Bounds for Permutation Rate-Distortion

Farzad Farnoud (Hassanzadeh), Moshe Schwartz, *Senior Member, IEEE*, and Jehoshua Bruck, *Fellow, IEEE*

***Abstract*— We study the rate-distortion relationship in the set of permutations endowed with the Kendall τ-metric and the Chebyshev metric (the ℓ_∞-metric). This paper is motivated by the application of permutation rate-distortion to the average-case and worst-case distortion analysis of algorithms for ranking with incomplete information and approximate sorting algorithms. For the Kendall τ-metric, we provide bounds for various distortion regimes, while for the Chebyshev metric, we present bounds that are valid for all distortions and are especially accurate for small distortions. In addition, for the Chebyshev metric, we provide a construction for covering codes.**

***Index Terms*— Rank modulation, rate-distortion, covering codes, permutations, Kendall τ-metric, Chebyshev metric, ℓ_∞-metric.**

I. Introduction

In the analysis of sorting and ranking algorithms, it is often assumed that complete information is available, that is, the answer to *every* question of the form “is $x > y$?” can be found, either by query or computation. A standard and straightforward result in this setting is that, on average, one needs at least $\log_2 n!$ pairwise comparisons to sort a randomly-chosen permutation of length n. In practice, however, it is usually the case that only partial information is available. One example is the learning-to-rank problem, where the solutions to pairwise comparisons are learned from data, which may be incomplete or difficult/expensive to collect [14], or in big-data settings, where the number of items may be so large as to make it impractical to query every pairwise comparison [12]. It may also be the case that only an approximately-sorted list is required, and thus one does not seek the solutions to all pairwise comparisons. In such cases, the question that arises is what is the quality of a ranking obtained from incomplete data, or an approximately-sorted list [12]–[14], [29].

One approach to quantify the quality of an algorithm that ranks with incomplete data is to find the relationship between the number of pairwise comparisons performed by the algorithm and the average, or worst-case, quality of the output ranking, as measured via a metric on the space of permutations. To explain, consider a deterministic algorithm for ranking n items that makes nR queries and outputs a ranking of length n. Suppose that the true ranking is ω. The information about ω is available to the algorithm only through the queries it makes. Since the algorithm is deterministic, the output, denoted $f(\omega)$, is uniquely determined by ω. The “distortion” of this output can be measured with a metric d as $\mathsf{d}(\omega, f(\omega))$. The goal is to find the relationship between R and $\mathsf{d}(\omega, f(\omega))$ when ω is chosen at random, and when it is chosen by an adversary.

A general way to quantify the best possible performance by such an algorithm is to use the rate-distortion theory on the space of permutations. In this context, the codebook is the set $\{f(\omega) : \omega \in \mathbb{S}_n\}$, where $\mathbb{S}_n$ is the set of permutations of length n, and the rate is determined by the number of queries. For a given rate, no algorithm can have a smaller distortion than what is dictated by rate-distortion.

For example, for the worst-case analysis, we are interested in studying codes $C \subseteq \mathbb{S}_n$, such that for any $\omega \in \mathbb{S}_n$ there exists $\omega' \in C$ which satisfies

$$\mathsf{d}(\omega, \omega') \leqslant D,$$

for some D. By setting $f(\omega) = \omega'$, we are guaranteed every point in space is distorted by f by no more than D. Such a code C is called a *covering code*, since balls of radius D that are centered around the codewords, cover the entire space.

An important ingredient is the choice of metric to use. A wide variety of metrics can be applied in various scenarios to permutations, including the Kendall τ-metric, Spearman’s footrule, the Chebyshev metric, and the Ulam metric [8]. In particular, the Kendall τ-metric is commonly used to compare and aggregate rankings [9], [10]. Recently, in coding theory, it was suggested that the rank-modulation scheme may alleviate some of the problems associated with reliable storage of information in non-volatile memories [16]. Subsequent works [3], [17], [21], [27], [28], focused on error-correcting codes, advocated the use of two metrics in particular in the context of rank modulation: the Kendall τ-metric, which counts the number of pairs that are ranked incorrectly, and the Chebyshev metric (also called the ℓ_∞-metric), which is the largest error in the rank of any item.

With this motivation, we study rate-distortion and covering codes in the space of permutations under the Kendall τ-metric and the Chebyshev metric. Rate-distortion and covering codes over permutations have only been recently studied in depth, starting with the work of [6], and followed by [18] and [24],

Manuscript received November 19, 2014; revised September 13, 2015; accepted November 16, 2015. Date of publication December 4, 2015; date of current version January 18, 2016. This work was supported in part by the National Science Foundation under Grant CIF-1218005 and in part by the Israel Science Foundation under Grant 130/14. This paper was presented in part at the 2014 IEEE International Symposium on Information Theory.

F. Farnoud and J. Bruck are with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: farnoud@caltech.edu; bruck@paradise.caltech.edu).

M. Schwartz is with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer Sheva 8410501, Israel (e-mail: schwartz@ee.bgu.ac.il).

Communicated by E. Tuncel, Associate Editor for Source Coding.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2015.2504521

all of which only use the Hamming distance over permutations. The paper [14] considers Spearman's footrule as a measure of distortion over permutations and studies comparison-based approximate sorting algorithms. Finally, the work of [29] considers the asymptotics of permutation covering codes using Kendall's τ-metric and the ℓ_1-metric of inversion vectors. The latest is the only work on rate-distortion and covering codes over permutations that studies metrics motivated by the application to non-volatile memories.

Our results on the Kendall τ-metric improve upon those presented in [29]. In particular, for the small distortion regime, $D = cn + O(1)$ (for $c > 0$), we eliminate the gap between the lower bound and the upper bound given in [29]; for the large distortion regime, $D = cn^2 + O(n)$, we provide a stronger lower bound; and for the medium distortion regime, $D = cn^{1+\alpha} + O(n)$ (for $0 < \alpha < 1$), we provide upper and lower bounds with error terms. Our study includes both worst-case and average-case distortions for the Kendall τ-metric and for the Chebyshev metric, as both measures are frequently used in the analysis of algorithms. Note that permutation rate-distortion results can also be applied to lossy compression of permutations, e.g., rank-modulation signals [16]. Finally, we also present covering codes for the Chebyshev metric, where covering codes for the Kendall τ-metric were already presented in [29]. The codes are the covering analog of the error-correcting codes already presented in [3], [17], [21], and [27].

The rest of the paper is organized as follows. In Section II, we present preliminaries and notation. Section III contains non-asymptotic results valid for both metrics under study. Section IV and Section V focus on the Kendall τ-metric and the Chebyshev metric, respectively. Finally, concluding remarks are presented in Section VI.

II. Preliminaries and Definitions

For a nonnegative integer n, let $[n]$ denote the set $\{1, \dots, n\}$, and let $\mathbb{S}_n$ denote the set of permutations of $[n]$. We denote a permutation $\sigma \in \mathbb{S}_n$ as $\sigma = [\sigma_1, \sigma_2, \dots, \sigma_n]$, where the permutation sets $\sigma(i) = \sigma_i$. We also denote the identity permutation by $\mathrm{Id} = [1, 2, \dots, n]$.

The Kendall τ-metric between two permutations $\omega, \sigma \in \mathbb{S}_n$ is the number of transpositions of adjacent elements needed to transform ω into σ, and is denoted by $\mathsf{d}_\mathsf{K}(\omega, \sigma)$. In contrast, the Chebyshev distance between ω and σ is defined as

$$\mathsf{d}_\mathsf{C}(\omega, \sigma) = \max_{i \in [n]} |\omega(i) - \sigma(i)|.$$

In the following we explore some properties of $\mathbb{S}_n$ under either d_K or d_C. Some of these properties are common to both d_K and d_C, and in these cases we shall use the notation $\mathsf{d}(\omega, \sigma)$ to denote the distance between ω and σ in either of the two metrics.

Both d_K and d_C are invariant; the former is left-invariant and the latter is right-invariant [8]. In other words, for all $f, g, h \in \mathbb{S}_n$,

$$\mathsf{d}_\mathsf{K}(f, g) = \mathsf{d}_\mathsf{K}(hf, hg) \quad \text{and} \quad \mathsf{d}_\mathsf{C}(f, g) = \mathsf{d}_\mathsf{C}(fh, gh).$$

Hence, the size of the ball of a given radius in either metric does not depend on its center. The size of a ball of radius r with respect to d_K, d_C, and d, is given, respectively, by $\mathsf{B}_\mathsf{K}(r)$, $\mathsf{B}_\mathsf{C}(r)$, and $\mathsf{B}(r)$. The dependence of the size of the ball on n is implicit.

