

Compression in the Space of Permutations

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

Abstract

In this paper, we investigate the problem of compression of data that are in the form of permutations. This problem has direct applications in the storage of rankings or ordinal data as well as in the analysis of sorting algorithms.

Lossy compression of permutations constitute new information theoretic challenges because of the uniqueness of the data format. Following the classical theory of compression, we propose the rate-distortion problem for the permutation space under uniform distribution and analyze the required rate of compression R to achieve recovery-distortion D with respect to different practical and useful distortion measures, including Kendall tau distance, Spearman's footrule, Chebyshev distance and inversion- ℓ_1 distance.

Under these distortion measures, we establish equivalences of the source code designs, which follow from the near isometry of the distances. Our results are non-asymptotic. Furthermore, for all permutation spaces of interest, we provide explicit code designs that incur low encoding/decoding complexities.

Apart from the uniform input distribution, we also comment on the compression of Mallows model, a popular nonuniform ranking model on the permutation space. We show that, for the Mallows model, both the entropy and the maximum distortion at zero rate are much lower than the uniform counterpart. This indicates greater compression ratio, and suggests that it would be worthwhile to solve the challenge of designing entropy-achieving compression schemes with low computational complexity for Mallows model.

I. INTRODUCTION

PERMUTATIONS are fundamental mathematical objects and *codes in permutations* is a well-studied subject in coding theory, with a variety of applications that correspond to different metric functions on the symmetric group on n elements \mathcal{S}_n . 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 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.

Storing ranking data: In applications such as recommendation systems, users rank products and the system analyze these rankings to provide new recommendations. To have personalized recommendation, it may be necessary to store the ranking of each user in the system, and hence the storage efficiency of ranking data is of interest. 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. 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. This poses the question of the number of bits needed for permutation storage when a certain amount of error can be tolerated.

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 permutation by pairwise comparisons, and since each comparison in sorting provides at most 1 bit of information, the log-size of the permutation set \mathcal{S}_n , $\log n!$, provides a lower bound to the required number of comparisons. Similarly, the lossy source coding of permutations provides a lower bound to the problem of comparison-based approximate sorting, which can be seen as searching a true permutation subject to a certain distortion. Again, the log-size of the code indicates the amount of information (in terms of bit) needed to specify the true permutation, which in turn provides a lower bound on the number of pairwise comparisons needed.

This work was supported, in part, by AFOSR under Grant No. FA9550-11-1-0183, and by NSF under Grant No. CCF-1017772. Arya Mazumdar's research was also supported in part by a startup grant from University of Minnesota.

D. Wang and G. W. Wornell are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 02139 (Email: {dawang,gww}@mit.edu). A. Mazumdar is with the Department of Electrical and Computer Engineering, University of Minnesota, MN, 55455 (Email: arya@umn.edu).

The material in this paper was presented in part at International Symposium of Information Theory, Istanbul, Turkey, 2013.

The problem of approximate sorting was first proposed in [7] as the problem of partial sorting, which is also called multiple selection. An algorithm that matches the information-theoretic lower bound of this paper is proposed in [8]. In another line of work, [9] derived both lower and upper bounds for approximate sorting in the moderate distortion regime with respect to the Spearman’s footrule metric [10] (see Definition 1 below).

Remark 1 (Compression is easier than sorting). *It is worth noting that compression is easier than sorting, in the sense that every comparison-based sorting algorithm corresponds to a compression scheme of the permutation space, while the reverse does not hold in general. In particular, the string of bits that represent comparison outcomes in any deterministic (approximate) sorting algorithm corresponds to a (lossy) representation of the permutation.*

Beyond the above applications, the rate-distortion theory on 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 rather fundamental problem of lossy compression in permutation spaces in this paper. Following the classical rate-distortion formulation, we aim to determine that, 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- ℓ_1 distance. As we shall see in Section II, each of these distortion measures 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 all the distortion measures of interest in this paper. We use these relations to establish the equivalence of source codes in permutation spaces with different distortion measures. Our results indicate that, while these distance measures usually have different intended applications, an optimal coding scheme for one distortion measure is effectively optimal for other 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 distribution over the permutation spaces. In some applications, we may have prior knowledge about the permutation data, which can be captured in a non-uniform distribution. There are a variety of distributional models, such as the Bradley-Terry model [11], the Luce-Plackett model [12], [13], and the Mallows model [14] that arise in different contexts. Among these, we choose the Mallows model due to its richness and applicability in various ranking applications [15]–[17]. 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.

Our analysis provides an information-theoretic lower bound on query complexity for *all* approximate sorting algorithms that achieve a certain distortion, while the lower bound in [18] only holds for the class of partial sorting algorithms that achieve the same distortion. The multiple selection algorithm proposed in [8] turns out to be optimal for the general approximate sorting problem as well and hence our information-theoretic lower bound is tight.

The rest of the paper is organized as follows. We first present the problem formulation in Section II. We then show that there exist close relationships between the 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. Finally we conclude in Section VII.

