Breakpoint Analysis and Permutation Codes in Generalized Kendall tau and Cayley Metrics

Yeow Meng Chee and Van Khu Vu
School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore
emails: {ymchee, vankhu001}@ntu.edu.sg

***Abstract*—Permutation codes under the Cayley, Kendall tau, and Ulam metrics have been studied recently due to applications in flash memories. We consider permutation codes under more general metrics. We use breakpoints in permutations to gain additional insights to distances in codes. As a result, we construct codes under these general metrics that are larger than those previously known under more restricted metrics.**

I. Introduction

Permutation codes are subsets of the symmetric group. Since their introduction by Blake [1], they have been actively studied for applications in communications [2]–[6] and have also stimulated mathematical interest [7]–[10]. For reliability in communication applications, distance constraints are normally imposed on codes. In the case of permutation codes, the metrics that have been considered are the Hamming metric, Cayley metric, Kendall tau metric, and Ulam metric. Recently, Farnoud et al. [11] considered permutation codes in the Hamming, Cayley, Kendall tau, and Ulam metrics.

One difficulty with designing permutation codes with respect to the Cayley, Kendall tau, and Ulam metrics is that although these metrics are intuitive, they are rather complex. Indeed, Bulteau et al. [12] have recently shown that determining the distance between two permutations under the generalized Kendall tau metric (see Section II for definition) is NP-hard. This complexity can make analysis difficult. To overcome this difficulty, we consider a simpler notion in permutations, known as *breakpoints*. The concept of breakpoints was introduced by Watterson et al. [13] and plays an important role in determining similarities of genomes (see [14]). Here, we show how transpositions affect the number of breakpoints in permutations and use the number of breakpoints as a proxy measure for Cayley, Kendall tau, and Ulam distances between permutations. The advantage of this approach is the simplicity in determining the number of breakpoints in a permutation, revealing features of permutation codes which are otherwise hard to observe. This enables us not only to improve on the size of permutation codes in the Ulam metric obtained by Farnoud et al. [11], but to also show that the larger permutation codes we constructed have the same minimum distance even under more generalized metrics.

II. Preliminaries

Let X be a finite set of n elements and let $\preceq$ be a total order on X (with $\prec$ as the induced strict total order). As we introduce the definitions and notations below, it is useful to think of X as a finite subset of $\mathbb{Z}$ and $\preceq$ (respectively, $\prec$) as the order $\leq$ (respectively, $<$). Two elements $a, b \in X$, $a \prec b$, are *consecutive* if there does not exist c such that $a \prec c \prec b$. Let $m, n \in X$ such that $m \preceq n$. The *interval* $\{i \in X : m \preceq i \preceq n\}$ is denoted by $[m, n]$. The interval $[a, b] \subseteq X$ is said to *precede* the interval $[c, d] \subseteq X$, written $[a, b] \prec [c, d]$, if $b \prec c$. Furthermore, if b and c are consecutive, then $[a, b]$ and $[c, d]$ are said to be *consecutive* intervals.

A *permutation* on X is a bijection $f : X \to X$. The set of all permutations on X forms a group under functional composition. This group is known as the *symmetric group* of degree n, and is denoted by $\mathrm{Sym}(X)$. We normally take $X = [1, n]$ and abbreviate $\mathrm{Sym}([1, n])$ to S_n. A permutation $f \in \mathrm{Sym}(X)$ is written as a vector $f = (f(x_1), f(x_2), \ldots, f(x_n))$, where $x_1, x_2, \ldots, x_n \in X$ and $x_1 \prec x_2 \prec \cdots \prec x_n$. The *identity* in $\mathrm{Sym}(X)$ is the permutation $\varepsilon_n = (x_1, x_2, \ldots, x_n)$.

The *support* of a permutation $f \in \mathrm{Sym}(X)$, denoted by $\mathrm{supp}(f)$, is the set of elements in X that are not fixed by f, that is, $\mathrm{supp}(f) = \{x \in X : f(x) \neq x\}$. A permutation whose support contains precisely two elements is called a *transposition*. Hence, a transposition whose support is $\{i, j\}$ is obtained from the identity by swapping the positions of i and j. A transposition whose support contains two consecutive elements of X is called an *adjacent transposition*.