A code C is a subset $C \subseteq \mathbb{S}_n$. For a code C and a permutation $\omega \in \mathbb{S}_n$, let

$$\mathsf{d}(\omega, C) = \min_{\sigma \in C} \mathsf{d}(\omega, \sigma)$$

be the (minimal) distance between ω and C.

We use $\hat{\mathsf{M}}(D)$ to denote the minimum number of codewords required for a worst-case distortion D. That is, $\hat{\mathsf{M}}(D)$ is the size of the smallest code C such that for all $\omega \in \mathbb{S}_n$, we have $\mathsf{d}(\omega, C) \leqslant D$. Similarly, let $\bar{\mathsf{M}}(D)$ denote the minimum number of codewords required for an average distortion D under the uniform distribution on $\mathbb{S}_n$, that is, the size of the smallest code C such that

$$\frac{1}{n!} \sum_{\omega \in \mathbb{S}_n} \mathsf{d}(\omega, C) \leqslant D.$$

Note that $\bar{\mathsf{M}}(D) \leqslant \hat{\mathsf{M}}(D)$. In what follows, we assume that the distortion D is an integer. For worst-case distortion (but not for average-case distortion), this assumption does not lead to a loss of generality as the metrics under study are integer valued.

We also define

$$\hat{\mathsf{R}}(D) = \frac{1}{n} \lg \hat{\mathsf{M}}(D), \quad \bar{\mathsf{R}}(D) = \frac{1}{n} \lg \bar{\mathsf{M}}(D),$$
$$\hat{\mathsf{A}}(D) = \frac{1}{n} \lg \frac{\hat{\mathsf{M}}(D)}{n!}, \quad \bar{\mathsf{A}}(D) = \frac{1}{n} \lg \frac{\bar{\mathsf{M}}(D)}{n!},$$

where we use lg as a shorthand for $\log_2$. It is clear that

$$\hat{\mathsf{R}}(D) = \hat{\mathsf{A}}(D) + \frac{\lg n!}{n}, \quad \bar{\mathsf{R}}(D) = \bar{\mathsf{A}}(D) + \frac{\lg n!}{n}.$$

The reason for defining $\hat{\mathsf{A}}$ and $\bar{\mathsf{A}}$ is that they sometimes lead to simpler expressions compared to $\hat{\mathsf{R}}$ and $\bar{\mathsf{R}}$. Furthermore, $\hat{\mathsf{A}}$ (resp. $\bar{\mathsf{A}}$) can be interpreted as the difference between the number of bits per symbol required to identify a codeword in a code of size $\hat{\mathsf{M}}$ (resp. $\bar{\mathsf{M}}$) and the number of bits per symbol required to identify a permutation in $\mathbb{S}_n$.

Throughout the paper, for $\hat{\mathsf{M}}$, $\bar{\mathsf{M}}$, $\hat{\mathsf{A}}$, $\bar{\mathsf{A}}$, $\hat{\mathsf{R}}$, and $\bar{\mathsf{R}}$, subscripts K and C denote that the subscripted quantity corresponds to the Kendall τ-metric and the Chebyshev metric, respectively. Lack of subscripts indicates that the result is valid for both metrics.

In the sequel, bounds on the binomial coefficient and Stirling's approximation (for example, see [7]) will be useful,

$$\frac{2^{nH(p)}}{\sqrt{8np(1-p)}} \leqslant \binom{n}{pn} \leqslant \frac{2^{nH(p)}}{\sqrt{2\pi np(1-p)}}, \tag{1}$$
$$\sqrt{2\pi n}(n/e)^n < n! < \sqrt{2\pi n}(n/e)^n e^{1/(12n)}, \tag{2}$$

where $H(\cdot)$ is the binary entropy function and $0 < p < 1$. Furthermore, to denote $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$, we use

$$f(x) \sim g(x) \text{ as } x \to \infty,$$

or if the variable x is clear from the context, we simply write $f \sim g$.

TABLE I

A SUMMARY OF THE ASYMPTOTIC BOUNDS ON $\hat{\mathsf{A}}_K(D)$ AND $\bar{\mathsf{A}}_K(D)$, WHERE $c > 0$, $0 < \alpha < 1$ ARE CONSTANTS

Regime	Bound	Location
$D = cn + O(1)$	$\bar{\mathsf{A}}_K(D), \hat{\mathsf{A}}_K(D) = -\lg \frac{(1+c)^{1+c}}{c^c} + O\left(\frac{\lg n}{n}\right)$	Lemmas 6 and 9
$D = cn^{1+\alpha} + O(n)$	$-\lg(ecn^{\alpha}) + O(n^{-\alpha}) \leqslant \bar{\mathsf{A}}_K(D) \leqslant \hat{\mathsf{A}}_K(D) \leqslant -\lg(cn^{\alpha}) + O(n^{-\alpha} + n^{\alpha-1})$	Lemma 10
$D = cn^2 + O(n)$	$-\lg(ecn) + O\left(\frac{1}{n}\right) \leqslant \hat{\mathsf{A}}_K(D) \leqslant -\lg(ecn) + (1+c)\lg e + O\left(\frac{\lg n}{n}\right)$ $-\lg(ecn) + O\left(\frac{\lg n}{n}\right) \leqslant \bar{\mathsf{A}}_K(D) \leqslant \hat{\mathsf{A}}_K(D)$	Lemma 11

III. NON-ASYMPTOTIC BOUNDS

In this section, we derive non-asymptotic bounds, that is, bounds that are valid for all positive integers n and D. The results in this section apply to both the Kendall τ-metric and the Chebyshev distance, as well as any other left-invariant or right-invariant distances on permutations.

The next lemma gives two basic lower bounds for $\hat{\mathsf{M}}(D)$ and $\bar{\mathsf{M}}(D)$.

Lemma 1: For all $n, D \in \mathbb{N}$,

$$\hat{\mathsf{M}}(D) \geqslant \frac{n!}{\mathsf{B}(D)}, \quad \bar{\mathsf{M}}(D) > \frac{n!}{\mathsf{B}(D)(D+1)}.$$

Proof: The first inequality follows from the fact that every codeword covers at most $\mathsf{B}(D)$ permutations of $\mathbb{S}_n$. For the second inequality, fix n and D. Consider a code $C \subseteq \mathbb{S}_n$ of size M and suppose the average distortion of this code is at most D. There are at most $M\mathsf{B}(D)$ permutations ω such that $\mathsf{d}(\omega, C) \leqslant D$ and at least $n! - M\mathsf{B}(D)$ permutations ω such that $\mathsf{d}(\omega, C) \geqslant D + 1$. Hence,

$$D > (D+1)\left(1 - \frac{M\mathsf{B}(D)}{n!}\right).$$

The second inequality then follows. ■

The following theorem by Stein [26] can be used to obtain existence results for covering codes (see, e.g., [7]), and thus provide upper bounds.

Theorem 2 [26, Th. 2]: Consider a finite set X of cardinality N, and a family $\{A_i\}_{i=1}^t$ of sets that cover X, with $|A_i| \leqslant a$ for all i. Suppose each element of X is in at least q of the sets A_i. Then there is subfamily of $\{A_i\}_{i=1}^t$ containing at most

$$\frac{N}{a} + \frac{t}{q}\left(\frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{a}\right) \leqslant \frac{N}{a} + \frac{t}{q}\ln a$$

sets that cover X.

In our context X is $\mathbb{S}_n$, A_i are the balls of radius D centered at each permutation, and therefore $N = t = n!$ and $a = q = \mathsf{B}(D)$. Hence, the theorem implies that

$$\hat{\mathsf{M}}(D) \leqslant \frac{n!}{\mathsf{B}(D)}(1 + \ln \mathsf{B}(D)).$$

The following theorem summarizes the results of this section.

Theorem 3: For all $n, D \in \mathbb{N}$,

$$\frac{n!}{\mathsf{B}(D)} \leqslant \hat{\mathsf{M}}(D) \leqslant \frac{n!}{\mathsf{B}(D)}(1 + \ln \mathsf{B}(D)), \quad (3)$$

$$\frac{n!}{\mathsf{B}(D)(D+1)} < \bar{\mathsf{M}}(D) \leqslant \hat{\mathsf{M}}(D). \quad (4)$$

IV. THE KENDALL τ-METRIC

The goal of this section is to consider the rate-distortion relationship for the permutation space endowed by the Kendall τ-metric. First, we find non-asymptotic upper and lower bounds on the size of the ball in the Kendall τ-metric. Then, in the following subsections, we consider asymptotic bounds for the small, medium, and large distortion regimes. To help the reader navigate the various asymptotic results in this metric, a summary is given in Table I.

