

# Optimal Index Assignment for Scalar Quantizers and M-PSK via a Discrete Convolution-Rearrangement Inequality

Yunxiang Yao and Wai Ho Mow

Department of Electronic and Computer Engineering  
The Hong Kong University of Science and Technology  
Email: yyaoaj@ust.hk, eewhmow@ust.hk

**Abstract**—This paper investigates the problem of finding an optimal nonbinary index assignment from  $M$  quantization levels of a maximum entropy scalar quantizer to  $M$ -PSK symbols transmitted over a symmetric memoryless channel with additive noise following decreasing probability density function (such as the AWGN channel) so as to minimize the channel mean-squared distortion. The so-called zigzag mapping under maximum-likelihood (ML) decoding was known to be asymptotically optimal, but the problem of determining the optimal index assignment for any given signal-to-noise ratio (SNR) is still open. Based on a generalized version of the Hardy-Littlewood convolution-rearrangement inequality, we prove that the zigzag mapping under ML decoding is optimal for all SNRs. It is further proved that the same optimality results also hold under minimum mean-square-error (MMSE) decoding. Numerical results are presented to verify our optimality results and to demonstrate the performance gain of the optimal  $M$ -ary index assignment over the state-of-the-art binary counterpart for the case of 8-PSK over the AWGN channel.

## I. INTRODUCTION

Index assignment (IA) is a low-complexity approach for joint source-channel coding design. The classical binary IA problem aims to assign bit labels to the quantization code vectors in a way to ensure that any two code vectors with a small Euclidean distance have their corresponding bit labels close in the Hamming space. Therefore, if the transmitted bit labels are corrupted by noise and decoded erroneously at the receiver, the resulting distortion may not be too large. Some important results on the optimal binary IA problem are well-known. For the maximum entropy scalar quantizers and the binary symmetric channel (BSC), the *natural binary code* (NBC) is an optimal IA for all crossover probabilities, in the sense of minimizing the channel mean-squared distortion (MSD) [1], [2]. For general quantizers and discrete memoryless channels (DMC) with nonbinary channel symbols, lower bounds for the channel MSD were studied in [3], [4]. Generally speaking, finding the optimal IA is known to be NP-hard [3]. To the best of our knowledge, there are few previous works on the optimal IA problem, except for the aforementioned binary case and the specific nonbinary case to be described next.

In [5], the nonbinary IA problem of mapping the  $M$  levels of a maximum entropy scalar quantizer to the  $M$ -ary phase

shift keying ( $M$ -PSK) symbols transmitted over the additive white Gaussian noise (AWGN) channel was addressed. It was proved that the so-called *zigzag mapping* constructed therein is an asymptotically optimal IA at a sufficiently high signal-to-noise ratio (SNR) when a maximum-likelihood (ML) decoder is used. As an extension of the work, in [6] the authors proposed a near-optimal index assignment scheme for  $M^L$ -level quantizers to  $M$ -PSK symbols at a sufficiently high SNR. Besides the setting under ML decoding, the performance of the zigzag mapping for  $M$ -PSK with minimum mean-square-error (MMSE) decoding at a sufficiently high SNR was investigated in [7]. However, the corresponding optimal IA problem for any given SNR is still open. In particular, it is unclear if different IA constructions are needed for different SNRs.

Rearrangement inequalities indicate permutations of two or more functions or sets that optimize an objective function involving them. They are powerful tools in function analysis and are widely used in the proof of other inequalities [8]. For instance, classical rearrangement inequalities have been applied to prove the existence and uniqueness of the ground states of the Schrödinger equation in quantum mechanics [9]. For a convolution involving three continuous functions, the *Riesz convolution-rearrangement inequality* characterizes the rearrangements of the three functions that maximize the convolution [10]. It has found applications in information theory and communication problems, such as the network optimization [11], power entropy inequality [8], and Cover's problem for Gaussian relay channels [12]. Its discrete version was first developed by Hardy and Littlewood [13, Theorem 371] for characterizing the rearrangement of two sets. The inequality was extended to prime cyclic groups in [14] and then it was applied to the proof of discrete entropy inequalities in [15]. In [16], the Hardy-Littlewood convolution-rearrangement inequality was also generalized on discrete metric spaces. Recently, another discrete version of the Riesz rearrangement inequality on Hamming sphere was derived and applied to solve Cover's problem for the binary symmetric relay channel [17]. Furthermore, many results in majorization theory which plays an important role in optimization are established by using rearrangement inequalities [18]. Therefore, it is very interesting to investigate rearrangement inequalities with applications to coding and information theory. In this work, we

The work was supported by the Hong Kong Research Grants Council under project no. GRF 16233816.

relate the IA problem for  $M$ -PSK under both ML decoding and MMSE decoding to a rearrangement inequality to settle the aforementioned problem for any given SNR.

The paper is structured as follows. In the next section, we state the problem formulation of index assignment for  $M$ -PSK under ML decoding. In Section III, we provide a discrete convolution-rearrangement inequality and apply it to the IA problem. In Section IV, we show that the optimal index assignment for  $M$ -PSK under MMSE decoding can also be proved by the inequality. Finally, simulation results are demonstrated to verify the optimality of the proposed IA under both ML decoding and MMSE decoding.