II. PROBLEM FORMULATION

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

A. Notation and facts

Let \mathcal{S}_n denote the symmetric group of n elements. We write the elements of \mathcal{S}_n as arrays 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. For a permutation σ , we denote its permutation inverse

by σ^{-1} , where $\sigma^{-1}(x) = i$ when $\sigma(i) = x$. and $\sigma(i)$ is the i -th element in array σ . For example, the permutation inverse of $\sigma = [2, 5, 4, 3, 1]$ is $\sigma^{-1} = [5, 1, 4, 3, 2]$. We denote the identity permutation by Id , i.e., $\text{Id} \triangleq [1, 2, \dots, n]$. 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 denote the set $\{1, \dots, n\}$ as $[n]$, and let $[a : b] \triangleq \{a, a + 1, \dots, b - 1, b\}$ for any two integers a and 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

For distortion measures, it is natural to use the distance measure on the permutation set \mathcal{S}_n , and there exist many possibilities [19]. In this paper we choose a few distortion measures of interest in a variety of application settings, including Spearman's footrule (ℓ_1 distance between two permutation vectors), Chebyshev distance (ℓ_∞ distance between two permutation vectors), Kendall tau distance and the inversion- ℓ_1 distance.

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 , then we say the permutation σ is the ranking of these 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 from pairwise comparisons.

Given two rankings σ_1 and σ_2 , we measure the total deviation of ranking and maximum deviation of ranking by Spearman's footrule and Chebyshev distance respectively.

Definition 1 (Spearman's footrule [10]). *Given two permutations $\sigma_1, \sigma_2 \in \mathcal{S}_n$, the Spearman's footrule between σ_1 and σ_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 σ_1 and σ_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. Lemma 20) and the Chebyshev distance in \mathcal{S}_n is upper bounded by $n - 1$.

Given two lists of items with ranking σ_1 and σ_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 π_1 that changes the ranking of π_1 to the ranking of π_2 as the Kendall tau distance.

Definition 3 (Kendall tau distance). *The Kendall tau distance $d_\tau(\sigma_1, \sigma_2)$ from one permutation σ_1 to another permutation σ_2 is defined as the minimum number of transpositions of pairwise adjacent elements required to change σ_1 into σ_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 σ_1 to σ_2 . For example,*

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

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 [20].

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). 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 $\mathbf{x}_\sigma \in \mathcal{G}_n \triangleq [0 : 1] \times [0 : 2] \times \cdots \times [0 : n - 1]$, where $\mathbf{x}_\sigma(i')$, $1 \leq i' \leq n - 1$ is the number of inversions in σ in which $i' + 1$ is the first element. Formally, for $i' = 2, \dots, n$,

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

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

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

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 $n - \mathbf{x}_\pi(n - 1) = 5 - 1 = 4$.

It is well known that mapping from \mathcal{S}_n to \mathcal{G}_n is one-to-one and straightforward [20].

With these, we define the inversion- ℓ_1 distance.

Definition 5 (inversion- ℓ_1 distance). Given two permutations $\sigma_1, \sigma_2 \in \mathcal{S}_n$, we define the inversion- ℓ_1 distance, ℓ_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- ℓ_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 to each other. While the operational significance of the inversion- ℓ_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 operates on the ranking domain, inversion vector and Kendall tau distance can be viewed as operating on the inverse of the ranking domain.

Remark 3. The ℓ_1, ℓ_∞ distortion measures above can be readily generalized to weighted versions to incorporate different emphasis on different parts of the ranking.

In particular, using a weighted version that only puts non-zero weight to the first k components of the permutation vector corresponds to the case that we only the distortion of the top- k items (top- k selection problem).

C. Rate-distortion problems

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

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 average-case distortion is a set of permutations such that for a σ 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 $\sigma \in \mathcal{S}_n$.

Definition 7 (Codebook for worst-case distortion). The codebook for permutations under worst-case distortion can be defined analogously to Definition 6, except (4) now becomes

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

We use $\hat{\mathcal{C}}_n$ to denote a (n, D_n) source code under the worst-case distortion.

In this paper we mostly focus on the case P is uniformly distributed over the symmetric group \mathcal{S}_n , except in Section VI, where a distribution rising from the Mallows model is used.

Definition 8 (Rate function). Given a source code \mathcal{C}_n and a distortion D_n , let $A(n, D_n)$ be the minimum size of \mathcal{C}_n , 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 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- ℓ_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- ℓ_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 $D = \Theta(n^2)$ ¹.

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 all four distortion measures defined in Section II-B are closely related to each other, which is summarized in (7) and (8). These relationships implies a variety of equivalence in the lossy compression schemes, which we exploit to derive the rate-distortion functions in Section IV.

¹In 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.

For any $\sigma_1 \in \mathcal{S}_n$ and σ_2 is randomly uniformly chosen from \mathcal{S}_n ,

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

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

where $x \lesssim y$ indicates $x < c \cdot y$ for some constant $c > 0$, and $\stackrel{w.h.p.}{\lesssim_\infty}$ indicates \lesssim with high probability.

The following sections provide detailed arguments for (7) and (8) by analyzing the relationship between different pairs of distortion measures.

