$\mathbf{a}_\sigma(i) = i - \tilde{\mathbf{x}}_\sigma(i),$$

where  $\tilde{\mathbf{x}}_\sigma$  is an extended inversion vector, which simply is an inversion vector  $\mathbf{x}_\sigma$  with 0 prepended.

$$\tilde{\mathbf{x}}_\sigma(i) = \begin{cases} 0 & i = 1 \\ \mathbf{x}_\sigma(i-1) & 2 \leq i \leq n \end{cases}$$

Therefore,

$$\begin{aligned} d_\tau(\sigma, \text{Id}) &= d_{\mathbf{x}, \ell_1}(\sigma, \text{Id}) \\ &= \sum_{i=1}^n (i - \mathbf{a}_\sigma(i)) = \binom{n+1}{2} - \sum_{i=1}^n \mathbf{a}_\sigma(i). \end{aligned}$$

**Example 5.** For  $n = 4$  and reference permutation  $\text{Id} = [1, 2, 3, 4]$ , if  $\mathbf{a} = [1, 1, 1, 1]$ , then  $\sigma = [4, 3, 2, 1]$ , which corresponds to  $\tilde{\mathbf{x}}_\sigma = [0, 1, 2, 3]$ .

**Theorem 8** (Mallows model via RIM [31], [21]). Given reference permutation  $\pi$  and

$$p_{i,j} = \frac{q^{i-j}}{1 + q + \dots + q^{i-1}}, 1 \leq j \leq i \leq n,$$

RIM induces the same distribution as the Mallows model  $\mathcal{M}(\pi, q)$ .

This observation allows us to convert compressing the Mallows model to a standard problem in source coding.

**Theorem 9.** Compressing a Mallows model is equivalent to compressing a vector source  $\mathbf{X} = [X_1, X_2, \dots, X_n]$ , where  $X_i$  is a geometric random variable truncated at  $i - 1$ ,  $1 \leq i \leq n$ , i.e.,

$$\begin{aligned} \mathbb{P}[X_i = j] &= \frac{q^j}{\sum_{j'=0}^{i-1} q^{j'}} \\ &= \frac{q^j(1-q)}{1-q^i}, 0 \leq j \leq i-1 \end{aligned}$$

*Proof:* This follows directly from Lemma 7 and Theorem 8. ■

### B. Lossless compression

We consider the lossless compression of Mallows model.

**Corollary 10.**

$$H(\mathcal{M}(q)) = H(\mathcal{M}(1/q))$$

*Proof:* This follows directly from Theorem 8. ■

**Lemma 11** (Entropy of Mallows model).

$$\begin{aligned} H(\mathcal{M}(q)) &= \sum_{k=1}^n H(X_k) \\ &= \begin{cases} \frac{H_b(q)}{1-q} n + g(n, q) & q \neq 1 \\ \log n! & q = 1 \end{cases}, \end{aligned}$$

where  $\{X_k\}$  are truncated geometric random variables defined in Theorem 9,  $H_b(\cdot)$  is the binary entropy function,  $g(n, q) = \Theta(1)$ , and  $\lim_{q \rightarrow 0} g(n, q) = 0$ .

The proof is presented in Appendix C-A. Fig. 6 shows plots of  $H(\mathcal{M}(q))$  for different values of  $n$  and  $q$ .

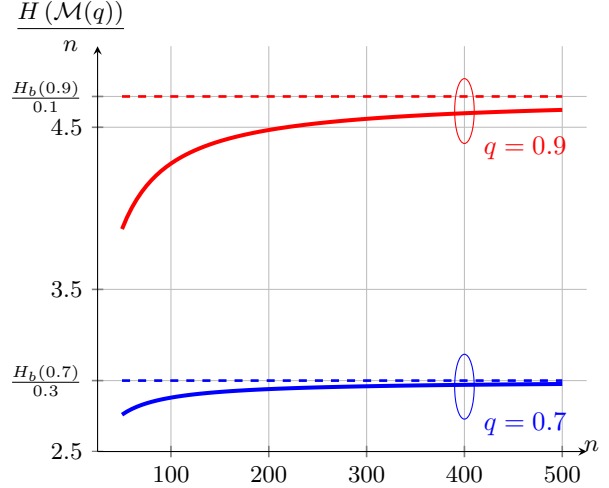


Fig. 6. Entropy of the Mallows model for  $q = 0.7$  and  $q = 0.9$ , where the dashed lines are the coefficients of the linear terms,  $H_b(q)/(1-q)$ .

**Remark 9.** Performing entropy-coding for each  $X_i$ ,  $1 \leq i \leq n$  is sub-optimal in general as the overhead is  $O(1)$  for each  $i$  and hence  $O(n)$  for  $\mathbf{X}$ , which is on the same order of the entropy  $H(\mathcal{M}(q))$  when  $q \neq 1$ .

### C. Lossy compression

By Theorem 9, the lossy compression of Mallows model is equivalent to the lossy compression of the independent non-identical source  $\mathbf{X}$ . However, it is unclear whether an analytical solution of the rate-distortion function for this source can be derived, and below we try to gain some insights via characterizing the typical set of the Mallows model in Lemma 12, which implies that at rate 0, the average-case distortion is  $\Theta(n)$ , while under the uniform distribution, Theorem 4 indicates that it takes  $n \log n + o(n \log n)$  bits to achieve average-case distortion of  $\Theta(n)$ .