## II. PROBLEM FORMULATION

In this section, we give the problem formulation for the index assignment under ML decoding. Consider an  $M$ -level maximum entropy scalar quantizer characterized by a set of quantization levels (i.e. the codebook)  $\mathcal{Q} = \{q_0, \dots, q_{M-1}\}$ , where  $q_i \in \mathbb{R}$  with  $0 \leq q_0 < q_1 < \dots < q_{M-1}$ . The maximum entropy quantizer outputs each quantization level with an equal probability of  $1/M$ . An  $M$ -PSK constellation is defined as

$$\mathcal{S} \triangleq \{s_k = e^{j2\pi k/M} | k = 0, 1, \dots, M-1\}, \quad (1)$$

where  $j \triangleq \sqrt{-1}$ . To describe the nonbinary index assignment for  $M$ -PSK, let us define a vector

$$\boldsymbol{\pi} = [\pi_0, \pi_1, \dots, \pi_{M-2}, \pi_{M-1}]. \quad (2)$$

It is a permutation of the quantization indices, i.e.,  $\pi_k \in \{0, \dots, M-1\}$ . The IA is a bijective mapping between the quantizer and the constellation in a way that each quantization level  $q_{\pi_k}$  is assigned to a distinct  $M$ -PSK symbol  $s_k$ .

Each quantization level is modulated as an  $M$ -PSK symbol following the bijective mapping  $\boldsymbol{\pi}$ , and the  $M$ -PSK symbol is then transmitted over a memoryless channel with an additive noise following a symmetrically decreasing probability density function (such as the AWGN channel). At the receiver, an  $M$ -PSK demodulator detects the most likely transmitted index based on the received signal and the quantizer decoder reconstructs the source symbol by producing the quantization level corresponding to the detected index. The channel MSD is defined as

$$D_C(\boldsymbol{\pi}) = \frac{1}{M} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P(s_j | s_i) (q_{\pi_i} - q_{\pi_j})^2, \quad (3)$$

where  $P(s_j | s_i)$  is the probability that  $M$ -PSK symbol  $s_j$  is detected conditioned on  $s_i$  is transmitted. The optimal IA problem is to find  $\boldsymbol{\pi}$  that minimizes  $D_C(\boldsymbol{\pi})$  over the set of all possible permutations. It is worth noting that the zigzag mapping [5] is produced by the permutation

$$\boldsymbol{\pi}_{zz} \triangleq [0, 1, 3, \dots, M-1, \dots, 6, 4, 2].$$

Note that transition probabilities of the channel satisfy

$$\sum_{i=0}^{M-1} P(s_j | s_i) = \sum_{j=0}^{M-1} P(s_j | s_i) = 1. \quad (4)$$

The channel MSD can be written as

$$D_C(\boldsymbol{\pi}) = \frac{2}{M} \sum_{i=0}^{M-1} q_i^2 - \frac{2}{M} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} q_{\pi_i} q_{\pi_j} P(s_j | s_i). \quad (5)$$

For a given quantizer, the first term of (5) is fixed. By defining  $p_{i,j} \triangleq P(s_j | s_i)$ , the nonbinary index assignment problem can be formulated as

$$\max_{\boldsymbol{\pi} \in \mathcal{P}} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} q_{\pi_i} q_{\pi_j} p_{i,j}, \quad (6)$$

where  $\mathcal{P}$  denotes the set of all possible  $M$ -ary permutations.

## III. OPTIMAL INDEX ASSIGNMENT BASED ON A DISCRETE CONVOLUTION-REARRANGEMENT INEQUALITY

For  $M$ -PSK transmission, the channel matrix  $P$  of the resulted  $M$ -ary DMC always satisfies the following conditions

$$\begin{cases} p_{i,j} \geq 0, & (7a) \\ p_{i,j} = p_{i',j'}, & d(i,j) = d(i',j'), & (7b) \\ p_{i,j} \geq p_{i',j'}, & d(i,j) < d(i',j'), & (7c) \end{cases}$$

where  $0 \leq i, j, i', j' \leq M-1$ , and  $d(i,j)$  is defined by

$$d(i,j) \triangleq \min\{(i-j) \bmod M, M - ((i-j) \bmod M)\}, \quad (8)$$

where the mod operation returns an integer between 0 and  $M-1$ .

Note that  $P$  is a non-negative circulant matrix, i.e.,  $p_{i,j}$  is a non-negative function involving two integers  $i, j$  and its value depends on  $d(i,j)$  only. For convenience of notation we define a monotonically decreasing function  $k(d(i,j)) \triangleq p_{i,j}$ . Let us define a vector  $\mathbf{x}$  by

$$\mathbf{x} \triangleq [x_{-m}, x_{-m+1}, \dots, x_0, \dots, x_{n-1}, x_n], \quad (9)$$

where  $m \triangleq \lfloor (M-1)/2 \rfloor$ ,  $n \triangleq \lfloor M/2 \rfloor$  and  $x_i \triangleq q_{\pi_{i+m}}$ . Then (6) can be rewritten in a discrete convolution form

$$\max_{\boldsymbol{\pi} \in \mathcal{P}_Q} \sum_{i=-m}^n \sum_{j=-m}^n x_i k(d(i,j)) x_j, \quad (10)$$

where  $\mathcal{P}_Q$  is the set of all orderings of quantizer codebook  $\mathcal{Q}$ .

To solve the problem (10), we give a discrete convolution-rearrangement inequality in the following theorem.