Transpositions and adjacent transpositions can be generalized to "interval" versions as follows. Instead of swapping two positions in an identity, we swap two intervals. More precisely, let $A, B \subseteq [1, n]$ be two intervals such that $A \prec B$. Suppose that $A = [i, j]$ and $B = [k, \ell]$. Then swapping the two intervals A and B in the identity of S_n as in (1) gives the permutation (2). We denote the permutation in (2) by $\phi(A, B)$ and call it a *generalized transposition*. We say that the generalized transposition $\phi(A, B)$ is an *adjacent* generalized transposition if A and B are consecutive intervals. Our usual notions of transpositions and adjacent transpositions are exactly the case when A and B are intervals of size one in a generalized transposition $\phi(A, B)$. *Translocations* defined by Farnoud et al. [11] are adjacent generalized transpositions $\phi(A, B)$ where at least one of A and B is an interval of size one. Adjacent transpositions are special cases of translocations.

Example 1. Let $f = (1, 3, 6, 2, 8, 4, 5, 7) \in S_8$. Then,

$$f\phi([2, 3], [6, 7, 8]) = (1, 4, 5, 7, 2, 8, 3, 6),$$
$$f\phi(\{3\}, \{6\}) = (1, 3, 4, 2, 8, 6, 5, 7).$$

$$(1, 2, \ldots, i-1, \boxed{i, i+1, \ldots, j}, j+1, j+2, \ldots, k-1, \boxed{k, k+1, \ldots, \ell}, \ell+1, \ell+2, \ldots, n) \tag{1}$$

$$(1, 2, \ldots, i-1, \boxed{k, k+1, \ldots, \ell}, j+1, j+2, \ldots, k-1, \boxed{i, i+1, \ldots, j}, \ell+1, \ell+2, \ldots, n) \tag{2}$$

The concept of *breakpoints* in permutations is crucial in this paper.

Definition 1 (Breakpoint)**.** A *breakpoint* in a permutation $f \in S_n$ is an $i \in [2, n]$ such that $f(i) - f(i-1) \neq 1$. The number of breakpoints in a permutation f is denoted by $b(f)$.

The definition of breakpoints we use above is from [12]. There are other definitions of breakpoints but these differ among each other in only boundary conditions, and do not quite affect the construction and proof techniques used here.

A. Metrics on Symmetric Groups

Let ρ be a metric on S_n. We say that ρ is *left-invariant* if $\rho(af, ag) = \rho(f, g)$ for all $a, f, g \in S_n$. Many interesting metrics can be defined on S_n (see, [15]). The following are three that have been studied for coding in flash memories.

Definition 2 (Hamming Metric)**.** The *Hamming distance* between two permutations $f, g \in S_n$ is the number of positions which f and g differ:

$$\rho_{\mathrm{H}}(f, g) = |\{i \in [1, n] : f(i) \neq g(i)\}|.$$

Definition 3 (Cayley Metric)**.** The *Cayley distance* between two permutations $f, g \in S_n$ is the minimum number of transpositions which are needed to change f to g by post-multiplication:

$$\rho_{\mathrm{C}}(f, g) = \min\{n : \ f\pi_1\pi_2 \ldots \pi_n = g \text{ and } \pi_1, \pi_2, \ldots, \pi_n \text{ are transpositions}\}.$$

Definition 4 (Kendall tau Metric)**.** The *Kendall tau distance* between two permutations $f, g \in S_n$ is the minimum number of adjacent transpositions which are needed to change f to g by post-multiplication:

$$\rho_{\mathrm{K}}(f, g) = \min\{n : \ f\pi_1\pi_2 \ldots \pi_n = g \text{ and } \pi_1, \pi_2, \ldots, \pi_n \text{ are adjacent transpositions}\}.$$

As for transpositions, the Cayley and Kendall tau metrics can be generalized to “interval" versions.