1) *Spearman's footrule and Chebyshev distance*: Let σ_1 and σ_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.

Theorem 1. For any $\pi \in \mathcal{S}_n$, let σ 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 $c_1 < 1/3$.

Proof: See Appendix B-A. ■

2) *Spearman's footrule and Kendall tau distance*: The following theorem is a well-known result on the relationship between Kendall tau distance and ℓ_1 distance of permutation vectors.

Theorem 2 ([10]). Let σ_1 and σ_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- ℓ_1 distance and Kendall tau distance*: We show that the inversion- ℓ_1 distance and the Kendall tau distance are closely related in Theorem 3, and Theorem 4, which helps to establish the equivalence of the rate-distortion problem later.

Theorem 3. Let σ_1 and σ_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 B-B. ■

Remark 4. The lower and upper bounds in Theorem 3 are tight in the sense that there exist permutations σ_1 and σ_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 3 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, Theorem 4 shows that Kendall tau distance scaled by a constant actually provides a lower bound to the inversion- ℓ_1 distance with high probability.

Theorem 4. For any $\pi \in \mathcal{S}_n$, let σ 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 $c_2 < 1/2$.

Proof: See Appendix B-C. ■

IV. TRADE-OFFS BETWEEN RATE AND DISTORTION

In this section we present 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 5, which shows the equivalence of lossy source codes under different distortion measures. This indicates that for all the distortion measures in this paper, the lossy compression scheme for one measure preserves distortion under other measures under average-case distortion, and hence all compression schemes can be used interchangeably, given appropriate transformation of the permutation representation and scaling the distortion. Then with these equivalence relationships, Theorem 6 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 $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ when the distortion leads to degenerate rates $R(D_n) = 0$ and $R(D_n) = 1$.

A. Rate distortion functions

Theorem 5 (Equivalence of lossy source codes). *Under both average-case and worst-case distortion, 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 $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$.

Furthermore, under average-case distortion, a following source code on the left hand side implies a source code on the right hand side:

- 5) (n, D_n) source code for $\mathcal{X}(\mathcal{S}_n, d_{\ell_1}) \Rightarrow (n, D_n/(nc_1) + O(1))$ source code for $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$ for any $c_1 < 1/3$,
- 6) (n, D_n) source code for $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1}) \Rightarrow (n, D_n/c_2 + O(n))$ source code for $\mathcal{X}(\mathcal{S}_n, d_\tau)$ for any $c_2 < 1/2$.

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

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

As we show below, for the uniform distribution on \mathcal{S}_n , the rate-distortion function is the same for both average- and worst-case, apart from the terms that are asymptotically negligible.

Theorem 6 (Rate distortion functions). *For permutation spaces $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$, $\mathcal{X}(\mathcal{S}_n, d_\tau)$, and $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$,*

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

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

$$\begin{aligned} \bar{R}(D_n) &= \hat{R}(D_n) \\ &= \begin{cases} 1 & \text{if } D_n = O(1), \\ 1 - \delta & \text{if } D_n = \Theta(n^\delta), \quad 0 < \delta \leq 1. \end{cases} \end{aligned} \quad (15)$$

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

Proof sketch: The achievability comes from the compression schemes proposed in Section V. The 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.

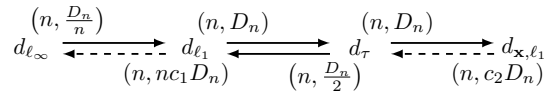


Fig. 1: Relationship between source codes. An arrow indicates a source code in one space implies a source in another space, where the solid arrow indicates for both average-case and worst-case distortions, and the dashed arrow indicates for average-case only.

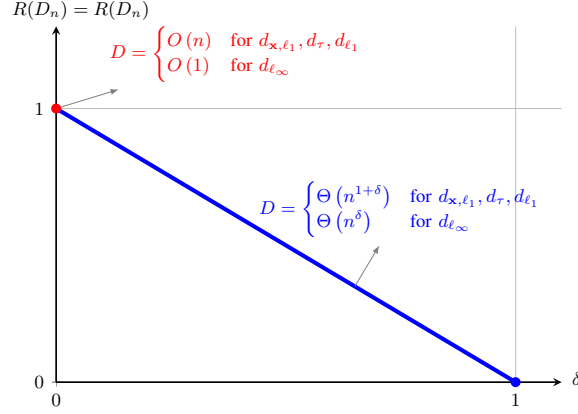


Fig. 2: Rate-distortion function for permutation spaces $\mathcal{X}(\mathcal{S}_n, d_{x, \ell_1})$, $\mathcal{X}(\mathcal{S}_n, d_\tau)$, $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$, and $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$.

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 C-B. ■

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

$$\lim_{n \rightarrow \infty} \mathbb{P}[d(f_n(\sigma), \sigma) > D_n] = 0 \quad (16)$$

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 δ via the asymptotic relationship $\log D / \log n - 1$ (for permutation spaces $\mathcal{X}(\mathcal{S}_n, d_{x, \ell_1})$, $\mathcal{X}(\mathcal{S}_n, d_\tau)$, 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 $(1 - \delta)n \log_2 n$.