**Theorem 1:** Suppose  $M$  is a positive integer. Let  $\mathbb{Z}_M = \{\lceil \frac{1-M}{2} \rceil, \lceil \frac{1-M}{2} \rceil + 1, \dots, \lfloor \frac{M}{2} \rfloor\}$ . Let  $k$  be a decreasing non-negative function on  $[0, \infty)$  and let  $d(i,j)$  be defined by (8). For any two non-negative functions  $f$  and  $g$  on  $\mathbb{Z}_M$ , we have

$$\sum_{i,j \in \mathbb{Z}_M} f(i) k(d(i,j)) g(j) \leq \sum_{i,j \in \mathbb{Z}_M} f^*(i) k(d(i,j)) g^*(j), \quad (11)$$

where  $f^*$  is a discrete symmetric decreasing rearrangement of  $f$  such that  $f^*$  is derived from a permutation on the set of function values of  $f$  and satisfies

$$\begin{cases} f^*(0) \geq f^*(1) \geq f^*(-1) \geq f^*(2) \geq f^*(-2) \geq \dots \\ \quad \geq f^*(\lfloor \frac{M}{2} \rfloor) \geq f^*(\lceil \frac{1-M}{2} \rceil), & M \text{ is odd,} \\ f^*(0) \geq f^*(1) \geq f^*(-1) \geq f^*(2) \geq f^*(-2) \geq \dots \\ \quad \geq f^*(\lceil \frac{1-M}{2} \rceil) \geq f^*(\lfloor \frac{M}{2} \rfloor), & M \text{ is even.} \end{cases}$$

For ease of understanding, we provide a simple example with  $M = 6$ . Let us consider a non-negative monotonically decreasing function  $k(d(i, j)) = e^{-d(i, j)}$  with integers  $i, j \in [-2, 3]$  and consider  $f$  and  $g$  as

$$\begin{cases} f(i) = i^2, \\ g(i) = M + 2i, \end{cases} \quad i = -2, -1, \dots, 2, 3.$$

Denotes by  $\mathbf{f}$  a vector whose  $i$ -th element is  $f(i)$ , in which  $i = -2, -1, \dots, 2, 3$ . Two vectors associated with  $f$  and  $g$  are

$$\mathbf{f} = [4, 1, 0, 1, 4, 9], \quad \mathbf{g} = [2, 4, 6, 8, 10, 12].$$

We want to find orderings (i.e., permutations) of  $\mathbf{f}$  and  $\mathbf{g}$  that maximize the discrete convolution on the left-hand side of (11). Note that there are  $M$  elements in  $\mathbf{f}$  and  $\mathbf{g}$ , and there are  $M!$  orderings of  $\mathbf{f}$  and  $\mathbf{g}$ , respectively. Among all of the  $M! \times M!$  orderings, the one following

$$\mathbf{f}^* = [1, 4, 9, 4, 1, 0], \quad \mathbf{g}^* = [4, 8, 12, 10, 6, 2],$$

gives the maximum discrete convolution value.

*Remark 1:* An anonymous reviewer points out that Theorem 1 we derived is a rediscovery of [16, Theorem 5.1]. Due to space limitation, readers can find our proof in the full version of this paper [19].

The major difference of Theorem 1 and the Hardy-Littlewood convolution-rearrangement inequality [13, Theorem 371] is the definition of  $d(i, j)$ . For Theorem 1, if we consider a graph composed of  $M$  points joined together in a circle, then  $d(i, j)$  agrees with the graphic distance between the  $i$ -th and the  $j$ -th vertices on the circle. Therefore,  $d(i, j)$  is a periodic function of  $i - j$  with period  $M$ . However, in [13, Theorem 371]  $d(i, j)$  is symmetrically increasing with  $i - j$ . It is worth noting that the Hardy-Littlewood convolution-rearrangement inequality [13, Theorem 371] is a special case of Theorem 1 and can be obtained by letting the number of points on the discrete circle graph tend to infinity.

According to Theorem 1, the objective function of (10) achieves its maximum when  $\mathbf{x}$  is ordered as

$$x_0 \geq x_1 \geq x_{-1} \geq x_2 \geq x_{-2} \geq \dots, \quad (12)$$

where  $\dots$  indicates that we keep on going until we exhaust all integers in  $\mathbb{Z}_M$ .

The optimal solution for (6) can be consequently found as

$$\begin{cases} \pi = [0, 2, \dots, M-1, M-2, \dots, 1], & \text{for odd } M, \\ \pi = [1, 3, \dots, M-1, M-2, \dots, 0], & \text{for even } M. \end{cases} \quad (13)$$

Finally, we have Theorem 2 for the optimal IA for  $M$ -PSK under ML decoding.

*Theorem 2:* For maximum entropy scalar quantizers and  $M$ -PSK transmission over a memoryless channel with additive noise following a symmetrically decreasing probability density function, the optimal IA under ML decoding for minimizing channel MSD is

$$\begin{cases} [q_0, q_2, \dots, q_{M-1}, q_{M-2}, \dots, q_1], & \text{for odd } M, \\ [q_1, q_3, \dots, q_{M-1}, q_{M-2}, \dots, q_0], & \text{for even } M. \end{cases} \quad (14)$$

In [5], a set of distortion-preserving transforms for  $M$ -PSK are introduced. Given any IA, cyclically shifting the indices to the right, i.e.,  $[q_1, q_3, \dots, q_2, q_0] \rightarrow [q_0, q_1, q_3, \dots, q_2]$ , does not change the channel MSD. Besides, a reflection operation  $[q_2, q_4, \dots, q_1, q_0] \rightarrow [q_0, q_1, \dots, q_4, q_2]$  does not influence the channel MSD. Note that the IA for even  $M$  in (14) can be transformed to the zigzag mapping [5] by a cyclic shift operation, and the IA for odd  $M$  in (14) can be transformed to the zigzag mapping by a reflection operation and a cyclic shift operation. Therefore, the zigzag mapping under ML decoding is proved to be optimal for all SNRs.

