Discrete Mathematics 349 (2026) 115253

Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc

Rank-modulation codes for DNA storage in shotgun sequencing: Structure and distance properties

A. Abdollahi [a,*], J. Bagherian [a], F. Jafari [a], M. Khatami [a], A. Orak [a], F. Parvaresh [b], R. Sobhani [c]

[a] *Department of Pure Mathematics, Faculty of Mathematics and Statistics, University of Isfahan, Isfahan 81746-73441, Iran*
[b] *Department of Electrical Engineering, Faculty of Engineering, University of Isfahan, Isfahan 81746-73441, Iran*
[c] *Department of Applied Mathematics and Computer Science, Faculty of Mathematics and Statistics, University of Isfahan, Isfahan 81746-73441, Iran*

ARTICLE INFO

Article history:
Received 30 July 2025
Received in revised form 31 May 2026
Accepted 1 June 2026
Available online xxxx

Keywords:
DNA storage
Permutation codes
de Bruijn graphs
Hamming metric
Kendall τ-metric
Ulam metric

ABSTRACT

Shotgun sequencing is a widely used method for decoding information stored in DNA molecules. This technique generates a histogram or profile vector displaying the occurrence frequencies of all substrings of a specified length ℓ within DNA molecules. To improve resilience against reading errors, a rank-modulation scheme has been proposed. In this scheme, information is encoded as a permutation based on the relative ranking of substring frequencies, rather than their absolute values in a profile vector. For a given alphabet Σ of size q, it is known that not all permutations in S_{Σ^ℓ} (the set of all permutations of strings of length ℓ over Σ) are feasible; some permutations do not correspond to a ranking of a profile vector for any given sequence.

In this paper, we study feasible permutations. We resolve an open problem on the minimum distance of feasible permutations and further examine their maximum distance. Additionally, by identifying a subgroup of order $2q!$ in S_{Σ^ℓ}, we prove that the set of feasible permutations forms a disjoint union of some right cosets of this subgroup. This structure restricts the search for feasible permutations to a set of right coset representatives of the subgroup in S_{Σ^ℓ}, rather than examining all permutations in S_{Σ^ℓ}.

1. Introduction

Recently, there has been a growing interest in coding for DNA storage devices, driven by successful proof of the possibility of storing information with an unparalleled information density in DNA molecules [4,8,15,25]. Shotgun sequencing, a technique discussed in [19,20], is a common technique for reading information stored in DNA molecules. Shotgun sequencing enables relatively accurate reading of lengthy DNA sequences. In this method, short substrings are read separately and reassembled together to form the original string.

The concept of the DNA storage channel, as introduced in [17], is motivated by the shotgun sequencing process (see Fig. 1). In such a channel with an alphabet Σ of size q and a window length ℓ, data is stored as a string over Σ, which is

* Corresponding author.
E-mail addresses: a.abdollahi@math.ui.ac.ir (A. Abdollahi), bagherian@sci.ui.ac.ir (J. Bagherian), math_fateme@yahoo.com (F. Jafari), m.khatami@sci.ui.ac.ir (M. Khatami), orakazadeh62@gmail.com (A. Orak), f.parvaresh@eng.ui.ac.ir (F. Parvaresh), r.sobhani@sci.ui.ac.ir (R. Sobhani).

https://doi.org/10.1016/j.disc.2026.115253
0012-365X/

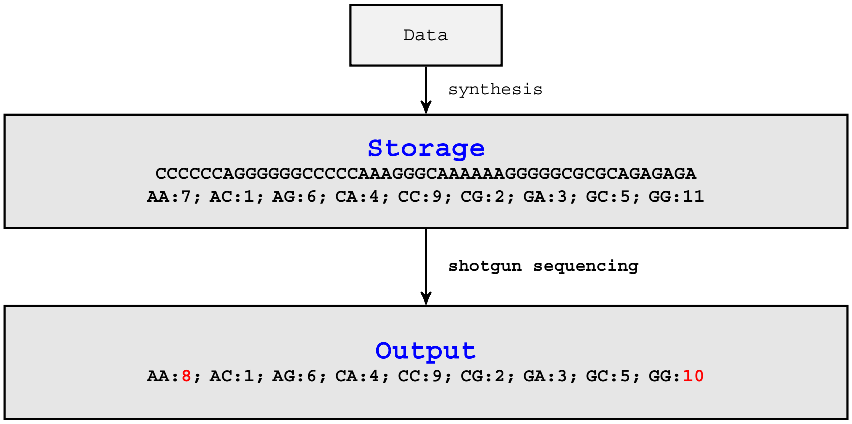

Fig. 1. An illustration of information transmission in a DNA storage channel with $\Sigma = \{A, C, G\}$ and $\ell = 2$. The circular DNA strand stored in the system is recovered using shotgun sequencing, producing a profile vector that may contain some errors but still preserve the rank order of the ℓ-gram frequencies.

treated as circular in this work. The output from the channel is a histogram, or profile vector, potentially containing errors. Each of the q^ℓ entries in this vector counts the occurrences of a specific ℓ-gram (a substring of length ℓ) within the DNA sequence. These ℓ-grams are drawn from Σ^ℓ, which represents all possible strings of length ℓ over the alphabet Σ. Potential errors in this channel model arise from substitutions during DNA synthesis (writing), from sequencing errors (reading), and from incomplete read coverage.

Developing robust coding schemes is essential for addressing the various error types inherent in DNA storage systems, and in [17], several approaches have been proposed for this purpose. Among these approaches is the rank modulation scheme. It was introduced in [17, Sec. VIII.B] and further explored in [5,22]. This approach has been studied in contexts such as vector digitization and signal detection under noisy conditions [3,7,23]. Additionally, the applications of rank modulation have expanded to include powerline communication [26] and, more recently, storage reliability enhancement in non-volatile memory systems [16].

In DNA storage, rank modulation schemes can handle error patterns that do not change the relative ordering of the elements in the profile vector. In particular, rank modulation schemes ignore the absolute values of the profile vector entries and use only their ranks when comparing entries with one another. Thus, each profile vector can be viewed as encoding the permutation induced by its entries. For a permutation to be well defined, the profile values must be distinct, which is a reasonable condition when the strings are long and the window size is small. For example, in Fig. 1, despite the noisy counts produced by shotgun sequencing, the ordering of the frequencies of the 2-grams remains unchanged. In particular, the rank ordering still follows

$$AC < CG < GA < CA < GC < AG < AA < CC < GG.$$

Therefore, the permutation induced by the profile vector obtained from the shotgun-sequencing output is identical to the permutation induced by the original string (see the corresponding permutation in Example 2.8). As a result, this error pattern in the profile vector is correctable using a rank modulation scheme.

A question addressed here is whether, for any arbitrary ranking of ℓ-gram frequencies, one can construct a string whose ℓ-gram counts precisely match this desired order. As stated in [22], the answer is generally no. For instance, consider the alphabet $\Sigma = \{A, C, G\}$ and $\ell = 2$. In this case, no string over Σ exists such that the frequency ranking of its 2-grams satisfies $CA < CG < AC < GC$. This is because achieving such a ranking would imply that the number of times 'C' appears as a prefix of a 2-gram is less than the number of times it appears as a suffix. But, for any string, these two counts must be equal (as the string is treated as circular), leading to a contradiction. Therefore, not every permutation in S_{Σ^ℓ} (the set of all permutations of strings of length ℓ over Σ) is feasible, meaning that some permutations cannot arise from ranking the profile vector of any sequence over Σ. The conditions for a permutation to be feasible have been examined in [5,22], where the authors establish a necessary and sufficient condition (see Lemma 2.6 and Examples 2.10 and 2.11 for feasible and infeasible cases).

A main problem in the field of DNA storage is to determine the number of feasible permutations in S_{Σ^ℓ} for given values of q and ℓ. This problem is important because the number of feasible permutations determines the size of the permutation space that can be used by rank modulation schemes in DNA storage. However, the computational cost of finding this number grows very fast as q and ℓ increase, making it difficult to calculate for larger parameters. For some small values of q and ℓ, the exact numbers have been found in [5,22]. As an alternative, [5,22] provide lower and upper bounds for the number of feasible permutations, which we review in Section 3. Despite these efforts, a general method to count these permutations for any q and ℓ is still missing.

In this paper, we first investigate the algebraic structure of feasible permutations in order to reduce the search space for determining their number. Specifically, we identify a subgroup of S_{Σ^ℓ} with order $2q!$ and show that the set of feasible permutations is a union of several right cosets of this subgroup. Consequently, the search for feasible permutations can be restricted to the set of coset representatives, reducing the search space by a factor of $2q!$. However, a complete classification of these feasible cosets that would fully determine these numbers is still not known.

As discussed, in the framework of rank modulation each DNA sequence induces a feasible permutation by ranking the components of its profile vector. Errors introduced during the shotgun sequencing process may affect the resulting profile vector and consequently modify the induced permutation. Therefore, to design schemes that are robust to such errors, it is natural to study sets of feasible permutations together with appropriate distance measures. Numerous distance measures for permutations have been investigated in the literature. While the Hamming metric is the most commonly used, in the context of rank modulation particular attention has been given to other metrics, such as the infinity metric (also known as the Chebyshev metric) [18,24], the Ulam metric [13], and the Kendall τ-metric [1,2,6,27,28]. For the reader's convenience, the definitions of these metrics are reviewed in Section 4 (see also [11]). One of the open problems raised in [5] is as follows:

Problem 1.1. [5, P. 11] Given a natural number $\ell \in \mathbb{N}$ and an alphabet Σ, what is the minimum pairwise distance between feasible permutations in S_{Σ^ℓ}?

In this paper, we answer this question by determining the minimum pairwise distance between feasible permutations (see Theorem 4.3). In addition, we obtain several further results concerning distances between feasible permutations, including their maximum possible distance (see Theorem 4.9).

The paper is organized as follows. Section 2 introduces the notation and the formal definitions required in the rest of the paper. Section 3 investigates the algebraic structure of the set of feasible permutations. It concludes with an application of our results to compute the number of feasible permutations for $q \in \{3, 4\}$ and $\ell = 2$, and with a review of existing lower and upper bounds for this number. Finally, Section 4 investigates feasible permutations under various metrics and presents several results, including a solution to Problem 1.1.

2. Preliminaries

Throughout this paper, we use the symbol Σ to represent an alphabet of size q. A string of length ℓ over Σ, referred to as an ℓ-gram, is a sequence of ℓ letters from Σ. The set of all such ℓ-grams is denoted by Σ^ℓ. Additionally, we define $\widetilde{\Sigma^\ell} = \{s_1 s_2 \cdots s_\ell \in \Sigma^\ell \mid |\{s_1, s_2, \dots, s_\ell\}| > 1\}$, which includes only those ℓ-grams that contain at least two distinct letters. Also, the set of all finite strings over Σ is denoted by Σ^*. Let's consider an example with an alphabet $\Sigma = \{a, b\}$ and $\ell = 2$. Then, $\Sigma^2 = \{aa, ab, ba, bb\}$ is the set of all 2-grams and $\widetilde{\Sigma^2} = \{ab, ba\}$.

A *directed graph* Γ consists of a non-empty set of vertices $V(\Gamma)$ and a (possibly empty) set of directed edges $E(\Gamma) \subseteq V(\Gamma) \times V(\Gamma)$. Each directed edge $(v_1, v_2) \in E(\Gamma)$ has an *initial vertex* v_1 and a *terminal vertex* v_2. To simplify notation and avoid ambiguity later in the paper, we denote such an edge by $v_1 \to v_2$ instead of the ordered pair (v_1, v_2). An edge $u \to u$ in $E(\Gamma)$, for some $u \in V(\Gamma)$, is called a *loop* in Γ. A *weighted directed graph* is a directed graph Γ, where each edge $e \in E(\Gamma)$ is assigned a numerical value called its weight, denoted by $wt(e)$. For a vertex $v \in V(\Gamma)$, let $\mathcal{I}(v)$ denote the set of edges whose terminal vertex is v, and $\mathcal{O}(v)$ the set of edges whose initial vertex is v. The *in-degree* of v is $|\mathcal{I}(v)|$, and the *out-degree* of v is $|\mathcal{O}(v)|$. A weighted directed graph Γ is called *balanced* if, for all $v \in V(\Gamma)$, $\sum_{e\in\mathcal{I}(v)} wt(e) = \sum_{e\in\mathcal{O}(v)} wt(e)$.

Let the vertices of the directed graph Γ be labeled as $v_1, v_2, \dots, v_n$ and the edges as $e_1, e_2, \dots, e_m$. The *incidence matrix* of Γ is the matrix $B = (B_{ij}) \in \{-1, 0, 1\}^{n\times m}$ defined by

$$B_{ij} = \begin{cases} 1, & \text{if } e_j \in \mathcal{I}(v_i) \setminus \mathcal{O}(v_i), \\ -1, & \text{if } e_j \in \mathcal{O}(v_i) \setminus \mathcal{I}(v_i), \\ 0, & \text{otherwise.} \end{cases}$$

The *null space* of the incidence matrix $B \in \mathbb{R}^{n\times m}$, denoted by null$(B)$, is defined as null$(B) := \{x \in \mathbb{R}^m \mid Bx = \mathbf{0}\}$, where $\mathbf{0}$ is the vector in $\mathbb{R}^n$ whose entries are all zero.