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 $\mathcal{X}(\mathcal{S}_n, d_{x, \ell_1})$.

Theorem 7. *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, $r_\tau^s(D_n) \leq r(D_n) \leq \overline{r_\tau^s(D_n)}$, where*

$$\underline{r_\tau^s(D_n)} = \begin{cases} -a(1 - \delta)n^\delta \log n + O(n^\delta), & 0 < \delta < 1 \\ -n \left[\log \frac{(1+a)^{1+a}}{a^a} \right] + o(n), & \delta = 1 \end{cases}, \quad (17)$$

$$\overline{r_\tau^s(D_n)} = \begin{cases} -n^\delta \frac{a \log 2}{2} + O(1), & 0 < a < 1 \\ -n^\delta \frac{\log [2a]!}{[2a]} + O(1), & a \geq 1 \end{cases}. \quad (18)$$

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

$$\underline{r_\tau^1(D_n)} = \max \{0, n \log 1/(2be^2)\}, \quad (19)$$

$$\overline{r_\tau^1(D_n)} = n \log \lceil 1/(2b) \rceil + O(\log n). \quad (20)$$

Remark 5. *Some of the results above for $\mathcal{X}(\mathcal{S}_n, d_\tau)$, since their first appearances in the conference version [21], have been improved subsequently by [22]. More specifically, for the small distortion regime, [22, Lemma 7, Lemma 10] provides an improved upper bound and show that $r_\tau^s(D_n) = \underline{r_\tau^s(D_n)}$ in (17). For the large distortion regime, [22, Lemma 11] shows a lower bound that is tighter than (19).*

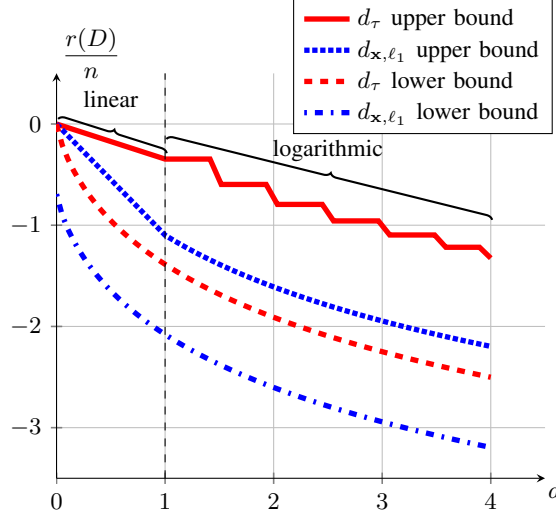


Fig. 3: Higher-order trade-off between rate and distortion in the small distortion regime with $D = an$. The slope discontinuities of the d_τ upper bound in the range of $a \geq 1$ is due to the flooring in (18).

Theorem 8. In the permutation space $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$, when $D_n = an^\delta, 0 < \delta \leq 1$,

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

where $\underline{r}_{\mathbf{x}, \ell_1}^s(D_n) = \underline{r}_\tau^s(D_n) - n^\delta \log 2$ (cf. (17)) and

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

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

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

where $\underline{r}_{\mathbf{x}, \ell_1}^l(D_n) = \underline{r}_\tau^l(D_n)$ (cf. (19)) and $\overline{r}_{\mathbf{x}, \ell_1}^l(D_n) = n \log \lceil 1/(4b) \rceil + O(1)$.

Proof: The achievability are presented in Section V-D and Section V-E. For converse, note that

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

where $N(D_n)$ is the maximum of the size of balls with radius D_n in the corresponding permutation space (cf. Appendix A for definitions), then a lower bound on $|\mathcal{C}_n|$ follows from the upper bound on $N(D_n)$ in Lemma 17 and Lemma 19. We omit the details as it is analogous to the proof of Theorem 6. ■

The bounds to $r(D_n)$ of both Kendall tau distance and inversion- ℓ_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, are optimal as they achieve the rate-distortion functions in Theorem 6. 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 building block algorithms with the proper parameters, we can achieve the corresponding trade-offs between rate and distortion shown in Section IV.

The equivalence relationships in Theorem 5 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

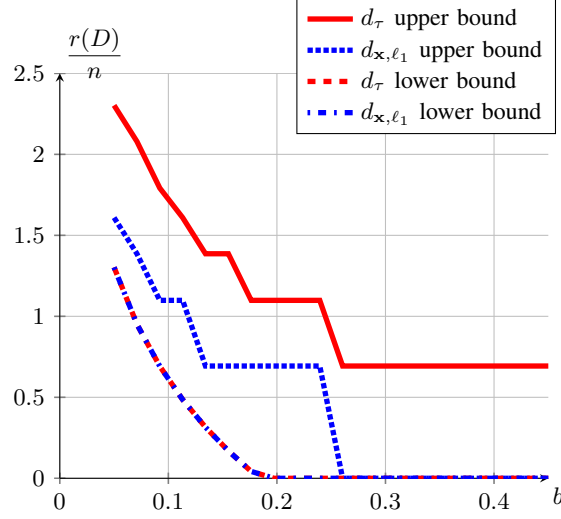


Fig. 4: Higher-order trade-off between rate and distortion in the large distortion regime with $D = bn^2$. The lower bounds for d_τ and $d_{\mathbf{x}, \ell_1}$ are identical.