#### IV. OPTIMAL INDEX ASSIGNMENT FOR $M$ -PSK UNDER MMSE DECODING

In this section, the optimal IA for  $M$ -PSK under MMSE decoding is investigated. Different from the ML decoder that maps the detected  $M$ -PSK symbol back to a quantization level following the IA, we can alternatively consider an MMSE decoder which computes and outputs  $y_j$  based on detected symbol  $s_j$  by

$$y_j = E(q|s_j) = \frac{\sum_{k=0}^{M-1} \frac{1}{M} q_{\pi_k} P(s_j|s_k)}{\sum_{k=0}^{M-1} \frac{1}{M} P(s_j|s_k)}, \quad (15)$$

and the channel MSD is

$$D_C(\pi) = \frac{1}{M} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} P(s_j|s_i) (q_{\pi_i} - y_j)^2. \quad (16)$$

According to (4), the channel MSD is formulated as

$$\begin{aligned} D_C(\pi) &= \frac{1}{M} \left( \sum_{i=0}^{M-1} q_i^2 - 2 \sum_{j=0}^{M-1} \sum_{i=0}^{M-1} P(s_j|s_i) q_{\pi_i} y_j + \sum_{j=0}^{M-1} y_j^2 \right) \\ &= \frac{1}{M} \sum_{i=0}^{M-1} q_i^2 - \frac{1}{M} \sum_{j=0}^{M-1} \left( \sum_{k=0}^{M-1} q_{\pi_k} P(s_j|s_k) \right)^2. \end{aligned} \quad (17)$$

Letting  $p_{k,j} = P(s_j|s_k)$ , then minimizing the channel MSD is equivalent to

$$\max_{\pi \in \mathcal{P}} \sum_{j=0}^{M-1} \left( \sum_{k=0}^{M-1} q_{\pi_k} p_{k,j} \right)^2. \quad (18)$$

And it can be simplified as

$$\max_{\pi \in \mathcal{P}} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} q_{\pi_i} q_{\pi_j} h_{i,j}, \quad (19)$$

where  $H = PP^T$ .

The similarity between (19) and (6) gives us the insight to solve (19) by Theorem 1. For this purpose, we investigate the property of  $H$  and get the following lemma.

*Lemma 1:* Suppose that the conditions in (7) hold for two square matrices  $Q$  and  $R$ . Then the conditions in (7) also hold for the matrix  $QR^T$ .

*Proof:* The condition (7a) holds for the matrix  $QR^T$  since the product of two non-negative matrices is also non-negative.

Let us define  $H \triangleq QR^T$ . Note that  $Q$  and  $R$  are symmetric circulant matrices since (7b) holds for them. According to [20],  $H$  is also a symmetric circulant matrix. This fact implies that (7b) also holds for  $H$ .

To prove the condition (7c), let us define two function compositions by  $k_q(d(i, j)) \triangleq q_{i,j}$  and  $k_r(d(i, j)) \triangleq r_{i,j}$  for  $1 \leq i, j \leq M$ , where  $k_r$  and  $k_q$  are two monotonically decreasing functions, and  $d(i, j)$  is defined by (8). Because of (7b), the two function compositions can represent all entries in  $Q$  and  $R$ . Entries of  $H$  are computed by

$$h_{i,j} = \sum_{k=1}^M q_{i,k} r_{j,k} = \sum_{k=1}^M k_q(d(i, k)) k_r(d(j, k)). \quad (20)$$

The condition (7b) holds iff the following condition

$$h_{i,j} \geq h_{i',j'}, \quad \text{if } d(i', j') = d(i, j) + 1 \quad (21)$$

holds. Defining  $\Delta_{ii'jj'} \triangleq h_{i,j} - h_{i',j'}$ . For the sake of convenience, subscripts of  $\Delta$  will be omitted in the remainder of the paper. Then based on (20) we have

$$\begin{aligned} \Delta &= \sum_{k=1}^M k_q(d(i, k)) k_r(d(j, k)) - \sum_{k=1}^M k_q(d(i', k)) k_r(d(j', k)) \\ &\stackrel{(a)}{=} \sum_{k=1}^M k_q(d(0, k)) k_r(d(j-i, k)) \\ &\quad - \sum_{k=1}^M k_q(d(0, k)) k_r(d(j'-i', k)), \end{aligned} \quad (22)$$

where equality (a) follows from the fact that

$$\sum_{k=1}^M k_q(d(i, k)) k_r(d(j, k)) = \sum_{k=1-i}^{M-i} k_q(d(0, k)) k_r(d(j-i, k)),$$

and  $d(i, k) = d(i, k + M)$ .

According to (8), if  $i$  and  $j$  are integers between 1 and  $M$ ,  $d(i, j)$  is equivalent to the following two cases

$$d(i, j) = \begin{cases} |j-i|, & |j-i| \leq M/2, \\ M-|j-i|, & |j-i| > M/2. \end{cases}$$

Then the condition  $d(i', j') = d(i, j) + 1$  in (21) is equivalent to the following four cases

$$\begin{cases} |j'-i'| = |j-i| + 1, & |j-i| < M/2, |j'-i'| \leq M/2, \\ M-|j'-i'| = |j-i| + 1, & |j-i| < M/2, |j'-i'| \geq M/2, \\ |j'-i'| = M-|j-i| + 1, & |j-i| > M/2, |j'-i'| \leq M/2, \\ |j'-i'| = |j-i| - 1, & |j-i| > M/2, |j'-i'| \geq M/2. \end{cases} \quad (23)$$

For a given  $|j-i|$ , the first two cases have the same value of  $d(i', j')$ . Because of (7b), the two cases also have the same value of  $h_{i',j'}$ . For proving (21), it is sufficient to examine one of them for  $|j-i| < M/2$ . Similarly, we just need to consider one of the last two cases for  $|j-i| > M/2$ .