Throughout this section, we assume $1 \leqslant D < \frac{1}{2}\binom{n}{2}$ and $n \geqslant 1$. Note that D is upper bounded by $\binom{n}{2}$, and the case of $\frac{1}{2}\binom{n}{2} \leqslant D \leqslant \binom{n}{2}$ leads to trivial codes, e.g., $\{\text{Id}, [n, n-1, \ldots, 1]\}$ and $\{\text{Id}\}$.

A. Non-Asymptotic Results

Let $\mathbb{X}_n$ be the set of integer vectors $x = x_1, x_2, \ldots, x_n$ of length n such that $0 \leqslant x_i \leqslant i - 1$ for all $i \in [n]$. It is well known (for example, see [17]) that there is a bijection between $\mathbb{X}_n$ and $\mathbb{S}_n$ such that for corresponding elements $x \in \mathbb{X}_n$ and $\omega \in \mathbb{S}_n$, we have

$$\mathsf{d}_{\mathsf{K}}(\omega, \text{Id}) = \sum_{i=2}^{n} x_i.$$

Hence

$$\mathsf{B}_{\mathsf{K}}(r) = \left|\left\{x \in \mathbb{X}_n : \sum_{i=2}^{n} x_i \leqslant r\right\}\right|, \quad (5)$$

for $1 \leqslant r \leqslant \binom{n}{2}$. Thus, the number of nonnegative integer solutions to the equation $\sum_{i=2}^n x_i \leqslant r$ is at least $\mathsf{B}_{\mathsf{K}}(r)$, i.e.,

$$\mathsf{B}_{\mathsf{K}}(r) \leqslant \binom{r+n-1}{r}. \quad (6)$$

This bound is already known, appearing as [29, Lemma 1].

Furthermore, for $\delta \geqslant 0$ such that δn is an integer, it can be shown that

$$\mathsf{B}_{\mathsf{K}}(\delta n) \geqslant \lfloor 1+\delta \rfloor! \lfloor 1+\delta \rfloor^{n - \lfloor 1+\delta \rfloor}, \quad (7)$$

by noting the facts that the right-hand side of (7) counts the elements of $\mathbb{X}_n$ such that

$$\begin{cases} 0 \leqslant x_i \leqslant i - 1, & \text{for } i \leqslant \lfloor 1+\delta \rfloor, \\ 0 \leqslant x_i \leqslant \lfloor \delta \rfloor, & \text{for } i > \lfloor 1+\delta \rfloor, \end{cases}$$

and that

$$\left(\sum_{i \leqslant \lfloor 1+\delta \rfloor} (i-1)\right) + (n - \lfloor 1+\delta \rfloor)\lfloor \delta \rfloor \leqslant \lfloor \delta \rfloor n \leqslant \delta n.$$

Next we find a lower bound on $\mathsf{B}_{\mathsf{K}}(r)$ for $r < n$. Let $I(n, r)$ denote the number of permutations in $\mathbb{S}_n$ that are at distance r from the identity. We have [4, p. 51] (or [20, p. 15])

$$I(n,r) = \binom{n+r-1}{r} - \left(\binom{n+r-2}{r-1} + \binom{n+r-3}{r-2}\right) + \sum_{j=2}^{\infty} (-1)^j f_j,$$

where

$$f_j = \binom{n+r-(u_j-j)-1}{r-(u_j-j)} + \binom{n+r-u_j-1}{r-u_j},$$

and $u_j = (3j^2 + j)/2$. For $j \geqslant 2$, we have $f_j \geqslant f_{j+1}$. Thus, for $r < n$,

$$I(n,r) \geqslant \binom{n+r-1}{r}\left(1 - \frac{r}{n+r-1}\left(1 + \frac{r-1}{n+r-2}\right)\right) \geqslant \frac{1}{4}\binom{n+r-1}{r}.$$

Hence, for $r < n$, we have

$$\mathsf{B}_{\mathsf{K}}(r) \geqslant \frac{1}{4}\binom{n+r-1}{r}. \tag{8}$$

In the next two theorems, we use the aforementioned bounds on $\mathsf{B}_{\mathsf{K}}(r)$ to derive lower and upper bounds on $\hat{\mathsf{A}}_K(D)$ and $\bar{\mathsf{A}}_K(D)$.

Theorem 4: For all $n, D \in \mathbb{N}$, and $\delta = D/n$,

$$\hat{\mathsf{A}}_K(D) \geqslant -\lg \frac{(1+\delta)^{1+\delta}}{\delta^\delta},$$
$$\bar{\mathsf{A}}_K(D) \geqslant -\lg \frac{(1+\delta)^{1+\delta}}{\delta^\delta} - \frac{\lg n}{n}.$$

Proof: For the worst-case distortion, we have

$$\mathsf{B}_{\mathsf{K}}(D) \overset{(a)}{\leqslant} \binom{n+\delta n-1}{\delta n} \leqslant \binom{(1+\delta)n}{\delta n} \overset{(b)}{\leqslant} \frac{2^{n(1+\delta)H\left(\frac{1}{1+\delta}\right)}}{\sqrt{2\pi n\delta/(1+\delta)}} \overset{(c)}{\leqslant} 2^{n(1+\delta)H\left(\frac{1}{1+\delta}\right)},$$

where (a) follows from (6), (b) follows from (1), and (c) follows from the facts that $\delta \geqslant 1/n$ and $n \geqslant 1$. The first result then follows from (3).

For the case of average distortion, we proceed as follows:

$$\begin{aligned}
\mathsf{B}_{\mathsf{K}}(D)(D+1) &= \mathsf{B}_{\mathsf{K}}(\delta n)(\delta n+1) \\
&\leqslant \binom{n+\delta n-1}{\delta n}(\delta n+1) \\
&= \binom{n+\delta n}{\delta n}\frac{\delta n+1}{1+\delta} \\
&\overset{(a)}{\leqslant} 2^{n(1+\delta)H\left(\frac{1}{1+\delta}\right)}\frac{\delta n+1}{\sqrt{2\pi n\delta(1+\delta)}} \\
&= 2^{n(1+\delta)H\left(\frac{1}{1+\delta}\right)}\sqrt{\frac{2\delta n}{\pi}}\,\frac{1+1/(\delta n)}{2\sqrt{1+\delta}} \\
&\overset{(b)}{\leqslant} 2^{n(1+\delta)H\left(\frac{1}{1+\delta}\right)}\sqrt{2\delta n/\pi},
\end{aligned}$$

where (a) follows from (1) and (b) is proved as follows. The expression $\frac{1+1/(\delta n)}{2\sqrt{1+\delta}}$ is decreasing in δ for positive δ and so it is maximized by letting $\delta = 1/n$. Hence,

$$\frac{1+1/(\delta n)}{2\sqrt{1+\delta}} \leqslant \frac{1}{\sqrt{1+1/n}} \leqslant 1.$$

Now, using (4) leads to (a stronger version of) the statement in the theorem. ■

Theorem 5: Assume $n, D \in \mathbb{N}$, and let $\delta = D/n$. We have

$$\bar{\mathsf{A}}_K(D) \leqslant \hat{\mathsf{A}}_K(D) \leqslant -\lg \frac{(1+\delta)^{1+\delta}}{\delta^\delta} + \frac{3\lg n + 12}{2n},$$

for $\delta < 1$, and

$$\bar{\mathsf{A}}_K(D) \leqslant \hat{\mathsf{A}}_K(D) \leqslant -\lg\lfloor 1+\delta\rfloor + \frac{1}{n}\lg\left(ne^{\lfloor 1+\delta\rfloor}\ln\lfloor 1+\delta\rfloor\right),$$

for $\delta \geqslant 1$.

Proof: For $\delta < 1$, we have

$$\begin{aligned}
\mathsf{B}_{\mathsf{K}}(D) = \mathsf{B}_{\mathsf{K}}(\delta n) &\geqslant \frac{1}{4}\binom{n+\delta n-1}{\delta n} \\
&\geqslant \frac{n}{4(n+\delta n)}\binom{n+\delta n}{\delta n} \\
&\geqslant \frac{1}{4(1+\delta)}\cdot\frac{2^{n(1+\delta)H\left(\frac{1}{1+\delta}\right)}}{\sqrt{8n\delta/(1+\delta)}} \\
&= \frac{1}{4}\cdot\frac{2^{n(1+\delta)H\left(\frac{1}{1+\delta}\right)}}{\sqrt{8n\delta(1+\delta)}} \geqslant \frac{2^{n(1+\delta)H\left(\frac{1}{1+\delta}\right)}}{16\sqrt{n}},
\end{aligned}$$

where the first inequality follows from (8) and the last step follows from the fact that $\delta \leqslant 1$, and so $\delta(1+\delta) \leqslant 2$.