A *subgraph* of a directed graph Γ is a directed graph Γ' such that $V(\Gamma') \subseteq V(\Gamma)$ and $E(\Gamma') \subseteq E(\Gamma) \cap (V(\Gamma') \times V(\Gamma'))$. An *undirected path* in Γ from a vertex u to a vertex v is a finite sequence $(v_1, e_1, v_2, e_2, \dots, e_{k-1}, v_k)$ such that $v_1 = u$, $v_k = v$, the vertices $v_1, \dots, v_k$ are all distinct, and for each $i = 1, \dots, k-1$, the edge $e_i \in E(\Gamma)$ is either $v_i \to v_{i+1}$ or $v_{i+1} \to v_i$. An *directed cycle* (respectively, *undirected cycle*) of length $k \geq 2$ in Γ is a sequence $(v_1, e_1, v_2, e_2, \dots, e_k, v_{k+1})$ such that $v_{k+1} = v_1$, the vertices $v_1, \dots, v_k$ and the edges $e_1, \dots, e_k$ are all distinct, and for each $i = 1, \dots, k$, the edge $e_i \in E(\Gamma)$ is $v_i \to v_{i+1}$ (respectively, either $v_i \to v_{i+1}$ or $v_{i+1} \to v_i$). A directed graph is said to be *weakly connected* if there exists an undirected path between every pair of vertices. A *spanning tree* in Γ is a weakly connected subgraph of Γ that includes all the vertices and contains no undirected cycles.

In the study of strings and sequences, the de Bruijn graph serves as an important analytical tool, defined as follows.

Definition 2.1. The de Bruijn graph of order $\ell \geq 1$ over Σ is the directed graph $G_{q,\ell}$, whose vertex set is $V(G_{q,\ell}) = \Sigma^\ell$, and whose edge set is

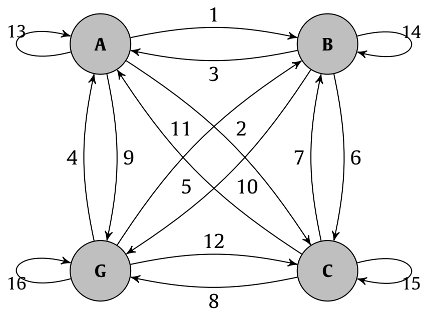

Fig. 2. The weighted de Bruijn graph $G_{4,1}$ of Example 2.8.

$$E(G_{q,\ell}) = \{s_0 s_1 \cdots s_{\ell-1} \to s_1 s_2 \cdots s_\ell \mid for\, all\, s_i \in \Sigma\}.$$

In the context of the de Bruijn graph $G_{g,\ell}$, the edge connecting $s_0 s_1 \cdots s_{\ell-1}$ to $s_1 s_2 \cdots s_\ell$ can be uniquely identified by the string $s_0 s_1 \cdots s_\ell \in \Sigma^{\ell+1}$. We denote by $\widetilde{G}_{q,\ell}$ the graph obtained from $G_{q,\ell}$ by removing all loops, i.e., edges corresponding to strings in $\Sigma^{\ell+1} \setminus \widetilde{\Sigma^{\ell+1}}$.

Let A and B be two non-empty sets. We denote by B^A the set of all functions from A to B. If A is finite, any function $f \in B^A$ can be identified as a vector indexed by A, where each entry at index $a \in A$ is $f(a)$, that is, the value of f at a. Throughout this paper, we write function composition from left to right; that is, for functions f and g, we write $fg(x) := g(f(x))$.

Taking inspiration from the principles of shotgun sequencing, prior research papers [17,22] have recommended the encoding of information within the profile vector of a DNA sequence. Subsequently, the definition of this profile vector will be explained.

Definition 2.2. Given a string $s = s_1 s_2 \cdots s_n$ in Σ^n and a positive integer ℓ, the profile vector of s with order ℓ, denoted by $p_{s,\ell} \in (\mathbb{N} \cup \{0\})^{\Sigma^\ell}$, is a non-negative integer vector indexed by Σ^ℓ. For each $w \in \Sigma^\ell$,

$$p_{s,\ell}(w) = |\{1 \le i \le n \mid s_i s_{i+1} \cdots s_{i+\ell-1} = w\}|,$$

where the indices are taken modulo n, that is, the string s is considered cyclic.

Example 2.3. The profile vector of order 2 for the string presented in Fig. 1 (which is a string in Σ^{48}), is defined as: $p_{s,2} = (7, 1, 6, 4, 9, 2, 3, 5, 11)$ where the indices of the profile vector are listed in lexicographic order: $AA, AC, AG, CA, CC, CG, GA, GC, GG$.

Definition 2.4. $\chi \in (\mathbb{N} \cup \{0\})^{\Sigma^\ell}$ is called feasible if there exists $s \in \Sigma^*$ whose profile vector of order ℓ is χ, namely, $p_{s,\ell} = \chi$.

Example 2.5. The vector $(7, 1, 6, 4, 9, 2, 3, 5, 11) \in \mathbb{N}^{\Sigma^2}$ for $\Sigma = \{A, C, G\}$ (with indices ordered lexicographically) is feasible, as it is the profile vector of order 2 for the string provided in Figure 1.

As discussed in the Introduction, not all vectors $\chi \in (\mathbb{N} \cup \{0\})^{\Sigma^\ell}$ are feasible. Consider the directed graph $G_{q,\ell-1}$ with vertices $\Sigma^{\ell-1}$. For each $w = w_0 \dots w_{\ell-1} \in \Sigma^\ell$, we place $\chi(w)$ parallel edges from $w_0 \dots w_{\ell-2}$ to $w_1 \dots w_{\ell-1}$, in particular, if $\chi(w) = 0$, the corresponding edge is omitted. As discussed in [5,22], the vector χ is the profile vector of order ℓ for some string s in Σ^*, if and only if the graph contains an Eulerian cycle (i.e., a directed cycle traversing each edge exactly once). This is equivalent to the condition that, for every vertex in $G_{q,\ell-1}$, the in-degree equals the out-degree. Hence, by replacing the $\chi(w)$ parallel edges discussed above with a single edge of weight $\chi(w)$, the following lemma is obtained.

Lemma 2.6. *[5, Lemma 1], the vector $\chi \in (\mathbb{N} \cup \{0\})^{\Sigma^\ell}$ is feasible if and only if the weighted de Bruijn graph $G_{q,\ell-1}$, with weights $wt(e) = \chi(e)$ for all $e \in \Sigma^\ell$, is balanced.*

Let A be a set of n elements, which can either be numbers or strings formed from a specified alphabet Σ. The set S_A represents the set of all permutations of the elements of A, and S_n denotes the symmetric group $S_{[n]}$, where $[n] := \{1, 2, \dots, n\}$. Suppose that the elements $a_1, a_2, \dots, a_n$ are ordered lexicographically, either numerically (if A contains numbers) or based on the order in Σ (if A contains strings). In this paper, for any permutation $\pi \in S_A$, we use the vector notation of π as $[\pi(a_1), \pi(a_2), \dots, \pi(a_n)]$. We also denote the identity permutation by $\pi_{id} := [a_1, a_2, \dots, a_n]$. The reverse of a permutation π is defined as $\pi^r := [\pi(a_n), \pi(a_{n-1}), \dots, \pi(a_1)]$. For distinct elements $i, j \in A$, (i, j), which is called transposition, is the permutation obtained from exchanging i and j in π_{id}.

Definition 2.7. Let A be a set, let $\pi \in S_A$, and let χ be a vector in $(\mathbb{N} \cup \{0\})^A$. We say that χ satisfies π, denoted by $\chi \models \pi$, if the entries of χ are distinct and, for any $a, a' \in A$, $\pi(a) < \pi(a')$ if and only if $\chi(a) < \chi(a')$. In this case, π is called the permutation induced by χ.

Recent studies, such as [17], proposed utilizing the rank-modulation scheme for DNA storage, and this approach was further investigated in [22]. In this framework, the information of a DNA sequence s over the alphabet Σ, for a given window length ℓ, is represented by a permutation π. This permutation is obtained by ranking the entries of the profile vector associated with the sequence; that is, $p_{s,\ell} \models \pi$.

Let $s_1, s_2, \ldots, s_{q^\ell}$ denote the elements of Σ^ℓ listed in lexicographic order. Suppose that the entries of the profile vector satisfy

$$p_{s,\ell}(s_{i_1}) < p_{s,\ell}(s_{i_2}) < \cdots < p_{s,\ell}(s_{i_{q^\ell}}).$$

Then the induced permutation $\pi \in S_{\Sigma^\ell}$ is defined by

$$\pi(s_{i_j}) = s_j, \qquad 1 \le j \le q^\ell.$$

In other words, each ℓ-gram is mapped to the ℓ-gram whose position in the lexicographic ordering coincides with its rank in the sorted list of profile values. The following example illustrates the permutation induced by the profile vector of the sequence shown in Fig. 1.

Example 2.8. Let $\Sigma = \{A, C, G\}$. Based on the ranking of the profile vector $p_{s,2}$ presented in Example 2.3, we obtain the following order:

$$\begin{aligned} p_{s,2}(AC) &< p_{s,2}(CG) < p_{s,2}(GA) < p_{s,2}(CA) < p_{s,2}(GC) \\ &< p_{s,2}(AG) < p_{s,2}(AA) < p_{s,2}(CC) < p_{s,2}(GG). \end{aligned}$$

This ranking induces the permutation $\pi = [GA, AA, CG, CA, GC, AC, AG, CC, GG]$ in S_{Σ^2}.

In the following definition, we introduce the feasible permutations.

Definition 2.9. We define a permutation π in S_{Σ^ℓ} as feasible if there exists a feasible vector χ that satisfies π. If no such vector exists, the permutation is called infeasible.

The following two examples provide instances of feasible and infeasible permutations.

Example 2.10. Consider the permutation

$$\sigma = [GA, AA, AB, CA, AC, GB, BB, CB, BA, BC, GC, BG, AG, CC, CG, GG]$$

in S_{Σ^2}, where $\Sigma = \{A, B, C, G\}$. The vector

$$(13, 1, 2, 9, 3, 14, 6, 10, 5, 7, 15, 8, 4, 11, 12, 16) \in (\mathbb{N} \cup \{0\})^{\Sigma^2},$$

with indices arranged in lexicographic order, satisfies σ. Furthermore, since the weighted de Bruijn graph $G_{4,1}$ (see Fig. 2), where weights are defined as $wt(e) = \chi(e)$ for all $e \in \Sigma^2$, is balanced, σ qualifies as a feasible permutation.

Example 2.11. Let $\Sigma = \{A, C, G\}$. Consider the permutation $\pi' = [AG, AA, AC, CA, GC, GA, CG, CC, GG]$ in S_{Σ^2}. We claim that π' is an infeasible permutation. To demonstrate this, suppose for the sake of contradiction that π' is feasible. Then, there must exist a vector χ' in $(\mathbb{N} \cup \{0\})^{\Sigma^2}$ such that $\chi' \models \pi'$ which implies $\chi'(AC) < \chi'(AG) < \chi'(AA) < \chi'(CA) < \chi'(GA)$. Hence, for the vertex A in $G_{3,1}$, $\chi'(AA) + \chi'(AC) + \chi'(AG) < \chi'(GA) + \chi'(CA) + \chi'(AA)$. Therefore, it follows from Lemma 2.6 that χ' is not a feasible vector that is a contradiction.

The following lemma, adapted from [22], gives a corresponding version of Lemma 2.6 that provides a necessary and sufficient condition for the feasibility of a permutation in terms of non-negative real vectors $\chi \in \mathbb{R}^{\Sigma^\ell}$, rather than integer vectors $\chi \in (\mathbb{N} \cup \{0\})^{\Sigma^\ell}$.

Lemma 2.12. *[22, Lemma 7] A permutation $\pi \in S_{\Sigma^\ell}$ is feasible if and only if there exists a non-negative vector $\chi \in \mathbb{R}^{\Sigma^\ell}$ such that $\chi \models \pi$ and the weighted de Bruijn graph $G_{q,\ell-1}$, with weights $wt(e) = \chi(e)$ for all $e \in \Sigma^\ell$, is balanced; that is, for all $v \in \Sigma^{\ell-1}$,*

$$\sum_{\sigma \in \Sigma} \chi(v\sigma) = \sum_{\sigma \in \Sigma} \chi(\sigma v). \tag{2.1}$$

The following definitions will be used in the next section.

Definition 2.13. Suppose $B \subseteq A$ are two finite sets, and π is a permutation in S_A. Then $\pi|_B$ is a unique permutation in S_B that keeps the relative order of the elements of B in π, namely, for all $b, b' \in B$, $\pi|_B(b) < \pi|_B(b')$ if and only if $\pi(b) < \pi(b')$. Furthermore, the permutation π is called an extension of the permutation $\pi|_B$ to the set A. Also, for each τ belonging to S_B, we define $[\tau]^A := \{\sigma \in S_A \mid \sigma|_B = \tau\}$.

Example 2.14. Let $M := \{AA, AC, CA, CC\}$, $D := \{AA, AC, CA\}$ and $\pi = [AC, CC, CA, AA]$ be a permutation in S_M. Then since $\pi(AA) < \pi(CA) < \pi(AC)$, $\sigma := \pi|_D = [AA, CA, AC]$ and

$$[\sigma]^M = \{\pi, [AA, CC, AC, CA], [AA, CC, CA, AC], [AA, CA, AC, CC]\}.$$

Definition 2.15. Let H be a subgroup of a finite group G. For $g \in G$, the set $Hg := \{hg \mid h \in H\}$ is called a right coset of H in G with representative g. Let $\mathbf{X}$ denote the set of all right cosets of H in G. Then $\mathbf{X}$ forms a partition of G, that is, $G = \bigcup_{X \in \mathbf{X}} X$ and $X \cap X' = \varnothing$ for all distinct $X, X' \in \mathbf{X}$, and $|\mathbf{X}| = |G|/|H|$. A subset $\mathcal{T} \subseteq G$ is called a right transversal of H in G if it contains exactly one representative from each right coset of H in G, that is, $G = \bigcup_{t \in \mathcal{T}} Ht$ and $Ht \neq Ht'$ whenever $t \neq t'$ in $\mathcal{T}$.

3. Algebraic structure of feasible permutation codes

In this section, we examine the algebraic structure of the set of feasible permutations in S_{Σ^ℓ}. First, in Lemma 3.5, we introduce a subgroup of S_{Σ^ℓ}. Then, in Corollary 3.9, we prove that the set of feasible permutations is a union of nontrivial cosets of this subgroup. Subsequently, to demonstrate the application of this result, we calculate the number of feasible permutations for $q = 3$ and $q = 4$ with $\ell = 2$.