**Lemma 12** (Typical set of Mallows model). There exists  $c_0(q)$ , a constant that depends on  $q$  only, such that for any  $r_0 \geq c_0(q)n$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}[d_\tau(\text{Id}, \sigma) \leq r_0; \mathcal{M}(\text{Id}, q)] = 1.$$

The proof is presented in Appendix C-B.

**Remark 10.** As pointed out in [31], Mallows model is only one specific distributional model that is induced by RIM. It is possible to generalize our analysis above to other distributional models that are also induced by RIM.

## VII. CONCLUDING REMARKS

In this paper, we first investigate the lossy compression of permutations under both worst-case distortion and

average-case distortions with uniform source distribution. We consider Kendall tau distance, Spearman's footrule, Chebyshev distance and inversion- $\ell_1$  distance as distortion measures. Regarding the lossy storage of ranking, our results provide the fundamental trade-off between storage and accuracy. Regarding approximate sorting, our results indicate that, given a moderate distortion  $D_n$  (see Section II for definition), an approximate sorting algorithm must perform *at least*  $\Theta(n \log n)$  pairwise comparisons, where constant implicitly in the  $\Theta$  term is exactly the rate-distortion function  $R(D_n)$ . As mentioned, this performance is indeed achieved by the multiple selection algorithm in [12]. This shows our information-theoretic lower bound for approximate sorting is tight.

In practical ranking systems where prior knowledge on the ranking is available, non-uniform model may be more appropriate. Our results on the Mallows model show that the entropy could be much lower ( $\Theta(n)$ ) than the uniform model ( $\Theta(n \log n)$ ). This greater compression ratio suggests that it would be worthwhile to solve the challenge of designing entropy-achieving compression schemes with low computational complexity for Mallows model. A deeper understanding on the rate-distortion trade-off of non-uniform models would be beneficial to the many areas that involves permutation model with a non-uniform distribution, such as the problem of learning to rank [21] and algorithm analysis [30].

#### ACKNOWLEDGMENT

The authors are grateful to an anonymous reviewer whose comment prompted an important correction to an earlier version of this paper.

#### APPENDIX A GEOMETRY OF PERMUTATION SPACES

In this section we provide results on the geometry of the permutation space that are useful in deriving rate-distortion bounds.

We first define  $D$ -balls centered at  $\sigma \in \mathcal{S}_n$  with radius  $D$  under distance  $d(\cdot, \cdot)$  and their maximum sizes:

$$B_d(\sigma, D) \triangleq \{\pi : d(\pi, \sigma) \leq D\}, \quad (28)$$

$$N_d(D) \triangleq \max_{\sigma \in \mathcal{S}_n} |B_d(\sigma, D)|. \quad (29)$$

Let  $B_\tau(\sigma, D)$ ,  $B_{\ell_1}(\sigma, D)$  and  $B_{\mathbf{x}, \ell_1}(\sigma, D)$  be the balls that correspond to the Kendall tau distance,  $\ell_1$  distance of the permutations, and  $\ell_1$  distance of the inversion vectors, and  $N_\tau(D)$ ,  $N_{\ell_1}(D)$ , and  $N_{\mathbf{x}, \ell_1}(D)$  be their maximum sizes respectively.

Note that (12) implies  $B_\tau(\sigma, D) \subset B_{\mathbf{x}, \ell_1}(\sigma, D)$  and thus  $N_\tau(D) \leq N_{\mathbf{x}, \ell_1}(D)$ . Below we establish upper bounds for  $N_{\mathbf{x}, \ell_1}(D)$  and  $N_\tau(D)$ , which are useful for establishing converse results later.

**Lemma 13.** For  $0 \leq D \leq n$ ,

$$N_\tau(D) \leq \binom{n+D-1}{D}. \quad (30)$$

*Proof:* Let the number of permutations in  $\mathcal{S}_n$  with at most  $k$  inversions be  $T_n(d) \triangleq \sum_{k=0}^d K_n(k)$ , where  $K_n(k)$  is defined in (2). Since  $\mathcal{X}(\mathcal{S}_n, d_\tau)$  is a regular metric space,

$$N_\tau(D) = T_n(D),$$

which is noted in several references such as [24]. An expression for  $K_n(D)$  (and thus  $T_n(D)$ ) for  $D \leq n$  appears in [24] (see [4] also). The following bound is weaker but sufficient in our context.

