LP Decodable Permutation Codes based on Linearly Constrained Permutation Matrices

Tadashi WADAYAMA†, Manabu HAGIWARA††*
† Department of Computer Science, Nagoya Institute of Technology
Email: wadayama@nitech.ac.jp
†† National Institute of Advanced Industrial Science and Technology,
Research Center for Information Security,
Email: hagiwara.hagiwara@aist.go.jp
* Center for Research and Development Initiative, Chuo University

Abstract—A set of linearly constrained permutation matrices are proposed for constructing a class of permutation codes. Making use of linear constraints imposed on the permutation matrices, we can formulate a minimum Euclidean distance decoding problem for the proposed class of permutation codes as a linear programming (LP) problem. The main feature of this novel class of permutation codes, called LP decodable permutation codes, is this LP decodability. It is demonstrated that the LP decoding performance of the proposed class of permutation codes is characterized by the vertices of the code polytope of an LP decodable permutation code. In addition, based on a probabilistic method, several theoretical results for randomly constrained permutation codes are derived.

I. INTRODUCTION

The origin of permutation codes dates back to the 1960s. Slepian [11] proposed a class of simple permutation codes, which is referred to as permutation modulation, and efficient soft decoding algorithms for these codes. This research has been extended and investigated by a number of researchers. Karlof [12] and Ingemarsson [13] investigated the optimization of the initial vector of the permutation modulation. Berger et al. [14] discussed applications of permutation codes to source coding problems.

There is another thread of research on a class of permutation codes of length n whose codewords contain exactly n-distinct symbols, i.e., any codeword can be obtained by applying a permutation to an initial vector, e.g., (0, 1, . . . , n − 1).

Some fundamental properties of such permutation codes were discussed by Blake et al. [1]. Vinck [9] proposed applications of permutation codes for power-line communication, which inspired subsequent research on permutation codes. Wadayama and Vinck [10] presented a multi-level construction of permutation codes with a large minimum distance. A number of constructions for permutation codes have been developed, including the construction given in [2] [4].

Recently, rank modulation codes for flash memory proposed by Jiang et al. [6] [7] generated renewed interest in permutation codes. For example, for flash memory coding, Kleeve et al. presented a new construction for permutation codes based on the Chebyshev distance [8], which is an appropriate distance measure for flash memory coding. Barg and Mazumdar [15] also investigated fundamental bounds on permutation codes in terms of the Kendall tau distance.

II. PRELIMINARIES

A. Notation and definition

In the present paper, matrices are represented by capital letters, and vectors are assumed to be column vectors. Let \( X \) be an \( n \times n \) real matrix. The notation \( X \geq 0 \) means that every
element in $X$ is non-negative. The notation $\text{vec}(X)$ represents a vectorization of $X$ given by

$$\text{vec}(X) \triangleq (X_{1,1} \cdots X_{1,n}, X_{2,1} \cdots X_{2,n}, X_{3,1} \cdots X_{n,n})^T.$$ 

The vector $1$ is an all-one vector, the length of which is determined by the context. The norm $|| \cdot ||$ denotes the Euclidean norm given by $||x|| = \sqrt{x^T x}$. The trace function $\text{trace}(X)$ returns the sum of the diagonal elements of $X$.

The symbol $\leq$ is defined by

$$\begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} \leq \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \iff \forall i \in [1,m], a_i \leq b_i,$$

where $\leq_i$ is either $= \leq$. For simplicity, the notation $\leq = (\leq_1, \ldots, \leq_m)^T$ is used to define $\leq$ (e.g., $\leq = (\leq_1, \ldots, \leq_m)^T$).

The following definition gives a class of matrices that is crucial to the arguments presented herein.

**Definition 1 (Permutation matrix):** An $n \times n$ binary real matrix $X \triangleq (X_{i,j})_{i,j \in [1,n]} \in \{0,1\}^{n \times n}$ is called a permutation matrix if and only if

$$\forall i,j \in [1,n], \sum_{j' \in [1,n]} X_{i,j'} = 1, \quad \sum_{i' \in [1,n]} X_{i',j} = 1. \quad (1)$$

The set of $n \times n$ permutation matrices is denoted by $\Pi_n$. \hfill $\Box$

It is also known that an $n \times n$ binary matrix is a permutation matrix if and only if the Hamming weights of every column and every row are exactly $1$. The cardinality of $\Pi_n$ is $n!$.