Definition 3.1. For fixed integers q and ℓ, let $\Phi_{q,\ell}$ denote the set of feasible permutations in S_{Σ^ℓ}, and let $\mathcal{F}_{q,\ell} := |\Phi_{q,\ell}|$ denote the number of such permutations.

As used in the proofs of the theorems in [5], to determine all feasible permutations in S_{Σ^ℓ}, the search space is reduced from the $q^\ell!$ permutations of S_{Σ^ℓ} to the $(q^\ell - q)!$ permutations of $S_{\widetilde{\Sigma^\ell}}$. This reduction follows from the observation that, for any sequence over Σ, the count of an ℓ-gram consisting of identical letters may increase or decrease without affecting the number of ℓ-grams in $\widetilde{\Sigma^\ell}$. For an exact formulation and the precise mathematical notation, see the following remark.

Remark 3.2. Let $\tilde{\Phi}_{q,\ell} := \{\pi|_{\widetilde{\Sigma^\ell}} \mid \pi \in \Phi_{q,\ell}\}$. Note that the weights of the loops in the de Bruijn graph $G_{q,\ell-1}$ correspond to the strings in $\Sigma^\ell \setminus \widetilde{\Sigma^\ell}$. Since removing these loop edges or assigning them zero weight in the weighted de Bruijn graph $G_{q,\ell-1}$ preserves the balance condition, it follows that $\sigma \in S_{\widetilde{\Sigma^\ell}}$ is an element of $\tilde{\Phi}_{q,\ell}$ if and only if there exists $\tilde{\chi} \in (\mathbb{N} \cup \{0\})^{\widetilde{\Sigma^\ell}}$ such that $\tilde{\chi} \models \sigma$, and the weighted graph $\widetilde{G}_{q,\ell-1}$, obtained by assigning weight $\tilde{\chi}(e)$ to each edge $e \in \widetilde{\Sigma^\ell}$, is balanced. Additionally, since every extension of a permutation $\sigma \in \tilde{\Phi}_{q,\ell}$ is an element of $\Phi_{q,\ell}$, we have $\Phi_{q,\ell} = \dot{\bigcup}_{\pi \in \tilde{\Phi}_{q,\ell}} [\pi]^{\Sigma^\ell}$, and $F_{q,\ell} = \tilde{F}_{q,\ell} \times \frac{(q^\ell)!}{(q^\ell - q)!}$, where $\tilde{F}_{q,\ell} := |\tilde{\Phi}_{q,\ell}|$.

In the following definition, we introduce certain elements of S_{Σ^ℓ} that are used in the construction of the subgroup established in Lemma 3.5.

Definition 3.3. Let $\ell \geq 2$. For each permutation $\rho \in S_\Sigma$, we define two permutations: $\pi_\rho^\ell \in S_{\widetilde{\Sigma^\ell}}$ by $\pi_\rho^\ell(x_1 \cdots x_\ell) = \rho(x_1) \cdots \rho(x_\ell)$ for all $x_1 \cdots x_\ell \in \widetilde{\Sigma^\ell}$; and $\bar{\pi}_\rho^\ell \in S_{\Sigma^\ell}$ by $\bar{\pi}_\rho^\ell(x_1 \cdots x_\ell) = \rho(x_1) \cdots \rho(x_\ell)$ for all $x_1 \cdots x_\ell \in \Sigma^\ell$. Also, we define two permutations: $\tau^\ell \in S_{\widetilde{\Sigma^\ell}}$ by $\tau^\ell(x_1 \cdots x_\ell) = x_\ell \cdots x_1$ for all $x_1 \cdots x_\ell \in \widetilde{\Sigma^\ell}$; and $\bar{\tau}^\ell \in S_{\Sigma^\ell}$ by $\bar{\tau}^\ell(x_1 \cdots x_\ell) = x_\ell \cdots x_1$ for all $x_1 \cdots x_\ell \in \Sigma^\ell$. It is clear that $\pi_\rho^\ell = \bar{\pi}_\rho^\ell|_{\widetilde{\Sigma^\ell}}$ and $\tau^\ell = \bar{\tau}^\ell|_{\widetilde{\Sigma^\ell}}$.

Example 3.4. Let $\Sigma = \{A, B, C\}$. Then $\Sigma^2 = \{AA, AB, AC, BA, BB, BC, CA, CB, CC\}$ and $\widetilde{\Sigma^2} = \{AB, AC, BA, BC, CA, CB\}$. Consider the permutation $\rho = [B, C, A] \in S_\Sigma$. Hence, the permutation $\bar{\pi}_\rho^2$ in S_{Σ^2} defined by $\bar{\pi}_\rho^2(x_1x_2) = \rho(x_1)\rho(x_2)$. For example, $\bar{\pi}_\rho^2(AB) = \rho(A)\rho(B) = BC$. Therefore, we get $\bar{\pi}_\rho^2 = [BB, BC, BA, CB, CC, CA, AB, AC, AA]$. Similarly, for the permutation π_ρ^2 in $S_{\widetilde{\Sigma^2}}$ by the same rule, we obtain $\pi_\rho^2 = [BC, BA, CB, CA, AB, AC]$. Also the permutation $\bar{\tau}^2$ in S_{Σ^2} defined by swapping the two symbols: $\bar{\tau}^2(x_1x_2) = x_2x_1$. So, we get $\bar{\tau}^2 = [AA, BA, CA, AB, BB, CB, AC, BC, CC]$. Similarly, we obtain $\tau^2 = [BA, CA, AB, CB, AC, BC]$.

Lemma 3.5. *Let Σ be an alphabet of size $q \geq 3$, and let $\ell \geq 2$ be a natural number. Define $\mathcal{H} := \bigcup_{\rho \in S_\Sigma} \{\pi_\rho^\ell, \pi_\rho^\ell \tau^\ell\}$. Then $\mathcal{H}$ forms a subgroup of size $2q!$ in $S_{\widetilde{\Sigma^\ell}}$.*

Proof. Let $H := \{\pi_\rho^\ell \mid \rho \in S_\Sigma\}$ and K be the subgroup generated by τ^ℓ in $S_{\widetilde{\Sigma^\ell}}$. Since τ^ℓ is an permutation of order 2 and according to definition of H, H and K are two subgroups of order $q!$ and 2 in $S_{\widetilde{\Sigma^\ell}}$, respectively. For each $\pi_\rho^\ell \in H$ and $x_1 \cdots x_\ell \in \widetilde{\Sigma^\ell}$, we have

$$\begin{aligned}\pi_\rho^\ell \tau^\ell(x_1 \cdots x_\ell) &= \tau^\ell(\rho(x_1)\cdots\rho(x_\ell)) = \rho(x_\ell)\cdots\rho(x_1)\\ &= \pi_\rho^\ell(x_\ell \cdots x_1) = \tau^\ell \pi_\rho^\ell(x_1 \cdots x_\ell).\end{aligned}$$

Therefore, $\mathcal{H} = HK$ is a subgroup of $S_{\widetilde{\Sigma^\ell}}$.
Now, we show that $H \cap K = \{\pi_{id}\}$ and therefore $|\mathcal{H}| = |K| \times |H| = 2q!$. Suppose, for a contradiction, that $\tau^\ell \in H$. Hence, there exists a permutation $\rho \in S_\Sigma$ such that for each $x_1 \cdots x_\ell \in \widetilde{\Sigma^\ell}$, $\tau^\ell(x_1 \cdots x_\ell) = x_\ell \cdots x_1 = \rho(x_1)\cdots\rho(x_\ell)$. Since $q \geq 3$, there exist distinct elements y and z in Σ. Consider two distinct elements $x_1 \cdots x_{\ell-1} y$ and $x_1 \cdots x_{\ell-1} z$ in $\widetilde{\Sigma^\ell}$. So, we have $y = \rho(x_1) = z$, that is a contradiction. Hence, $|\mathcal{H}| = 2q!$. This completes the proof. $\square$

Theorem 3.6. *Let Σ be an alphabet of size $q \geq 3$, and let $\ell \geq 2$ be a natural number. Then, $\sigma\delta \in \tilde{\Phi}_{q,\ell}$ for all $\delta \in \tilde{\Phi}_{q,\ell}$ and $\sigma \in \mathcal{H}$, where $\mathcal{H}$ is the subgroup defined in Lemma 3.5.*

Proof. Since $\delta \in \tilde{\Phi}_{q,\ell}$, there exists $\mu \in \Phi_{q,\ell}$ such that $\mu|_{\widetilde{\Sigma^\ell}} = \delta$. Hence, there exists $\chi \in \mathbb{N}^{\Sigma^\ell}$ such that $\chi \models \mu$ and for each $\upsilon \in \Sigma^{\ell-1}$,

$$\sum_{\sigma\in\Sigma} \chi\,(\sigma\upsilon) = \sum_{\sigma\in\Sigma} \chi\,(\upsilon\sigma)\,.$$

Suppose $\rho \in S_\Sigma$ and let $\bar{\chi} := \bar{\pi}_\rho^\ell \chi$. Therefore, for each $\upsilon \in \Sigma^{\ell-1}$, we have:

$$\begin{aligned}\sum_{\sigma\in\Sigma} \bar{\chi}\,(\sigma\upsilon) &= \sum_{\sigma\in\Sigma} \bar{\pi}_\rho^\ell \chi\,(\sigma\upsilon) = \sum_{\sigma\in\Sigma} \chi\,(\rho(\sigma)\rho(\upsilon)) = \sum_{\sigma\in\Sigma} \chi\,(\sigma\rho(\upsilon))\\ &= \sum_{\sigma\in\Sigma} \chi\,(\rho(\upsilon)\sigma) = \sum_{\sigma\in\Sigma} \chi\,(\rho(\upsilon)\rho(\sigma)) = \sum_{\sigma\in\Sigma} \bar{\pi}_\rho^\ell \chi\,(\upsilon\sigma) = \sum_{\sigma\in\Sigma} \bar{\chi}\,(\upsilon\sigma)\,,\end{aligned}$$

where $\rho(\upsilon)$ refers to $\rho(x_1)\ldots\rho(x_{\ell-1})$, for all $\upsilon = x_1 \ldots x_{\ell-1} \in \Sigma^{\ell-1}$. Therefore, $\bar{\chi}$ is also a feasible vector. Moreover, if $\upsilon = x_1 \ldots x_\ell$ and $\omega = y_1 \ldots y_\ell$ are two distinct elements in Σ^ℓ such that $\bar{\pi}_\rho^\ell \mu(\upsilon) < \bar{\pi}_\rho^\ell \mu(\omega)$, then we have $\mu\,(\rho(x_1)\cdots\rho(x_\ell)) < \mu\,(\rho(y_1)\cdots\rho(y_\ell))$. Since $\chi \models \mu$, $\chi\,(\rho(x_1)\cdots\rho(x_\ell)) < \chi\,(\rho(y_1)\cdots\rho(y_\ell))$. Therefore, $\bar{\pi}_\rho^\ell \chi(\upsilon) < \bar{\pi}_\rho^\ell \chi(\omega)$. Hence, $\bar{\chi} \models \bar{\pi}_\rho^\ell \mu$ and therefore $\bar{\pi}_\rho^\ell \mu \in \Phi_{q,\ell}$. It is easy to see that $\pi_\rho^\ell \delta = (\bar{\pi}_\rho^\ell \mu)|_{\widetilde{\Sigma^\ell}}$ and therefore $\pi_\rho^\ell \delta \in \tilde{\Phi}_{q,\ell}$.
In the following, we show that $\tau^\ell \delta \in \tilde{\Phi}_{q,\ell}$. Let $\tilde{\chi} := \bar{\tau}^\ell \chi$. Then for each$\upsilon = x_1 \cdots x_{\ell-1} \in \Sigma^{\ell-1}$, we have:

$$\begin{aligned}\sum_{\sigma\in\Sigma} \tilde{\chi}\,(\sigma\upsilon) &= \sum_{\sigma\in\Sigma} \bar{\tau}^\ell \chi\,(\sigma\upsilon) = \sum_{\sigma\in\Sigma} \bar{\tau}^\ell \chi\,(\sigma x_1 \cdots x_{\ell-1}) = \sum_{\sigma\in\Sigma} \chi\,(x_{\ell-1}\cdots x_1\sigma)\\ &= \sum_{\sigma\in\Sigma} \chi\,(\sigma x_{\ell-1}\cdots x_1) = \sum_{\sigma\in\Sigma} \bar{\tau}^\ell \chi\,(x_1 \cdots x_{\ell-1}\sigma)\\ &= \sum_{\sigma\in\Sigma} \bar{\tau}^\ell \chi\,(\upsilon\sigma) = \sum_{\sigma\in\Sigma} \tilde{\chi}\,(\upsilon\sigma)\,.\end{aligned}$$

Then, $\tilde{\chi}$is a feasible vector. Suppose that $\upsilon = x_1 \cdots x_\ell$ and $\omega = y_1 \cdots y_\ell$are two distinct elements in Σ^ℓ such that $\bar{\tau}^\ell \mu(\upsilon) < \bar{\tau}^\ell \mu(\omega)$. Then we have $\mu(x_\ell \cdots x_1) < \mu(y_l \cdots y_1)$. Since $\chi \models \mu$, we have $\chi(x_\ell \cdots x_1) < \chi(y_\ell \cdots y_1)$ and so $\bar{\tau}^\ell \chi(x_1 \cdots x_\ell) < \bar{\tau}^\ell \chi(y_1 \cdots y_\ell)$. Hence, $\tilde{\chi} \models \bar{\tau}^\ell \mu$ and therefore $\bar{\tau}^\ell \mu \in \Phi_{q,\ell}$. Now the result follows from $\tau^\ell \delta = (\bar{\tau}^\ell \mu)|_{\widetilde{\Sigma^\ell}}$ and this completes the proof. $\square$

Definition 3.7. Let G be a group with identity element e and Θ a non-empty set. Recall that we say G acts (on the left) on Θ if there exists a function $\phi : G \times \Theta \to \Theta$, denoted by $(g, \theta) \mapsto g \cdot \theta$, such that for all $g, h \in G$ and $\theta \in \Theta$, we have $(gh)\cdot\theta = g\cdot(h\cdot\theta)$ and $e\cdot\theta = \theta$. The *orbit* of an element $\theta \in \Theta$ under the action of G is defined as $\mathrm{Orb}_G(\theta) := \{g\cdot\theta \mid g \in G\}$.