By induction, or [32],  $T_n(D) = K_{n+1}(D)$  when  $D \leq n$ . Then noting that for  $k < n$ ,  $K_n(k) = K_n(k-1) + K_{n-1}(k)$  [24, Section 5.1.1] and for any  $n \geq 2$ ,

$$K_n(0) = 1, \quad K_n(1) = n-1, \quad K_n(2) = \binom{n}{2} - 1,$$

by induction, we can show that when  $1 \leq k < n$ ,

$$K_n(k) \leq \binom{n+k-2}{k}. \quad (31)$$

The product structure of  $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$  leads to a simpler analysis of the upper bound of  $N_{\mathbf{x}, \ell_1}(D)$ . ■

**Lemma 14.** For  $0 \leq D \leq n(n-1)/2$ ,

$$N_{\mathbf{x}, \ell_1}(D) \leq 2^{\min\{n, D\}} \binom{n+D}{D}. \quad (32)$$

*Proof:* For any  $\sigma \in \mathcal{S}_n$ , let  $\mathbf{x} = \mathbf{x}_\sigma \in \mathcal{G}_n$ , then

$$|B_{\mathbf{x}, \ell_1}(D)| = \sum_{r=0}^D |\{\mathbf{y} \in \mathcal{G}_n : d_{\ell_1}(\mathbf{x}, \mathbf{y}) = r\}|.$$

Let  $\mathbf{d} \triangleq |\mathbf{x} - \mathbf{y}|$ , and  $Q(n, r)$  be the number of integer solutions of the equation  $z_1 + z_2 + \dots + z_n = r$  with  $z_i \geq 0$ ,  $0 \leq i \leq n$ , then it is well known [33, Section 1.2] that

$$Q(n, r) = \binom{n+r-1}{r},$$

and it is not hard to see that the number of such  $\mathbf{d} = [d_1, d_2, \dots, d_{n-1}]$  that satisfies  $\sum_{i=1}^{n-1} d_i = r$  is upper bounded by  $Q(n-1, r)$ . Given  $\mathbf{x}$  and  $\mathbf{d}$ , at most  $m \triangleq \min\{D, n\}$  elements in  $\{y_i, 0 \leq i \leq n\}$  correspond to  $y_i = x_i \pm d_i$ . Therefore, for any  $\mathbf{x}$ ,  $|\{\mathbf{y} \in \mathcal{G}_n : d_{\ell_1}(\mathbf{x}, \mathbf{y}) = r\}| \leq 2^m Q(n, r)$  and hence

$$|B_{\ell_1}(\mathbf{x}, D)| \leq \sum_{r=0}^D 2^m Q(n, r) = 2^m \binom{n+D}{D}. \quad \blacksquare$$

Below we upper bound  $\log N_\tau(D)$  and  $\log N_{\mathbf{x}, \ell_1}(D)$  for small, moderate and large  $D$  regimes in Lemmas 15 to 17 respectively.

**Lemma 15** (Small distortion regime). When  $D = an^\delta$ ,  $0 < \delta \leq 1$  and  $a > 0$  is a constant,

$$\begin{aligned} & \log N_\tau(D) \\ & \leq \begin{cases} a(1-\delta)n^\delta \log n + O(n^\delta), & 0 < \delta < 1 \\ n \left[ \log \frac{(1+a)^{1+a}}{a^a} \right] + o(n), & \delta = 1 \end{cases}, \end{aligned} \quad (33)$$

$$\begin{aligned} & \log N_{\mathbf{x}, \ell_1}(D) \\ & \leq \begin{cases} a(1-\delta)n^\delta \log n + O(n^\delta), & 0 < \delta < 1 \\ n \left[ 2 + \log \frac{(1+a)^{1+a}}{a^a} \right] + o(n), & \delta = 1 \end{cases}. \end{aligned} \quad (34)$$

*Proof:* To upper bound  $N_\tau(D)$ , when  $0 < \delta < 1$ , we apply Stirling's approximation to (30) to have

$$\begin{aligned} & \log \binom{n+D-1}{D} \\ & = n \log \frac{n-1+D}{n-1} + D \log \frac{n-1+D}{D} + O(\log n). \end{aligned}$$

Substituting  $D = an^\delta$ , we obtain (33). When  $\delta = 1$ , the result follows from (9) in [34, Section 4]. The upper bound on  $N_{\mathbf{x}, \ell_1}(D)$  can be obtained similarly via (32). ■

**Lemma 16** (Moderate distortion regime). Given  $D = \Theta(n^{1+\delta})$ ,  $0 < \delta \leq 1$ , then

$$\log N_\tau(D) \leq \log N_{\mathbf{x}, \ell_1}(D) \leq \delta n \log n + O(n). \quad (35)$$

*Proof:* Apply Stirling's approximation to (32) and substitute  $D = \Theta(n^{1+\delta})$ . ■

**Remark 11.** It is possible to obtain tighter lower and upper bounds for  $\log N_\tau(D)$  and  $\log N_{\mathbf{x}, \ell_1}(D)$  based on results in [4].

**Lemma 17** (Large distortion regime). Given  $D = bn(n-1) \in \mathbb{Z}^+$ , then

$$\log N_\tau(D) \leq \log N_{\mathbf{x}, \ell_1}(D) \leq n \log(2ben) + O(\log n). \quad (36)$$

*Proof:* Substitute  $D = bn(n-1)$  into (32). ■

A. Proof of (10)

**Lemma 18.** For any  $\pi \in \mathcal{S}_n$ , let  $\sigma$  be a permutation chosen uniformly from  $\mathcal{S}_n$ , and  $X_{\ell_1} \triangleq d_{\ell_1}(\pi, \sigma)$ , then

$$\mathbb{E}[X_{\ell_1}] = \frac{n^2-1}{3} \quad \text{Var}[X_{\ell_1}] = \frac{2n^3}{45} + O(n^2). \quad (37)$$