Then  $\Delta$  can be sufficiently examined by the first and the last cases in (23). We give the proof of the first case. The last case can be proved similarly. Since  $H$  is symmetric, let us

assume  $i-j \geq 0$  and  $i'-j' \geq 0$  and get  $j'-i' = j-i-1$ . Then (22) is equivalent to

$$\begin{aligned} \Delta &= \sum_{k=1}^M k_q(d(0, k)) k_r(d(j-i, k)) \\ &\quad - \sum_{k=1}^M k_q(d(0, k-1)) k_r(d(j-i, k)), \end{aligned} \quad (24)$$

which follows from the fact that

$$\begin{aligned} &\sum_{k=1}^M k_q(d(0, k)) k_r(d(j-i-1, k)) \\ &= \sum_{k=2}^{M+1} k_q(d(0, k-1)) k_r(d(j-i, k)), \end{aligned}$$

and  $d(i, k) = d(i, k + M)$ .

Assume  $M$  is even. Note that the case of odd  $M$  can be proved similarly. From (24) we have

$$\begin{aligned} \Delta &= \sum_{k=1}^M \left( k_q(d(0, k)) - k_q(d(0, k-1)) \right) k_r(d(j-i, k)) \\ &= \sum_{k=1}^{\frac{M}{2}} \left( k_q(d(0, k)) - k_q(d(0, k-1)) \right) k_r(d(j-i, k)) \\ &\quad + \sum_{k=\frac{M}{2}+1}^M \left( k_q(d(0, k)) - k_q(d(0, k-1)) \right) k_r(d(j-i, k)) \\ &\stackrel{(b)}{=} \sum_{k=1}^{\frac{M}{2}} \left( k_q(d(0, k)) - k_q(d(0, k-1)) \right) k_r(d(j-i, k)) \\ &\quad + \sum_{k'=\frac{M}{2}}^{\frac{M}{2}} \left( k_q(d(0, k'-1)) - k_q(d(0, k')) \right) k_r(d(j-i, 1-k')) \\ &= \sum_{k=1}^{\frac{M}{2}} \left( k_q(d(0, k)) - k_q(d(0, k-1)) \right) \\ &\quad \times \left( k_r(d(j-i, k)) - k_r(d(j-i, 1-k)) \right), \end{aligned}$$

where equality (b) is obtained by letting  $k' = M - k + 1$ . It is obvious that

$$k_q(d(0, k)) - k_q(d(0, k-1)) \leq 0, \quad 1 \leq k \leq M/2.$$

Beside, we have

$$k_r(d(j-i, k)) - k_r(d(j-i, 1-k)) \leq 0, \quad 1 \leq k \leq M/2,$$

which follows from the fact that

$$d(j-i, k) \geq d(j-i, 1-k), \quad 0 \leq i-j < M/2.$$

Therefore, we have  $\Delta \geq 0$ . The condition (7c) is proved consequently. ■

According to Lemma 1, the optimal IA problem for  $M$ -PSK under MMSE decoding can also be solved by the Theorem 1. We have Theorem 3 for the optimal IA.

*Theorem 3:* For maximum entropy scalar quantizers and  $M$ -PSK transmission over a memoryless channel with additive noise following a symmetrically decreasing probability density function, the zigzag mapping is the optimal IA under MMSE decoding.

## V. NUMERICAL RESULTS

To verify the optimality of the zigzag mapping, numerical results are demonstrated to compare the MSD performance of the zigzag mapping with the optimal mapping by the exhaustive search. Real-valued source symbols following a uniform distribution over  $[0, 1]$  are generated. The source symbols are then quantized by an  $M$ -level uniform scalar quantizer. The  $M$ -level quantized symbols are mapped to  $M$ -PSK symbols following an index assignment.  $M$ -PSK symbols are transmitted over the AWGN channel. After ML or MMSE decoding, the source data are reconstructed. To show that the zigzag mapping is optimal at all SNRs, we perform an exhaustive search for each simulated SNR separately.

To show the gain of the nonbinary index assignment, we also compare its performance with the state-of-the-art binary counterpart. To modulate the bits as  $M$ -PSK symbols, Gray code is used to minimize the bit error rate of the resultant BSC. In binary index assignment design for maximum entropy scalar quantizers and the BSC, the NBC is known to be optimal under both ML decoding [1] and MMSE decoding [2]. Therefore, the NBC-Gray mapping is considered as the binary counterpart. To make the equivalent BSC from the  $M$ -PSK transmission memoryless, we assume an ideal bit interleaver to eliminate the correlation among all Gray coded bits.

Figure 1 and Figure 2 show channel MSD performances under ML decoding and MMSE decoding, respectively. To verify the optimality of the zigzag mapping at all SNRs, we also provide results at low SNRs. Here we only provide the results for  $M = 8$  since the search space (which consists of  $M!$  candidates) has a size that grows exponentially with  $M$ .

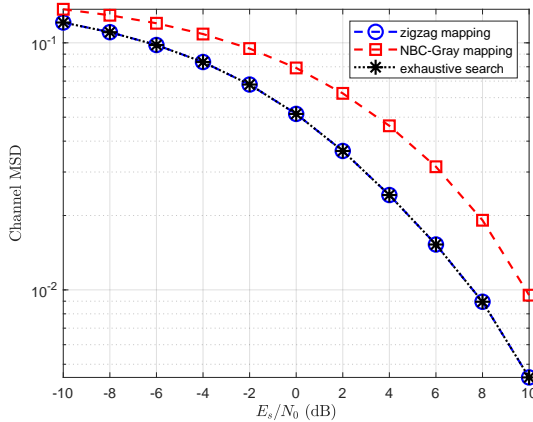


Fig. 1. Channel MSD of the zigzag mapping, the NBC-Gray mapping and the optimal mapping exhaustively search for each SNR for 8-PSK (i.e.,  $M = 8$ ) under ML decoding in AWGN channel.