Definition 5 (generalized Cayley Metric)**.** The *generalized Cayley distance* between two permutations $f, g \in S_n$ is the minimum number of generalized transpositions which are needed to change f to g by post-multiplication:

$$\rho_{\mathrm{gC}}(f, g) = \min\{n : \ f\pi_1\pi_2 \ldots \pi_n = g \text{ and } \pi_1, \pi_2, \ldots, \pi_n \text{ are generalized transpositions}\}.$$

Definition 6 (generalized Kendall tau Metric)**.** The *generalized Kendall tau distance* between two permutations $f, g \in S_n$ is the minimum number of adjacent generalized transpositions which are needed to change f to g by post-multiplication:

$$\rho_{\mathrm{gK}}(f, g) = \min\{n : \ f\pi_1\pi_2 \ldots \pi_n = g \text{ and } \pi_1, \pi_2, \ldots, \pi_n \text{ are adjacent generalized transpositions}\}.$$

Definition 7 (Ulam Metric)**.** The *Ulam distance* between two permutations $f, g \in S_n$ is the minimum number of translocations which are needed to change f to g by post-multiplication:

$$\rho_{\mathrm{U}}(f, g) = \min\{n : \ f\pi_1\pi_2 \ldots \pi_n = g \text{ and } \pi_1, \pi_2, \ldots, \pi_n \text{ are translocations}\}.$$

Since adjacent transpositions are special cases of translocations, which are in turn special cases of generalized adjacent transpositions, we have the inequality

$$\rho_{\mathrm{gC}}(f, g) \leq \rho_{\mathrm{gK}}(f, g) \leq \rho_{\mathrm{U}}(f, g) \leq \rho_{\mathrm{K}}(f, g),$$

for all $f, g \in S_n$.

All the metrics introduced here are left-invariant.

B. Permutation Codes

A *permutation code* of length n is a set $\mathcal{C} \subseteq S_n$. The elements of $\mathcal{C}$ are called *codewords*.

Let S_n be endowed with metric ρ. A permutation code $\mathcal{C} \subseteq S_n$ is said to have *(minimum) distance* d (under the metric ρ) if $\rho(f, g) \geq d$, for all distinct $f, g \in \mathcal{C}$. The maximum size of a permutation code of length n and distance d under the metric ρ is denoted by $A_\rho(n, d)$.

III. Breakpoints, Distances, and Bounds

The following is a simple observation.

Proposition 1. *Let $f, \pi \in S_n$. Then we have*

$$|b(f\pi) - b(f)| \leq \begin{cases} 3, & \text{if } \pi \text{ is an adjacent generalized transposition} \\ 4, & \text{if } \pi \text{ is a generalized transposition.} \end{cases}$$

What Proposition 1 says is that the number of breakpoints in a permutation can change by at most three if an adjacent generalized transposition is applied, and can change by at most four if a generalized transposition is applied. Consequently, we have:

Corollary 1. *Let $f, g \in S_n$. Then*
(i) $\rho_{\rm gK}(f,g) \geq \frac{1}{3}|b(f)-b(g)|$; *and*
(ii) $\rho_{\rm gC}(f,g) \geq \frac{1}{4}|b(f)-b(g)|$.
In particular, $\rho_{\rm gK}(f,\varepsilon_n) \geq \frac{1}{3}b(f)$ *and* $\rho_{\rm gC}(f,\varepsilon_n) \geq \frac{1}{4}b(f)$.

Farnoud et al. [11] provided the following bounds on the size of permutation codes in the Ulam metric.

Theorem 1 (Farnoud et al. [11]).

$$\frac{(n-d)!}{\binom{n}{d}} \leq A_{\rho_{\rm U}}(n,d+1) \leq (n-d)! \qquad \textit{and}$$
$$\frac{(n-2d-1)!}{2d+1} \leq A_{\rho_{\rm C}}(n,d+1) \leq \frac{n!}{d!}.$$

We provide the corresponding bounds in the generalized Kendall tau metric below. First, we give an upper bound on the number of permutations in S_n with at most k breakpoints.