Removing the binary constraint from the definition of the permutation matrices, we obtain the definition of doubly stochastic matrices.

**Definition 2 (Doubly stochastic matrix):** An $n \times n$ non-negative real matrix $X \triangleq (X_{i,j})_{i,j \in [1,n]}$ is called a doubly stochastic matrix if and only if (1) holds. \hfill $\Box$

The following theorem for a doubly stochastic matrix indicates that the set of doubly stochastic matrices is a convex polytope.\hfill $\Box$

**Theorem 1 (Birkhoff–von Neumann theorem [17] [18]):** Every doubly stochastic matrix is a convex combination of permutation matrices.

The set of $n \times n$ doubly stochastic matrices is a polytope called the Birkhoff polytope $B_n$ [17]. The Birkhoff–von Neumann theorem implies that any vertex (i.e., extreme point) of the Birkhoff polytope is a permutation matrix.

**B. LP decoding for permutation vectors**

Assume that $s \in \mathbb{R}^n$, which is referred to as the initial vector, is given.\hfill $^*$

The set of images of $s$ by left action of $X \in \Pi_n$ is referred to as the permutation vectors of $s$, which are given by $\Lambda(s) \triangleq \{X s \mid X \in \Pi_n \}$.

\hfill $^*$The elements in $s$ are not necessarily distinct.

We next consider the situation whereby a vector of $\Lambda(s)$ is transmitted to a receiver over an AWGN channel. In such a case, it is desirable to use an ML decoding algorithm to estimate the transmitted vector. The ML decoding rule can be described as $\hat{x} = \arg\min_{x \in \Lambda(s)} || y - x ||^2$, where $y$ is a received word.

The following theorem states that the ML decoding for $\Lambda(s)$ can be formulated as the following LP problem.

**Theorem 2 (LP decoding and ML certificate property):** Assume that a vector in $\Lambda(s)$ is transmitted over an AWGN channel and that $y \in \mathbb{R}^n$ is received at the receiver side. Let $X^*$ be the solution of the following LP problem:

$$\text{maximize} \quad \text{trace}(C^T X)$$

subject to

$$X \in \mathbb{R}^{n \times n}, X 1 = 1, 1^T X = 1^T, \quad X \geq 0, \quad (2)$$

where $C \triangleq y s^T$. If $X^*$ is integral, then $\hat{x} = X^* s$ holds.

**Proof:** The linear constraints in the above LP problem imply that $X$ is constrained to be a doubly stochastic matrix. On the other hand, the ML decoding rule can be rewritten as follows:

$$\hat{x} = \arg\min_{s \in \Lambda(s)} || y - s ||^2 = \arg\max_{X \in \Pi_n} \text{trace}(C^T X)s$$

Note that

$$\text{trace}(C^T X) = \sum_{i=1}^{n} \sum_{j=1}^{n} C_{i,j} X_{i,j}. \quad (3)$$

Since the vertices of the Birkhoff polytope form a permutation matrix, the ML decoding can be formulated as an integer LP (ILP) problem:

$$\text{maximize} \quad \text{trace}(C^T X)$$

subject to

$$X \in \Pi_n, \quad X \text{ is an integral matrix.}$$

By removing the integral constraint ($X$ is an integral matrix), we obtain the LP problem (2). If the solution of this LP problem is integral, it must coincide with the solution of the above ILP problem. \hfill $\blacksquare$

**III. LINEARLY CONSTRAINED PERMUTATION MATRICES AND LP DECODABLE PERMUTATION CODES**

It is natural to consider an extension of the LP decoding presented in the previous section. Additional linear constraints imposed on $\Pi_n$ produce a restricted set of $\Lambda(s)$. A decoding problem of such a set can be formulated as an LP problem, as in the case of the ML decoding of $\Lambda(s)$.

**A. Definitions**

The following definition for linearly constrained permutations gives an LP decodable subset of $\Lambda(s)$.