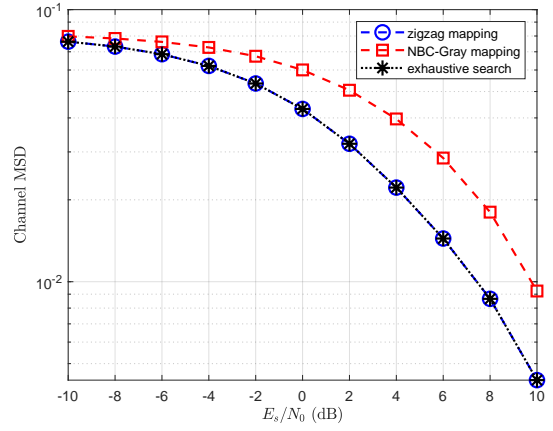


Fig. 2. Channel MSD of the zigzag mapping, the NBC-Gray mapping and the optimal mapping exhaustively search for each SNR for 8-PSK (i.e.,  $M = 8$ ) under MMSE decoding in AWGN channel.

From the figures, the performances of the zigzag mapping and the exhaustive search mapping are always the same. We also checked exhaustive search IAs for all simulated SNRs. They are the same as the zigzag mapping or can be transformed to the zigzag mapping by distortion-preserving transforms in [5]. Besides, the performance gain of the optimal zigzag mapping over that of the NBC-Gray mapping is significant. The SNR gains are up to around 3 dB under both ML and MMSE decoding. It is also worth noting that the channel MSD under MMSE decoding outperforms the one under ML decoding, especially at low SNRs.

## VI. CONCLUSION

A discrete convolution-rearrangement inequality was re-discovered. The inequality was applied to settle the optimal nonbinary index assignment problem for the  $M$ -level maximum entropy scalar quantizer and the  $M$ -PSK over a memoryless channel. For both ML and MMSE decoding, the zigzag mapping has been proved to be optimal for all SNRs. Simulation results were provided to verify the optimality of the zigzag mapping and to show the gain over the conventional binary index assignment. Further research on the application of rearrangement inequalities to address the IA and other problems in coding theory is a promising direction.

## APPENDIX

The discrete decreasing rearrangement  $f^*$  and  $g^*$  is equivalent to the fact that

$$\begin{cases} f^*(r) - f^*(r') \geq 0, & \text{if } |r'| > |r|, \text{ or if } r' = -r < 0, \\ g^*(s) - g^*(s') \geq 0, & \text{if } |s'| > |s|, \text{ or if } s' = -s < 0. \end{cases} \quad (25)$$

To prove the ordering following (25) gives the maximum, we define an operation  $\Omega_p(f, g)$  that swaps all pairs

$$(f(p-i), f(p+i)), (g(p-j), g(p+j)), \quad i, j = 1, 2, 3, \dots, \quad (26)$$

or in pairs

$$(f(p-i), f(p+i+1)), (g(p-j), g(p+j+1)), \quad i, j = 0, 1, 2, \dots, \quad (27)$$

which do not satisfy conditions in (25). Note that  $f(r), g(s)$  are set as 0 when  $r, s$  out of the scope of  $[-m, n]$ . It is worth pointing out that starting from any  $f$  and  $g$ , the operation  $\Omega_p(f, g)$  can be applied for different  $p$  iteratively, until we get  $f^*$  and  $g^*$ . Therefore, a sufficient condition of Theorem 1 is that  $\Omega_p$  always introduces non-negative increment.

To prove Theorem 1, we give the following lemma as the sufficient condition of Theorem 1.

**Lemma 2:** For any  $f$  and  $g$ , the operation  $\Omega_p(f, g)$  for arbitrary  $p$  can always introduce non-negative increment to the discrete convolution on the left hand side of (11).

*Proof:* Note that (26) and (27) always involve different pairs. We can prove Lemma 2 for them separately. Here we give the proof for pairs in (27) and the proof for pairs in (26) can be done following a similar procedure. In the proof we also assume  $M$  is an even number, the case for odd  $M$  also can be done similarly.

Given any  $p$ , then  $i, j$  in (27) satisfy that

$$\begin{cases} |p-i| \leq |p+i+1|, & |p-j| \leq |p+j+1|, & p \geq 0, \\ |p-i| \geq |p+i+1|, & |p-j| \geq |p+j+1|, & p < 0. \end{cases} \quad (28)$$

For given  $f$  and  $g$ , let us denote by  $I_w$  and  $J_w$  the set of  $i$  and  $j$  for which

$$\begin{cases} f(p-i) < f(p+i+1), & g(p-j) < g(p+j+1), & p \geq 0, \\ f(p-i) > f(p+i+1), & g(p-j) > g(p+j+1), & p < 0. \end{cases}$$

Note that the pairs corresponding to  $i \in I_w$  and  $j \in J_w$  do not satisfy (25). We also define the set of  $i, j$  who satisfy (25) as  $I_r$  and  $J_r$ . Note that the union of  $I_w$  and  $I_r$  is the set of all possible values that  $i$  can be, let us define it by  $I$ . Similarly, the union of  $J_w$  and  $J_r$  can also be defined as  $J$ .

