A Mutual Information Invariance Approach to Symmetry in Discrete Memoryless Channels

Bike Xie and Richard Wesel
Electrical Engineering Department
UCLA Henry Samueli School of Engineering and Applied Science
Los Angeles, CA 90095, USA
Email: xbk@ee.ucla.edu, wesel@ee.ucla.edu

***Abstract*— There are numerous notions of symmetry for discrete memoryless channels. A common goal of these various definitions is that the capacity may be easily computed once the channel is declared to be symmetric. In this paper we focus on a class of definitions of symmetry characterized by the invariance of the channel mutual information over a group of permutations of the input distribution. For definitions of symmetry within this class, we give a simple proof of the optimality of the uniform distribution. The fundamental channels are all symmetric with a general enough definition of symmetry. This paper provides a definition of symmetry that covers these fundamental channels along with a proof that is simple enough to find itself on the chalkboard of even the most introductory class in information theory.**

I. Introduction

There are many different definitions of symmetry for discrete memoryless channels in the literature including at least [1] [2] [3] [4] [5] [6]. This paper focuses on channels for which the mutual information doesn't change when the input distribution is permuted in certain ways. This approach does not produce a larger class of symmetric channels. The largest class appears to be the "GCT" symmetric channels defined by Chen and Yang [5]. However, the information-invariance approach requires only a simple convexity argument while still producing a simple-to-identify definition of symmetry that covers all the basic channels.

II. A Mutual Information Invariance Property

A discrete memoryless channel (n input, m output) with input X and output Y is determined by its transition probability matrix $T_{n\times m}$. For any input distribution p, denote the mutual information between X and Y as $I_p(X;Y)$.

Definition 1 *Input-invariance Symmetry: A discrete memoryless channel has input-invariance symmetry if for every input distribution p_1, there are $K-1$ complementary distributions $p_2,\cdots,p_K$ such that these distributions are permutations of p_1, $I_{p_i}(X;Y) = I_{p_1}(X;Y)$, and $\frac{1}{K}\sum_{i=1}^{K} p_i = u$, where u denotes uniform distribution.*

Theorem 1 *If a discrete memoryless channel with input X and output Y has input-invariance symmetry then the uniform distribution achieves capacity.*

Proof of Theorem 1: For any input distribution p_1,

$$I_{p_1}(X;Y) = \frac{1}{K}\sum_{i=1}^{K} I_{p_i}(X;Y) \tag{1}$$

$$\leq I_{\frac{1}{K}\sum_{i=1}^{K} p_i}(X;Y) \tag{2}$$

$$= I_u(X;Y), \tag{3}$$

where (2) follows from Jensen's inequality with the inequality reversed because $I_p(X;Y)$ is concave in p. Thus, uniform input distribution achieves capacity. Q.E.D.

As a simple (but powerful) example of input-invariance symmetry consider cyclic-shift symmetry. Suppose that

$$p(x) = \begin{cases} p_1 & \text{if } x = a \\ p_2 & \text{if } x = b \\ p_3 & \text{if } x = c \end{cases}. \tag{4}$$

This probability mass function (PMF) may be written as the vector $p(x) = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix}$. Let $p^{(1)}(x)$ indicate a cyclic shift by one of $p(x)$, which produces

$$p^{(1)}(x) = \begin{cases} p_3 & \text{if } x = a \\ p_1 & \text{if } x = b \\ p_2 & \text{if } x = c \end{cases}, \tag{5}$$

or $p^{(1)}(x) = \begin{bmatrix} p_3 & p_1 & p_2 \end{bmatrix}$. Similarly, let $p^{(i)}(x)$ indicate the cyclic shift by i. For example $p^{(2)}(x) = \begin{bmatrix} p_2 & p_3 & p_1 \end{bmatrix}$.

Definition 2 *Cyclic-shift Symmetry ([7]): A discrete memoryless channel has cyclic-shift symmetry if its mutual information is invariant to cyclic shifts in the input distribution.*

The binary symmetric channel, the binary erasure channel, and the noisy typewriter channel all have cyclic-shift symmetry. Later in this paper we show that Wang-Kulkarni-Poor cyclic symmetry [2] implies cyclic-shift symmetry.

Theorem 2 *A cyclic-shift symmetric channel has input-invariance symmetry with $K = n$ and $p_i = p_1^{(i-1)}, i = 2,\cdots,K$. Thus, the uniform input distribution achieves capacity for a cyclic-shift symmetric channel.*

III. Matrix Conditions for Input-invariance Symmetry

This section gives a sufficient condition for recognizing input-invariance symmetry from a channel transition probability matrix T. These sufficient conditions turn out to be equivalent to the Witsenhausen-Wyner definition of input-symmetry, as shown in the section after this.