Remark 3.8. It is well known that the orbits of a group action form a partition of the set Θ, that is, each element of Θ belongs to exactly one orbit. See [10, Chapter 1, p. 45].

Corollary 3.9. *Let Σ be an alphabet of size $q \geq 3$ and $\ell \geq 2$ be a natural number. Then $\tilde{\Phi}_{q,\ell}$ is a disjoint union of some non-trivial right cosets of $\mathcal{H}$ in $S_{\widetilde{\Sigma^\ell}}$, where $\mathcal{H}$ is the subgroup defined in Lemma 3.5. Moreover, if $\mathcal{T}$ is a right transversal of $\mathcal{H}$ in $S_{\widetilde{\Sigma^\ell}}$ and $F_\mathcal{T}$ denotes the number of feasible permutations in $\mathcal{T}$, then $\widetilde{F}_{q,\ell} = 2q!\,F_\mathcal{T}$.*

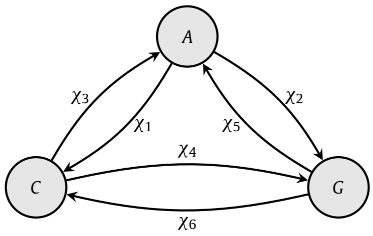

Fig. 3. The de Bruijn graph $\widetilde{G}_{3,1}$ with edge weights $\chi_1, \ldots, \chi_6$. The balance conditions are: $\chi_3 + \chi_5 = \chi_1 + \chi_2$ (at A), $\chi_1 + \chi_6 = \chi_3 + \chi_4$ (at C), and $\chi_2 + \chi_4 = \chi_5 + \chi_6$ (at G).

Proof. By Theorem 3.6, the action of the subgroup $\mathcal{H}$ on $\tilde{\Phi}_{q,\ell}$ is well-defined via the function $\mathcal{H} \times \tilde{\Phi}_{q,\ell} \to \tilde{\Phi}_{q,\ell}$, defined as $(\sigma, \delta) \mapsto \sigma\delta$. For each $\delta \in \tilde{\Phi}_{q,\ell}$, the orbit $\mathrm{Orb}_{\mathcal{H}}(\delta)$ coincides with the right coset $\mathcal{H}\delta$ of $\mathcal{H}$ in $S_{\widetilde{\Sigma^\ell}}$. Hence, the orbits of this action correspond to the right cosets of $\mathcal{H}$ in $S_{\widetilde{\Sigma^\ell}}$, and thus, $\tilde{\Phi}_{q,\ell}$ can be expressed as a disjoint union of these right cosets.

Now, it remains to show that $\pi_{id} \notin \tilde{\Phi}_{q,\ell}$. Suppose, for the sake of contradiction, that $\pi_{id} \in \tilde{\Phi}_{q,\ell}$. Then, by Remark 3.2, we have $\pi_{id} \in \Phi_{q,\ell}$. Hence, there exists $\chi \in (\mathbb{N} \cup \{0\})^{\Sigma^\ell}$ such that $\chi \models \pi_{id}$, and for all $v \in \Sigma^{\ell-1}$, $\sum_{\sigma\in\Sigma} \chi(\sigma v) = \sum_{\sigma\in\Sigma} \chi(v\sigma)$. Let $s_1, \ldots, s_q$ be the elements of Σ arranged in lexicographic order and let $v = \underbrace{s_1 \cdots s_1}_{\ell-1}$. Since $\chi \models \pi_{id}$, for all $\sigma \in \Sigma \setminus \{s_1\}$, we have $\chi(v\sigma) < \chi(\sigma v)$, which implies $\sum_{\sigma\in\Sigma} \chi(\sigma v) > \sum_{\sigma\in\Sigma} \chi(v\sigma)$, which is a contradiction. This proves the first statement. The second statement follows from the first statement together with the fact that $|\mathcal{H}| = 2q!$. $\square$

Note that a similar argument can be applied to show that if we define $\bar{\mathcal{H}} := \bigcup_{\rho\in S_\Sigma} \{\bar{\pi}_\rho^\ell, \bar{\pi}_\rho^\ell \bar{\tau}^\ell\}$ as a subset of S_{Σ^ℓ}, then $\Phi_{q,\ell}$ is a disjoint union of some non-trivial right cosets of $\bar{\mathcal{H}}$ in S_{Σ^ℓ}.

3.1. Applications of Corollary 3.9

Determining the number $\mathcal{F}_{q,\ell}$ of feasible permutations is a central problem in the study of rank modulation schemes for DNA storage. It is known that $\mathcal{F}_{q,1} = q!$ for all q, while $\mathcal{F}_{2,\ell} = 0$ for all $\ell \geq 2$ (see [22, p. 4]). In [22], a linear programming (LP) approach was developed to decide the feasibility of a given permutation. Specifically, a permutation $\pi \in S_{\Sigma^\ell}$ is feasible if and only if the following linear program is feasible:

$$\begin{aligned}
\min \quad & \sum_{w\in\Sigma^\ell} \chi(w) \\
\text{s.t.} \quad & \chi(w) \geq 1, \quad \forall w \in \Sigma^\ell, \\
& \chi(w_1) - \chi(w_0) \geq 1, \quad \forall w_0, w_1 \in \Sigma^\ell \text{ with } \pi(w_0) < \pi(w_1), \\
& \sum_{\sigma\in\Sigma} \chi(\sigma w) = \sum_{\sigma\in\Sigma} \chi(w\sigma), \quad \forall w \in \Sigma^{\ell-1}.
\end{aligned} \tag{3.1}$$

This algorithm is an immediate consequence of Lemma 2.12. Since the weights of loops do not affect the balance of the de Bruijn graph, the algorithm can be reformulated for permutations $\pi \in S_{\widetilde{\Sigma^\ell}}$. In particular, a permutation π belongs to $\widetilde{\Phi}_{q,\ell}$ if and only if the corresponding LP is feasible.

The reformulated algorithm for permutations in $S_{\widetilde{\Sigma^\ell}}$ involves $q^\ell - q$ variables and $2q^\ell - q - 1$ constraints. For example, let $\Sigma = \{A, C, G\}$ and consider the permutation $\pi = [AC, GC, CG, AG, CA, GA] \in S_{\widetilde{\Sigma^2}}$. Let $\chi = (\chi_1, \ldots, \chi_6) \in \mathbb{R}^{\widetilde{\Sigma^2}}$ correspond to the edges (AC, AG, CA, CG, GA, GC), respectively. Based on the balance equations in Fig. 3 and the fact that $\pi(AC) < \pi(CG) < \pi(GA) < \pi(CA) < \pi(GC) < \pi(AG)$, the LP formulation for π is:

$$\begin{aligned}
\min \quad & \sum_{i=1}^{6} \chi_i \\
\text{s.t.} \quad & \chi_i \geq 1 \quad (i = 1, \ldots, 6), \\
& \chi_4 - \chi_1 \geq 1,\ \chi_5 - \chi_4 \geq 1,\ \chi_3 - \chi_5 \geq 1,\ \chi_6 - \chi_3 \geq 1,\ \chi_2 - \chi_6 \geq 1, \\
& \chi_3 + \chi_5 = \chi_1 + \chi_2,\ \chi_1 + \chi_6 = \chi_3 + \chi_4,\ \chi_2 + \chi_4 = \chi_5 + \chi_6.
\end{aligned}$$

Since $\chi = (1, 6, 4, 2, 3, 5)$ is a feasible solution, π is a feasible permutation.

Using this approach, the value of $\mathcal{F}_{3,2}$ was previously computed as 30′240 (cf. [22, p. 7]). However, for larger q or ℓ, this method becomes computationally intractable due to the factorial growth of $|S_{\widetilde{\Sigma^\ell}}|$. By Corollary 3.9, the set of feasible permutations is the union of certain right cosets of the subgroup $\mathcal{H}$. Consequently, it suffices to examine a right transversal

Table 1
Computational results obtained using the reduction of Corollary 3.9. All computations were performed on a machine with Xeon E5-2680 v3 CPUs and 128 GB RAM.

q	$\lvert S_{\widetilde{\Sigma^2}}\rvert$	$\lvert\mathcal{T}\rvert$	$F_{\mathcal{T}}$	$\widetilde{F}_{q,2}$	Runtime (red.)	Runtime (full)
3	720	60	5	60	1.34s	17.12 s
4	479′001′600	9′979′200	727′481	34′919′088	115.8 h	≈ 231 d

for $\mathcal{H}$ in $S_{\widetilde{\Sigma^\ell}}$. Such a transversal has size $\frac{(q^\ell-q)!}{2q!}$, which represents a substantial reduction from the original $(q^\ell-q)!$ permutations. In the following, we apply Corollary 3.9 to calculate $\mathcal{F}_{q,2}$ for $q\in\{3,4\}$.

- $q=3$ **and** $\ell=2$**:**
Let $\Sigma=\{a,b,c\}$. Then $\widetilde{\Sigma^2}=\{ab,ac,ba,bc,ca,cb\}$, which we index by $\{1,2,\dots,6\}$ according to the lexicographic order:
$$ab\mapsto 1,\ ac\mapsto 2,\ ba\mapsto 3,\ bc\mapsto 4,\ ca\mapsto 5,\ cb\mapsto 6.$$
Moreover, $S_\Sigma=\{[a,b,c],[a,c,b],[b,a,c],[b,c,a],[c,a,b],[c,b,a]\}$. For each $\rho\in S_\Sigma$, the permutation $\pi_\rho^2\in S_{\widetilde{\Sigma^2}}$ is defined by $\pi_\rho^2(x_1x_2)=\rho(x_1)\rho(x_2)$. For instance, let $\rho=[b,a,c]$. Then
$$ab\mapsto ba,\quad ac\mapsto bc,\quad ba\mapsto ab,\quad bc\mapsto ac,\quad ca\mapsto cb,\quad cb\mapsto ca,$$
and hence $\pi^2_{[b,a,c]}=[3,4,1,2,6,5]$. Similarly, for the six permutations in S_Σ, the corresponding permutations in S_6 are
$$\begin{aligned}\pi^2_{[b,a,c]}&=[3,4,1,2,6,5], & \pi^2_{[c,b,a]}&=[6,5,4,3,2,1],\\ \pi^2_{[a,c,b]}&=[2,1,5,6,3,4], & \pi^2_{[b,c,a]}&=[4,3,6,5,1,2],\\ \pi^2_{[c,a,b]}&=[5,6,2,1,4,3], & \pi^2_{[a,b,c]}&=[1,2,3,4,5,6].\end{aligned}$$
Also, by definition of τ^ℓ, we have $\tau^2(x_1x_2)=x_2x_1$, which yields $\tau^2=[3,5,1,6,2,4]$ under the above indexing. According to Lemma 3.5, if $\mathcal{H}:=\bigcup_{\rho\in S_\Sigma}\{\pi_\rho^2,\pi_\rho^\ell\tau^2\}$, then $\mathcal{H}$ forms a subgroup of size $2\times 3!=12$ in $S_{\widetilde{\Sigma^2}}$.
According to Corollary 3.9, instead of examining all 720 elements of $S_{\widetilde{\Sigma^2}}$, it suffices to consider a right transversal of $\mathcal{H}$ in $S_{\widetilde{\Sigma^2}}$, which consists of $\frac{720}{12}=60$ permutations. Hence, it is enough to determine which representatives of these 60 right cosets of $\mathcal{H}$ are feasible. By solving the corresponding linear programming problem for each of these 60 permutations, only 5 elements were found to be feasible (the GAP [14] and Python [21] codes used for this computation are provided in the Appendix):
$$\pi_1=[1,4,2,5,3,6],\quad \pi_2=[1,4,3,5,2,6],\quad \pi_3=[1,5,2,4,3,6],$$
$$\pi_4=[1,5,3,4,2,6],\quad \pi_5=[1,5,4,3,2,6].$$
Thus, by Corollary 3.9, we have $\tilde{\Phi}_{3,2}=\dot{\bigcup}_{i=1}^5\mathcal{H}\pi_i$ and therefore, $\tilde{F}_{3,2}=60$, and by Remark 3.2, $F_{3,2}=30'240$.
- $q=4$ **and** $\ell=2$**:**
Similar to the case $q=3$ and $\ell=2$, we first determine the elements of the subgroup $\mathcal{H}$. Using the computational algebra system GAP [14], we computed a right transversal of $\mathcal{H}$ in $S_{\widetilde{\Sigma^2}}\cong S_{12}$. The transversal contains $\frac{(4^2-4)!}{2\cdot 4!}=\frac{12!}{48}=9'979'200$ representatives. For each representative in the transversal, we solved the corresponding linear programming problem using Python [21]. The GAP and Python codes used in our computations are provided in the Appendix. Among the 9′979′200 examined permutations, 727′481 were found to be feasible. Therefore, $\widetilde{F}_{4,2}=2\times 4!\times 727'481=34'919'088$, and consequently, by Remark 3.2, $F_{4,2}=1'525'265'763'840$.

Table 1 summarizes the computational results obtained using the reduction provided by Corollary 3.9. The second column shows the total number of permutations in $S_{\widetilde{\Sigma^2}}$, while the third column gives the size of a right transversal $\mathcal{T}$ of $\mathcal{H}$ in $S_{\widetilde{\Sigma^2}}$. The column $F_{\mathcal{T}}$ records the number of feasible representatives in the transversal, and the corresponding value of $\widetilde{F}_{q,2}=2q!F_{\mathcal{T}}$ is reported in the next column.