Proposition 2. *The number of permutations in S_n with at most k breakpoints is at most $\binom{n}{k}k!$.*

Proof. Every permutation $f = (f(1), f(2), \ldots, f(n)) \in S_n$ with at most k breakpoints has k special positions $i_1, i_2, \ldots, i_k \in [n]$ which contains all the breakpoints. The set $\{f(i_1), f(i_2), \ldots, f(i_k)\}$ can be selected in $\binom{n}{k}$ ways and elements ordered in the positions $i_1, i_2, \ldots, i_k$ in $k!$ ways. The remaining elements in $[1,n] \setminus \{f(i_1), f(i_2), \ldots, f(i_k)\}$ take their respective positions as in f. The number of permutations with at most k breakpoints is therefore at most $\binom{n}{k}k!$. This is an overestimate because some of the orderings of $f(i_1), f(i_2), \ldots, f(i_k)$ may give rise to more than k breakpoints in f. ■

Theorem 2. *Let $1 \leq d \leq n$. Then*

$$(n-3d)! \leq A_{\rho_{\rm gK}}(n,d+1) \leq (n-d)! \qquad \textit{and}$$
$$(n-4d)! \leq A_{\rho_{\rm gC}}(n,d+1) \leq (n-d)!.$$

Proof. We first prove the lower bound under $\rho_{\rm gK}$. The proof for the lower bound under $\rho_{\rm gC}$ proceeds similarly.

Consider the ball $B_{\rm gK}(f,r)$ with center f and radius r in the metric space $(S_n, \rho_{\rm gK})$, that is,

$$B(f,r) = \{g \in S_n : \rho_{\rm gK}(f,g) \leq r\}.$$

Since $\rho_{\rm gK}$ is left-invariant, we have $|B(f,r)| = |B(\varepsilon_n, r)| = |\{g \in S_n : \rho_{\rm gK}(g,\varepsilon_n) \leq r\}|$, for any $f \in S_n$. Since $\rho_{\rm gK}(g,\varepsilon_n) \geq \frac{1}{3}b(g)$ by Corollary 1, we have

$$\begin{aligned} |\{g \in S_n : \rho_{\rm gK}(g,\varepsilon_n) \leq r\}| &\leq |\{g \in S_n : b(g) \leq 3r\}| \\ &\leq \binom{n}{3r}(3r)!, \end{aligned}$$

with the last inequality following from Proposition 2. The Gilbert-Varshamov bound then gives

$$\begin{aligned} A_{\rho_{\rm gK}}(n,d+1) &\geq \frac{n!}{B(\varepsilon_n,d)} \\ &\geq \frac{n!}{\binom{n}{3d}(3d)!} \\ &= (n-3d)!. \end{aligned}$$

The upper bound under $\rho_{\rm gK}$ follows from the inequality $\rho_{\rm gK} \leq \rho_{\rm U}$ and Theorem 1, while the upper bound under $\rho_{\rm gC}$ follows from the inequality $\rho_{\rm gC} \leq \rho_{\rm gK}$. ■

IV. Permutation Codes in Generalized Kendall tau Metric

Let $f = (f(1), f(2), \ldots, f(n))$ and $g = (g(1), g(2), \ldots, g(n))$ be two vectors of length n. The *interleave vector* $f \mathbin{|||} g$ is the vector $(f(1), g(1), f(2), g(2), \ldots, f(n), g(n))$. By canonical extension, $|||$ also operates on sets of vectors: if A and B are sets of vectors of length n, then $A \mathbin{|||} B = \{a \mathbin{|||} b : a \in A \text{ and } b \in B\}$.

The following result was obtained by Farnoud et al. [11].

Proposition 3 (Farnoud et al. [11]). *If $\mathcal{C} \subseteq \mathrm{Sym}([p+1, 2p-1])$ is a permutation code of length $p-1$ and Hamming distance $3d/2$, then $\{\varepsilon_p\} \mathbin{|||} \mathcal{C}$ is a permutation code of length $2p-1$ and Ulam distance d.*

In fact, the exact same construction gives a permutation code that has distance d even under the generalized Kendall tau metric. Our proof mimics the proof of Farnoud et al. [11] with the following lemma replacing the role of [11, Lemma 23].