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 $\Theta(n \log n)$ [20, Exercise 6 in Section 5.1.1]. Therefore, the cost of representation transformation of permutations 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 the subsequences, either of the given permutation σ or of its inverse σ^{-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 first consider the space $\mathcal{X}(\mathcal{S}_n, d_\tau)$ and 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

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

and $\sigma[a : b]$ is a shorthand notation for the vector $[\sigma(a), \sigma(a+1), \dots, \sigma(b)]$. This procedure is illustrated in Fig. 5.

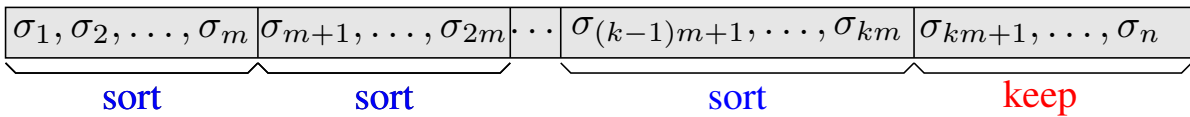


Fig. 5: Quantization by sorting subsequences.

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 then calculate the worst-case and average-case distortions:

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

$$\bar{D}_{d_\tau}(k, m) = k \frac{m(m-1)}{4} \leq km^2/4 \quad (22)$$

where (21) is from (39).

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

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

$$\bar{D}_{\ell_\infty}(k, m) \leq m - 1 \quad (24)$$

$$\hat{D}_{\ell_1}(k, m) = k \lfloor m^2 \rfloor / 2 \leq km^2/2 \quad (25)$$

$$\bar{D}_{\ell_1}(k, m) = k(m^2 - 1)/3, \quad (26)$$

where (25) comes from Lemma 20 and (26) comes from (38).

Remark 6. Due to the close relationship between Kendall tau distance and Spearman's footrule shown in (11), there exists an equivalent construction via the inverse permutation σ^{-1} of a permutation $\sigma \in \mathcal{S}_n$:

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

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 $km^2/2$.

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 inversion vector space \mathcal{G}_n .

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

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

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

$$m = \left\lceil k / \left(2\hat{D}_{\mathbf{x}, \ell_1} + 1 \right) \right\rceil \quad (28)$$

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 5 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 α depends on the distortion measure and whether we are considering worst-case or average-case distortion, as shown in (21) and (22) and (25) and (26), and is summarized in Table I.

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

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

TABLE I: The values of α for different compression scenarios.

α	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 $0 < \delta < 1$, $D_n = an^\delta$.

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

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

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

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

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

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

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

E. Compression in the large distortion regime

In this section we provide compression schemes in the 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 $0 < \delta < 1$, $D_n = bn^2$.

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

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

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 $m_k = \lceil k/(4b(k-1) + 1) \rceil$, $k = 2, \dots, n$. The distortion $D^{(k)}$ for each coordinate k satisfies

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

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

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

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

VI. COMPRESSION OF PERMUTATION SPACE WITH MALLOWS 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 [14], a model with a wide range of applications such ranking, partial ranking, and even algorithm analysis (see [23, Section 2e] and the references therein). In the context of storing user ranking data, 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, 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) π 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 $Z_{q, \pi} = \sum_{\sigma \in \mathcal{S}_n} q^{d_\tau(\sigma, \pi)}$. In particular, when the mode $\pi = \text{Id}$, $Z_q \triangleq Z_{q, \text{Id}} = [n]_q!$ [23, (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 Mallows model is in general $\Theta(n)$, implying lower space for storage 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 to average-case distortion.

Noting the Kendall tau distance is right-invariant [19], for the purpose of compression, without loss of generality, we can assume the mode $\pi = \text{Id}$, 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 [24] and later applied in [17].

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 σ by repeatedly inserting $\pi(i)$ before the j -th element in σ with probability $p_{i,j}$ (when $j = i$, we append $\pi(i)$ at the end of σ).*

Remark 7. Note that the insertion probabilities at step i is independent of the realization 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 9. Given a RIM with reference permutation $\pi = \text{Id}$ and insertion vector \mathbf{a}_σ , then the corresponding permutation σ satisfies

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

where $\tilde{\mathbf{x}}_\sigma$ is an extended inversion vector that simply prepends a 0 to an inversion vector \mathbf{x}_σ :

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

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 4. 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 10 (Mallows model via RIM [17], [24]). Given reference permutation π 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 11. 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 9 and Theorem 10. ■

B. Lossless compression

We consider the lossless compression of Mallows model.

Lemma 12.

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

Proof: This follows directly from Theorem 10. ■

Lemma 13 (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 $H_b(\cdot)$ is the binary entropy function, $g(n, q) = \Theta(1)$, and $\lim_{q \rightarrow 0} g(n, q) = 0$.