Since $(1+\ln x)/x$ is a decreasing function for $x \geqslant 1$, we can substitute the above lower bound on $\mathsf{B}_{\mathsf{K}}(D)$ in (3) to obtain

$$\begin{aligned}
\hat{\mathsf{M}}(D) &\leqslant \frac{16n!\sqrt{n}}{2^{n(1+\delta)H\left(\frac{1}{1+\delta}\right)}}\ln\left(\frac{e2^{n(1+\delta)H\left(\frac{1}{1+\delta}\right)}}{16\sqrt{n}}\right) \\
&\overset{(a)}{\leqslant} \frac{16n!n^{3/2}}{2^{n(1+\delta)H\left(\frac{1}{1+\delta}\right)}}(1+\delta)H\left(\frac{1}{1+\delta}\right)\ln 2 \\
&\overset{(b)}{\leqslant} \frac{64n!n^{3/2}}{2^{n(1+\delta)H\left(\frac{1}{1+\delta}\right)}},
\end{aligned}$$

where (a) follows from the fact that $e \leqslant 16\sqrt{n}$ and (b) from the fact that for $\delta \leqslant 1$, we have $(1+\delta)H(1/(1+\delta))\ln 2 \leqslant 2H(1/2)\ln 2 \leqslant 4$. Thus

$$\hat{\mathsf{A}}_K(D) \leqslant -\lg\frac{(1+\delta)^{1+\delta}}{\delta^\delta} + \frac{3\lg n + 12}{2n}.$$

For $\delta \geqslant 1$, by (7) and (2) we have

$$\mathsf{B}_{\mathsf{K}}(D) = \mathsf{B}_{\mathsf{K}}(\delta n) \geqslant \lfloor 1+\delta\rfloor!\lfloor 1+\delta\rfloor^{n-\lfloor 1+\delta\rfloor} \geqslant \frac{\lfloor 1+\delta\rfloor^n}{e^{\lfloor 1+\delta\rfloor}},$$

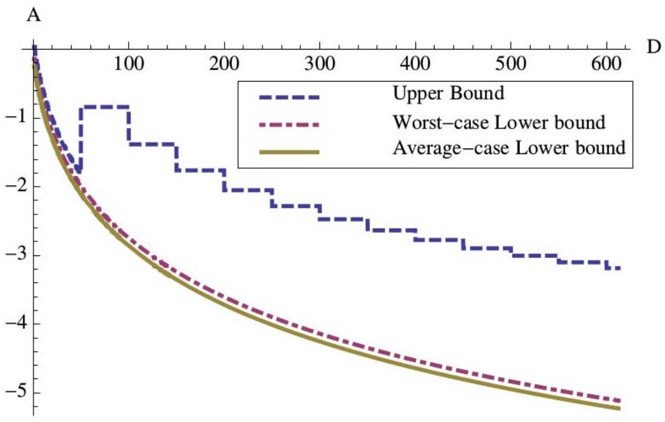

Fig. 1. Upper bound and lower bounds for $n = 50$ from Theorems 4 and 5.

implying

$$\begin{aligned}\hat{\mathsf{A}}_K(D) &\leqslant \frac{1}{n}\lg\frac{1+\ln \mathsf{B}_\mathsf{K}(\delta n)}{\mathsf{B}_\mathsf{K}(\delta n)}\\ &\leqslant \frac{1}{n}\lg\frac{e^{\lfloor 1+\delta\rfloor}}{\lfloor 1+\delta\rfloor^n}+\frac{1}{n}\lg\left(1+n\ln\lfloor 1+\delta\rfloor-\lfloor 1+\delta\rfloor\right)\\ &\leqslant \frac{1}{n}\lg\frac{e^{\lfloor 1+\delta\rfloor}}{\lfloor 1+\delta\rfloor^n}+\frac{1}{n}\lg\left(n\ln\lfloor 1+\delta\rfloor\right)\\ &\leqslant -\lg\lfloor 1+\delta\rfloor+\frac{1}{n}\lg\left(ne^{\lfloor 1+\delta\rfloor}\ln\lfloor 1+\delta\rfloor\right).\end{aligned}$$

■

The plots for the expressions given in Theorems 4 and 5 are given in Figure 1.

B. Small Distortion

In this subsection, we consider small distortions, that is, $D = O(n)$. First, suppose $D < n$, or equivalently, $\delta = D/n < 1$.

Lemma 6: For $\delta = D/n < 1$, we have that

$$\hat{\mathsf{A}}_K(D) = -\lg\frac{(1+\delta)^{1+\delta}}{\delta^\delta}+O\left(\frac{\lg n}{n}\right), \tag{9}$$

and that $\bar{\mathsf{A}}_K(D)$ satisfies the same equation.

Proof: The lemma is an immediate consequence of Theorems 4 and 5. ■

Next, let us consider the case of $D = \Theta(n)$. We introduce the following notation. Assume

$$f(z) = \sum_{i=0}^{\infty} a_i z^i,$$

is a formal power series. We denote the coefficient of z^i as $[z^i]f(z)$, i.e.,

$$[z^i]f(z) = a_i.$$

As was already mentioned in [17], from (5), it follows that

$$\mathsf{B}_\mathsf{K}(k) = [z^k]\frac{1}{1-z}\prod_{i=2}^{n}\frac{1-z^i}{1-z} = [z^k]\frac{\prod_{i=2}^{n}\left(1-z^i\right)}{(1-z)^n}.$$

Let

$$\begin{aligned} g(k,n) &= \binom{n+k-1}{k}^{-1}\mathsf{B}_\mathsf{K}(k),\\ \gamma(z,n) &= \sum_{i=0}^{\infty}\Gamma_i(n)z^i = \prod_{i=2}^{n}\left(1-z^i\right), \end{aligned} \tag{10}$$

and

$$f(z,n) = \sum_{i=0}^{\infty} F_i(n)z^i = \frac{1}{(1-z)^n},$$

where

$$F_i(n) = \binom{n+i-1}{i},$$

so that

$$g(k,n) = \frac{1}{F_k(n)}[z^k]f(z,n)\gamma(z,n).$$

We use the following theorem to find the asymptotics of $g(k,n)$ and $\mathsf{B}_\mathsf{K}(k)$ using the asymptotics of $\gamma(z,n)$ in Theorem 8.

Theorem 7 [22, Th. 3.1]: Let $f(z,n)$ and $\gamma(z,n)$ be two functions with Taylor series for all n,

$$f(z,n) = \sum_{i=0}^{\infty} F_i(n)z^i, \quad \gamma(z,n) = \sum_{i=0}^{\infty}\Gamma_i(n)z^i,$$

where $F_i(n) > 0$ for all sufficiently large n. Suppose

$$g(k,n) = \frac{1}{F_k(n)}[z^k]f(z,n)\gamma(z,n),$$

and let $n = n(k)$ be a function of k such that the limit $\rho = \lim_{k\to\infty}\frac{F_{k-1}(n(k))}{F_k(n(k))}$ exists. We have

$$g(k,n(k)) \sim \gamma(\rho,n(k)) \text{ as } k\to\infty,$$

provided that

1) *for all sufficiently large k and for all i,*
$$\left|\frac{\Gamma_i(n(k))}{\gamma(\rho,n(k))}\right| \leqslant p_i,$$
where $\sum_{i=0}^{\infty} p_i\rho^i < \infty$, and
2) *there exists a constant b, such that for all sufficiently large $i \leqslant k$ and large k,*
$$\left|\frac{F_{k-i}(n(k))}{F_k(n(k))}\right| \leqslant b\rho^i.$$

Theorem 8: Let $n = n(k) = k/c + O(1)$ for a constant $c > 0$. Then

$$\mathsf{B}_\mathsf{K}(k) \sim K_c\binom{n+k-1}{k} \tag{11}$$

as $k,n\to\infty$, where K_c is a positive constant,

$$K_c = \lim_{n\to\infty}\gamma\left(c/(1+c),n\right).$$

Proof: To prove the theorem, we use Theorem 7. To do this, we first let

$$\rho = \lim_{k\to\infty}\frac{\binom{n(k)+k-2}{k-1}}{\binom{n(k)+k-1}{k}} = \lim_{k\to\infty}\frac{k}{n(k)+k-1} = \frac{c}{1+c}.$$

We now turn our attention to Condition 1 of Theorem 7. First, we show that $\gamma\ (\rho, n\,(k))$ is bounded away from 0. We have

$$\ln \gamma\ (\rho, n\,(k)) \geqslant \sum_{i=2}^{\infty} \ln\left(1-\rho^i\right) \geqslant -\sum_{i=2}^{\infty} \frac{\rho^i}{1-\rho^i} \geqslant -\sum_{i=2}^{\infty} \frac{\rho^i}{1-\rho} = -\frac{\rho^2}{(1-\rho)^2},$$

where the second inequality follows from the fact that for $0 < x < 1$,

$$\ln(1-x) = -\sum_{i=1}^{\infty} \frac{x^i}{i} \geqslant -\sum_{i=1}^{\infty} x^i = \frac{-x}{1-x}.$$

Hence,

$$\gamma\ (\rho, n\,(k)) \geqslant e^{-\left(\frac{\rho}{1-\rho}\right)^2} > 0.$$

To satisfy Condition 1 of Theorem 7, it thus suffices to find p_i' such that $|\Gamma_i\ (n\,(k))| \leqslant p_i'$ and $\sum_{i=0}^{\infty} p_i' \rho^i < \infty$ and then let $p_i = p_i' e^{(\rho/(1-\rho))^2}$.