Lemma 1. *Let $f, g \in S_n$ be two permutations such that $\rho_{\rm gK}(f,g) = 1$, that is f and g can be transformed to each other through an adjacent generalized transposition. Then there exist at most three positions $i \in [1, n-1]$ such that for some $j \in [1, n-1]$, we have $f(i) = g(j)$ and $f(i+1) \neq g(j+1)$.*

Proof. For each $i \in [1, n-1]$, there is exactly one $j \in [1, n-1]$ such that $f(i) = g(j)$. Since $\rho_{\rm gK}(f,g) = 1$, there are at most three positions i such that $f(i) = g(j)$ and $f(i+1) \neq g(j+1)$. If the adjacent generalized transposition transforming f to g is $\phi([a,b],[b+1,c])$, these three positions are given by $a-1$, b, and c. ■

Theorem 3. *Let $\pi_1, \pi_2 \in \mathrm{Sym}([p+1, 2p-1])$. If $\rho_{\rm H}(\pi_1, \pi_2) \geq d$, then*

$$\rho_{\rm gK}(\varepsilon_p \mathbin{|||} \pi_1, \varepsilon_p \mathbin{|||} \pi_2) \geq \left\lceil \frac{2d}{3} \right\rceil.$$

Proof. Follow exactly the proof of [11, Theorem 25], with Lemma 1 above replacing the role of [11, Lemma 23]. ■

Corollary 2. *Let p, and $\mathcal{C}$ as in the hypothesis of Proposition 3. Then $\{\varepsilon_p\} \mathbin{|||} \mathcal{C}$ is a permutation code of length $2p-1$ and generalized Kendall tau distance d.*

It follows from Corollary 2 that there exists a permutation code of length n and generalized Kendall tau distance d, whose size is $A_{\rho_{\rm H}}(\lceil n/2 \rceil - 1, \lceil 3d/2 \rceil)$.

A. Recursive Construction

The following result, established by Farnoud et al. [11, Lemma 28] for the Ulam metric, can also be proved in the same manner as Corollary 2.

Theorem 4. *Let $\mathcal{C}_1 \subseteq S_p$ and $\mathcal{C}_2 \subseteq \mathrm{Sym}([p+1, 2p-1])$. If $\mathcal{C}_1$ has generalized Kendall tau distance d and $\mathcal{C}_2$ has Hamming distance $3d/2$, then $\mathcal{C}_1 \,|||\, \mathcal{C}_2$ has generalized Kendall tau distance d.*

Following Farnoud et al. [11], we can apply Theorem 4 recursively, with a permutation code of size one as base case (as in Corollary 2) to obtain

$$A_{\rho_{\mathrm{gK}}}(n, d) \geq \prod_{i=1}^{t} A_{\rho_{\mathrm{H}}}\left(\left\lceil \frac{n}{2^i} \right\rceil - 1, \frac{3d}{2}\right), \tag{3}$$

where

$$t = \left\lfloor \log_2 \frac{n}{3d/2+1} \right\rfloor.$$

We call this recursion of Farnoud, Skachek, and Milenkovic [11], which the required ingredients are really just permutation codes under the Hamming metric, as the *FSM-recursion*. If the ingredient permutation codes used at the steps of the recursion are permutation codes of Hamming distance δ, then we denote the resulting permutation code by $\mathrm{FSM}(n, \delta)$. Hence, what we have shown thus far is that $\mathrm{FSM}(n, 3d/2)$ is a permutation code of length n having generalized Kendall tau distance d.

V. Permutation Codes in generalized Cayley Metric

Permutation Codes in the generalized Cayley metric can be constructed along lines similar to that for the generalized Kendall tau metric.