For $q=3$, we also performed the computation without applying the reduction of Corollary 3.9. In this case, the LP formulation had to be solved for all 720 permutations in $S_{\widetilde{\Sigma^2}}$, resulting in a runtime of 17.12 seconds, compared to 1.34 seconds when restricting the search to the transversal. For $q=4$, a direct computation would require solving the LP for all 12! permutations. Assuming an approximately linear scaling with the number of LP instances, this corresponds to an estimated runtime of about 231 days.

The bounds for $F_{4,2}$ have been previously reported in [5]. In Table 1 of that paper, the best known lower bound is given as $F_{4,2}>1'296'453'150'720$, and the same table also reports the value $F_{4,2}=1'540'034'496'000$, which is stated to have been obtained via an exhaustive computer search. Our computation yields $F_{4,2}=1'525'265'763'840$. This value improves upon the previously reported lower bound but differs from the value listed in [5]. The discrepancy suggests that the reported computation of $F_{4,2}$ may require further verification.

We note that using computational verification for $q=3$ and $\ell=2$, we confirmed that $\mathcal{H}$ is the largest subgroup of $S_{\widetilde{\Sigma^\ell}}$ such that, for every $\sigma \in \mathcal{H}$ and $\pi \in \tilde{\Phi}_{q,\ell}$, $\sigma\pi$ is also in $\tilde{\Phi}_{q,\ell}$. This suggests that the same property may hold for all $q \geq 3$ and $\ell \geq 2$.

In the sequel, we provide an overview of the existing known bounds on the value of $\mathcal{F}_{q,\ell}$. In [22, Corollary 2], a general lower bound valid for all $q \geq 3$ and $\ell \geq 2$ is given. This bound was later improved in [5, Corollary 3] for all $q \geq 3$ and $\ell \geq 2$, except for the specific case of $q=3$ and $\ell=2$. The improved bound takes the following form:

$$\mathcal{F}_{q,\ell} \geq \left(q^\ell - q^{\ell-1} + q - 2\right)! \cdot \frac{q-1}{q+1} \cdot \frac{q^\ell !}{(q^\ell ! - q)!}.$$

Furthermore, the best known upper bounds for $\mathcal{F}_{q,\ell}$ are presented in [22, Theorems 1 and 2]. For $q \geq 3$ and $\ell \geq 3$, the upper bound is given by:

$$\mathcal{F}_{q,\ell} \leq q^\ell ! \cdot \left(\frac{q-1}{q+1}\right)^{\frac{1}{4}(q^{\ell-1}-q^{\ell-3})}.$$

In particular, for $\ell=2$, a specific upper bound is provided: $\mathcal{F}_{q,2} \leq q^2! \cdot \left(\frac{q-2}{q}\right)$.

4. Distance properties of feasible permutations

In the previous section we investigated the structural properties of the set $\Phi_{q,\ell}$ of feasible permutations. In this section we turn to another natural problem concerning this set, namely the study of distances between feasible permutations and the construction of large subsets with a prescribed minimum distance.

Let A be a set of elements and let d be a distance metric on permutations of A. A permutation code $\mathcal{C} \subseteq S_A$ with minimum distance δ is a subset such that $d(\pi,\rho) \geq \delta$ for all distinct $\pi,\rho \in \mathcal{C}$. The elements of $\mathcal{C}$ are called codewords.

Permutation codes provide robustness against errors in settings where information is represented by permutations. In the rank modulation framework considered in this paper, a DNA sequence induces a permutation through its ℓ-gram profile vector. Errors introduced during the sequencing process may alter this permutation. By selecting permutations from a code with sufficiently large minimum distance, the observed permutation can be decoded to the closest codeword with respect to the chosen distance metric, thereby allowing small errors to be corrected.

Since only feasible permutations can arise from strings, our goal is to construct permutation codes contained in $\Phi_{q,\ell}$ and to determine large subsets of $\Phi_{q,\ell}$ with prescribed minimum distance.

We begin by reviewing several distance metrics on permutations that are relevant to our setting.

Definition 4.1. Let $A = \{a_1,\dots,a_n\}$ be a set of elements ordered lexicographically, and let S_A denote the set of all permutations of A. Given two permutations $\pi,\rho \in S_A$, we define the following distances:

- **Kendall τ-distance:** $d_K(\pi,\rho)$ is the minimum number of adjacent transpositions required to transform π into ρ. An adjacent transposition is a swap of two consecutive entries in a permutation. More precisely, for $1 \leq i \leq n-1$, the adjacent transposition (a_i,a_{i+1}) transforms $[\pi(a_1),\dots,\pi(a_i),\pi(a_{i+1}),\dots,\pi(a_n)]$ into $[\pi(a_1),\dots,\pi(a_{i+1}),\pi(a_i),\dots,\pi(a_n)]$.
- **Hamming distance:** $d_H(\pi,\rho) = |\{i \in [n] \mid \pi(a_i) \neq \rho(a_i)\}|$, which counts the number of positions where the two permutations differ.
- **Infinity distance:** $d_\infty(\pi,\rho) = \max_{i\in[n]} |\mathrm{pos}(\pi(a_i)) - \mathrm{pos}(\rho(a_i))|$, where $\mathrm{pos}(x)$ denotes the position of x in the fixed lexicographic ordering of A.
- **Ulam distance:** $d_U(\pi,\rho) = n - \mathrm{LCS}(\pi,\rho)$, where $\mathrm{LCS}(\pi,\rho)$ is the length of the longest common subsequence between π and ρ.

Example 4.2. Let $A = \{a,b,c,d\}$ and consider $\pi = [c,a,d,b]$ and $\rho = [a,d,b,c]$.

- $d_H(\pi,\rho) = 4$, as they differ at every position.
- $d_K(\pi,\rho) = 3$, corresponding to the minimum number of swaps.
- $d_U(\pi,\rho) = 4-3 = 1$, since the longest common subsequence is $[a,d,b]$, which has length 3.
- $d_\infty(\pi,\rho) = 3$, since $\max\{|\mathrm{pos}(c)-\mathrm{pos}(a)|, |\mathrm{pos}(a)-\mathrm{pos}(d)|, |\mathrm{pos}(d)-\mathrm{pos}(b)|, |\mathrm{pos}(b)-\mathrm{pos}(c)|\} = \max\{2,3,2,1\} = 3$.

The following theorem answers Problem 1.1.

Theorem 4.3. *For a natural number $\ell \in \mathbb{N}$ and an alphabet Σ, the minimum distance between feasible permutations in S_{Σ^ℓ} is 1 when using Kendall τ-distance, infinity distance, or Ulam distance and is equal 2 when using Hamming distance.*

Proof. Let π be an feasible permutation in S_{Σ^ℓ}. Then it follows from Lemma 2.6 that there exists a vector $\chi \in (\mathbb{N} \cup \{0\})^{\Sigma^\ell}$ such that $\chi \models \pi$ and the weighted de Bruijn graph, $G_{q,\ell-1}$, with weights $wt(e) = \chi(e)$ for all $e \in \Sigma^\ell$, is balanced. Since, in the de Bruijn graph, the out-degree and in-degree of every vertex are equal, we may assume, without loss of generality, that all entries of χ are strictly positive, because we can add the same positive integer to every entry of χ. Let $u \in \Sigma$ and let $s_1, s_2, \dots, s_{i-1}, s_i, s_{i+1}, \dots, s_{q^\ell}$ be all lexicographically ordered elements of Σ^ℓ. Suppose that $s_i = \underbrace{uu\cdots u}_{\ell}$. Note that $\chi(s_i)$ is the weight of the loop (v, v) in $G_{q,\ell-1}$, where $v = \underbrace{uu\cdots u}_{\ell-1}$. By choosing a new positive integer weight $\chi'(s_i)$ for this loop (different from existing weights of χ), and keeping all other weights unchanged, we obtain a new vector $\chi' \in \mathbb{N}^{\Sigma^\ell}$. Since changing the loop weights does not affect the graph's balance, the weighted de Bruijn graph $G_{q,\ell-1}$, with weights $wt(e) = \chi'(e)$ for all $e \in \Sigma^\ell$, then becomes balanced. Thus the permutation induced by χ' is feasible. We will leverage this fact. Suppose that $\chi(s_{k_1}) < \chi(s_{k_2}) < \chi(s_{k_3}) < \cdots < \chi(s_{k_{q^\ell}})$, where $\{k_1, \dots, k_{q^\ell}\} = [q^\ell]$. Further suppose that $k_t = i + 1$. Note that since $q \geq 3$, we may assume that $t \notin \{q^\ell, 1\}$. Consider two vectors χ_0 and χ_1 defined as follows: $\chi_0(s_j) = \chi_1(s_j) = \chi(s_j)$ for all $j \neq i$ and $\chi_0(s_i) = \frac{\chi(s_{k_t}) + \chi(s_{k_{t+1}})}{2}$ and $\chi_1(s_i) = \frac{\chi(s_{k_t}) + \chi(s_{k_{t-1}})}{2}$. Note that if $\chi_0(s_i)$ or $\chi_1(s_i)$ are not integer, then without loss of generality we may replace χ with 2χ. According to the fact established above, if π_0 and π_1 are the permutations induced by χ_0 and χ_1, respectively, then π_0 and π_1 are two feasible permutations in S_{Σ^ℓ}. Clearly, $\pi_0(s_j) = \pi_1(s_j)$ for all $j \in [q^\ell] \setminus \{i, i + 1\}$ and if $\pi_0(s_i) = s_m$, then $\pi_0(s_{i+1}) = s_{m-1}$, $\pi_1(s_{i+1}) = s_m$ and $\pi_1(s_i) = s_{m-1}$. Therefore, $d_H(\pi_0, \pi_1) = |\{s_i, s_{i+1}\}| = 2$, $\pi_0\pi_1^{-1} = (s_i, s_{i+1})$ and so $d_K(\pi_0, \pi_1) = 1$. Also in view of the definitions of the infinity distance and Ulam distance, $d_\infty(\pi_0, \pi_1) = 1$ and $d_U(\pi_0, \pi_1) = 1$. This completes the proof. □

At this point, a natural question arises: *What is the minimum distance in $\widetilde{\Phi}_{q,\ell}$ with respect to the Hamming, Kendall, Ulam, and infinity distances?* We address this question in the special case $\ell = 2$, as presented in Theorem 4.6. Using a suitable basis for the null space of the de Bruijn graph $\widetilde{G}_{q,1}$, we construct a pair of permutations in $\widetilde{\Phi}_{q,2}$ that attain the minimum value of each of these distances. However, determining the minimum distance in the general case $\ell > 2$ remains an open problem. In the following, we review the necessary concepts used in the proof.
Let Γ be a directed graph and let C be an undirected cycle in Γ. The *cycle incidence vector* of C, denoted by η_C, is a vector in $\{-1, 0, 1\}^{E(\Gamma)}$ defined as follows. Fix an arbitrary traversal direction on C, and for each edge $e \in E(\Gamma)$, let

$$\eta_C(e) = \begin{cases} 1 & \text{if } e \in C \text{ and has the same orientation as the traversal direction,} \\ -1 & \text{if } e \in C \text{ but has the opposite orientation,} \\ 0 & \text{if } e \notin C. \end{cases}$$

The following proposition is derived from [9, P. 7].

Proposition 4.4. *Let Γ be a weakly connected directed graph, and let T be a spanning tree of Γ. For each edge $e \in E(\Gamma) \setminus E(T)$, adding e to T creates a unique undirected cycle C_e^T. Furthermore, the set of cycle incidence vectors*

$$\left\{ \eta_{C_e^T} \;\middle|\; e \in E(\Gamma) \setminus E(T) \right\}$$

forms a basis for the null space of the incidence matrix of Γ.

The feasibility of a permutation π can be expressed in terms of the null space of the incidence matrix of the de Bruijn graph, as stated in the following remark.

Remark 4.5. Let Σ be an alphabet of size q, and $\ell \geq 2$. Let $e_1, \dots, e_{q^\ell - q}$ denote the edges of the graph $\widetilde{G}_{q,\ell-1}$. Suppose that $\mathcal{M}$ is the incidence matrix of $\widetilde{G}_{q,\ell-1}$, where the i-th column of $\mathcal{M}$ corresponds to the edge e_i. Then any vector $\chi \in \mathbb{R}^{q^\ell - q}$ in the null space of $\mathcal{M}$ can be interpreted as a function in $\mathbb{R}^{\widetilde{\Sigma^\ell}}$, where the entry χ_i represents the value assigned to the string corresponding to edge e_i. It is clear that a vector χ lies in the null space of the incidence matrix $\mathcal{M}$ if and only if the graph $\widetilde{G}_{q,\ell-1}$, weighted by assigning each edge e_i the value χ_i, is balanced. Thus, a permutation $\pi \in S_{\widetilde{\Sigma^\ell}}$ belongs to $\tilde{\Phi}_{q,\ell}$ if and only if there exists a vector $\chi \in \mathrm{null}(\mathcal{M}) \cap (\mathbb{N} \cup \{0\})^{\widetilde{\Sigma^\ell}}$ such that $\chi \models \pi$.