For all positive integers m, we have

$$|\Gamma_i\ (m)| = \left|[z^i] \prod_{j=2}^{m} \left(1-z^j\right)\right| \leqslant \left|[z^i] \prod_{j=2}^{m} \left(1+z^j\right)\right| \leqslant \left|[z^i] \prod_{j=1}^{\infty} \left(1+z^j\right)\right| < e^{\pi\sqrt{2/3}\sqrt{i}},$$

where the last inequality follows from the facts that $\prod_{j=1}^{\infty}\left(1+z^j\right)$ is the generating function for the number of partitions of a positive integer into distinct parts and that the number of partitions of a positive integer i is bounded by $e^{\pi\sqrt{2/3}\sqrt{i}}$ [2, p. 316].

We let $p_i' = e^{\pi\sqrt{2/3}\sqrt{i}}$ and apply the root test to the sum $\sum_{i=0}^{\infty} p_i' \rho^i$ to prove its convergence. Since

$$\lim_{i\to\infty} \left(p_i'\rho^i\right)^{1/i} = \lim_{i\to\infty} e^{\pi\sqrt{2/3}/\sqrt{i}} \rho < 1,$$

the sum converges and Condition 1 of Theorem 7 is satisfied. Condition 2 of Theorem 7 is proved in [22, Th. 3.1]. Hence,

$$\frac{\mathsf{B}_{\mathsf{K}}(k)}{\binom{n+k-1}{k}} \sim \gamma\ \left(\frac{c}{1+c}, n\right).$$

To complete the proof, we must show that the limit $\lim_{n\to\infty} \gamma\ (c/\,(1+c)\,, n)$ exists and is positive. This is evident as $\gamma\ (c/\,(1+c)\,, n)$ is decreasing and, as shown before, bounded away from 0. ■

For $D = cn + O\,(1)$ with c a positive constant, we have

$$\begin{aligned}\frac{1}{n}\lg \mathsf{B}_{\mathsf{K}}(D) &= \frac{1}{n}\lg\binom{n+D-1}{D} + O\left(\frac{1}{n}\right) \\ &= \frac{n+cn+O\,(1)}{n} H\left(\frac{c}{1+c} + O\left(\frac{1}{n}\right)\right) + O\left(\frac{\lg n}{n}\right) \\ &= (1+c)\,H\left(\frac{c}{1+c}\right) + O\left(\frac{1}{n}\right) + O\left(\frac{\lg n}{n}\right) \\ &= (1+c)\,H\left(\frac{c}{1+c}\right) + O\left(\frac{\lg n}{n}\right),\end{aligned}$$

where we have used (11) for the first step. Using (3), for $D = cn + O\,(1)$, we find

$$\hat{\mathsf{A}}_K(D) \geqslant -\frac{1}{n}\lg \mathsf{B}_{\mathsf{K}}(D) = -\,(1+c)\,H\left(\frac{c}{1+c}\right) + O\left(\frac{\lg n}{n}\right)$$

and

$$\hat{\mathsf{A}}_K(D) \leqslant -\frac{1}{n}\lg \mathsf{B}_{\mathsf{K}}(D) + \frac{1}{n}\lg\left(1+\ln \mathsf{B}_{\mathsf{K}}(D)\right) = -\,(1+c)\,H\left(\frac{c}{1+c}\right) + O\left(\frac{\lg n}{n}\right)$$

The derivation for $\bar{\mathsf{A}}_K(cn + O(1))$ is similar. We thus have the following lemma.

Lemma 9: For a constant $c > 0$, we have

$$\hat{\mathsf{A}}_K\ (cn + O\,(1)) = -\lg\frac{(1+c)^{1+c}}{c^c} + O\left(\frac{\lg n}{n}\right). \quad (12)$$

Furthermore, $\bar{\mathsf{A}}_K(cn + O(1))$ satisfies the same equation.

The results given in (9) and (12) are given as lower bounds in [29, eq. (14)]. We have thus shown that these lower bounds in fact match the quantity under study. Furthermore, we have shown that $\bar{\mathsf{A}}_K(D)$ satisfies the same relations.

C. Medium Distortion

We next consider the medium distortion regime, that is, $D = cn^{1+\alpha} + O\,(n)$ for constants $c > 0$ and $0 < \alpha < 1$. For this case, from [29], we have

$$\hat{\mathsf{A}}_K(D) \sim -\lg n^{\alpha}.$$

In this subsection, we improve upon this result by providing upper and lower bounds with error terms. We note that the improvement in the upper bound comes at the cost of a non-constructive proof, compared with the constructive approach of [29].

Lemma 10: For $D = cn^{1+\alpha} + O(n)$, where α and c are constants such that $0 < \alpha < 1$ and $c > 0$, we have

$$-\lg\left(ecn^{\alpha}\right) + O\left(n^{-\alpha}\right) \leqslant \hat{\mathsf{A}}_K(D) \leqslant -\lg\left(cn^{\alpha}\right) + O\left(n^{-\alpha} + n^{\alpha-1}\right).$$

Furthermore, $\bar{\mathsf{A}}_K(D)$ satisfies the same inequalities.

Proof: From Theorem 4, we have

$$\hat{\mathsf{A}}_K(D) \geqslant -\lg\frac{(1+\delta)^{1+\delta}}{\delta^{\delta}} = -\lg(1+\delta) - \lg\left(1+\frac{1}{\delta}\right)^{\delta} \geqslant -\lg(e(1+\delta)).$$

Note that $\delta = D/n = cn^{\alpha} + O\,(1)$. We find

$$\hat{\mathsf{A}}_K(D) \geqslant -\lg\left(ecn^{\alpha} + O\,(1)\right) = -\lg\left(ecn^{\alpha}\right) + O\left(n^{-\alpha}\right).$$

From Theorem 4, it also follows that the same holds for $\bar{\mathsf{A}}_K(D)$, as $\lg n/n = O(n^{-\alpha})$.

On the other hand, from Theorem 5,

$$\bar{\mathsf{A}}_K(D) \leqslant \hat{\mathsf{A}}_K(D) \leqslant -\lg\left(cn^{\alpha} + O\,(1)\right) + \frac{1}{n}\lg e^{O(n^{\alpha})} = -\lg\left(cn^{\alpha}\right) + O\left(n^{-\alpha} + n^{\alpha-1}\right).$$

■

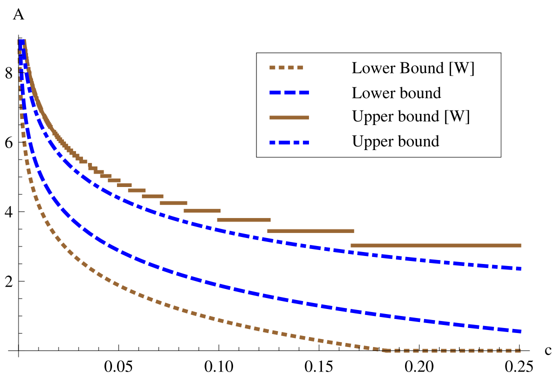

Fig. 2. Bounds on $\hat{\mathsf{A}}_K(D)+\lg n$ for $D = cn^2 + O(n)$ where the error terms are ignored. The bounds denoted by [W] are those from [29].

D. Large Distortion

In the large distortion regime, we have $D = cn^2 + O(n)$ and $\delta = cn + O(1)$.