*Proof:*

$$\begin{aligned} \mathbb{E}[X_{\ell_1}] & = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n |i-j| = \frac{2}{n} \sum_{i=1}^n \sum_{j=1}^i |i-j| \\ & = \frac{2}{n} \sum_{i=1}^n \sum_{j'=0}^{i-1} j' = \frac{1}{n} \sum_{i=1}^n (i^2 - i) \\ & = \frac{1}{n} \left( \sum_{i=1}^n i^2 - \sum_{i=1}^n i \right) \\ & = \frac{1}{n} \left( \frac{2n^3 + 3n^2 + n}{6} - \frac{n^2 + n}{2} \right) \\ & = \frac{n^2 - 1}{3}. \end{aligned}$$

And  $\text{Var}[X_{\ell_1}]$  can be derived similarly [8, Table 1]. ■

*Proof for (10):* For any  $c > 0$ ,  $cn \cdot d_{\ell_\infty}(\pi, \sigma) \leq cn(n-1)$ , and for any  $c_1 < 1/3$ , (37) and Chebyshev inequality indicate  $\mathbb{P}[d_{\ell_1}(\pi, \sigma) < c_1 n(n-1)] = O(1/n)$ . Therefore,

$$\begin{aligned} & \mathbb{P}[d_{\ell_1}(\pi, \sigma) \geq c_1 n \cdot d_{\ell_\infty}(\pi, \sigma)] \\ & \geq \mathbb{P}[d_{\ell_1}(\pi, \sigma) \geq c_1 n(n-1)] \\ & = 1 - \mathbb{P}[d_{\ell_1}(\pi, \sigma) < c_1 n(n-1)] \\ & = 1 - O(1/n). \end{aligned}$$

B. Proof of Theorem 2

**Lemma 19.** For any two permutations  $\pi, \sigma$  in  $\mathcal{S}_n$  such that  $d_{\mathbf{x}, \ell_1}(\pi, \sigma) = 1$ ,  $d_\tau(\pi, \sigma) \leq n-1$ .

*Proof:* Let  $\mathbf{x}_\pi = [a_2, a_3, \dots, a_n]$  and  $\mathbf{x}_\sigma = [b_2, b_3, \dots, b_n]$ , then without loss of generality, we have for a certain  $2 \leq k \leq n$ ,

$$a_i = \begin{cases} b_i & i \neq k \\ b_i + 1 & i = k. \end{cases}$$

Let  $\pi'$  and  $\sigma'$  be permutations in  $\mathcal{S}_{n-1}$  with element  $k$  removed from  $\pi$  and  $\sigma$  correspondingly, then  $\mathbf{x}_{\pi'} = \mathbf{x}_{\sigma'}$ , and hence  $\pi' = \sigma'$ . Therefore, the Kendall tau distance between  $\sigma$  and  $\pi$  is determined only by the location of element  $k$  in  $\sigma$  and  $\pi$ , which is at most  $n-1$ . ■

*Proof of Theorem 2:* It is known that (see, e.g., [35, Lemma 4])

$$d_{\ell_1}(\mathbf{x}_{\pi_1}, \mathbf{x}_{\pi_2}) \leq d_\tau(\pi_1, \pi_2).$$

Furthermore, the proof of [35, Lemma 4] indicates that for any two permutation  $\pi_1$  and  $\pi_2$  with  $k = d_{\mathbf{x}, \ell_1}(\pi_1, \pi_2)$ , let  $\sigma_0 \triangleq \pi_1$  and  $\sigma_k \triangleq \pi_2$ , then there exists a sequence of permutations  $\sigma_1, \sigma_2, \dots, \sigma_{k-1}$  such

that  $d_{\mathbf{x}, \ell_1}(\sigma_i, \sigma_{i+1}) = 1, 0 \leq i \leq k-1$ . Then

$$\begin{aligned} d_\tau(\pi_1, \pi_2) &\leq \sum_{i=0}^{k-1} d_\tau(\sigma_i, \sigma_{i-1}) \\ &\stackrel{(a)}{\leq} \sum_{i=0}^{k-1} (n-1) = (n-1)d_{\mathbf{x}, \ell_1}(\pi_1, \pi_2), \end{aligned}$$

where (a) is due to Lemma 19.  $\blacksquare$

### C. Proof of (13)

To prove (13), we analyze the mean and variance of the Kendall tau distance and inversion- $\ell_1$  distance between a permutation in  $\mathcal{S}_n$  and a randomly selected permutation, in (39) and Lemma 21 respectively.

**Lemma 20.** *For any  $\pi \in \mathcal{S}_n$ , let  $\sigma$  be a permutation chosen uniformly from  $\mathcal{S}_n$ , and  $X_\tau \triangleq d_\tau(\pi, \sigma)$ , then*

$$\mathbb{E}[X_\tau] = \frac{n(n-1)}{4}, \quad (38)$$

$$\text{Var}[X_\tau] = \frac{n(2n+5)(n-1)}{72}. \quad (39)$$

*Proof:* Let  $\sigma'$  be another permutation chosen independently and uniformly from  $\mathcal{S}_n$ , then we have both  $\pi\sigma^{-1}$  and  $\sigma'\sigma^{-1}$  are uniformly distributed over  $\mathcal{S}_n$ .