Definition 3 *A column-preserving row permutation is a row permutation that preserves the set of columns (although not necessarily their positions).*

Theorem 3 *The channel described by transition probability matrix $T_{n\times m}$ has input-invariance symmetry if the following two conditions are satisfied:*

- *The columns of the transition probability matrix can be partitioned into subsets such that each subset has rows that are permutations of each other and columns that are permutation of each other.*
- *There exists a set of column preserving row permutation's $\{G_i, i = 1, \cdots, n\}$ including $G_1 = I$ such that for any distribution p, $\frac{1}{n}\sum_{i=1}^{n} pG_i = u$.*

As an example, for the binary erasure channel the columns of

$$T = \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \end{bmatrix} \tag{6}$$

may be partitioned into the subsets described above as follows:

$$S_1 = \left\{ \begin{bmatrix} \alpha \\ \alpha \end{bmatrix} \right\}, \qquad S_2 = \left\{ \begin{bmatrix} 1-\alpha \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1-\alpha \end{bmatrix} \right\}. \tag{7}$$

The set of column-preserving row permutations is

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\}.$$

Another example is the 4-input, 4-output channel

$$T = \begin{bmatrix} a & b & c & d \\ b & a & d & c \\ c & d & a & b \\ d & c & b & a \end{bmatrix}. \tag{8}$$

In this case, partitioning into subsets is not necessary. The rows of T are permutations of each other. The columns of T are permutations of each other. It has the following four column-preserving row permutations:

$$G_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, G_2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{9}$$

$$G_3 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, G_4 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}. \tag{10}$$

Furthermore, $\frac{1}{4}\sum_{i=1}^{4} pG_i = u$ for any distribution p.

Proof of Theorem 3: For any input distribution p_1, let $p_i = p_1 G_i$ for $i = 2, \cdots, n$. In order to prove the theorem, it suffices to show that $I_{p_i}(X;Y) = I_{p_1}(X;Y)$.

The first condition, that all the rows of P are permutations of each other, forces $H(Y|X = x)$ to be the same for all x. This means that

$$H(Y|X) = \sum_x p(x) H(Y|X = x) \tag{11}$$

is the same regardless of the input distribution $p(x)$. In particular, $H_{p_i}(Y|X) = H_{p_1}(Y|X)$.

Now we use the second condition to show that $H_{p_i}(Y) = H_{p_1}(Y)$. The $G_i, i \in \{1, \cdots, n\}$, are column-preserving row permutations for the transition probability matrix T, Thus there exist permutation matrices $\Pi_i, i = 1, \cdots, n$ such that $G_i T = T\Pi_i$. Hence the output distribution corresponding to the input distribution p_i is $p_i T = p_1 G_i T = p_1 T \Pi_i$, and so $H_{p_i}(Y) = H_{p_1}(Y)$.

Since $H(Y)$ and $H(Y|X)$ are invariant to these column-preserving row permutation's, $I(X;Y)$ is also invariant to these column preserving row permutation's in the input distribution, and the channel has input-invariance symmetry.

IV. Relationships with other definitions of symmetry

There are many different definitions of symmetric channels in the literature [1] [2] [3] [4] [5] [6]. This section discusses how input-invariance symmetry relates to other notions of symmetry. We begin with the Witsenhausen-Wyner [1] definitions of symmetry, one of which is closely related to input-invariance symmetry.

A. Witsenhausen-Wyner symmetry

Let Φ_n denote the representation of the symmetric group of permutations of n objects by the $n \times n$ permutation matrices. For an $n \times m$ stochastic matrix T (an n input, m output channel), let $\mathcal{G}_{\rm i}$ be the set $\{G \in \Phi_n : \exists \pi \in \Phi_m, \text{s.t. } GT = T\Pi\}$ and $\mathcal{G}_{\rm o}$ be the set $\{\Pi \in \Phi_m : \exists G \in \Phi_n, \text{s.t. } GT = T\Pi\}$. If $G_1 T = T\Pi_1$, $G_2 T = T\Pi_2$, then $G_1 G_2 T = T\Pi_1\Pi_2$, which shows that $\mathcal{G}_{\rm i}$ and $\mathcal{G}_{\rm o}$ are subgroups of the finite groups Φ_n and Φ_m respectively [1].