Theorem 4.6. *Let Σ be an alphabet with size $q \geq 3$. The minimum distance between permutations in $\tilde{\Phi}_{q,2}$ is 1 when using the Kendall τ-distance, infinity distance, or Ulam distance, and is equal to 2 when using the Hamming distance.*

Proof. Let $s_1, \dots, s_q$ be the elements of the alphabet Σ, ordered lexicographically. Note that The graph $\widetilde{G}_{q,1}$ has the vertex set $V(\widetilde{G}_{q,1}) = \{s_1, \dots, s_q\}$ and for each pair of distinct vertices s_i and s_j in $V(\widetilde{G}_{q,1})$, $\{s_i \to s_j, s_j \to s_i\} \subseteq E(\widetilde{G}_{q,1})$. This property will be used throughout the proof.
We consider the proof in two cases:

Case 1: $q = 3$. Define two vectors $x := (20, 4, 14, 12, 10, 6)$ and $y := (30, 4, 28, 12, 6, 10)$ in $\mathbb{N}^{\widetilde{\Sigma^2}}$, where the coordinates correspond to the strings in $\widetilde{\Sigma}^2$ ordered lexicographically. Define

$$\sigma_1 := [s_3s_2, s_1s_2, s_3s_1, s_2s_3, s_2s_1, s_1s_3], \quad \sigma_2 := [s_3s_2, s_1s_2, s_3s_1, s_2s_3, s_1s_3, s_2s_1].$$

Then $x \models \sigma_1$ and $y \models \sigma_2$. Also the weighted de Bruijn graph $\widetilde{G}_{q,1}$, with weights given by $wt(e) = x(e)$ and with weights given by $wt(e) = y(e)$ for all $e \in \widetilde{\Sigma^2}$, is balanced. Hence, it follows from Remark 3.2 that $\sigma_1, \sigma_2 \in \tilde{\Phi}_{q,2}$. Furthermore, the distances between them are: $d_H(\sigma_1, \sigma_2) = 2$, $d_K(\sigma_1, \sigma_2) = 1$, $d_\infty(\sigma_1, \sigma_2) = 1$ and $d_U(\sigma_1, \sigma_2) = 1$.

Case 2: $q \geq 4$. We present the proof of this case in four steps.
Step 1: In this step, using Proposition 4.4, we determine a basis for the null space of the incidence matrix of $\widetilde{G}_{q,1}$. Consider the spanning tree T in $\widetilde{G}_{q,1}$ with vertex set $V(T) = V(\widetilde{G}_{q,1})$ and $E(T) = \{s_q \to s_{q-1}, s_{q-1} \to s_{q-2}, \cdots, s_2 \to s_1\}$ (See Fig. 4, which shows the spanning tree T highlighted in blue in $\widetilde{G}_{4,1}$).

Let $\mathcal{P} := \{(i, j) \in [q] \times [q] \,|\, s_i \to s_j \in E(\widetilde{G}_{q,1}) \setminus E(T)\}$. For each $(i, j) \in \mathcal{P}$, let $\mathcal{C}_{i,j}$ denote the unique undirected cycle obtained by adding $s_i \to s_j$ to T; that is, the cycle consisting of the edge $s_i \to s_j$ together with the unique path between s_i and s_j in T. We fix the cyclic orientation of $\mathcal{C}_{i,j}$ as follows: we traverse from s_i to s_j along the directed edge $s_i \to s_j$, and then follow the (undirected) path in T from s_j back to s_i. For each edge $e \in E(\widetilde{G}_{q,1})$, let $w_e \in \Sigma^2$ denote the string associated with e. According to the chosen traversal direction, we have: if $i < j$, then $\eta_{\mathcal{C}_{i,j}}(w_e) = 1$ for all edge $e \in E(\widetilde{G}_{q,1})$ that appears in $\mathcal{C}_{i,j}$, and $\eta_{\mathcal{C}_{i,j}}(w_e) = 0$ otherwise; if $i > j$, then $\eta_{\mathcal{C}_{i,j}}(s_is_j) = 1$, for each edge e in T lying on the undirected path from s_j to s_i, $\eta_{\mathcal{C}_{i,j}}(w_e) = -1$; and for all other edges, $\eta_{\mathcal{C}_{i,j}}(w_e) = 0$.

By Proposition 4.4, the set $\left\{\eta_{\mathcal{C}_{i,j}} \,\middle|\, (i, j) \in \mathcal{P}\right\}$ forms a basis for the null$(\mathcal{M})$, where $\mathcal{M}$ denotes the incidence matrix of $\widetilde{G}_{q,1}$.

Step 2: We now describe the general form of a vector in the null space as a linear combination of the basis vectors obtained in Step 1.

Let $\{m_{i,j} \,|\, (i, j) \in \mathcal{P}\}$ be arbitrary real coefficients, and define

$$x := \sum_{(i,j)\in P} m_{i,j}\, \eta_{\mathcal{C}_{i,j}} \in \mathrm{null}(\mathcal{M}).$$

For a directed edge $s_t \to s_{t'}$ in $E(\widetilde{G}_{q,1})$, there are two cases to consider: either $s_t \to s_{t'}$ belongs to the spanning tree T, or it does not. By construction of T, we have $s_t \to s_{t'} \in E(T)$ if and only if $t - t' = 1$.

Define $\mathcal{D} := \{(i, j) \in \mathcal{P} \,|\, i < j,\ j - i > 1\}$ and for each $1 \leq t \leq q$, let $\mathcal{D}_t := \{(l, r) \in \mathcal{D} \,|\, l < t \leq r\}$ which has size 0 when $t = 1$, and $(q - i + 1)(i - 2) + (q - i)$ when $2 \leq t \leq q$. Note that if $s_t \to s_{t'} \notin E(T)$, then the cycle $\mathcal{C}_{t,t'}$ is the only cycle in $\{\mathcal{C}_{i,j} \,|\, (i, j) \in \mathcal{P}\}$ that contains this edge. Conversely, if $s_t \to s_{t'} \in E(T)$, then the edge belongs to $\mathcal{C}_{t',t}$ as well as to both cycles $\mathcal{C}_{i,j}$ and $\mathcal{C}_{j,i}$ for every $(i, j) \in \mathcal{D}_t$. Hence, using Step 1, we obtain

$$x(s_ts_{t'}) = \begin{cases} m_{t',t} + \sum_{(l,r)\in\mathcal{D}_t}(m_{l,r} - m_{r,l}), & \text{if } t - t' = 1, \\ m_{t,t'} & \text{if } t - t' \neq 1, \end{cases} \tag{4.1}$$

Step 3. In this step, we construct two vectors, $\hat{x}$ and $\bar{x}$, in $\mathrm{null}(\mathcal{M}) \cap \mathbb{N}^{\widetilde{\Sigma^2}}$, each having pairwise distinct entries.

Consider the vector $\hat{x} := \sum_{(i,j)\in\mathcal{P}} \hat{m}_{i,j}\, \eta_{\mathcal{C}_{i,j}}$, where the coefficients are defined as

$$\hat{m}_{1,2} = 1, \quad \hat{m}_{1,3} = 2, \quad \hat{m}_{3,1} = 3, \quad \hat{m}_{2,3} = 4, \quad \hat{m}_{3,4} = 5, \ \ldots, \ \hat{m}_{q-1,q} = q + 1.$$

Moreover, for each $(i, j) \in \mathcal{D} \setminus \{(1, 3)\}$ we set $\hat{m}_{i,j} = \gamma_{i,j}(q + 2)$ and $\hat{m}_{j,i} = (\gamma_{i,j} - 1)(q + 2)$, where $\gamma_{i,j}$ is an odd integer greater than 1, distinct across the pairs in $\mathcal{D} \setminus \{(1, 3)\}$, and different from $q - 3$ and $q - 2$. In other words, we need to choose $|\mathcal{D} \setminus \{(1, 3)\}| = \frac{q^2-3q}{2}$ pairwise distinct odd integers satisfying the above conditions. Note that $\mathcal{P} = \mathcal{D} \cup \{(i, j) \,|\, (j, i) \in \mathcal{D}\} \cup \{(i, i + 1) \,|\, 1 \leq i \leq q - 1\}$, so every coefficient $\hat{m}_{i,j}$ with $(i, j) \in \mathcal{P}$ has been fully specified.

From Step 2, if $j - i \neq 1$ then $\hat{x}(s_is_j) = \hat{m}_{i,j}$. Furthermore, since $\{\mathcal{D}_i \,|\, 2 \leq i \leq q,\ (1, 3) \in \mathcal{D}_i\} = \{\mathcal{D}_2, \mathcal{D}_3\}$ and using Equation (4.1), for the tree edges $s_i \to s_{i-1}$, $2 \leq i \leq q$, we obtain:

- If $i = 2$:

$$\begin{aligned}\hat{x}(s_2s_1) &= \hat{m}_{1,2} + \hat{m}_{1,3} - \hat{m}_{3,1} + \sum_{(l,r)\in\mathcal{D}_2\setminus\{(1,3)\}} (\hat{m}_{l,r} - \hat{m}_{r,l}) \\ &= (q + 2) \cdot |\mathcal{D}_2 \setminus \{(1, 3)\}| = (q - 3)(q + 2).\end{aligned}$$

- If $i = 3$:

$$\begin{aligned}\hat{x}(s_3s_2) &= \hat{m}_{2,3} + \hat{m}_{1,3} - \hat{m}_{3,1} + \sum_{(l,r)\in\mathcal{D}_3\setminus\{(1,3)\}} (\hat{m}_{l,r} - \hat{m}_{r,l}) \\ &= 3 + (q + 2) \cdot |\mathcal{D}_3 \setminus \{(1, 3)\}| = 3 + (2q - 6)(q + 2).\end{aligned}$$

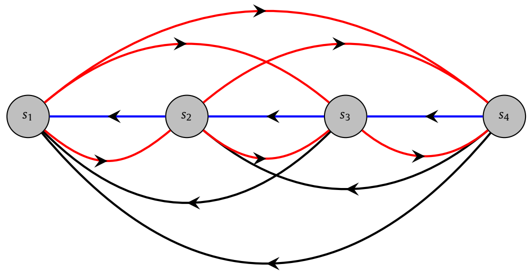

Fig. 4. The graph corresponding to Example 4.7. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

- If $4 \le i \le q$:

$$\begin{aligned}\hat{x}(s_i s_{i-1}) &= \hat{m}_{i-1,i} + \sum_{(l,r)\in\mathcal{D}_i} (\hat{m}_{l,r} - \hat{m}_{r,l}) \\ &= \hat{m}_{i-1,i} + \big((q-i+1)(i-2) + (q-i)\big)(q+2).\end{aligned}$$

Observe that, by the definition of $\hat{x}$, for every non-tree edge the value $\hat{x}(e)$ is either less than $q+2$ or a multiple of $q+2$, but never equal to $(q+3)(q+2)$. Moreover, for each $4 \le i \le q$ we have $5 \le \hat{m}_{i-1,i} \le q+1$ and $(q-i+1)(i-2)+(q-i) \ne 0$. Since $q \ge 4$, it follows that the values $\hat{x}(w_e)$ on all edges $e \in E(\widetilde{G}_{q,1})$ are pairwise distinct.

Now define the vector $\bar{x} := \sum_{(i,j)\in\mathcal{P}} \bar{m}_{i,j}\, \eta_{\mathcal{C}_{i,j}} \in \mathbb{N}^{\widetilde{\Sigma^2}}$, where $\bar{m}_{1,2} = 2$, $\bar{m}_{1,3} = 1$, and $\bar{m}_{i,j} = \hat{m}_{i,j}$ for all $(i,j) \in \mathcal{P} \setminus \{(1,2),(1,3)\}$. With this definition, in addition to the strings $s_1 s_2$ and $s_1 s_3$, on which the $\bar{x}$ values are clearly different from the corresponding $\hat{x}$ values, the only strings on which $\bar{x}$ may possibly differ from $\hat{x}$ are $s_2 s_1$ and $s_3 s_2$.

For these edges we have:

$$\begin{aligned}\bar{x}(s_2 s_1) &= \bar{m}_{1,2} + \bar{m}_{1,3} - \bar{m}_{3,1} + \sum_{(i,j)\in\mathcal{D}_2\setminus\{(1,3)\}} (\bar{m}_{i,j} - \bar{m}_{j,i}) \\ &= (q-3)(q+2) = \hat{x}(s_2 s_1),\end{aligned}$$

and

$$\begin{aligned}\bar{x}(s_3 s_2) &= \bar{m}_{2,3} + \bar{m}_{1,3} - \bar{m}_{3,1} + \sum_{(i,j)\in\mathcal{D}_3\setminus\{(1,3)\}} (\bar{m}_{i,j} - \bar{m}_{j,i}) \\ &= 2 + (2q-6)(q+2) = \hat{x}(s_3 s_2) - 1.\end{aligned}$$

Since $\hat{x}(w_e) \ne 2 + (2q-6)(q+2)$ for all edges $e \in E(\widetilde{G}_{q,1})$, the vector $\bar{x}$ also has pairwise distinct entries in $\mathrm{null}(\mathcal{M}) \cap \mathbb{N}^{\widetilde{\Sigma^2}}$.

Step 4: In view of Remark 4.5 and Step 3, if $\hat{\sigma}$ and $\bar{\sigma}$ are the permutations induced by the rankings of the vectors $\hat{x}$ and $\bar{x}$, respectively, then $\hat{\sigma}$ and $\bar{\sigma}$ belong to $\widetilde{\Phi}_{q,2}$. By Step 3, for every edge $e \in E(\widetilde{G}_{q,1}) \setminus \{s_3 \to s_2\}$ we have either $\hat{x}(w_e) < q+2$ or $\hat{x}(w_e) = r + k(q+2)$ for some $r \in \{0,5,\dots,q+1\}$ and $k \ge 1$. Since $\hat{x}(s_3 s_2) = 3 + (2q-6)(q+2)$ and $\bar{x}(s_3 s_2) = \hat{x}(s_3 s_2) - 1$, whenever $\hat{x}(w_{e_1}) < \hat{x}(s_3 s_2) < \hat{x}(w_{e_2})$ for some $e_1, e_2 \in E(\widetilde{G}_{q,1}) \setminus \{s_3 \to s_2\}$, we also have $\bar{x}(w_{e_1}) < \bar{x}(s_3 s_2) < \bar{x}(w_{e_2})$. Hence $\hat{\sigma} = (s_1 s_2, s_1 s_3)\,\bar{\sigma}$, and consequently $d_H(\hat{\sigma},\bar{\sigma}) = 2$, $d_K(\hat{\sigma},\bar{\sigma}) = 1$, $d_\infty(\hat{\sigma},\bar{\sigma}) = 1$, and $d_U(\hat{\sigma},\bar{\sigma}) = 1$. This completes the proof. $\square$

The following example presents the permutations obtained from the construction used in the proof of Theorem 4.6 that achieve the minimum distance when $q = 4$.

