

Lossy Compression of Permutations

Da Wang
EECS Dept., MIT
Cambridge, MA, USA
Email: dawang@mit.edu

Arya Mazumdar
ECE Dept., Univ. of Minnesota
Twin Cities, MN, USA
Email: arya@umn.edu

Gregory W. Wornell
EECS Dept., MIT
Cambridge, MA, USA
Email: gww@mit.edu

Abstract—We investigate the lossy compression of permutations by analyzing the trade-off between the size of a source code and the distortion with respect to Kendall tau distance, Spearman’s footrule, Chebyshev distance and ℓ_1 distance of inversion vectors. We show that given two permutations, Kendall tau distance upper bounds the ℓ_1 distance of inversion vectors and a scaled version of Kendall tau distance lower bounds the ℓ_1 distance of inversion vectors with high probability, which indicates an equivalence of the source code designs under these two distortion measures. Similar equivalence is established for all the above distortion measures, every one of which has different operational significance and applications in ranking and sorting. These findings show that an optimal coding scheme for one distortion measure is effectively optimal for other distortion measures above.

I. INTRODUCTION

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.

In a variety of applications such as college admission and recommendation systems (e.g., `Yelp.com` and `IMDb.com`), ranking, or the relative ordering of data, is the key object of interest. As a ranking of n items can be represented as a permutation of 1 to n , storing a ranking is equivalent to storing a permutation. In general, to store a permutation of n elements, we need $\log_2(n!) \approx n \log_2 n - n \log_2 e$ bits. In applications such as recommendation systems, it may be necessary to store the ranking of all users in the system, and hence the storage efficiency of ranking data is of interest. Furthermore, in many use cases a rough knowledge of the ranking (e.g., finding one of the top five elements instead of the top element) is sufficient. This pose the question of the number of bits needed for storage when a certain amount error can be tolerated.

In addition to application on compression, source coding of the permutation space is also related to the analysis of comparison-based sorting algorithms. Given a group of elements of distinct values, comparison-based sorting can be viewed as the process of finding 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 provides a lower bound to the required number of comparisons, i.e., $\log n! = n \log n - O(n)$. 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 certain distortion. Again, the log-size of the code indicates the amount of information (in terms of bit) needed to specify the true permutation subject to certain distortion, which in turn provides a lower bound on the number of pairwise comparisons needed.

The problem of approximate sorting has been investigated in [1], where results for the moderate distortion regime are derived with respect to the Spearman’s footrule metric [2] (see below for definition).

On the other hand, every comparison-based sorting algorithm corresponds to a compression scheme of the permutation space, as we can treat the outcome of each comparison as 1 bit. This string of bits is a (lossy) representation of the permutation that is being (approximately) sorted. However, reconstructing the permutation from the compressed representation may not be straightforward.

In our earlier work [3], a rate-distortion theory for permutation space is developed, with the *worst-case distortion* as the parameter. The rate-distortion functions and source code designs for two different distortion measures, Kendall tau distance and the ℓ_1 distance of the inversion vectors, are derived. In Section III of this paper we show that under *average-case distortion*, the rate-distortion problem under Kendall tau distance and ℓ_1 distance of the inversion vectors are equivalent and hence the code design could be used interchangeably, leading to simpler coding schemes for the Kendall tau distance case (than developed in [3]), as discussed in Section IV.

Moreover, the rate-distortion problem under Chebyshev distance is also considered and its equivalence to the cases above is established. Operational meaning and importance of all these distance measures is discussed in Section II. While these distance measures usually have different intended applications, our findings show that an optimal coding scheme for one distortion measure is effectively optimal for other distortion measures.

II. PROBLEM FORMULATION

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

A. Notation

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

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.

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 $\sigma^{-1} = [5, 1, 4, 3, 2]$. 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 .

B. Rate-distortion problem

In this section we define the rate-distortion problems under both average-case distortion and worst-case distortion.

Definition 1 (Codebook for average-case distortion). *An (n, D) 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 a encoding mapping $f_n : \mathcal{S}_n \rightarrow \bar{\mathcal{C}}_n$ that*

$$\mathbb{E}_P [d(f_n(\sigma), \sigma)] \leq D. \quad (1)$$

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 2 (Codebook for worst-case distortion). *The codebook for permutations under worst-case distortion can be defined analogously to Definition 1, except (1) now becomes*

$$\max_{\sigma \in \mathcal{S}_n} d(f_n(\sigma), \sigma) \leq D. \quad (2)$$

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

Throughout the paper we focus on the case that P is uniformly distributed over the symmetric group \mathcal{S}_n .