Lemma 2. *Let $f, g \in S_n$ be two permutations such that $\rho_{\mathrm{gC}}(f, g) = 1$, that is f and g can be transformed to each other through a generalized transposition. Then there exist at most four positions $i \in [1, n-1]$ such that for some $j \in [1, n-1]$, we have $f(i) = g(j)$ and $f(i+1) \neq g(j+1)$.*

Proof. For each $i \in [1, n-1]$, there is exactly one $j \in [1, n-1]$ such that $f(i) = g(j)$. Since $\rho_{\mathrm{gK}}(f, g) = 1$, there are at most three positions i such that $f(i) = g(j)$ and $f(i+1) \neq g(j+1)$. If the adjacent generalized transposition transforming f to g is $\phi([a, b], [c, d])$, these four positions are given by $a-1$, b, $c-1$, and d. ∎

Theorem 5. *Let $\pi_1, \pi_2 \in \mathrm{Sym}([p+1, 2p-1])$. If $\rho_{\mathrm{H}}(\pi_1, \pi_2) \geq d$, then*

$$\rho_{\mathrm{gC}}(\varepsilon_p \,|||\, \pi_1, \varepsilon_p \,|||\, \pi_2) \geq \left\lceil \frac{d}{2} \right\rceil.$$

Proof. Follow exactly the proof of [11, Theorem 25], with Lemma 2 above replacing the role of [11, Lemma 23]. ∎

Corollary 3. *Let $\mathcal{C} \subseteq \mathrm{Sym}([p+1, 2p-1])$ be a permutation code of length $p-1$ and Hamming distance $2d$. Then $\{\varepsilon_p\} \,|||\, \mathcal{C}$ is a permutation code of length $2p-1$ and generalized Cayley distance d.*

It follows from Corollary 3 that there exists a permutation code of length n and generalized Cayley distance d, whose size is $A_{\rho_{\mathrm{H}}}(\lceil n/2 \rceil - 1, \lceil 2d \rceil)$. Corollary 3 can be generalized as in the previous section to give the following.

Theorem 6. *Let $\mathcal{C}_1 \subseteq S_p$ and $\mathcal{C}_2 \subseteq \mathrm{Sym}([p+1, 2p-1])$. If $\mathcal{C}_1$ has generalized Cayley distance d and $\mathcal{C}_2$ has Hamming distance $2d$, then $\mathcal{C}_1 \,|||\, \mathcal{C}_2$ has generalized Cayley distance d.*

Based on Theorem 6, it is now evident that we can apply the FSM-recursion with ingredient permutation codes of Hamming distance $2d$ to obtain $\mathrm{FSM}(n, 2d)$, which is a permutation code of length n and generalized Cayley distance d. This gives

$$A_{\rho_{\mathrm{gC}}}(n, d) \geq \prod_{i=1}^{t} A_{\rho_{\mathrm{H}}}\left(\left\lceil \frac{n}{2^i} \right\rceil - 1, 2d\right), \tag{4}$$

where

$$t = \left\lfloor \log_2 \frac{n}{2d+1} \right\rfloor.$$

VI. Code Enlargement via Breakpoint Analysis

So far, what we have done is observe that the Ulam metric is similar enough to the more general generalized Kendall tau and generalized Cayley metrics for us to prove that the permutation codes constructed by Farnoud et al. [11] under the Ulam metric have similar distance properties under the generalized Kendall tau and generalized Cayley metrics.

In this section, we use the number of breakpoints in the codewords of $\mathrm{FSM}(n, \delta)$ to suggest additional codewords that can be added to $\mathrm{FSM}(n, \delta)$ without affecting distance properties.

First, observe that each codeword in $\mathrm{FSM}(n, \delta)$ has $n-1$ breakpoints, that is, $b(f) = n-1$, for all $f \in \mathrm{FSM}(n, \delta)$. By Proposition 1, $d-1$ adjacent generalized transpositions applied to a permutation can lower its number of breakpoints by at most $3(d-1)$, and $d-1$ generalized transpositions applied to a permutation can lower its number of breakpoints by at most $4(d-1)$. Hence, if we can find a permutation code $\mathcal{C} \subseteq S_n$ of generalized Kendall tau distance (respectively, generalized Cayley distance) d such that each of its codewords have at most $n-3d+1$ (respectively, $n-4d+2$) breakpoints, then $\mathrm{FSM}(n, 3d/2) \cup \mathcal{C}$ (respectively, $\mathrm{FSM}(n, 2d) \cup \mathcal{C}$) is a permutation code of length n and generalized Kendall tau distance (respectively, generalized Cayley distance) d. To construct such a code $\mathcal{C}$, we perform the following:

When under the generalized Kendall tau metric,

take an $\mathrm{FSM}(n-3d+1, 3d/2) \subseteq \mathrm{Sym}([3d, n])$ and append to every vector in $\mathrm{FSM}(n-3d+1, 3d/2)$ the sequence $(1, 2, \ldots, 3d-1)$. The resulting permutation code $\mathcal{C}$ is of length n, generalized Kendall tau distance d, with each codeword having $n-3d+1$ breakpoints.