Let us define  $d$  as the value of the discrete convolution, i.e.,

$$\chi = \sum_{r, s \in \mathbb{Z}_M} f(r)k(d(r, s))g(r). \quad (29)$$

According to the aforementioned definition of  $I$  and  $J$ , (29) can be divided into four partial sums

$$\begin{aligned} d &= \sum_{i \in I} \sum_{j \in J} \left( k(d(i, j)) (f(p-i)g(p-j) + f(p+i+1)g(p+j+1)) \right. \\ &\quad \left. + k(d(i+1, -j)) (f(p-i)g(p+j+1) + f(p+i+1)g(p-j)) \right) \\ &= \sum_{i \in I_r} \sum_{j \in J_r} \left( k(d(i, j)) (f(p-i)g(p-j) + f(p+i+1)g(p+j+1)) \right. \\ &\quad \left. + k(d(i+1, -j)) (f(p-i)g(p+j+1) + f(p+i+1)g(p-j)) \right) \\ &+ \sum_{i \in I_w} \sum_{j \in J_w} \left( k(d(i, j)) (f(p-i)g(p-j) + f(p+i+1)g(p+j+1)) \right. \\ &\quad \left. + k(d(i+1, -j)) (f(p-i)g(p+j+1) + f(p+i+1)g(p-j)) \right) \\ &+ \sum_{i \in I_w} \sum_{j \in J_r} \left( k(d(i, j)) (f(p-i)g(p-j) + f(p+i+1)g(p+j+1)) \right. \\ &\quad \left. + k(d(i+1, -j)) (f(p-i)g(p+j+1) + f(p+i+1)g(p-j)) \right) \\ &= \sum_{i \in I_r} \sum_{j \in J_w} \left( k(d(i, j)) (f(p-i)g(p-j) + f(p+i+1)g(p+j+1)) \right. \\ &\quad \left. + k(d(i+1, -j)) (f(p-i)g(p+j+1) + f(p+i+1)g(p-j)) \right). \end{aligned} \quad (30)$$

Let us define the four partial terms corresponding to the four convolution ranges as  $\chi_1, \chi_2, \chi_3$ , and  $\chi_4$ .

To show the increment of  $\Omega_p$  on  $\chi$ , its effect on each of the four terms need to be examined separately. It is clear that  $\Omega_p$  has no influence on  $\chi_1$ . It is also trivial to check that  $\chi_2$  is not affected by  $\Omega_p$ . Therefore, the only two partial sums need to be considered are  $\chi_3$  and  $\chi_4$ .

Let us define  $d_3$  to be the increment produced by  $\Omega_p$  on  $\chi_3$ , i.e.,

$$\begin{aligned} d_3 &\triangleq \sum_{i \in I_w} \sum_{j \in J_r} \left( k(d(i, j)) (f(p-i)g(p+j+1) + f(p+i+1)g(p-j)) \right. \\ &\quad \left. + k(d(i+1, -j)) (f(p-i)g(p-j) + f(p+i+1)g(p+j+1)) \right. \\ &\quad \left. - \sum_{i \in I_w} \sum_{j \in J_r} \left( k(d(i, j)) (f(p-i)g(p-j) + f(p+i+1)g(p+j+1)) \right. \right. \\ &\quad \left. \left. + k(d(i+1, -j)) (f(p-i)g(p+j+1) + f(p+i+1)g(p-j)) \right) \right). \end{aligned} \quad (31)$$

Now the task becomes proving that (31) is non-negative.

Let us simplify  $d_3$  in (31) as

$$\begin{aligned} d_3 &= \sum_{i \in I_r} (f(p+i+1) - f(p-i)) \\ &\quad \times \sum_{j \in J_w} (k(d(i, j)) - k(d(-i, j+1))) (g(p-j) - g(p+j+1)). \end{aligned} \quad (32)$$

Recall that  $f(r), g(s)$  are 0 if  $r, s$  outside of the scope of  $[-m, n]$ . If  $j \in J_w$ , both  $p-j$  and  $p+j+1$  should be in the range of  $[-m, n]$ . Therefore, the range of  $I_w, J_w$  should be

$$0 \leq i, j \leq m - |p|, \quad \text{if } i \in I_w, j \in J_w. \quad (33)$$

Similarly, for  $i \in I_r$ , at least one of  $p-i$  and  $p+i+1$  should be in the  $[-m, n]$ . The range of  $I_w, J_w$  should be

$$0 \leq i, j \leq m + |p|, \quad \text{if } i \in I_r, j \in J_r. \quad (34)$$

Then  $d_3$  in (32) can be further divided into two terms

$$\begin{aligned} d_3 &= \sum_{\substack{0 \leq i \leq m - |p|, \\ i \in I_r}} (f(p+i+1) - f(p-i)) \\ &\quad \times \sum_{j \in J_w} (k(d(i, j)) - k(d(-i, j+1))) (g(p-j) - g(p+j+1)) \\ &+ \sum_{m - |p| + 1 \leq i \leq m + |p|} (f(p+i+1) - f(p-i)) \\ &\quad \times \sum_{j \in J_w} (k(d(i, j)) - k(d(-i, j+1))) (g(p-j) - g(p+j+1)). \end{aligned} \quad (35)$$

Note that the second term does not exist if  $p = 0$ . And all  $i$  in the range of the second term satisfy  $i \in I_r$  because of (33).

It is trivial that the first term in (35) is non-negative. When  $p \neq 0$ , the second term of (35) also need to be considered, in which  $m - |p| + 1 \leq i \leq m + |p|$ . For each  $m + 1 \leq i \leq m + |p|$ , there is  $m - |p| + 1 \leq (2m + 1) - i \leq m$  such that

$$\begin{cases} d(i, j) = d(-(2m + 1) - i, j + 1), \\ d(-i, j + 1) = d((2m + 1) - i, j). \end{cases} \quad (36)$$

Therefore, the second term of (35) can be written as

$$\begin{aligned} &\sum_{m+1 \leq i \leq m+|p|} \left( (f(p+i+1) - f(p-i)) - (f(p+i'+1) - f(p-i')) \right) \\ &\quad \times \sum_{j \in J_w} (k(d(i, j)) - k(d(-i, j+1))) (g(p-j) - g(p+j+1)), \end{aligned} \quad (37)$$

where  $i' = (2m + 1) - i$  for notational convenience.