Lemma 11: Suppose $D = cn^2 + O(n)$ *for a constant* $0 < c < 1/2$. *We have*

$$-\lg(ecn)+O\left(\frac{1}{n}\right) \leqslant \hat{\mathsf{A}}_K(D) \leqslant -\lg(ecn) + (1+c)\lg e + O\left(\frac{\lg n}{n}\right).$$

Furthermore,

$$-\lg(ecn) + O\left(\frac{\lg n}{n}\right) \leqslant \bar{\mathsf{A}}_K(D) \leqslant \hat{\mathsf{A}}_K(D).$$

Proof: Let $\delta = D/n = cn + O(1)$. Similar to the proof of the lower bound in Lemma 10, we have $\hat{\mathsf{A}}_K(D) \geqslant -\lg(e(1+\delta))$, and thus

$$\hat{\mathsf{A}}_K(D) \geqslant -\lg(ecn + O(1)) \geqslant -\lg(ecn) + O\left(\frac{1}{n}\right).$$

Similarly,

$$\bar{\mathsf{A}}_K(D) \geqslant -\lg(e(1+\delta)) + O\left(\frac{\lg n}{n}\right) \geqslant -\lg(ecn) + O\left(\frac{\lg n}{n}\right).$$

On the other hand, from Theorem 5,

$$\bar{\mathsf{A}}_K(D) \leqslant \hat{\mathsf{A}}_K(D) \leqslant -\lg(cn) + c\lg e + O\left(\frac{\lg n}{n}\right).$$

■

From [29], we have

$$-\lg(ecn) - 1 + O\left(\frac{\lg n}{n}\right) \leqslant \hat{\mathsf{A}}_K(D) \leqslant -\lg\frac{n}{e\lceil 1/(2c)\rceil} + O\left(\frac{\lg n}{n}\right). \quad (13)$$

These bounds are compared in Figure 2, where we added the term $\lg n$ to remove dependence on n.

V. The Chebyshev Metric

We now turn to consider the rate-distortion function for the permutation space under the Chebyshev metric. We start by stating lower and upper bounds on the size of the ball in the Chebyshev metric, and then construct covering codes.

A. Bounds

For an $n \times n$ matrix A, the permanent of $A = (A_{i,j})$ is defined as,

$$\mathrm{per}(A) = \sum_{\omega\in\mathbb{S}_n}\prod_{i=1}^{n} A_{i,\omega(i)}.$$

It is well known [19], [25] that $\mathsf{B}_{\mathsf{C}}(r)$ can be expressed as the permanent of the $n \times n$ binary matrix A for which

$$A_{i,j} = \begin{cases} 1 & |i-j| \leqslant r \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

According to Brégman's Theorem (see [5]), for any $n \times n$ binary matrix A with r_i 1's in the i-th row

$$\mathrm{per}(A) \leqslant \prod_{i=1}^{n}(r_i!)^{\frac{1}{r_i}}.$$

Using this bound we can state the following lemma (partially given in [19] and extended in [27]).

Lemma 12 [27]: For all $0 \leqslant r \leqslant n-1$,

$$\mathsf{B}_{\mathsf{C}}(r) \leqslant \begin{cases} ((2r+1)!)^{\frac{n-2r}{2r+1}}\prod_{i=r+1}^{2r}(i!)^{\frac{2}{i}}, & 0 \leqslant r \leqslant \frac{n-1}{2}, \\ (n!)^{\frac{2r+2-n}{n}}\prod_{i=r+1}^{n-1}(i!)^{\frac{2}{i}}, & \frac{n-1}{2} \leqslant r \leqslant n-1. \end{cases}$$

The following lower bound was given in [19].

Lemma 13 [19]: For all $0 \leqslant r \leqslant (n-1)/2$,

$$\mathsf{B}_{\mathsf{C}}(r) \geqslant \frac{(2r+1)^n}{2^{2r}}\frac{n!}{n^n}.$$

We extend this lemma to the full range of parameters.

Lemma 14: For all $0 \leqslant r \leqslant n-1$,

$$\mathsf{B}_{\mathsf{C}}(r) \geqslant \begin{cases} \frac{(2r+1)^n}{2^{2r}}\frac{n!}{n^n}, & 0 \leqslant r \leqslant \frac{n-1}{2}, \\ \frac{n!}{2^{2(n-r)}}, & \frac{n-1}{2} \leqslant r \leqslant n-1. \end{cases}$$

Proof: Only the second claim requires proof, so suppose that $(n-1)/2 \leqslant r \leqslant n-1$. The proof follows the same lines as the one appearing in [19]. Let A be defined as in (14), and let B be an $n \times n$ matrix with

$$B_{i,j} = \begin{cases} 2, & i+j \leqslant n-r, \\ 2, & i+j \geqslant n+r+2, \\ A_{i,j}, & \text{otherwise.} \end{cases}$$

We observe that B/n is doubly stochastic. It follows that

$$\mathsf{B}_{\mathsf{C}}(r) = \mathrm{per}(A) \geqslant \frac{\mathrm{per}(B)}{2^{2(n-r)}} \geqslant \frac{n^n}{2^{2(n-r)}}\mathrm{per}\left(\frac{B}{n}\right) \geqslant \frac{n!}{2^{2(n-r)}},$$

where the last inequality follows from Van der Waerden's Theorem [23]. ■

Theorem 15: Let $n \in \mathbb{N}$, and let $0 < \delta < 1$ be a constant rational number such that $D = \delta n$ is an integer. Then

$$\hat{\mathsf{R}}_C(D) \geqslant \begin{cases} \lg \frac{1}{2\delta} + 2\delta \lg \frac{e}{2} + O(\lg n/n), & 0 < \delta \leqslant \frac{1}{2} \\ 2\delta \lg \delta + 2(1-\delta)\lg e + O(\lg n/n), & \frac{1}{2} \leqslant \delta \leqslant 1 \end{cases}$$

and

$$\hat{\mathsf{R}}_C(D) \leqslant \begin{cases} \lg \frac{1}{2\delta} + 2\delta + O(\lg n/n), & 0 < \delta \leqslant \frac{1}{2} \\ 2(1-\delta) + O(\lg n/n), & \frac{1}{2} \leqslant \delta \leqslant 1 \end{cases}$$

Furthermore, the same bounds also hold for $\bar{\mathsf{R}}_C(D)$.

Proof: First, we prove the lower bound for $\hat{\mathsf{R}}_C(D)$ using Theorem 3, which implies $\hat{\mathsf{R}}_C(D) \geqslant \frac{1}{n}\lg n! - \frac{1}{n}\lg \mathsf{B}_\mathsf{C}(D)$, and Lemma 12. Let

$$T_1 = ((2D+1)!)^{(n-2D)/(2D+1)},$$
$$T_2 = \prod_{i=D+1}^{2D} (i!)^{2/i},$$

so that $\mathsf{B}_\mathsf{C}(D) \leqslant T_1 T_2$ for $0 < D < (D-1)/2$. We have

$$\begin{aligned} \lg T_1 &= \frac{n-2\delta n}{2\delta n+1}\lg(2\delta n+1)! \\ &= \frac{n-2\delta n}{2\delta n+1}\left((2\delta n+1)\lg\left(\frac{2\delta n+1}{e}\right) + O(\lg n)\right) \\ &= (n-2\delta n)\lg\left(\frac{2\delta n+1}{e}\right) + O(\lg n) \\ &= (n-2\delta n)\lg(2\delta n/e) + O(\lg n), \end{aligned}$$

and

$$\begin{aligned} \lg T_2 &= 2\sum_{i=\delta n+1}^{2\delta n} \frac{1}{i}\lg i! = 2\sum_{i=\delta n+1}^{2\delta n}\left(\lg\frac{i}{e} + O\left(\frac{\lg i}{i}\right)\right) \\ &= 2\sum_{i=\delta n+1}^{2\delta n} \lg i - 2\delta n \lg e + O(\lg n) \\ &= 2\lg\frac{(2\delta n)!}{(\delta n)!} - 2\delta n\lg e + O(\lg n) \\ &= 2\delta n + 2\delta n\lg(2\delta n/e) - 2\delta n \lg e + O(\lg n). \end{aligned}$$

From these expressions and Lemma 12, it follows that

$$\frac{1}{n}\lg \mathsf{B}_\mathsf{C}(D) \leqslant \lg(2\delta n/e) + 2\delta\lg(2/e) + O(\lg n/n).$$

The lower bound for $0 < \delta \leqslant 1/2$ then follows from Theorem 3. The proof of the lower bound for $1/2 < \delta \leqslant 1$ is similar.

Next, we prove the upper bound for $\hat{\mathsf{R}}_C(D)$. From Theorem 3, we have

$$\hat{\mathsf{M}}_C(D) \leqslant \frac{n!}{\mathsf{B}_\mathsf{C}(D)}(1 + \ln \mathsf{B}_\mathsf{C}(D)) \leqslant \frac{n!}{\mathsf{B}_\mathsf{C}(D)}(1 + \ln n!).$$

While the last inequality seems crude, it will not change the asymptotic result. Hence, for $0 \leqslant D \leqslant (n-1)/2$,

$$\begin{aligned} \hat{\mathsf{R}}_C(D) &\leqslant \frac{1}{n}\lg\left(\frac{n!(1+\ln n!)}{\mathsf{B}_\mathsf{C}(\delta n)}\right) \\ &\leqslant \frac{1}{n}\lg\left(\frac{2^{2\delta n}n^n}{(2\delta n+1)^n}\right) + O\left(\frac{\lg n}{n}\right) \\ &\leqslant \lg\frac{1}{2\delta} + 2\delta + O\left(\frac{\lg n}{n}\right), \end{aligned} \tag{15}$$

where we have used Lemma 14 for the second inequality.