Note that Kendall tau distance is right-invariant [22], then  $d_\tau(\pi, \sigma) = d_\tau(\pi\sigma^{-1}, \text{Id})$  and  $d_\tau(\sigma', \sigma) = d_\tau(\sigma'\sigma^{-1}, \text{Id})$  are identically distributed, and hence the result follows [8, Table 1] and [24, Section 5.1.1].  $\blacksquare$

**Lemma 21.** *For any  $\pi \in \mathcal{S}_n$ , let  $\sigma$  be a permutation chosen uniformly from  $\mathcal{S}_n$ , and  $X_{\mathbf{x}, \ell_1} \triangleq d_{\mathbf{x}, \ell_1}(\pi, \sigma)$ , then*

$$\begin{aligned} \mathbb{E}[X_{\mathbf{x}, \ell_1}] &> \frac{n(n-1)}{8}, \\ \text{Var}[X_{\mathbf{x}, \ell_1}] &< \frac{(n+1)(n+2)(2n+3)}{6}. \end{aligned}$$

*Proof:* It is not hard to see that when  $\sigma$  is a permutation chosen uniformly from  $\mathcal{S}_n$ ,  $\mathbf{x}_\sigma(i)$  is uniformly distributed in  $[0 : i], 1 \leq i \leq n-1$ . Therefore,  $X_{\mathbf{x}, \ell_1} = \sum_{i=1}^{n-1} |a_i - U_i|$ , where  $U_i \sim \text{Unif}([0 : i])$  and  $a_i \triangleq \mathbf{x}_\pi(i)$ . Let  $V_i = |a_i - U_i|$ ,  $m_1 = \min\{i - a_i, a_i\}$  and  $m_2 = \max\{i - a_i, a_i\}$ , then

$$\mathbb{P}[V_i = d] = \begin{cases} 1/(i+1) & d=0 \\ 2/(i+1) & 1 \leq d \leq m_1 \\ 1/(i+1) & m_1+1 \leq d \leq m_2 \\ 0 & \text{otherwise.} \end{cases}$$

Hence,

$$\begin{aligned} \mathbb{E}[V_i] &= \sum_{d=1}^{m_1} d \frac{2}{i+1} + \sum_{d=m_1+1}^{m_2} d \frac{1}{i+1} \\ &= \frac{2(1+m_1)m_1 + (m_2+m_1+1)(m_2-m_1)}{2(i+1)} \\ &= \frac{1}{2(i+1)}(m_1^2 + m_2^2 + i) \\ &\geq \frac{1}{2(i+1)} \left( \frac{(m_1+m_2)^2}{2} + i \right) = \frac{i(i+2)}{4(i+1)} > \frac{i}{4}, \\ \text{Var}[V_i] &\leq \mathbb{E}[V_i^2] \leq \frac{2}{i+1} \sum_{d=0}^i d^2 \leq (i+1)^2. \end{aligned}$$

Then,

$$\begin{aligned} \mathbb{E}[X_{\mathbf{x}, \ell_1}] &= \sum_{i=1}^{n-1} \mathbb{E}[V_i] > \frac{n(n-1)}{8}, \\ \text{Var}[X_{\mathbf{x}, \ell_1}] &= \sum_{i=1}^{n-1} \text{Var}[V_i] < \frac{(n+1)(n+2)(2n+3)}{6}. \end{aligned}$$

With (39) and Lemma 21, now we show that the event that a scaled version of the Kendall tau distance is larger than the inversion- $\ell_1$  distance is unlikely.

*Proof for (13):* Let  $c_2 = 1/3$ , let  $t = n^2/7$ , then noting

$$\begin{aligned} t &= \mathbb{E}[c \cdot X_\tau] + |\Theta(\sqrt{n})| \text{Std}[X_\tau] \\ &= \mathbb{E}[X_{\mathbf{x}, \ell_1}] - |\Theta(\sqrt{n})| \text{Std}[X_{\mathbf{x}, \ell_1}], \end{aligned}$$

by Chebyshev inequality,

$$\begin{aligned} \mathbb{P}[c \cdot X_\tau > X_{\mathbf{x}, \ell_1}] &\leq \mathbb{P}[c \cdot X_\tau > t] + \mathbb{P}[X_{\mathbf{x}, \ell_1} < t] \\ &\leq O(1/n) + O(1/n) = O(1/n). \end{aligned}$$

The general case of  $c_2 < 1/2$  can be proved similarly.  $\blacksquare$

## APPENDIX B

### PROOFS ON RATE-DISTORTION FUNCTIONS

#### A. Proof of Theorem 3

*Proof:* Statement 1 follows from (9).

Statement 2 and 3 follow from Theorem 1. For statement 2, let the encoding mapping for the  $(n, D_n)$  source code in  $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$  be  $f_n$  and the encoding mapping in  $\mathcal{X}(\mathcal{S}_n, d_\tau)$  be  $g_n$ , then

$$g_n(\pi) = [f_n(\pi^{-1})]^{-1}$$

is a  $(n, D_n)$  source code in  $\mathcal{X}(\mathcal{S}_n, d_\tau)$ . The proof for Statement 3 is similar.