**Definition 3 (linearly constrained permutation matrix):** Let $m$ and $n$ be positive integers. Assume that

$$A \in \mathbb{Z}^{m \times n^2}, \quad b \in \mathbb{Z}^m$$
and $\leq$ are given. A set of linearly constrained permutation matrices is defined by $\Pi(A, b, \leq) \triangleq \{ X \in \mathbb{R}^n | A \text{ vec}(X) \leq b \}$.

Note that $A \text{ vec}(X) \leq b$ formally represents $m$ additional equalities and inequalities. These additional constraints provide a restriction on permutation matrices.

From the linearly constrained permutation matrices, LP decodable permutation codes are naturally defined as follows.

**Definition 4 (LP decodable permutation code):** Assume the same set up as in Definition 3. Suppose also that $s \in \mathbb{R}^n$ is given. The set of vectors $\Lambda(A, b, \leq, s)$ given by

$$\Lambda(A, b, \leq, s) \triangleq \{ X \in \mathbb{R}^n | X \in \Pi(A, b, \leq) \}$$

is referred to as an LP decodable permutation code.

The following example shows a case in which an additional linear constraint imposes a restriction on permutation matrices.

**Example 1:** Consider the set of linearly constrained permutation matrices that consists of $4 \times 4$ permutation matrices satisfying the linear constraint $\text{tr}(X) = 0$. This constraint implies that the diagonal elements of the permutation matrices are constrained to be zero. This means that such permutation matrices correspond to permutations without fixed points, which are referred to as derangements. For $n = 4$, there exist nine derangement permutation matrices. In this case, the triple $(A, b, \leq)$ is defined by

$$A = \text{vec}(I), \quad b = 0, \quad \leq \triangleq (\approx),$$

where $I$ is the $4 \times 4$ identity matrix. Multiplying these matrices by the initial vector $s = (0, 1, 2, 3)^T$ from the left, we immediately obtain the members of $\Lambda(A, b, \leq, (0, 1, 2, 3)^T)$:

$$(1, 0, 3, 2)^T, \quad (1, 2, 3, 0)^T, \quad (1, 3, 0, 2)^T,$$

$$(2, 0, 3, 1)^T, \quad (2, 3, 0, 1)^T, \quad (2, 3, 1, 0)^T,$$

$$(3, 0, 1, 2)^T, \quad (3, 2, 0, 1)^T, \quad (3, 2, 1, 0)^T.$$  

**B. LP decoding for LP decodable permutation codes**

The LP decoding of $\Lambda(A, b, \leq, s)$ is a natural extension of the LP decoding for $\Lambda(A)$. Assume that a vector in $\Lambda(A, b, \leq, s)$ is transmitted over an AWGN channel and $y$ is $\mathbb{R}^n$ is given. The procedure for the LP decoding of $\Lambda(A, b, \leq, s)$ is given as follows.

**LP decoding for an LP decodable permutation code**

1) Solve the following LP problem, and let $X^*$ be the solution.

maximize $\text{trace}(C^TX)$

subject to $X \in \mathbb{R}^{nxn}$,

$$X \geq 0, \quad X1 = 1, \quad 1^TX = 1^T,$$

$$A \text{ vec}(X) \leq b,$$

where $C = yb^T$.  

2) Output $X^*$ if $X^*$ is integral. Otherwise, declare a decoding failure.

**C. Remarks**

Several remarks should be made regarding the LP decoding for $\Lambda(A, b, \leq, s)$.

The feasible set of (6) is a subset of the feasible set of (2). All of the matrices in $\Pi(A, b, \leq)$ are feasible and permutation matrices that do not belong to $\Pi(A, b, \leq)$ are infeasible. This implies that all of the integral points of the feasible set (6) coincide with $\Pi(A, b, \leq)$.

The LP problem (6) is a relaxed problem of the ML decoding problem over AWGN channels:

minimize $||y - z||^2$ subject to $z \in \Lambda(A, b, \leq, s)$.  

This can be easily shown, as in the case (2). As a consequence of the above properties on integral points and on the relaxation, it can be concluded that the LP decoding for $\Lambda(A, b, \leq, s)$ also has the ML-certificate property. Namely, if the output of LP decoding is not decoding failure (i.e., $X^*$ is integral), then the output is exactly the same as the solution of the minimum distance decoding problem (7).