Similarly, for $(n-1)/2 < D \leqslant n$,

$$\begin{aligned} \hat{\mathsf{R}}_C(D) &\leqslant \frac{1}{n}\lg 2^{2n(1-\delta)} + O\left(\frac{\lg n}{n}\right) \\ &\leqslant 2(1-\delta) + O\left(\frac{\lg n}{n}\right). \end{aligned} \tag{16}$$

The proof of the lower bound for $\bar{\mathsf{R}}_C(D)$ is similar to that of $\hat{\mathsf{R}}_C(D)$ except that we use $\bar{\mathsf{M}}(D) > n!/(\mathsf{B}(D)(D+1))$ from Theorem 3. The proof of the upper bound for $\bar{\mathsf{R}}_C(D)$ follows from the fact that $\bar{\mathsf{R}}_C(D) \leqslant \hat{\mathsf{R}}_C(D)$. ■

In the Chebyshev metric we define the *small-distortion regime* as the regime in which the covering radius of the code, D, satisfies $D = o(n)$, or alternatively, δ tends to 0. If we examine Theorem 15, we note that in the small-distortion regime, the ratio of the upper bound to the lower bound tends to 1 as δ tends to 0. Thus, the bounds are in particular accurate in the small-distortion regime.

B. Code Construction

Let $A = \{a_1, a_2, \dots, a_m\} \subseteq [n]$ be a subset of indices, $a_1 < a_2 < \cdots < a_m$. For any permutation $\sigma \in \mathbb{S}_n$ we define $\sigma|_A$ to be the permutation in $\mathbb{S}_m$ that preserves the relative order of the sequence $\sigma(a_1), \sigma(a_2), \dots, \sigma(a_m)$. Intuitively, to compute $\sigma|_A$ we keep only the *coordinates* of σ from A, and then relabel the entries to $[m]$ while keeping relative order. In a similar fashion we define

$$\sigma|^A = \left(\sigma^{-1}|_A\right)^{-1}.$$

This time, however, to calculate $\sigma|^A$ we keep only the *values* of σ from A, and then relabel the entries to $[m]$ while keeping relative order.

Example 16: Let $n = 6$ and consider the permutation

$$\sigma = [6, 1, 3, 5, 2, 4].$$

We take $A = \{3, 5, 6\}$. We then have

$$\sigma|_A = [2, 1, 3],$$

since we keep positions 3, 5, and 6, of σ, giving us $[3, 2, 4]$, and then relabel these to get $[2, 1, 3]$.

Similarly, we have

$$\sigma|^A = [3, 1, 2],$$

since we keep the values 3, 5, and 6, of σ, giving us $[6, 3, 5]$, and then relabel these to get $[3, 1, 2]$.

Construction 1: Let n and d be positive integers, $1 \leqslant d \leqslant n-1$. Furthermore, we define the sets

$$A_i = \{i(d+1) + j : 1 \leqslant j \leqslant d+1\} \cap [n],$$

for all $0 \leqslant i \leqslant \lfloor (n-1)/(d+1) \rfloor$. We now construct the code C defined by

$$C = \left\{\sigma \in \mathbb{S}_n : \sigma|^{A_i} = \mathrm{Id} \text{ for all } i\right\}.$$

We note that this construction already appears in [29, Remark 4], however there it is given for the ℓ_1-metric over permutations, and thus, it has a different minimum distance.

Theorem 17: Let n and d be positive integers, $1 \leqslant d \leqslant n-1$*. Then the code* $C \subseteq \mathbb{S}_n$ *of Construction 1 has covering radius exactly d and size*

$$M = \frac{n!}{(d+1)!^{\lfloor n/(d+1) \rfloor}(n \bmod (d+1))!}. \tag{17}$$

Proof: Let $\sigma \in \mathbb{S}_n$ be any permutation. We let I_i denote the indices in which the elements of A_i appear in σ. Let us now construct a new permutation σ' in which the elements of A_i appear in indices I_i, but they sorted in ascending order. Thus

$$\sigma'|^{A_i} = \mathrm{Id},$$

for all i, and so σ' is a codeword in C.

We observe that if $\sigma(j) \in A_i$, then $\sigma'(j) \in A_i$ as well. It follows that

$$\left|\sigma(j) - \sigma'(j)\right| \leqslant d$$

and so

$$\mathsf{d}_{\mathsf{C}}(\sigma, \sigma') \leqslant d.$$

Finally, we contend that the permutation $\sigma = [n, n-1, \ldots, 1]$ is at distance exactly d from the code C. Note that we already know that there is a codeword $\sigma' \in C$ such that $\mathsf{d}_{\mathsf{C}}(\sigma, \sigma') \leqslant d$. We now show there is no closer codeword in C. Let us attempt to build such a permutation σ''. Consider $\sigma(n) = 1$. The value of $\sigma''(n)$ is in A_i for some i, and since σ'' is a codeword, $\sigma''(n)$ must be the largest in A_i. Thus

$$\sigma''(n) \in \{\max(A_i) : 1 \leqslant i \leqslant \lceil n/(d+1) \rceil\} \geqslant d+1.$$

It follows that

$$\left|\sigma''(n) - \sigma(n)\right| \geqslant d$$

and so

$$\mathsf{d}_{\mathsf{C}}(\sigma, \sigma'') \geqslant d.$$

■

The next theorem presents the asymptotic rate of the construction.

Theorem 18: The code from Construction 1 has the following asymptotic rate,

$$R = H\left(\delta \left\lfloor \frac{1}{\delta} \right\rfloor\right) + \delta \left\lfloor \frac{1}{\delta} \right\rfloor \lg \left\lfloor \frac{1}{\delta} \right\rfloor - o(1).$$

Proof: We note that

$$(n \bmod d+1) = n - (d+1)\left\lfloor \frac{n}{d+1} \right\rfloor.$$

We then rewrite (17) and get

$$2^{Rn} = \frac{n!}{(\delta n+1)!^{\lfloor n/(\delta n+1) \rfloor}\left(n - (\delta n+1)\lfloor n/(\delta n+1) \rfloor\right)!}.$$

We recall Stirling's approximation from (2), stating that

$$m! = \left(\frac{m}{e}\right)^m 2^{o(m)},$$

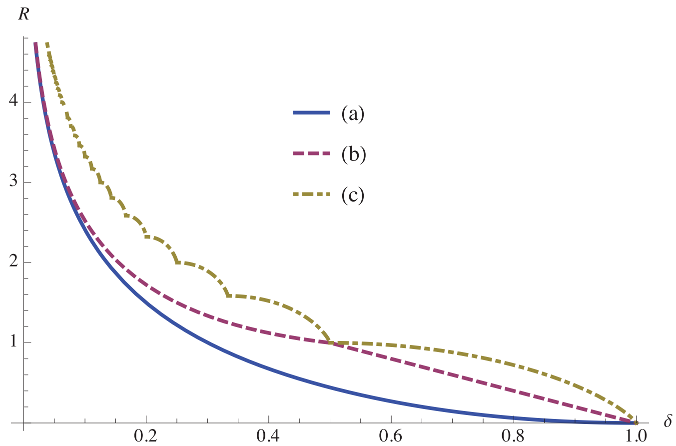

Fig. 3. Rate-distortion in the Chebyshev metric: The lower and upper bounds of Theorem 15, (a) and (b), and the rate of the code construction, given in Theorem 18, (c).

and use it to obtain

$$\begin{aligned} 2^{Rn} &= \frac{\left(\frac{n}{e}\right)^n}{\left(\frac{\delta n+1}{e}\right)^{(\delta n+1)\left\lfloor \frac{n}{\delta n+1} \right\rfloor}} \\ &\quad \cdot \frac{1}{\left(\frac{n-(\delta n+1)\left\lfloor \frac{n}{\delta n+1} \right\rfloor}{e}\right)^{n-(\delta n+1)\left\lfloor \frac{n}{\delta n+1} \right\rfloor}} \cdot 2^{o(n)} \\ &= \frac{1}{\delta^{n\delta\left\lfloor \frac{1}{\delta} \right\rfloor}\left(1 - \delta\left\lfloor \frac{1}{\delta} \right\rfloor\right)^{n(1-\delta)\left\lfloor \frac{1}{\delta} \right\rfloor}} \cdot 2^{o(n)}. \end{aligned}$$

If we now take $\log_2$ of both sides, divide by n, and rearrange, we arrive at the desired form. ■

The bounds given in Theorem 15 and the rate of the code construction, given in Theorem 18, are shown in Figure 3.