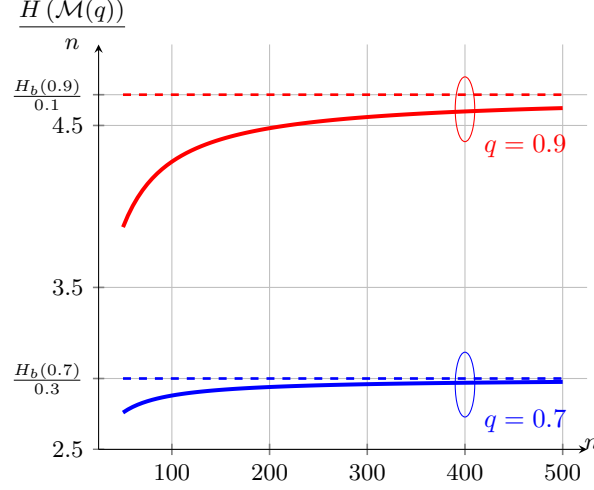


Fig. 6: Entropy of the Mallows model for $q = 0.7$ and $q = 0.9$, where the dashed lines are the coefficients of the linear terms, $H_b(q)/(1-q)$.

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

Remark 8. Doing 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 11, 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 Mallows model in Lemma 14, which implies that at rate 0, the average-case distortion is $\Theta(n)$, while under the uniform distribution, Theorem 6 indicates that it takes $n \log n + o(n \log n)$ bits to achieve average-case distortion of $\Theta(n)$.

Lemma 14 (Typical set of Mallows model). *There exists $c_0(q)$, a constant that depends on q only, such that for any $r_0 \geq c_0(q)n$,*

$$\lim_{n \rightarrow \infty} \mathbb{P}[B_r(r_0)] = 1.$$

The proof is presented in Appendix D-B.

Remark 9. As pointed out in [24], 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 is also induced by RIM.

VII. CONCLUDING REMARKS

In this paper, we first consider the lossy compression of permutations under the uniform distribution for both average-case and worst-case distortions, in terms of Kendall tau distance, Spearman's footrule, Chebyshev distance and inversion- ℓ_1 distance. Regarding the lossy storage of ranking, our results provide the fundamental trade-off between storage and accuracy. Regarding approximate sorting, our results indicate that it is the constant in front of the $n \log n$ term that matters. Therefore, given a distortion D_n , an approximate sorting algorithm should aim to achieve the optimal constant in front of the $n \log n$ term, in the number of pairwise comparisons, and this constant is exactly rate $R(D_n)$. As mentioned, this is indeed achieved by the multiple selection algorithm in [8], showing the 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 learning to rank [17] and algorithm analysis [23].

APPENDIX A GEOMETRY OF PERMUTATION SPACES

In this section we provide results on the geometry of the permutation space that are useful in deriving the 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:

$$B_d(\sigma, D) \triangleq \{\pi : d(\pi, \sigma) \leq D\}, \quad (29)$$

$$N_d(D) \triangleq \max_{\sigma \in \mathcal{S}_n} |B_d(\sigma, D)|. \quad (30)$$

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, ℓ_1 distance of the permutations, and ℓ_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 15. For $0 \leq D \leq n$,

$$N_\tau(D) \leq \binom{n+D-1}{D}. \quad (31)$$

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 [20]. An expression for $K_n(D)$ (and thus $T_n(D)$) for $D \leq n$ appears in [20] (see [4] also). The following bound is weaker but sufficient in our context.

By induction, or [25], $T_n(D) = K_{n+1}(D)$ when $D \leq n$. Then noting that for $k < n$, $K_n(k) = K_n(k-1) + K_{n-1}(k)$ [20, 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 $1 \leq k < n$,

$$K_n(k) \leq \binom{n+k-2}{k}. \quad (32)$$

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 16. For $0 \leq D \leq n(n-1)/2$,

$$N_{\mathbf{x}, \ell_1}(D) \leq 2^{\min\{n, D\}} \binom{n+D}{D}. \quad (33)$$

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 [26, 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} , $|\{y \in \mathcal{G}_n : d_{\ell_1}(\mathbf{x}, y) = r\}| \leq 2^m Q(n, r)$ and hence

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

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 17 to 19 respectively. ■

Lemma 17 (Small distortion regime). *When $D = an^\delta, 0 < \delta \leq 1$ and $a > 0$ is a constant,*

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

$$\begin{aligned} \log N_{\mathbf{x}, \ell_1}(D) &\leq \begin{cases} a(1-\delta)n^\delta \log n + O(n^\delta), & 0 < \delta < 1 \\ n \left[2 + \log \frac{(1+a)^{1+a}}{a^a} \right] + o(n), & \delta = 1 \end{cases}. \end{aligned} \quad (35)$$

Proof: To upper bound $N_\tau(D)$, when $0 < \delta < 1$, we apply Stirling's approximation to (31) 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 (34). When $\delta = 1$, the result follows from (9) in [27, Section 4]. The upper bound on $N_{\mathbf{x}, \ell_1}(D)$ can be obtained similarly via (33). ■