Statement 4 follow directly from (12).  $\blacksquare$



## B. Proof of Theorem 4

We prove Theorem 4 by achievability and converse.

1) *Achievability*: The achievability for all permutation spaces of interest under both *worst-case distortion* and *average-case distortion* are established via the explicit code constructions in Section V.

2) *Converse*: For the converse, we show by contradiction that under average-case distortion, if the rate is less than  $1 - \delta$ , then the average distortion is larger than  $D_n$ . Therefore,  $\bar{R} \geq 1 - \delta$ , and hence  $\hat{R} \geq \bar{R} \geq 1 - \delta$ .

When  $\delta = 1$ ,  $\bar{R} = \hat{R} = 0$ . When  $0 \leq \delta < 1$ , for any  $0 < \varepsilon < 1 - \delta$  and any codebook  $\bar{\mathcal{C}}_n$  with size such that

$$\log |\bar{\mathcal{C}}_n| = (1 - \delta - \varepsilon)n \log n + O(n), \quad (40)$$

from (7), when  $D_n = \Theta(n^{1+\delta})$  or  $D_n = O(n)$ ,

$$\begin{aligned} N_{\ell_1}(2D_n) |\bar{\mathcal{C}}_n| &\leq N_{\tau}(2D_n) |\bar{\mathcal{C}}_n| \leq \\ N_{\mathbf{x}, \ell_1}(2D_n) |\bar{\mathcal{C}}_n| &\stackrel{(a)}{\leq} n!/2; \end{aligned}$$

when  $D_n = \Theta(n^\delta)$  or  $D_n = O(1)$ ,

$$N_{\ell_\infty}(2D_n) |\bar{\mathcal{C}}_n| \leq N_{\ell_1}(2D_n n) |\bar{\mathcal{C}}_n| \leq n!/2$$

when  $n$  sufficiently large, where (a) follows from (35).

Therefore, given  $\bar{\mathcal{C}}_n$ , there exists at least  $n!/2$  permutations in  $\mathcal{S}_n$  that has distortion larger than  $2D_n$ , and hence the average distortion w.r.t. uniform distribution over  $\mathcal{S}_n$  is larger than  $D_n$ .

Therefore, for any codebook with size indicated in (40), we have average distortion larger than  $D_n$ . Therefore, any  $(n, D_n)$  code must satisfy  $\hat{R} \geq \bar{R} \geq 1 - \delta$ .

## APPENDIX C

### PROOFS ON MALLOWS MODEL

#### A. Proof of Lemma 11

*Proof*: When  $q = 1$  the Mallows model reduces to the uniform distribution on the permutation space. When  $q \neq 1$ , let  $X^n = [X_1, X_2, \dots, X_n]$  be the inversion vector, and denote a geometric random variable by  $G$  and a geometric random variable truncated at  $k$  by  $G_k$ . Define

$$E_k = \begin{cases} 0 & G \leq k \\ 1 & \text{o.w.} \end{cases},$$

then  $\mathbb{P}[E_k = 0] = Q_k = 1 - q^{k+1}$ . Note

$$\begin{aligned} H(G_k, E) &= H(G|E_k) + H(E_k) \\ &= H(E_k|G) + H(G) \\ &= H(G) \end{aligned}$$

and

$$\begin{aligned} H(G|E_k) &= H(G|E_k = 0) Q_k \\ &\quad + H(G|E_k = 1) (1 - Q_k) \\ &= H(G_k) Q_k + H(G) (1 - Q_k), \end{aligned}$$

we have

$$H(G_k) = H_b(q)/(1-q) - H_b(Q_k)/Q_k.$$

Then

$$\begin{aligned} H(\mathcal{M}(q)) &= \sum_{k=0}^{n-1} H(G_k) \\ &= \frac{nH_b(q)}{1-q} - \sum_{k=1}^n \frac{H_b(q^k)}{1-q^k}. \end{aligned}$$

It can be shown via algebraic manipulations that

$$\sum_{k=1}^n H_b(q^k) \leq \frac{2q - q^2}{(1-q)^2} = \Theta(1),$$

therefore

$$H(\mathcal{M}(q)) = \frac{nH_b(q)}{1-q} - \Theta(1). \quad \blacksquare$$

#### B. Proof of Lemma 12

We first show an upper bound  $K_n(k)$  (cf. (2) for definition), the number of permutations with  $k$  inversion in  $\mathcal{S}_n$ .