Definition 3 (Rate function). *Given a source code \mathcal{C}_n and a sequence of distortions $\{D_n, n \in \mathbb{Z}^+\}$, 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.

As 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.

C. 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 [4]. 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 4 (Spearman's footrule [2]). *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 5 (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 $\lfloor n^2/2 \rfloor$ and the Chebyshev distance in \mathcal{S}_n is upper bounded by $n - 1$.

Given two list 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 6 (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 natural connection to sorting algorithms. In particular, given a list of items with values v_1, v_2, \dots, v_n such that $v_{\sigma^{-1}(1)} \succ v_{\sigma^{-1}(2)} \succ \dots \succ v_{\sigma^{-1}(n)}$, $d_\tau(\sigma^{-1}\sigma, \text{Id})$ is the number of swaps needed to sort this list of items in a bubble-sort algorithm [5].

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

Definition 7 (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

$$K_n(k) \triangleq |\{\sigma \in \mathcal{S}_n : I_n(\sigma) = k\}| \quad (3)$$

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. Mathematically, 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 [5].

With these, we define the inversion- ℓ_1 distance.

Definition 8 (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

$$d_{\mathbf{x}, \ell_1}(\sigma_1, \sigma_2) \triangleq \sum_{i=1}^{n-1} |\mathbf{x}_{\sigma_1}(i) - \mathbf{x}_{\sigma_2}(i)|. \quad (4)$$

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.

Remark 1. 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).

III. RELATIONSHIPS BETWEEN DISTORTION MEASURES

In this section we show all four distortion measures defined in Section II-C are closely related to each other.

A. Spearman's footrule and Kendall tau distance

Theorem 1 (Relationship of Kendall tau distance and ℓ_1 distance of permutation vectors [2]). Let σ_1 and σ_2 be any permutations in \mathcal{S}_n , then

$$d_{\ell_1}(\sigma_1, \sigma_2)/2 \leq d_\tau(\sigma_1^{-1}, \sigma_2^{-1}) \leq d_{\ell_1}(\sigma_1, \sigma_2). \quad (5)$$

B. ℓ_1 distance of inverse vectors and Kendall tau distance

We show that the ℓ_1 distance of inversion vectors and the Kendall tau distance are closely related in Theorem 2, and Theorem 3, which helps to establish the equivalence of the rate-distortion problem later.

The Kendall tau distance between two permutation vectors provides upper and lower bounds to the ℓ_1 distance between the inversion vectors of the corresponding permutations, as indicated by the following theorem.

Theorem 2. Let σ_1 and σ_2 be any permutations in \mathcal{S}_n , then for $n \geq 2$,

$$\frac{1}{n-1} d_\tau(\sigma_1, \sigma_2) \leq d_{\mathbf{x}, \ell_1}(\mathbf{x}_{\sigma_1}, \mathbf{x}_{\sigma_2}) \leq d_\tau(\sigma_1, \sigma_2) \quad (6)$$

The proof of this theorem is relatively straight-forward and hence omitted due to space constraint.

Remark 2. The lower bound in Theorem 2 is tight as there exists permutations σ_1 and σ_2 that satisfy the equality. For example, when $n = 2m$, let

$$\begin{aligned} \sigma_1 &= [1, 3, \dots, 2m-3, 2m-1, 2m, 2m-2, \dots, 4, 2], \\ \sigma_2 &= [2, 4, \dots, 2m-2, 2m, 2m-1, 2m-3, \dots, 3, 1], \end{aligned}$$

then $d_\tau(\sigma_1, \sigma_2) = n(n-1)/2$ and $d_{\mathbf{x}, \ell_1}(\sigma_1, \sigma_2) = n/2$. For another instance, let

$$\begin{aligned} \sigma_1 &= [1, 2, \dots, n-2, n-1, n], \\ \sigma_2 &= [2, 3, \dots, n-1, n, 1], \end{aligned}$$

then $d_\tau(\sigma_1, \sigma_2) = n-1$ and $d_{\mathbf{x}, \ell_1}(\sigma_1, \sigma_2) = 1$.