The feasible set of the LP problem (6) is the intersection of the Birkhoff polytope and a (possibly unbounded) convex set defined by the additional constraints. The intersection becomes a polytope, which is called a code polytope. The decoding performance of LP decoding is closely related to the code polytope given by the following definition.

**Definition 5 (Code polytope):** The polytope $\mathcal{P}(A, b, \leq)$ defined by

$$\mathcal{P}(A, b, \leq) \triangleq B_n \cap \{ X \in \mathbb{R}^{nxn} | A \text{ vec}(X) \leq b \}$$

is called the code polytope for $\Pi(A, b, \leq)$, where $B_n$ is the Birkhoff polytope corresponding to $\Pi_n$.

In an LP decoding process, these fractional vertices becomes a possible candidate of an LP solution. Thus, these fractional vertices can be considered to be pseudo permutation matrices, which degrade the decoding performance of the LP decoding.

**IV. ANALYSIS FOR LP DECODING PERFORMANCE**

In this section, an upper bound on decoding error probability for LP decoding is presented.

**A. Upper bound on LP decoding error probability**

An advantage of the LP formulation of a decoding algorithm is its simplicity for detailed decoding performance analysis. The geometrical properties of a code polytope are closely related to its decoding performance for the LP decoding. We can evaluate the block error probability of the proposed scheme with reasonable accuracy if we have sufficient information on a set of vertices of a code polytope. The bound presented in this section has a close relationship to the pseudo codeword analysis on LDPC codes [3].

In this section, a set of parameters $A, b, \leq, s$ are assumed to be given. Let $V$ be the set of vertices of the code polytope $\mathcal{P}(A, b, \leq)$. In general, $V$ contains fractional vertices.

The following lemma provides a bridge between a code polytope and the corresponding decoding error probability.
Lemma 1 (Upper bound on block error rate for LPD):
Assume that a codeword $Xs$ is transmitted to a receiver via an AWGN channel, where $X \in \Pi(A,b,\mathcal{E})$. The additive white Gaussian noise with mean 0 and variance $\sigma^2$ is assumed. The receiver uses the LP decoding algorithm presented in the previous section. In this case, the block error probability $P_{LP}(X)$ is upper bounded by

$$P_{LP}(X) \leq \sum_{\tilde{X} \in \mathcal{V}(X)} Q \left( \frac{||Xs||^2 - (\tilde{X}s)^T Xs}{\sigma ||\tilde{X}s - Xs||} \right),$$

where the $Q$-function is the tail probability of the normal Gaussian distribution.

Example 2: We have performed the following computer experiment using the following two codes:

1) LP decodable permutation code corresponding to the derangements of length 5. The additional linear constraint is $\omega(X) = 0$. A transmitted word $(1,0,4,2,3)^T$ is assumed. The code polytope has 44 vertices, which are all integral vertices.

2) LP decodable permutation code of length 5 corresponding to an additional linear constraint $X_{1,1} + X_{5,5} = 1$. A transmitted word $(0,4,3,2,1)^T$ is assumed. The code polytope has 330 vertices. The set of vertices contains 35 integral vertices and 294 fractional vertices.

An AWGN channel with noise variance $\sigma^2$ is assumed. The signal-to-noise ratio is defined by $SNR = 10 \log_{10} (1/\sigma^2)$. The LP decoding described in the previous section was used for decoding.

Figure 1 shows the upper bounds and simulation results on the block error probability of these permutation codes. It is readily observed that the upper bounds presented in this section exhibit reasonable agreement with the simulation results.

Although both codes have similar cardinalities (44 and 36), the derangement code provides much better block error probabilities than those of the code with the constraint $X_{1,1} + X_{5,5} = 1$. This is because the existence of fractional vertices (i.e., 294 fractional vertices) severely degrades the decoding performance of the code with the constraint $X_{1,1} + X_{5,5} = 1$, as compared with the derangement code.

Acknowledgement

The first author appreciates Prof. Han Vinck for directing the author's attention to the field of permutation codes. The first author also wishes to thank Dr. Jun Muramatsu for inspiring discussions on permutations.

References