VI. Conclusion

In this paper, we presented rate-distortion results for the space of permutations endowed by the Kendall τ-metric and the Chebyshev metric. For the former, we improved upon the previously known results and for the latter we established new results. These findings can be further improved by providing tighter bounds and better constructions. Indeed, in the case of the Chebyshev distance the construction only attains the bound at two points. A different approach for constructing such codes may be needed, perhaps employing deeper combinatorial reasoning. Additionally, there remains a gap between the lower and upper bounds on the size of a ball in the Chebyshev metric, resulting in bounds which are not tight.

It would also be interesting to study another classical distance metric on permutations in the context of rate-distortion, namely the Ulam distance, also known as the edit distance. The Ulam distance [1] is defined as the number of edits required to take one permutation to another and has been studied in coding theory in the context of rank modulation codes [11] and for measuring sortedness of data streams [15].

ACKNOWLEDGMENTS

The authors would like to thank the associate editor and anonymous reviewers, whose comments helped improve the presentation of the paper.

REFERENCES

[1] D. Aldous and P. Diaconis, "longest increasing subsequences: From patience sorting to the Baik–Deift–Johansson theorem," *Bulletin*, vol. 36, no. 4, pp. 413–432, 1999.

[2] T. M. Apostol, *Introduction to Analytic Number Theory*. New York, NY, USA: Springer-Verlag, 1976.

[3] 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, Jul. 2010.

[4] M. Bóna, *Combinatorics of Permutations*. Boca Raton, FL, USA: CRC Press, 2012.

[5] L. M. Brégman, "Some properties of nonnegative matrices and their permanents," *Sov. Math. Dokl.*, vol. 14, no. 4, pp. 945–949, 1973.

[6] P. J. Cameron and I. M. Wanless, "Covering radius for sets of permutations," *Discrete Math.*, vol. 293, pp. 91–109, Apr. 2005.

[7] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein, *Covering Codes*. Amsterdam, The Netherlands: North Holland, 1997.

[8] M. Deza and T. Huang, "Metrics on permutations, a survey," *J. Combinat., Inf. Syst. Sci.*, vol. 23, pp. 173–185, 1998.

[9] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, "Rank aggregation methods for the Web," in *Proc. 10th Int. Conf. World Wide Web*, 2001, pp. 613–622.

[10] F. Farnoud and O. Milenkovic, "An axiomatic approach to constructing distances for rank comparison and aggregation," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6417–6439, Oct. 2014.

[11] F. Farnoud, V. Skachek, and O. Milenkovic, "Error-correction in flash memories via codes in the Ulam metric," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3003–3020, May 2013.

[12] F. Farnoud, E. Yaakobi, and J. Bruck, "Approximate sorting of data streams with limited storage," in *Computing and Combinatorics* (Lecture Notes in Computer Science), vol. 8591, Z. Cai, A. Zelikovsky, and A. Bourgeois, Eds. Springer International Publishing, 2014, pp. 465–476.

[13] F. Farnoud, E. Yaakobi, and J. Bruck, "Approximate sorting of data streams with limited storage," *J. Combinat. Optim.*, pp. 1–32, Aug. 2015. [Online]. Available: http://dx.doi.org/10.1007/s10878-015-9930-6

[14] J. Giesen, E. Schuberth, and M. Stojakovi, "Approximate sorting," in *LATIN 2006: Theoretical Informatics*, vol. 3887. Springer Berlin Heidelburg, Jan. 2006, pp. 524–531.

[15] P. Gopalan, T. S. Jayram, R. Krauthgamer, and R. Kumar, "Estimating the sortedness of a data stream," in *Proc. ACM-SIAM Symp. Discrete Algorithms (SODA)*, New Orleans, LA, USA, Jan. 2007, pp. 318–327.

[16] A. Jiang, R. Mateescu, M. Schwartz, and J. Bruck, "Rank modulation for flash memories," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2659–2673, Jun. 2009.

[17] A. Jiang, M. Schwartz, and J. Bruck, "Correcting charge-constrained errors in the rank-modulation scheme," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2112–2120, May 2010.

[18] P. Keevash and C. Y. Ku, "A random construction for permutation codes and the covering radius," *Designs, Codes Cryptogr.*, vol. 41, no. 1, pp. 79–86, 2006.

[19] T. Kløve, "Spheres of permutations under the infinity norm—Permutations with limited displacement," Dept. Inform., Univ. Bergen, Bergen, Norway, Tech. Rep. 376, Nov. 2008.

[20] D. E. Knuth, *The Art of Computer Programming: Sorting and Searching*, vol. 3, 2nd ed. Reading, MA, USA: Addison-Wesley, 1998.

[21] A. Mazumdar, A. Barg, and G. Zémor, "Constructions of rank modulation codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 1018–1029, Feb. 2013.

[22] O. Milenkovic and K. J. Compton, "Probabilistic transforms for combinatorial urn models," *Combinat., Probab., Comput.*, vol. 13, nos. 4–5, pp. 645–675, Jul. 2004.

[23] H. Minc, *Permanents* (Encyclopedia of Mathematics and Its Applications), vol. 6. Cambridge, U.K.: Cambridge Univ. Press, 1978.

[24] J. Quistorff, "A survey on packing and covering problems in the Hamming permutation space," *Electron. J. Combinat.*, vol. 13, pp. 1–13, Apr. 2006.

[25] M. Schwartz, "Efficiently computing the permanent and Hafnian of some banded Toeplitz matrices," *Linear Algebra Appl.*, vol. 430, no. 4, pp. 1364–1374, Feb. 2009.

[26] S. K. Stein, "Two combinatorial covering theorems," *J. Combinat. Theory A*, vol. 16, no. 3, pp. 391–397, 1974.

[27] I. Tamo and M. Schwartz, "Correcting limited-magnitude errors in the rank-modulation scheme," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2551–2560, Jun. 2010.

[28] I. Tamo and M. Schwartz, "On the labeling problem of permutation group codes under the infinity metric," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6595–6604, Oct. 2012.

[29] D. Wang, A. Mazumdar, and G. W. Wornell, "A rate-distortion theory for permutation spaces," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 2562–2566.

Farzad Farnoud (Hassanzadeh) is a postdoctoral scholar at the California Institute of Technology. He received his MS degree in Electrical and Computer Engineering from the University of Toronto in 2008. From the University of Illinois at Urbana- Champaign, he received his MS degree in mathematics and his PhD in Electrical and Computer Engineering in 2012 and 2013, respectively. His research interests include the information-theoretic and algorithmic analysis of genomic evolutionary processes, ranking-based information processing, and coding for flash memory. He is a recipient of the Robert T. Chien Memorial Award for demonstrating excellence in research in electrical engineering from the University of Illinois at Urbana-Champaign.

Moshe Schwartz (M'03–SM'10) is an associate professor at the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Israel. His research interests include algebraic coding, combinatorial structures, and digital sequences.

Dr. Schwartz received the B.A. (*summa cum laude*), M.Sc., and Ph.D. degrees from the Technion - Israel Institute of Technology, Haifa, Israel, in 1997, 1998, and 2004 respectively, all from the Computer Science Department. He was a Fulbright post-doctoral researcher in the Department of Electrical and Computer Engineering, University of California San Diego, and a post-doctoral researcher in the Department of Electrical Engineering, California Institute of Technology. While on sabbatical 2012-2014, he was a visiting scientist at the Massachusetts Institute of Technology (MIT).

Dr. Schwartz received the 2009 IEEE Communications Society Best Paper Award in Signal Processing and Coding for Data Storage, and the 2010 IEEE Communications Society Best Student Paper Award in Signal Processing and Coding for Data Storage.

Jehoshua Bruck (S'86–M'89–SM'93–F'01) is the Gordon and Betty Moore Professor of computation and neural systems and electrical engineering at the California Institute of Technology (Caltech). His current research interests include information theory and systems and the theory of computation in nature.

Dr. Bruck received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion-Israel Institute of Technology, in 1982 and 1985, respectively, and the Ph.D. degree in electrical engineering from Stanford University, in 1989. His industrial and entrepreneurial experiences include working with IBM Research where he participated in the design and implementation of the first IBM parallel computer; cofounding and serving as Chairman of Rainfinity (acquired in 2005 by EMC), a spin-off company from Caltech that created the first virtualization solution for Network Attached Storage; as well as cofounding and serving as Chairman of XtremIO (acquired in 2012 by EMC), a start-up company that created the first scalable all-flash enterprise storage system.

Dr. Bruck is a recipient of the Feynman Prize for Excellence in Teaching, the Sloan Research Fellowship, the National Science Foundation Young Investigator Award, the IBM Outstanding Innovation Award and the IBM Outstanding Technical Achievement Award.