When under the generalized Cayley metric,

take an $\mathrm{FSM}(n-4d+2, 2d) \subseteq \mathrm{Sym}([4d, n])$ and append to every vector in $\mathrm{FSM}(n-4d+2, 2d)$ the sequence $(1, 2, \ldots, 4d-1)$. The resulting permutation code $\mathcal{C}$ is of length n, generalized Cayley distance d, with each codeword having $n-4d+2$ breakpoints.

We can enlarge the $\mathrm{FSM}(n-3d+1, 3d/2)$ or $\mathrm{FSM}(n-4d+2, 2d)$ used in the construction of $\mathcal{C}$ above in a similar manner. This forms the basis for a recursion which gives the following improvements on (3) and (4):

(i)

$$\begin{aligned}A_{\rho_{\mathrm{gK}}}(n,d) \geq &\prod_{i=1}^{t_0} A_{\rho_{\mathrm{H}}}\left(\left\lceil\frac{n}{2^i}\right\rceil - 1, \frac{3d}{2}\right)\\ &+\prod_{i=1}^{t_1} A_{\rho_{\mathrm{H}}}\left(\left\lceil\frac{n-3d+1}{2^i}\right\rceil - 1, \frac{3d}{2}\right)\\ &+\prod_{i=1}^{t_2} A_{\rho_{\mathrm{H}}}\left(\left\lceil\frac{n-6d+2}{2^i}\right\rceil - 1, \frac{3d}{2}\right)\\ &+\cdots\\ &+\prod_{i=1}^{t_s} A_{\rho_{\mathrm{H}}}\left(\left\lceil\frac{n-s(3d+1)}{2^i}\right\rceil - 1, \frac{3d}{2}\right)\\ =&\sum_{j=0}^{s}\prod_{i=1}^{t_j} A_{\rho_{\mathrm{H}}}\left(\left\lceil\frac{n-j(3d+1)}{2^i}\right\rceil - 1, \frac{3d}{2}\right),\end{aligned}$$

where

$$\begin{aligned}s &= \left\lfloor\frac{n}{3d+1}\right\rfloor, \qquad \text{and}\\ t_j &= \left\lfloor\log_2\frac{n-j(3d+1)}{3d/2+1}\right\rfloor, \qquad \text{for } j\in[0,s].\end{aligned}$$

(ii)

$$\begin{aligned}A_{\rho_{\mathrm{gC}}}(n,d) \geq &\prod_{i=1}^{t_0} A_{\rho_{\mathrm{H}}}\left(\left\lceil\frac{n}{2^i}\right\rceil - 1, 2d\right)\\ &+\prod_{i=1}^{t_1} A_{\rho_{\mathrm{H}}}\left(\left\lceil\frac{n-4d+2}{2^i}\right\rceil - 1, 2d\right)\\ &+\prod_{i=1}^{t_2} A_{\rho_{\mathrm{H}}}\left(\left\lceil\frac{n-8d+4}{2^i}\right\rceil - 1, 2d\right)\\ &+\cdots\\ &+\prod_{i=1}^{t_s} A_{\rho_{\mathrm{H}}}\left(\left\lceil\frac{n-s(4d+2)}{2^i}\right\rceil - 1, 2d\right)\\ =&\sum_{j=0}^{s}\prod_{i=1}^{t_j} A_{\rho_{\mathrm{H}}}\left(\left\lceil\frac{n-j(4d+2)}{2^i}\right\rceil - 1, 2d\right),\end{aligned}$$

where

$$\begin{aligned}s &= \left\lfloor\frac{n}{4d+2}\right\rfloor, \qquad \text{and}\\ t_j &= \left\lfloor\log_2\frac{n-j(4d+2)}{2d+1}\right\rfloor, \qquad \text{for } j\in[0,s].\end{aligned}$$

We note that while the codes constructed above are strictly larger than those constructed by Farnoud et al. [11], they do not improve on the rates asymptotically.