Lemma 18 (Moderate distortion regime). *Given $D = \Theta(n^{1+\delta}), 0 < \delta \leq 1$, then*

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

Proof: Apply Stirling's approximation to (33) and substitute $D = \Theta(n^{1+\delta})$. ■

Remark 10. *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 19 (Large distortion regime). *Given $D = bn(n-1) \in \mathbb{Z}^+$, then*

$$\log N_\tau(D) \leq \log N_{\mathbf{x}, \ell_1}(D) \leq n \log(2ben) + O(\log n). \quad (37)$$

Proof: Substitute $D = bn(n-1)$ into (33). ■

APPENDIX B PROOFS ON THE RELATIONSHIPS AMONG DISTORTION MEASURES

Lemma 20.

$$\max_{\sigma \in \mathcal{S}_n, \sigma' \in \mathcal{S}_n} d_{\ell_1}(\sigma, \sigma') = \lfloor n^2/2 \rfloor.$$

Proof: Note that $d_{\ell_1}(\sigma, \sigma')$ is maximized when $\sigma = [1, 2, \dots, n]$ and $\sigma' = [n, n-1, \dots, 1]$. Therefore, when n is even,

$$\begin{aligned} &\max_{\sigma \in \mathcal{S}_n, \sigma' \in \mathcal{S}_n} d_{\ell_1}(\sigma, \sigma') \\ &= \begin{cases} \sum_{k=1}^{n/2} (2k-1) = n^2/2 & n \text{ is even,} \\ \sum_{k=1}^{(n-1)/2} 2k = (n^2-1)/2 & n \text{ is odd.} \end{cases} \end{aligned}$$

A. Proof of Theorem 1

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

$$\mathbb{E}[X_{\ell_1}] = \frac{n^2 - 1}{3} \quad \text{Var}[X_{\ell_1}] = \frac{2n^3}{45} + O(n^2). \quad (38)$$

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 [10, Table 1]. ■

Proof for Theorem 1: For any $c > 0$, $cn \cdot d_{\ell_\infty}(\pi, \sigma) \leq cn(n-1)$, and for any $c_1 < 1/3$, Lemma 21 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 3

Lemma 22. For any two permutations π, σ 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 π' and σ' be permutations in \mathcal{S}_{n-1} with element k removed from π and σ correspondingly, then $\mathbf{x}_{\pi'} = \mathbf{x}_{\sigma'}$, and hence $\pi' = \sigma'$. Therefore, the Kendall tau distance between σ and π is determined only by the location of element k in σ and π , which is at most $n-1$. ■

Proof of Theorem 3: It is known that (see, e.g., [28, Lemma 4])

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

Furthermore, the proof of [28, Lemma 4] indicates that for any two permutation π_1 and π_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

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

where (a) is due to Lemma 22. ■

C. Proof of Theorem 4

To prove Theorem 4, we analyze the mean and variance of the Kendall tau distance and inversion- ℓ_1 distance between a permutation in \mathcal{S}_n and a randomly selected permutation, in Lemma 23 and Lemma 24 respectively.

Lemma 23. *For any $\pi \in \mathcal{S}_n$, let σ be a permutation chosen uniformly from \mathcal{S}_n , and $X_\tau \triangleq d_\tau(\pi, \sigma)$, then*

$$\mathbb{E}[X_\tau] = \frac{n(n-1)}{4}, \quad (39)$$

$$\text{Var}[X_\tau] = \frac{n(2n+5)(n-1)}{72}. \quad (40)$$

Proof: Let σ' 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 [19], 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 [10, Table 1] and [20, Section 5.1.1]. ■

Lemma 24. *For any $\pi \in \mathcal{S}_n$, let σ 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 σ 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 Lemma 23 and Lemma 24, now we show that the event that a scaled version of the Kendall tau distance is larger than the inversion- ℓ_1 distance is unlikely.

Proof for Theorem 4: 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. ■

APPENDIX C PROOFS ON THE RATE DISTORTION FUNCTIONS

A. Proof of Theorem 5

Proof: Statement 1 follows from (9).

Statement 2 and 3 follow from Theorem 2. 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).

For Statement 5, define

$$\mathcal{B}_n(\pi) \triangleq \{\sigma : c_1 \cdot n \cdot d_{\ell_\infty}(\sigma, \pi) \leq d_{\ell_1}(\sigma, \pi)\},$$

then Theorem 1 indicates that

$$|\mathcal{B}_n(\pi)| = (1 - O(1/n))n!.$$

Let $\bar{\mathcal{C}}'_n$ be the (n, D_n) source code for $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$, π_σ be the codeword for σ in \mathcal{C}'_n , then by Theorem 1,

$$\begin{aligned} \mathbb{E}[d_{\ell_\infty}(\pi_\sigma, \sigma)] &= \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} d_{\ell_\infty}(\sigma, \pi_\sigma) \\ &= \frac{1}{n!} \left[\sum_{\sigma \in \mathcal{B}_n(\pi_\sigma)} d_{\ell_\infty}(\sigma, \pi_\sigma) + \sum_{\sigma \in \mathcal{S}_n \setminus \mathcal{B}_n(\pi_\sigma)} d_{\ell_\infty}(\sigma, \pi_\sigma) \right] \\ &\leq \frac{1}{n!} \left[\sum_{\sigma \in \mathcal{B}_n(\pi_\sigma)} d_{\ell_1}(\sigma, \pi_\sigma) + \sum_{\sigma \in \mathcal{S}_n \setminus \mathcal{B}_n(\pi_\sigma)} n \right] \\ &\leq D_n/(nc_1) + O(1/n)n = D_n/(nc_1) + O(1). \end{aligned}$$