Example 4.7. Let $\Sigma = \{s_1, s_2, s_3, s_4\}$. In Fig. 4, the spanning tree T of the graph $\widetilde{G_{4,1}}$ is highlighted in blue. Consistent with the convention used in the proof of Theorem 4.6, we define $\mathcal{P} = \{(i,j) \in [4]\times[4] \mid s_i \to s_j \in E(\widetilde{G_{4,1}}) \setminus E(T)\}$. For each edge $s_i \to s_j \in E(\widetilde{G_{4,1}}) \setminus E(T)$, there is a unique fundamental cycle $\mathcal{C}_{i,j}$. In the table below, the column associated with $(i,j) \in \mathcal{P}$ lists the corresponding vector $\eta_{\mathcal{C}_{i,j}}$. For each edge $e \in E(\widetilde{G_{4,1}})$, the entry in that row equals $\eta_{\mathcal{C}_{i,j}}(w_e)$. In Fig. 4, edges $s_i \to s_j$ for which the vector $\eta_{\mathcal{C}_{i,j}}$ takes only values in $\{0,1\}$ are marked in red, while those whose vectors contain a -1 entry are marked in black. These vectors form a basis for the null space of the incidence matrix of $\widetilde{G_{4,1}}$. Now set

$$\hat{m}_{1,2} = 1, \qquad \hat{m}_{1,3} = 2, \qquad \hat{m}_{3,1} = 3, \qquad \hat{m}_{2,3} = 4, \qquad \hat{m}_{3,4} = 5,$$

and choose $\gamma_{1,4}=5$ and $\gamma_{2,4}=7$, which, according to the proof of Theorem 4.6, must be distinct odd positive integers different from 1. Since $q+2=6$, this gives

$$\hat{m}_{1,4}=5\cdot 6=30, \qquad \hat{m}_{4,1}=4\cdot 6=24, \qquad \hat{m}_{2,4}=7\cdot 6=42, \qquad \hat{m}_{4,2}=6\cdot 6=36.$$

Hence, $\hat{X}=\sum_{(i,j)\in\mathcal{P}}\hat{m}_{i,j}\,\eta_{C_{i,j}}=(1,\,2,\,30,\,6,\,4,\,42,\,3,\,15,\,5,\,24,\,36,\,17)$. Next, define $\bar{m}_{1,2}=2$, $\bar{m}_{1,3}=1$ and $\bar{m}_{i,j}=\hat{m}_{i,j}$ for all $(i,j)\in\mathcal{P}\setminus\{(1,2),(1,3)\}$. Then $\bar{X}=\sum_{(i,j)\in\mathcal{P}}\bar{m}_{i,j}\,\eta_{C_{i,j}}=(2,\,1,\,30,\,6,\,4,\,42,\,3,\,14,\,5,\,24,\,36,\,17)$.

According to the proof of Theorem 4.6, the permutations in $S_{\widetilde{\Sigma^2}}$ induced by the rankings of $\hat{X}$ and $\bar{X}$ are feasible, and they are given by

$$\hat{\sigma}=[s_1s_2,\,s_1s_3,\,s_4s_1,\,s_2s_4,\,s_2s_1,\,s_4s_3,\,s_1s_4,\,s_3s_1,\,s_2s_3,\,s_3s_4,\,s_4s_2,\,s_3s_2],$$

and

$$\bar{\sigma}=[s_1s_3,\,s_1s_2,\,s_4s_1,\,s_2s_4,\,s_2s_1,\,s_4s_3,\,s_1s_4,\,s_3s_1,\,s_2s_3,\,s_3s_4,\,s_4s_2,\,s_3s_2].$$

Also, we have $d_H(\hat{\sigma},\bar{\sigma})=2$, $d_K(\hat{\sigma},\bar{\sigma})=1$, $d_\infty(\hat{\sigma},\bar{\sigma})=1$, and $d_U(\hat{\sigma},\bar{\sigma})=1$.

	$\eta_{C_{1,2}}$	$\eta_{C_{1,3}}$	$\eta_{C_{1,4}}$	$\eta_{C_{2,3}}$	$\eta_{C_{2,4}}$	$\eta_{C_{3,1}}$	$\eta_{C_{3,4}}$	$\eta_{C_{4,1}}$	$\eta_{C_{4,2}}$
$s_1\to s_2$	1	0	0	0	0	0	0	0	0
$s_1\to s_3$	0	1	0	0	0	0	0	0	0
$s_1\to s_4$	0	0	1	0	0	0	0	0	0
$s_2\to s_1$	1	1	1	0	0	-1	0	-1	0
$s_2\to s_3$	0	0	0	1	0	0	0	0	0
$s_2\to s_4$	0	0	0	0	1	0	0	0	0
$s_3\to s_1$	0	0	0	0	0	1	0	0	0
$s_3\to s_2$	0	1	1	1	1	-1	0	-1	-1
$s_3\to s_4$	0	0	0	0	0	0	1	0	0
$s_4\to s_1$	0	0	0	0	0	0	0	1	0
$s_4\to s_2$	0	0	0	0	0	0	0	0	1
$s_4\to s_3$	0	0	1	0	1	0	1	-1	-1

Lemma 4.8. *Let Σ be an alphabet with size $q\geq 3$ and let $\ell\geq 2$. If $\delta\in\Phi_{q,\ell}$ (respectively, $\delta\in\tilde{\Phi}_{q,\ell}$), then $\delta^r\in\Phi_{q,\ell}$ (respectively, $\delta^r\in\tilde{\Phi}_{q,\ell}$).*

Proof. Let $a_1,a_2,\ldots,a_q$ and $s_1,s_2,\ldots,s_{q^\ell}$ (respectively, $s_1,s_2,\ldots,s_{q^\ell-q}$) be the elements of Σ and Σ^ℓ (respectively, $\widetilde{\Sigma^\ell}$) arranged in lexicographic order. Define the permutation $\rho\in S_\Sigma$ by $\rho(a_i)=a_{q-i+1}$ for all $1\leq i\leq q$. Then, by the definition of $\bar{\pi}^\ell_\rho$ and π^ℓ_ρ, it follows that $\bar{\pi}^\ell_\rho(s_i)=s_{q^\ell-i+1}$ (respectively, $\pi^\ell_\rho(s_i)=s_{q^\ell-q-i+1}$) for all relevant i. Hence, by the proof of Theorem 3.6, we have $\delta^r=\bar{\pi}^\ell_\rho\delta\in\Phi_{q,\ell}$ (respectively, $\delta^r=\pi^\ell_\rho\delta\in\tilde{\Phi}_{q,\ell}$), completing the proof. $\square$

Theorem 4.9. *For a natural number $\ell\geq 2$ and an alphabet Σ of size $q\geq 3$, the maximum distance between permutations in $\Phi_{q,\ell}$ is $\binom{q^\ell}{2}$ under the Kendall τ-metric, $q^\ell-1$ under the infinity and Ulam metrics, and q^ℓ under the Hamming metric.*

Proof. It follows from Lemma 4.8 that if $\delta\in\Phi_{q,\ell}$, then $\delta^r\in\Phi_{q,\ell}$. Since $d_K(\delta,\delta^r)=\binom{q^\ell}{2}$, the maximum distance between feasible permutations in $\Phi_{q,\ell}$ equals $\binom{q^\ell}{2}$ under the Kendall τ-metric and equals to $q^\ell-1$ under Ulam metric. Let $s_1,s_2,\ldots,s_q$ be all lexicographically ordered elements of Σ. Consider the permutation $\rho\in S_\Sigma$ defined by $\rho(s_i)=s_{i+1}$ for all $1\leq i\leq q-1$ and $\rho(s_q)=s_1$. Then it is clear that $\bar{\pi}^\ell_\rho(s)\neq s$, for all $s\in\Sigma^\ell$. Therefore, $d_H(\bar{\pi}_\rho\delta,\delta)=q^\ell$, for all $\delta\in\Phi_{q,\ell}$. So, it follows from the proof of Theorem 3.6 that the maximum distance between feasible permutations in $\Phi_{q,\ell}$ equals q^ℓ under the Hamming metric.

Suppose that π is a feasible permutation in $\Phi_{q,\ell}$, and let χ be a vector in $(\mathbb{N}\cup\{0\})^{\Sigma^\ell}$ such that $\chi\models\pi$. Without loss of generality, we may assume that the minimum entries of χ are greater than zero. Assume that $u\in\Sigma^\ell\setminus\widetilde{\Sigma^\ell}$. As mentioned earlier, changing the value of $\chi(u)$ to any natural number that is distinct from all other entries of χ still results in a feasible vector. We define a new vector $\chi'\in(\mathbb{N}\cup\{0\})^{\Sigma^\ell}$ as follows. If $\chi(u)=\min\{\chi(v)\mid v\in\Sigma^\ell\}$, then we set $\chi'=\chi$. Otherwise, we let $\chi'(v)=\chi(v)$ for all $v\in\Sigma^\ell\setminus\{u\}$, and define $\chi'(u)=\min\{\chi(v)\mid v\in\Sigma^\ell\}-1$. Similarly, we define a vector $\tilde{\chi}\in(\mathbb{N}\cup\{0\})^{\Sigma^\ell}$ as follows. If $\chi(u)=\max\{\chi(v)\mid v\in\Sigma^\ell\}$, then we set $\tilde{\chi}=\chi$. Otherwise, we let $\tilde{\chi}(v)=\chi(v)$ for all $v\in\Sigma^\ell\setminus\{u\}$, and define $\tilde{\chi}(u)=\max\{\chi(v)\mid v\in\Sigma^\ell\}+1$. Let π' and $\tilde{\pi}$ denote the permutations induced by the ordering of the components of χ' and $\tilde{\chi}$ in Σ^ℓ, respectively. Then it is clear that $d_\infty(\pi',\tilde{\pi})=\left|\mathrm{pos}(\pi'(u))-\mathrm{pos}(\tilde{\pi}(u))\right|=|1-q^\ell|=q^\ell-1$. This completes the proof. $\square$

Using an argument similar to the proof of Theorem 4.9, the following theorem can also be proved.

Theorem 4.10. *For a natural number $\ell \geq 2$ and an alphabet Σ of size $q \geq 3$, the maximum distance between permutations in $\tilde{\Phi}_{q,\ell}$ is $\binom{q^\ell - q}{2}$ under the Kendall τ-metric, $q^\ell - q - 1$ under the infinity and Ulam metrics, and $q^\ell - q$ under the Hamming metric.*

We now introduce the notation for the largest size of permutation codes under the Hamming and Kendall τ-metrics. Let $M(n,d)$ denote the maximum size of a permutation code in S_n with minimum Hamming distance at least d, and let $P(n,d)$ denote the maximum size of a permutation code in S_n with minimum Kendall τ-distance at least d. Here $1 \leq d \leq n$ in the Hamming metric, whereas $1 \leq d \leq \binom{n}{2}$ in the Kendall τ-metric.

A basic problem is to determine these quantities. It is known that $M(n,1) = M(n,2) = n!$, $M(n,3) = \frac{n!}{2}$, $M(n,n) = n$ (see [12, Lemma 1.1]), and $P(n,1) = n!$ and $P(n,2) = \frac{n!}{2}$ and if $\frac{2}{3}\binom{n}{2} < d \leq \binom{n}{2}$, then $P(n,d) = 2$ (see [6, Theorem 10]).

Restricting attention to feasible permutations, we denote by $M_{\Phi_{q,\ell}}(d)$ and $P_{\Phi_{q,\ell}}(d)$ the maximum sizes of permutation codes contained in $\Phi_{q,\ell}$ with minimum Hamming distance at least d and minimum Kendall τ-distance at least d, respectively. Since any two distinct feasible permutations differ in at least two positions, it follows that $P_{\Phi_{q,\ell}}(1) = M_{\Phi_{q,\ell}}(1) = M_{\Phi_{q,\ell}}(2) = \mathcal{F}_{q,\ell}$. The following theorem establishes several results on these quantities.

Theorem 4.11. *Let Σ be an alphabet with size $q \geq 3$ and $\ell \in \mathbb{N}$. Then $M_{\Phi_{q,\ell}}(3) = \frac{\mathcal{F}_{q,\ell}}{2}$, $P_{\Phi_{q,\ell}}(2) \geq \frac{\mathcal{F}_{q,\ell}}{2}$ and $P_{\Phi_{q,\ell}}(2) = 2$, for all $\frac{2}{3}\binom{q^\ell}{2} < d \leq \binom{q^\ell}{2}$.*

Proof. Let π be an feasible permutation in S_{Σ^ℓ}. Then there exists a non-negative vector $\chi \in (\mathbb{N} \cup \{0\})^{q^\ell}$ such that $\chi \models \pi$ and the weighted de Bruijn graph, $G_{q,\ell-1}$, with weights $wt(e) = \chi(e)$ for all $e \in \Sigma^\ell$, is balanced. Let $u, v \in \Sigma$, $\lambda = \underbrace{uu \cdots u}_{\ell}$, $\lambda_0 = \underbrace{vv \cdots v}_{\ell}$, $\bar{\lambda} = \underbrace{uu \cdots u}_{\ell-1}$ and $\bar{\lambda_0} = \underbrace{vv \cdots v}_{\ell-1}$. Consider the vector $\bar{\chi} \in (\mathbb{N} \cup \{0\})^{q^\ell}$ defined by $\bar{\chi}(\lambda) = \chi(\lambda_0)$, $\bar{\chi}(\lambda_0) = \chi(\lambda)$ and $\bar{\chi}(t) = \chi(t)$, for all $t \in \Sigma^\ell \setminus \{\lambda, \lambda_0\}$. Since $\chi(\lambda)$ and $\chi(\lambda_0)$ are the weight of the loops $(\bar{\lambda}, \bar{\lambda})$ and $(\bar{\lambda_0}, \bar{\lambda_0})$ in $G_{q,\ell-1}$, respectively, the weighted de Bruijn graph, $G_{q,\ell-1}$, with weights $wt(e) = \bar{\chi}(e)$ for all $e \in \Sigma^\ell$, is balanced. Therefore, since $\bar{\chi} \models (\lambda, \lambda_0)\pi$, where ρ is the transposition (λ, λ_0) in S_{Σ^ℓ}, $\rho\pi$ is a feasible permutation in S_{Σ^ℓ} that has the opposite parity to π. Define $\mathcal{A}_{q,\ell} := \Phi_{q,\ell} \cap \mathrm{Alt}(\Sigma^\ell)$ and $\mathcal{B}_{q,\ell} := \Phi_{q,\ell} \backslash \mathcal{A}_{q,\ell}$, where $\mathrm{Alt}(\Sigma^\ell)$ denotes the alternating group on Σ^ℓ, i.e., the set of all even permutations in S_{Σ^ℓ}. Now, consider the map $\varphi : \mathcal{A}_{q,\ell} \longrightarrow \mathcal{B}_{q,\ell}$ defined by $\varphi(\pi) = \rho\pi$ for all $\pi \in \mathcal{A}_{q,\ell}$. It is easy to see that φ is bijection and therefore $|\mathcal{A}_{q,\ell}| = \frac{\mathcal{F}_{q,\ell}}{2}$.