VII. Conclusion

Breakpoint analysis has been a useful technique in the study of the complexity of sorting problems under various allowable data movements (see, for example, [16]–[18]). In this paper, we use breakpoint analysis to obtain enlargement of codes constructed by Farnoud et al. [11], even under more general metrics, such as the generalized Kendall tau and generalized Cayley metrics. Our ability to do this comes from the relative simplicity of counting breakpoints and their usefulness as a proxy measure for distances under transposition-type metrics. We are hopeful that breakpoint analysis will be useful for designing codes under other metrics not considered here.

References

[1] I. F. Blake, "Permutation codes for discrete channels," *IEEE Trans. Inform. Theory*, vol. 20, no. 1, pp. 138–140, 1974.
[2] I. F. Blake, G. Cohen, and M. Deza, "Coding with permutations," *Inform. and Control*, vol. 43, no. 1, pp. 1–19, 1979.
[3] A. J. H. Vinck, "Coded modulation for power line communications," *AEÜ - Int J. Electron. Commun.*, vol. 54, no. 1, pp. 45–49, 2000.
[4] W. Chu, C. J. Colbourn, and P. Dukes, "Constructions for permutation codes in powerline communications," *Des. Codes Cryptogr.*, vol. 32, no. 1-3, pp. 51–64, 2004.
[5] C. J. Colbourn, T. Kløve, and A. C. H. Ling, "Permutation arrays for powerline communication and mutually orthogonal Latin squares," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1289–1291, 2004.
[6] A. Jiang, R. Mateescu, M. Schwartz, and J. Bruck, "Rank modulation for flash memories," *IEEE Trans. Inform. Theory*, vol. 55, no. 6, pp. 2659–2673, 2009.
[7] M. Deza, "Matrices dont deux lignes quelconque coïncident dans un nombre donné de positions communes," *J. Combin. Theory Ser. A*, vol. 20, no. 3, pp. 306–318, 1976.
[8] P. Frankl and M. Deza, "On the maximum number of permutations with given maximal or minimal distance," *J. Comb. Theory Ser. A*, vol. 22, no. 3, pp. 352–360, 1977.
[9] M. Deza and S. A. Vanstone, "Bounds for permutation arrays," *J. Statist. Plann. Inference*, vol. 2, no. 2, pp. 197–209, 1978.
[10] G. Cohen and M. Deza, "Some metrical problems on s_n," *Ann. Discrete Math.*, vol. 8, pp. 211–216, 1980.
[11] F. Farnoud, V. Skachek, and O. Milenkovic, "Error-correction in flash memories via codes in the Ulam metric," *IEEE Trans. Inform. Theory*, vol. 59, no. 5, pp. 3003–3020, 2013.
[12] L. Bulteau, G. Fertin, and I. Rusu, "Sorting by transpositions is difficult," *SIAM J. Discrete Math.*, vol. 26, no. 3, pp. 1148–1180, 2012.
[13] G. A. Watterson, W. J. Ewens, T. E. Hall, and A. Morgan, "The chromosome inversion problem," *J. Theoret. Biol.*, vol. 99, pp. 1–7, 1982.
[14] R. C. Deonier, S. Tavaré, and M. Waterman, *Computational Genome Analysis: An Introduction*. New York: Springer, 2005.
[15] P. Diaconis, *Group Representations in Probability and Statistics*, ser. Lecture Notes - Monograph Series. Institute of Mathematical Statistics, 1988, vol. 11.
[16] J. Kececioglu and D. Sankoff, "Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement," *Algorithmica*, vol. 13, pp. 180–210, 1995.
[17] V. Bafna and A. Pevzner, "Sorting by transpositions," *SIAM J. Discrete Math.*, vol. 11, no. 2, pp. 224–240, 1998.
[18] L. S. Heath and J. P. C. Vergara, "Sorting by bounded block-moves," *Discrete Appl. Math.*, vol. 88, no. 1–3, pp. 181–206, 1998.