**Lemma 22** (Bounds on  $K_n(k)$ ). *For  $k = cn$ ,*

$$K_n(k) \leq \frac{1}{\sqrt{2\pi nc/(1+c)}} 2^{n(1+c)H_b(1/(1+c))}.$$

*Proof*: By definition,  $K_n(k)$  equals to the number of non-negative integer solutions of the equation  $z_1 + z_2 + \dots + z_{n-1} = k$  with  $0 \leq z_i \leq i, 1 \leq i \leq n-1$ . Then similar to the derivations in the proof of Lemma 14,

$$K_n(k) < Q(n-1, k) = \binom{n+k-2}{k}.$$

Finally, applying the bound [27]

$$\binom{n}{pn} \leq \frac{2^{nH_b(p)}}{\sqrt{2\pi np(1-p)}}$$

completes the proof.  $\blacksquare$

*Proof of Lemma 12*: Note

$$d_\tau(\sigma, \text{Id}) = d_{\mathbf{x}, \ell_1}(\sigma, \mathbf{0}).$$

Therefore,

$$\sum_{\sigma \in \mathcal{S}_n, d_\tau(\sigma, \text{Id}) \geq r_0} \mathbb{P}[\sigma] = \frac{1}{Z_q} \sum_{r=r_0}^{\binom{n}{2}} q^r K_n(r).$$

And Lemma 22 indicates for any  $r = cn$ ,

$$q^r K_n(r) \leq \frac{2^{n[(1+c)H_b(\frac{1}{1+c}) - c \log_2 \frac{1}{q}]}}{\sqrt{2\pi nc/(1+c)}}.$$

Define

$$E(c, q) \triangleq \left[ (1+c)H_b\left(\frac{1}{1+c}\right) - c \log_2 \frac{1}{q} \right],$$

then for any  $\varepsilon > 0$ , there exists  $c_0$  such that for any  $c \geq c_0(q)$ ,  $E(c, q) < -\varepsilon$ . Therefore, let  $r_0 \geq c_0 n$ ,

$$\sum_{\sigma \in \mathcal{S}_n, d_r(\sigma, \text{Id}) \geq r_0} \mathbb{P}[\sigma] \leq \frac{1}{\sqrt{2\pi n c / (1+c)}} \frac{1}{Z_q} \sum_{r=r_0}^{\binom{n}{2}} 2^{-n\varepsilon} \rightarrow 0$$

as  $n \rightarrow \infty$ . ■

## REFERENCES

- [1] I. Blake, G. Cohen, and M. Deza, "Coding with permutations," *Information and Control*, vol. 43, no. 1, pp. 1–19, 1979.
- [2] C. Colbourn, T. Klove, and A. Ling, "Permutation arrays for powerline communication and mutually orthogonal latin squares," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1289–1291, 2004.
- [3] T. Klove, T. Lin, S. Tsai, and W. Tzeng, "Permutation arrays under the Chebyshev distance," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2611–2617, 2010.
- [4] A. Barg and A. Mazumdar, "Codes in permutations and error correction for rank modulation," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3158–3165, 2010.
- [5] J. Barbay and G. Navarro, "Compressed representations of permutations, and applications," in *Proc. 26th International Symposium on Theoretical Aspects of Computer Science (STACS)*, 2009, pp. 111–122.
- [6] J. Barbay, J. Fischer, and G. Navarro, "LRM-Trees: compressed indices, adaptive sorting, and compressed permutations," *Theoretical Computer Science*, vol. 459, no. 0, pp. 26–41, 2012.
- [7] J. Giesen, E. Schuberth, and M. Stojakovi, "Approximate sorting," *Fundamenta Informaticae*, vol. 90, no. 1, pp. 67–72, 2009.
- [8] P. Diaconis and R. L. Graham, "Spearman's footrule as a measure of disarray," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 2, pp. 262–268, 1977.
- [9] J. M. Chambers, "Algorithm 410: Partial sorting," *Commun. ACM*, vol. 14, no. 5, pp. 357–358, 1971.
- [10] D. Wang, "Computing with unreliable resources: design, analysis and algorithms," PhD Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2014.
- [11] M. Fredman, "How good is the information theory bound in sorting?" *Theoretical Computer Science*, vol. 1, no. 4, pp. 355–361, 1976.
- [12] K. Kaligosi, K. Mehlhorn, J. I. Munro, and P. Sanders, "Towards optimal multiple selection," in *Automata, Languages and Programming*, ser. Lecture Notes in Computer Science, Luís Caires, G. F. Italiano, Luís Monteiro, C. Palamidessi, and M. Yung, Eds. Springer Berlin Heidelberg, Jan. 2005, no. 3580, pp. 103–114.
- [13] J. Barbay, "From time to space: Fast algorithms that yield small and fast data structures," in *Space-Efficient Data Structures, Streams, and Algorithms*. Springer, 2013, pp. 97–111.
- [14] A. Mazumdar and O. Milenkovic, "Smooth representation of rankings," in *Proc. Conf. on Inform. Sci. & Sys. (CISS)*, Mar. 2014, pp. 1–4.
- [15] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. the method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, 1952.
- [16] R. D. Luce, *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.
- [17] R. L. Plackett, "The analysis of permutations," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 24, no. 2, pp. 193–202, 1975.
- [18] C. L. Mallows, "Non-null ranking models. i," *Biometrika*, vol. 44, no. 1-2, pp. 114–130, 1957.
- [19] W. Cheng and E. Hüllermeier, "A new Instance-Based label ranking approach using the mallows model," in *Advances in Neural Networks - ISNN 2009*, ser. Lecture Notes in Computer Science, W. Yu, H. He, and N. Zhang, Eds. Springer Berlin Heidelberg, Jan. 2009, no. 5551, pp. 707–716.
- [20] A. Klementiev, D. Roth, and K. Small, "Unsupervised rank aggregation with distance-based models," in *Proceedings of the 25th International Conference on Machine Learning*, ser. ICML '08. New York, NY, USA: ACM, 2008, pp. 472–479.
- [21] T. Lu and C. Boutilier, "Learning mallows models with pairwise preferences," *ICML-11, Bellevue*, WA, 2011.
- [22] M. Deza and T. Huang, "Metrics on permutations, a survey," *Journal of Combinatorics, Information and System Sciences*, vol. 23, no. 1-4, pp. 173–185, 1998.
- [23] M. G. Kendall, "A new measure of rank correlation," *Biometrika*, pp. 81–93, 1938.
- [24] D. E. Knuth, *Art of Computer Programming, Volume 3: Sorting and Searching*, 2nd ed. Addison-Wesley Professional, 1998.
- [25] —, *Art of Computer Programming, Volume 3: Sorting and Searching*, 1st ed. Addison-Wesley, 1973.
- [26] D. Wang, A. Mazumdar, and G. Wornell, "Lossy compression of permutations," in *Proc. IEEE Int. Symp. Inform. Th. (ISIT)*, Honolulu, HI, Jun. 2014, pp. 11–15.
- [27] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein, *Covering Codes*. Elsevier, Apr. 1997.
- [28] D. Wang, A. Mazumdar, and G. Wornell, "A rate-distortion theory for permutation spaces," in *Proc. IEEE Int. Symp. Inform. Th. (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 2562–2566.
- [29] F. Farnoud, M. Schwartz, and J. Bruck, "Rate-Distortion for ranking with incomplete information," *arXiv:1401.3093 [cs, math]*, Jan. 2014. [Online]. Available: <http://arxiv.org/abs/1401.3093>
- [30] P. Diaconis and A. Ram, "Analysis of systematic scan metropolis algorithms using Iwahori-Hecke algebra techniques," Department of Statistics, Stanford University, Tech. Rep. 2000-11, Mar. 2000.
- [31] J. Doignon, A. Peke, and M. Regenwetter, "The repeated insertion model for rankings: Missing link between two subset choice models," *Psychometrika*, vol. 69, no. 1, pp. 33–54, 2004.
- [32] R. Shreevatsa, "The On-Line Encyclopedia of Integer Sequences," 2013. [Online]. Available: <http://oeis.org/A161169>
- [33] R. P. Stanley, *Enumerative Combinatorics, Vol. 1*. Cambridge University Press, Apr. 1997.
- [34] G. Louchard, H. Prodinger, and P. O. Wits, "The number of inversions in permutations: a saddle point approach," *Journal of Integer Sequences*, vol. 6, no. 2, pp. 1–19, 2003.
- [35] A. Mazumdar, A. Barg, and G. Zemor, "Constructions of rank modulation codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 1018–1029, 2013.