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, Theorem 3 shows that it provides a tight lower bound with high probability.

Theorem 3. For any $\pi \in \mathcal{S}_n$, let σ be a permutation chosen uniformly from \mathcal{S}_n , then

$$\mathbb{P}[c_1 \cdot d_\tau(\pi, \sigma) \leq d_{\mathbf{x}, \ell_1}(\pi, \sigma)] = 1 - O(1/n) \quad (7)$$

for any positive constant $c_1 < 1/2$.

Proof: See Section V-A. ■

C. Spearman's footrule and Chebyshev distance

Let σ_1 and σ_2 be any permutations in \mathcal{S}_n , then

$$d_{\ell_1}(\sigma_1, \sigma_2) \leq n \cdot d_{\ell_\infty}(\sigma_1, \sigma_2), \quad (8)$$

and additionally, the scaled Chebyshev distance lower bounds the Spearman's footrule with high probability.

Theorem 4. For any $\pi \in \mathcal{S}_n$, let σ be a permutation chosen uniformly from \mathcal{S}_n , then

$$\mathbb{P}[c_2 \cdot n \cdot d_{\ell_\infty}(\pi, \sigma) \leq d_{\ell_1}(\pi, \sigma)] = 1 - O(1/n) \quad (9)$$

for any positive constant $c_2 < 1/3$.

Proof: See Section V-B. ■

IV. RATE DISTORTION FUNCTIONS

In this section we build upon the results in Section III and prove the equivalence of lossy source codes under different distortion measures in Theorem 5, which lead to the rate distortion functions in Theorem 6.

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 proof is based on the relationships between various distortion measures investigated in Section III and we present the details in Section V-C.

We obtain Theorem 6 as a direct consequence of Theorem 5.

Theorem 6 (Rate distortion functions for distortion measures). *For permutation spaces $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$, $\mathcal{X}(\mathcal{S}_n, d_\tau)$, and $\mathcal{X}(\mathcal{S}_n, d_{\ell_1})$, and for $0 < \delta \leq 1$,*

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

For the permutation space $\mathcal{X}(\mathcal{S}_n, d_{\ell_\infty})$ and $0 < \delta \leq 1$,

$$\bar{R}(D_n) = \hat{R}(D_n) = \begin{cases} 1 & \text{if } D_n = O(1) \\ 1 - \delta & \text{if } D_n = \Theta(n^\delta). \end{cases}$$

Proof: For achievability, we note that the achievability for permutation spaces $\mathcal{X}(\mathcal{S}_n, d_\tau)$ and $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ under worst-case distortion is provided in [3, Theorem 6 and 8], which state that

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

The achievability for other permutation spaces then follows from Theorem 5.

For converse, we observe observation that for uniform distribution over \mathcal{S}_n , the rate-distortion functions for $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ is the same under average-case and worst-case distortions, as pointed out in [3, Remark 2]. Then the converse for other permutation spaces follows from Theorem 5. \blacksquare

Remark 3. *Because the rate distortion functions under average-case and worst-case distortion coincides, if we require*

$$\lim_{n \rightarrow \infty} \mathbb{P}[d(f_n(\sigma), \sigma) > D_n] = 0 \quad (10)$$

instead of $\mathbb{E}[d(f_n(\sigma), \sigma)] \leq D_n$ in Definition 1, then the asymptotic rate-distortion trade-off remains the same.

Theorem 5 indicates that for all the distortion measures in this paper, the lossy compression scheme for one measure preserves distortion under other measures, and hence all compression schemes can be used interchangeably under average-case distortion, after transforming the permutation representation and scaling the distortion correspondingly.

For the vector representation of permutation, compression based on Kendall tau distance is essentially optimal, which can be achieved by partitioning each permutation vector into subsequences with proper sizes and sorting them accordingly [3]. For the inversion vector representation of permutation, a simple component-wise scalar quantization achieves the optimal rate distortion trade-off, as shown in [3]. In particular, given $D = cn^{1+\delta}$, $0 < \delta < 1$, for the $(k-1)$ -th component of the inversion vector ($k = 2, \dots, n$), we quantize k points in $[0 : k-1]$ uniformly with $m_k = \lceil kn/(2D) \rceil$ points, resulting component-wise average distortion $D_k = D/n$ and overall average distortion $= \sum_{k=2}^n D_k \leq D$, and log of codebook size $\log M_n = \sum_{k=2}^n \log m_k = \sum_{k=2}^n \log \lceil kn/(2D) \rceil = (1-\delta)n \log n - O(n)$.