Note that  $f(p - i) = f(p - i') = 0$  when  $p < 0$  since both  $p - i$  and  $p - i'$  are outside the range  $[-m, n]$ . Similarly,  $f(p + i + 1) = f(p + i' + 1) = 0$  when  $p > 0$ . Then for any  $p \neq 0$  there is

$$((f(p+i+1)-f(p-i))-(f(p+i'+1)-f(p-i')))(g(p-j)-g(p+j+1)) \leq 0$$

We also can prove

$$k(d(i, j)) \leq k(d(-i, j+1)), \quad m+1 \leq i \leq m+|p|, 0 \leq j \leq m-|p|$$

by trivially check that

$$d(i, j) \geq d(-i, j+1), \quad m+1 \leq i \leq m+|p|, 0 \leq j \leq m-|p|.$$

Therefore, (37) is non-negative for an arbitrary  $p$ . Consequently, (35) is proved to be non-negative for any  $p$ , i.e.,  $d_3$  is always non-negative. Similarly, the increment produced by  $\Omega_p$  on  $\chi_4$  can be proved to be non-negative. Finally, the increment introduced by  $\Omega_p$  to all of  $\chi_1, \chi_2, \chi_3, \chi_4$  are non-negative. The increment produced by  $\Omega_p$  on  $\chi$  for any  $p$  is always non-negative. ■

## REFERENCES

- [1] S. W. McLaughlin, D. L. Neuhoff, and J. J. Ashley, "Optimal binary index assignments for a class of equiprobable scalar and vector quantizers," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 2031–2037, Nov. 1995.
- [2] B. Farber and K. Zeger, "Quantizers with uniform encoders and channel optimized decoders," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 62–77, Jan. 2004.
- [3] G. Ben-David and D. Malah, "Bounds on the performance of vector-quantizers under channel errors," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2227–2235, Jun. 2005.
- [4] X. Wu, H. D. Mittelmann, X. Wang, and J. Wang, "On computation of performance bounds of optimal index assignment," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3229–3233, Dec. 2011.
- [5] H. Chan, W. H. Mow, and C. Leung, "Index assignment for scalar quantization with M-ary phase shift keying," in *Proc. IEEE Int. Symp. Inf. Theory*, Adelaide, SA, Sep. 2005, pp. 357–361.
- [6] Y. Yao and W. H. Mow, "Near-optimal nonbinary index assignment for equiprobable uniform quantizers and M-PSK transmission," in *Proc. IEEE Int. Conf. Commun.*, Dublin, Ireland, Ireland, Jul. 2020, pp. 1–6.
- [7] D. Qiao, W. H. Mow, and C. Leung, "Scalar quantizers with uniform encoders and channel-optimized decoders for M-PSK schemes," in *Proc. IEEE GLOBECOM*, Miami, FL, USA, Dec. 2010, pp. 1–5.
- [8] L. Wang and M. Madiman, "Beyond the entropy power inequality, via rearrangements," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5116–5137, Sep. 2014.
- [9] P. Choquard and J. Stubbe, "The one-dimensional Schrödinger–Newton equations," *Lett. Math. Phys.*, vol. 81, no. 2, pp. 177–184, Jul. 2007.
- [10] F. Riesz, "Sur une inégalité intégrale," *J. London Math. Soc.*, vol. 5, pp. 162–168, Jul. 1930.
- [11] B. Hajek, K. Mitzel, and S. Yang, "Paging and registration in cellular networks: Jointly optimal policies and an iterative algorithm," *IEEE Trans. Inf. Theory*, vol. 54, no. 2, pp. 608–622, Feb. 2008.
- [12] X. Wu, L. P. Barnes, and A. Özgür, "The capacity of the relay channel: Solution to Cover's problem in the Gaussian case," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 255–275, Oct. 2019.
- [13] G. H. Hardy, J. E. Littlewood, G. Pólya *et al.*, *Inequalities*. Cambridge, U.K.: Cambridge University Press, 1952.
- [14] V. Lev, "Linear equations over  $\mathbb{F}_p$  and moments of exponential sums," *Duke mathematical journal*, vol. 107, no. 2, pp. 239–263, 2001.
- [15] M. Madiman, L. Wang, and J. O. Woo, "Entropy inequalities for sums in prime cyclic groups," *arXiv preprint arXiv:1710.00812*, 2017.
- [16] A. R. Pruss, "Discrete convolution-rearrangement inequalities and the Faber-Krahn inequality on regular trees," *Duke mathematical journal*, vol. 91, no. 3, pp. 463–514, 1998.
- [17] L. P. Barnes, X. Wu, and A. Özgür, "A solution to Cover's problem for the binary symmetric relay channel: Geometry of sets on the Hamming sphere," in *Proc. 55th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2017, pp. 844–851.
- [18] T. Ando, "Majorization, doubly stochastic matrices, and comparison of eigenvalues," *Linear Algebra and its Applications*, vol. 118, pp. 163–248, 1989.
- [19] Y. Yao and W. H. Mow, "Optimal index assignment for scalar quantizers and M-PSK via a discrete convolution-rearrangement inequality," *arXiv preprint arXiv:2010.10300*, 2021.
- [20] R. Horn and C. R. Johnson, *Matrix analysis*. New York: Cambridge university press, 2012.