**Da Wang** received the B.A.Sc. degree with honors in electrical engineering from the University of Toronto, Toronto, ON, Canada, and the S.M. and Ph.D. degrees in electrical engineering and computer science (EECS) from the Massachusetts Institute of Technology (MIT), Cambridge, in 2008, 2010 and 2014, respectively.

Dr. Wang is a receipt for several awards or fellowships, including Jacobs Fellowship in 2008, Claude E. Shannon Research Assistantship in 2011–2012, and Wellington and Irene Loh Fund Fellowship in 2014. His research interests include information theory, distributed computing and statistical inference.

**Arya Mazumdar** (S'05-M'13) is an assistant professor in University of Minnesota-Twin Cities (UMN) since January 2013. Before coming to UMN, he was a postdoctoral scholar at the Massachusetts Institute of Technology (MIT). He received his Ph.D. degree from University of Maryland, College Park, in 2011.

Arya is a recipient of 2014-15 NSF CAREER award and the 2010 IEEE ISIT Student Paper Award. He is also the recipient of the Distinguished Dissertation Fellowship Award, 2011, at the University of Maryland. He spent the summers of 2008 and 2010 at the Hewlett-Packard Laboratories, Palo Alto, CA, and IBM Almaden Research Center, San Jose, CA, respectively. Arya's research interests include error-correcting codes, information theory and their applications.

**Gregory W. Wornell** (S'83-M'91-SM'00-F'04) received the B.A.Sc. degree in electrical engineering from the University of British Columbia, Vancouver, BC, Canada, and the S.M. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, in 1985, 1987, and 1991, respectively.

Since 1991, he has been on the faculty at MIT, where he is the Sumitomo Professor of Engineering in the department of Electrical Engineering and Computer Science (EECS). He leads the Signals, Information, and Algorithms Laboratory in the Research Laboratory of Electronics, and co-chairs the EECS department graduate program. He has held visiting appointments at the former AT&T Bell Laboratories, Murray Hill, NJ, the University of California, Berkeley, CA, and Hewlett-Packard Laboratories, Palo Alto, CA.

His research interests and publications span the areas of information theory, digital communication, statistical inference, and signal processing, and include algorithms and architectures for wireless networks, sensing and imaging systems, multimedia applications, and aspects of computational biology and neuroscience.

Dr. Wornell has been involved in the Information Theory and Signal Processing Societies of the IEEE in a variety of capacities, and maintains a number of close industrial relationships and activities. He has won a number of awards for both his research and teaching.