Remark 4. *This scheme is slightly different from the one in [3] as it is designed for average distortion, while the latter for worst-case distortion.*

Remark 5. *While the compression algorithm in $\mathcal{X}(\mathcal{S}_n, d_{\mathbf{x}, \ell_1})$ is conceptually simple and has time complexity $\Theta(n)$, it takes $\Theta(n \log n)$ runtime to convert a permutation from its vector representation to its inversion vector representation [5, 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.*

V. PROOFS

A. Proof of Theorem 3

To prove Theorem 3, we analyze the mean and variance of the Kendall tau distance and ℓ_1 distance of inversion vectors between a permutation in \mathcal{S}_n and a randomly selected permutation, in Lemma 8 and Lemma 9 respectively.

We first state the following fact without proof.

Lemma 7. *Let σ be a permutation chosen uniformly from \mathcal{S}_n , then $\mathbf{x}_\sigma(i)$ is uniformly distributed in $[0 : i]$, $1 \leq i \leq n-1$.*

Lemma 8. *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*

$$\mathbb{E}[X_\tau] = \frac{n(n-1)}{4}, \quad (11)$$

$$\text{Var}[X_\tau] = \frac{n(2n+5)(n-1)}{72}. \quad (12)$$

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 [4], then $d_\tau(\pi, \sigma) = d_\tau(\pi\sigma^{-1}, e)$ and $d_\tau(\sigma', \sigma) = d_\tau(\sigma'\sigma^{-1}, e)$ are identically distributed, and hence the result follows [2, Table 1] and [5, Section 5.1.1]. ■

Lemma 9. 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: By Lemma 7, we have $X_{\mathbf{x}, \ell_1} = \sum_{i=1}^{n-1} |a_i - U_i|$, 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 8 and Lemma 9, now we show that the event that a scaled version of the Kendall tau distance is larger than the ℓ_1 distance of inversion vectors is unlikely.

Proof for Theorem 3: Let $c_1 = 1/3$, 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 $c_1 < 1/2$ can be proved similarly. ■

B. Proof for Theorem 4

Lemma 10. 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

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

Proof: See [2, Table 1]. ■

Proof for Theorem 4: For any $c > 0$, $cn \cdot d_{\ell_\infty}(\pi, \sigma) \leq cn(n-1)$, and for any $c_2 < 1/3$, Lemma 10 and Chebyshev inequality indicate $\mathbb{P}[d_{\ell_1}(\pi, \sigma) < c_2 n(n-1)] = O(1/n)$. Therefore,

$$\begin{aligned} &\mathbb{P}[d_{\ell_1}(\pi, \sigma) \geq c_2 n \cdot d_{\ell_\infty}(\pi, \sigma)] \\ &\geq \mathbb{P}[d_{\ell_1}(\pi, \sigma) \geq c_2 n(n-1)] \\ &= 1 - \mathbb{P}[d_{\ell_1}(\pi, \sigma) < c_2 n(n-1)] \\ &= 1 - O(1/n). \end{aligned}$$

C. Proof for Theorem 5

Proof: Statement 1 follows from (8).

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 (6).

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 4 indicates that $|\mathcal{B}_n(\pi)| = (1 - O(1/n))n!$. Let \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 4,

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

The proof of Statement 6 is analogous to Statement 5. ■

REFERENCES

- [1] J. Giesen, E. Schubert, and M. Stojakovi, "Approximate sorting," in *LATIN 2006: Theoretical Informatics*. Berlin, Heidelberg: Springer, 2006, vol. 3887, pp. 524–531.
- [2] 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.
- [3] D. Wang, A. Mazumdar, and G. W. Wornell, "A rate-distortion theory for permutation spaces," in *Proc. IEEE Int. Symp. Inform. Th. (ISIT)*, 2013, pp. 2562–2566.
- [4] M. Deza and T. Huang, "Metrics on permutations, a survey," *Journal of Combinatorics, Information and System Sciences*, vol. 23, pp. 173–185, 1998.
- [5] D. E. Knuth, *Art of Computer Programming, Volume 3: Sorting and Searching*, 2nd ed. Addison-Wesley Professional, 1998.