For Statement 6, similar to the proof of statement 5, define

$$\mathcal{A}_n(\pi) \triangleq \{\sigma : c_2 \cdot d_\tau(\sigma, \pi) \leq d_{\mathbf{x}, \ell_1}(\sigma, \pi)\}$$

then Theorem 4 indicates that $|\mathcal{A}_n(\pi)| = (1 - O(1/n))n!$. Let $\bar{\mathcal{C}}'_n$ be the (n, D_n) source code for $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ and σ be a permutation chosen uniformly from \mathcal{S}_n , then let π_σ be the codeword for σ in $\bar{\mathcal{C}}'_n$,

$$\begin{aligned} & \mathbb{E}[d_\tau(\pi_\sigma, \sigma)] \\ &= \frac{1}{n!} \sum_{\sigma \in \mathcal{S}_n} d_\tau(\sigma, \pi_\sigma) \\ &= \frac{1}{n!} \left[\sum_{\sigma \in \mathcal{A}_n(\pi_\sigma)} d_\tau(\sigma, \pi_\sigma) + \sum_{\sigma \in \mathcal{S}_n \setminus \mathcal{A}_n(\pi_\sigma)} d_\tau(\sigma, \pi_\sigma) \right] \\ &\leq \frac{1}{n!} \left[\sum_{\sigma \in \mathcal{A}_n(\pi_\sigma)} d_{\mathbf{x}, \ell_1}(\sigma, \pi_\sigma)/c_2 + \sum_{\sigma \in \mathcal{S}_n \setminus \mathcal{A}_n(\pi_\sigma)} n^2/2 \right] \\ &\leq D_n/c_2 + O(1/n)n^2 = D_n/c_2 + O(n). \end{aligned}$$

■

B. Proof of Theorem 6

We prove Theorem 6 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

$$\log |\bar{\mathcal{C}}_n| = (1 - \delta - \varepsilon)n \log n + O(n), \quad (41)$$

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 $D_n = O(1)$,

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

when n sufficiently large, where (a) follows from (36).

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 (41), 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 D PROOFS ON MALLOWS MODEL

A. Proof of Lemma 13

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

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

■

B. Proof of Lemma 14

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 25 (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 16,

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

Finally, applying the bound [29]

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

completes the proof.

■

Proof of Lemma 14: 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 25 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 $r_0 \geq c_0 n$,

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

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. Klve, 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. M. Chambers, "Algorithm 410: Partial sorting," *Commun. ACM*, vol. 14, no. 5, pp. 357–358, 1971.
- [8] 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.
- [9] J. Giesen, E. Schuberth, and M. Stojakovi, "Approximate sorting," *Fundamenta Informaticae*, vol. 90, no. 1, pp. 67–72, 2009.
- [10] 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.
- [11] 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.
- [12] R. D. Luce, *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.
- [13] 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.
- [14] C. L. Mallows, "Non-null ranking models. i," *Biometrika*, vol. 44, no. 1-2, pp. 114–130, 1957.
- [15] 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.
- [16] 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.
- [17] T. Lu and C. Boutilier, "Learning mallows models with pairwise preferences," *ICML-11, Bellevue, WA*, 2011.
- [18] M. Fredman, "How good is the information theory bound in sorting?" *Theoretical Computer Science*, vol. 1, no. 4, pp. 355–361, 1976.
- [19] M. Deza and T. Huang, "Metrics on permutations, a survey," *Journal of Combinatorics, Information and System Sciences*, vol. 23, pp. 173–185, 1998.
- [20] D. E. Knuth, *Art of Computer Programming, Volume 3: Sorting and Searching*, 2nd ed. Addison-Wesley Professional, 1998.
- [21] 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.
- [22] 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>
- [23] P. Diaconis and A. Ram, *Analysis of systematic scan Metropolis algorithms using Iwahori-Hecke algebra techniques*. Department of Statistics, Stanford University, 2000.
- [24] 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.
- [25] R. Shreevatsa, "The On-Line Encyclopedia of Integer Sequences," 2013. [Online]. Available: <http://oeis.org/A161169>
- [26] R. P. Stanley, *Enumerative Combinatorics, Vol. 1*. Cambridge University Press, Apr. 1997.
- [27] 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.
- [28] A. Mazumdar, A. Barg, and G. Zemor, "Constructions of rank modulation codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 1018–1029, 2013.
- [29] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein, *Covering Codes*. Elsevier, Apr. 1997.