It is known that $\mathrm{Alt}(\Sigma^\ell)$ is a permutation code with largest size and minimum distance 2 and 3 under the Kendall τ-distance and the Hamming distance, respectively. So, $M_{\Phi_{q,\ell}}(3) \geq \frac{\mathcal{F}_{q,\ell}}{2}$ and $P_{\Phi_{q,\ell}}(2) \geq \frac{\mathcal{F}_{q,\ell}}{2}$. Now, suppose for a contrary that there is a subset $\mathcal{C} \subset \Phi_{q,\ell}$ with minimum Hamming distance 3 such that $|\mathcal{C}| > \frac{\mathcal{F}_{q,\ell}}{2}$. Then there exists an element $\tau \in \mathcal{C}$ such that $\rho\tau \in \mathcal{C}$. This is while $d_H(\tau, \rho\tau) = 2$, that is a contradiction. So, $M_{\Phi_{q,\ell}}(3) = \frac{\mathcal{F}_{q,\ell}}{2}$. In view of [6, Theorem 10], $P_{\Phi_{q,\ell}}(d) \leq 2$ for all $\frac{2}{3}\binom{n}{2} < d \leq \binom{q^\ell}{2}$. Since $d_K(\sigma, \sigma^r) = \binom{q^\ell}{2}$, the result follows from Lemma 4.8. This completes the proof. $\square$

5. Conclusion

In this work, we examined the structure and distance properties of feasible permutations arising from ℓ-gram profile vectors in DNA storage systems. Building on the algebraic reductions developed in Section 3, we demonstrated that feasible permutations necessarily form a collection of right cosets of a subgroup of size $2q!$, leading to substantial reductions in the search space. We then analyzed feasible permutations under several permutation metrics and provided a complete answer to Problem 1.1 by proving that their minimum distance is inherently the smallest possible. Furthermore, we established additional results on the maximum attainable distance between feasible permutations. These findings contribute to a deeper understanding of rank-modulation-based coding for DNA storage and offer structural tools that may facilitate future efforts to compute or bound $\mathcal{F}_{q,\ell}$ for larger parameter values.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Computational codes

In this appendix, we provide the GAP and Python codes used to compute feasible permutations for the cases $q = 4$ and $q = 3$.

—

A.1. Case $q = 4$

A.1.1. GAP code

```
K:=[ (), (2,3)(5,6)(7,10)(8,11)(9,12), (1,2)(4,7)(5,8)(6,9)(11,12),
(1,2,3)(4,7,10)(5,9,11)(6,8,12),
(1,3,2)(4,10,7)(5,11,9)(6,12,8), (1,3)(4,10)(5,12)(6,11)(8,9),
(1,4)(2,5)(3,6)(7,8)(10,11),
(1,4)(2,6)(3,5)(7,11)(8,10)(9,12), (1,4)(2,7)(3,10)(5,8)(6,11)(9,12),
(1,5,7)(2,4,8)(3,6,9)(10,11,12),
(1,5,9,10)(2,6,7,11)(3,4,8,12), (1,6,12,7)(2,4,11,9)(3,5,10,8),
(1,6,10)(2,5,12)(3,4,11)(7,8,9),
(1,7,5)(2,8,4)(3,9,6)(10,12,11), (1,7,12,6)(2,9,11,4)(3,8,10,5),
(1,8)(2,7)(3,9)(4,5)(10,12),
(1,8,11)(2,9,10)(3,7,12)(4,5,6), (1,9)(2,7)(3,8)(4,12)(5,10)(6,11),
(1,9,4,12)(2,8,6,10)(3,7,5,11),
(1,10,9,5)(2,11,7,6)(3,12,8,4), (1,10,6)(2,12,5)(3,11,4)(7,9,8),
(1,11,8)(2,10,9)(3,12,7)(4,6,5),
(1,11)(2,12)(3,10)(4,6)(7,9), (1,12,4,9)(2,10,6,8)(3,11,5,7),
(1,12)(2,11)(3,10)(4,9)(5,8)(6,7) ];;
H:=Group(K);;
G:=SymmetricGroup(12);;
O:=RightTransversal(G,H);;
M := List(O,i->ListPerm(i,12));;
PrintTo("M.txt", M);
```

A.1.2. Python code

```
import ast
import pulp
import time

def check_lp(pi):
    n = 12

    prob = pulp.LpProblem("PermutationLP", pulp.LpMinimize)

    x = [pulp.LpVariable(f"x{i+1}", lowBound=1) for i in range(n)]

    prob += pulp.lpSum(x)

    prob += -x[0] - x[1] - x[2] + x[3] + x[6] + x[9]   == 0
    prob +=  x[0] - x[3] - x[4] - x[5] + x[7] + x[10]  == 0
    prob +=  x[1] + x[4] - x[6] - x[7] - x[8] + x[11]  == 0
    prob +=  x[2] + x[5] + x[8] - x[9] - x[10] - x[11] == 0

    sorted_indices = sorted(range(n), key=lambda i: pi[i])

    for k in range(n - 1):
        i = sorted_indices[k]
        j = sorted_indices[k + 1]
        prob += x[j] - x[i] >= 1

    prob.solve(pulp.PULP_CBC_CMD(msg=False))

    return pulp.LpStatus[prob.status] == "Optimal"


with open("M.txt", "r") as f:
    M = ast.literal_eval(f.read())

print("Total permutations in M:", len(M))
```

```
start_time = time.time()

Feasible_M = []

for pi in M:
    if check_lp(pi):
        Feasible_M.append(pi)

end_time = time.time()
total_time = end_time - start_time

print("Number of feasible permutations:", len(Feasible_M))
print(f"Total runtime: {total_time:.2f} seconds")
```

A.1.3. Output

```
Total permutations in M: 9979200
Number of feasible permutations: 727481
Total runtime: 417007.26 seconds
```

A.2. Case $q = 3$

A.2.1. GAP code

```
K := [(),(1,3)(2,4)(5,6),(1,6)(2,5)(3,4),(1,2)(3,5)(4,6),(1,4,5)(2,3,6),
(1,5,4)(2,6,3),(1,3)(2,5)(4,6)];;
H := Group(K);;
G := SymmetricGroup(6);;
O := RightTransversal(G, H);;
M6 := List(O, i -> ListPerm(i, 6));;
PrintTo("M6.txt", M6);
```

A.2.2. Python code

```
import ast
import pulp
import time

def check_lp(pi):
    n = 6
    prob = pulp.LpProblem("PermutationLP", pulp.LpMinimize)
    x = [pulp.LpVariable(f"x{i+1}", lowBound=1) for i in range(n)]
    prob += pulp.lpSum(x)

    prob += -x[0] - x[1] + x[2] + x[4] == 0
    prob +=  x[0] - x[2] - x[3] + x[5] == 0
    prob +=  x[1] + x[3] - x[4] - x[5] == 0

    s = sorted(range(n), key=lambda i: pi[i])
    for k in range(n-1):
        prob += x[s[k+1]] - x[s[k]] >= 1

    prob.solve(pulp.PULP_CBC_CMD(msg=False))
    return pulp.LpStatus[prob.status] == "Optimal"


with open("M6.txt") as f:
    M6 = ast.literal_eval(f.read())

start = time.time()
F6 = [pi for pi in M6 if check_lp(pi)]
end = time.time()

print("Total permutations in M6:", len(M6))
```

```
print("Number of feasible permutations:", len(F6))
print(f"Total runtime: {end-start:.2f} seconds")
```

A.2.3. Output

```
Total permutations in M6: 60
Feasible permutations:
[1, 4, 2, 5, 3, 6]
[1, 4, 3, 5, 2, 6]
[1, 5, 2, 4, 3, 6]
[1, 5, 3, 4, 2, 6]
[1, 5, 4, 3, 2, 6]
Number of feasible permutations: 5
Total runtime: 1.34 seconds
```

Data availability

No data was used for the research described in the article.

References

[1] A. Abdollahi, J. Bagherian, F. Jafari, M. Khatami, F. Parvaresh, R. Sobhani, New upper bounds on the size of permutation codes under Kendall τ-metric, Cryptogr. Commun. 15 (2023) 891–903.
[2] A. Barg, A. Mazumdar, Codes in permutations and error correction for rank modulation, IEEE Trans. Inf. Theory 56 (7) (2010) 3158–3165.
[3] T. Berger, F. Jelinek, J.K. Wolf, Permutation codes for sources, IEEE Trans. Inf. Theory 18 (1) (1972) 160–169.
[4] J. Bornholt, R. Lopez, D.M. Carmean, L. Ceze, G. Seelig, K. Strauss, A DNA-based archival storage system, ACM SIGOPS Operat. Syst. Rev. 50 (2) (2016) 637–649.
[5] N. Beeri, M. Schwartz, Improved rank-modulation codes for dna storage with shotgun sequencing, IEEE Trans. Inf. Theory 68 (6) (2022) 3719–3730.
[6] S. Buzaglo, T. Etzion, Bounds on the size of permutation codes with the Kendall τ-metric, IEEE Trans. Inf. Theory 61 (6) (2015) 3241–3250.
[7] H.D. Chadwick, L. Kurz, Rank permutation group codes based on Kendall's correlation statistic, IEEE Trans. Inf. Theory 15 (2) (1969) 306–315.
[8] G.M. Church, Y. Gao, S. Kosuri, Next-generation digital information storage in DNA, Science 337 (6102) (2012) 16–28.
[9] T.F. Coleman, A. Pothen, The null space problem I. Complexity, SIAM J. Algebr. Discrete Methods 7 (4) (1986) 527–537.
[10] D. David, R. Foote, Abstract Algebra, 3rd ed., Wiley, 2003.
[11] M. Deza, H. Huang, Metrics on permutations, a survey, J. Comb. Inf. Syst. Sci. 23 (1998) 173–185.
[12] P. Dukes, N. Sawchuck, Bounds on permutation codes of distance four, J. Algebr. Comb. 31 (2010) 143–158.
[13] F. Farnoud, V. Skachek, O. Milenkovic, Error-correction in flash memories via codes in the Ulam metric, IEEE Trans. Inf. Theory 59 (2013) 3003–3020.
[14] The GAP Group, GAP-groups, algorithms, and programming, version 4.11.1, http://www.gap-system.org, 2021.
[15] N. Goldman, et al., Towards practical, high-capacity, low-maintenance information storage in synthesized DNA, Nature 494 (2013) 77–80.
[16] A. Jiang, R. Mateescu, M. Schwartz, J. Bruck, Rank modulation for flash memories, IEEE Trans. Inf. Theory 55 (6) (2009) 2659–2673.
[17] H.M. Kiah, G.J. Puleo, O. Milenkovic, Codes for DNA sequence profiles, IEEE Trans. Inf. Theory 62 (6) (2016) 3125–3146.
[18] T. Kløve, T.T. Lin, D.C. Tsai, W.G. Tzeng, Permutation arrays under the Chebychev distance, IEEE Trans. Inf. Theory 56 (2010) 2611–2617.
[19] J.R. Miller, S. Koren, G. Sutton, Assembly algorithms for next generation sequencing data, Genomics 95 (6) (2010) 315–327.
[20] A. Motahari, G. Bresler, D. Tse, Information theory for DNA sequencing: Part I: A basic model, in: Proc. IEEE Int. Symp. Inf. Theory, Cambridge, MA, USA, 2012, pp. 2741–2745.
[21] Python Software Foundation, Python Language Reference, https://www.python.org/.
[22] N. Raviv, M. Schwartz, E. Yaakobi, Rank-modulation codes for DNA storage with shotgun sequencing, IEEE Trans. Inf. Theory 65 (1) (2019) 50–64.
[23] D. Slepian, Permutation modulation, Proc. IEEE 53 (3) (1965) 228–236.
[24] M.Z. Shieh, S.C. Tsai, Computing the ball size of frequency permutations under Chebychev distance, in: Proc. of IEEE Int. Symp. on Inform. Theory, St. Petersburg, Russia, 2011, pp. 2100–2104.
[25] S.M.H.T. Yazdi, Y. Yuan, J. Ma, H. Zhao, O. Milenkovic, A rewritable, random-access DNA-based storage system, Sci. Rep. 5 (1) (2015) 1–10.
[26] H. Vinck, J. Haering, T. Wadayama, Coded M-FSK for power line communications, in: Proc. of IEEE Int. Symp. on Inform. Theory, Sorrento, Italy, 2000, p. 137.
[27] X. Wang, Y. Zhang, Y. Yang, G. Ge, New bounds of permutation codes under Hamming metric and Kendall's τ-metric, Des. Codes Cryptogr. 85 (3) (2017) 533–545.
[28] X. Wang, Y. Wang, W. Yin, F-W. Fu, Nonexistence of perfect permutation codes under the Kendall τ-metric, Des. Codes Cryptogr. 89 (11) (2021) 2511–2531.