Definition 4 *Witsenhausen-Wyner (WW) Input Symmetry [1]: A discrete memoryless channel T is WW input symmetric if the set $\mathcal{G}_{\rm i}$ is transitive, i.e., each element of $\{1, \cdots, n\}$ can be mapped to every other element of $\{1, \cdots, n\}$ by some member of $\mathcal{G}_{\rm i}$.*

Definition 5 *WW Output Symmetry [1]: A discrete memoryless channel T is WW output symmetric if the set $\mathcal{G}_{\rm o}$ is transitive, i.e., each element of $\{1, \cdots, m\}$ can be mapped to every other element of $\{1, \cdots, m\}$ by some member of $\mathcal{G}_{\rm o}$.*

Definition 6 *WW Symmetry [1]: A discrete memoryless channel T is WW-symmetric if both $\mathcal{G}_{\rm i}$ and $\mathcal{G}_{\rm o}$ are transitive.*

Theorem 4 *If a channel T is WW input symmetric, then rows of T are permutations of each other. Similarly, if a channel T is output symmetric, then columns of T are permutations of each other.*

Proof of Theorem 4: Suppose the channel T is WW input symmetric. Let $G_{i,j} \in \mathcal{G}_{\rm i}$ be a permutation matrix that maps the $i^{\rm th}$ row to the $j^{\rm th}$ row. There exists a permutation matrix Π such that $G_{i,j}T = T\Pi$. Thus, the $i^{\rm th}$ row of T is the same as the $j^{\rm th}$ row of $T\Pi$, and hence is a permutation of the $j^{\rm th}$ row of T. Since $\mathcal{G}_{\rm i}$ is transitive, all rows of T are permutations of each other. Similarly, all columns of a output symmetric channel T are permutations of each other. Q.E.D.

Theorem 5 *Any WW-input-symmetric channel T also has input-invariance symmetry. Particularly, suppose $|\mathcal{G}_{\rm o}| = K$, $\mathcal{G}_{\rm i} = \{G_1 = I, G_2, \cdots, G_K\}$, for every input distribution p_1, there are $K-1$ complementary input distribution $p_2, \cdots, p_K$ with $p_i = p_1 G_i$ for $i \in \{2, \cdots, K\}$, $I_{p_i}(X;Y) = I_{p_1}(X;Y)$ and $\frac{1}{K}\sum_{i=1}^{K} p_i = u$.*

Proof of Theorem 5: Since $p_i = p_1 G_i$ for $i \in \{2, \cdots, K\}$, gives the construction of the complementary distributions and shows that these distributions are permutation of p_1, it suffices to show that $I_{p_i}(X;Y) = I_{p_1}(X;Y)$ and $\frac{1}{K}\sum_{i=1}^{K} p_i = u$. Note that by Theorem 4 the rows of T are permutations of each other, so $H(Y|X)$ is independent of the input distribution.

$$\begin{aligned} I_{p_i}(X;Y) &= H_{p_i}(Y) - H(Y|X) && (12)\\ &= H(p_i T) - H(Y|X) && (13)\\ &= H(p_1 G_i T) - H(Y|X) && (14)\\ &= H(p_1 T) - H(Y|X) && (15)\\ &= H_{p_1}(Y) - H(Y|X) && (16)\\ &= I_{p_1}(X;Y). && (17) \end{aligned}$$

Let $q = \frac{1}{K}\sum_{i=1}^{K} p_i$. $\forall G \in \mathcal{G}_{\rm i}$,

$$\begin{aligned} qG &= \frac{1}{K}\sum_{i=1}^{K} p_i G && (18)\\ &= \frac{1}{K}\sum_{i=1}^{K} p_1 G_i G && (19)\\ &= \frac{1}{K}\sum_{i=1}^{K} p_1 G_i && (20)\\ &= q, && (21) \end{aligned}$$

where (20) is true because $\mathcal{G}_{\rm i}$ is a group. Since $\mathcal{G}_{\rm i}$ is transitive, q must be equal to u. Q.E.D.

It should be clear from the above proof that any WW-input-symmetric channel will satisfy the matrix conditions for input-invariance symmetry given in the previous section. Furthermore, any channel that satisfies those matrix conditions must also be WW input symmetric since the requirement that $\frac{1}{n}\sum_{i=1}^{n} pG_i = u$ guarantees that the set $\mathcal{G}_{\rm i}$ is transitive.

Note that in Theorem 5 and its proof $K \geq n$ because there are at least n distinct elements $G_{1,1}, \cdots, G_{1,n} \in \mathcal{G}_{\rm i}$.

As a consequence of Theorem 5, a WW-symmetric channel is also a input-invariance symmetric channel. However, a WW-output-symmetric channel doesn't have to be input-invariance symmetric. For instance, a 3-input 2-output channel

$$T = \begin{pmatrix} \alpha & 1-\alpha \\ 1/2 & 1/2 \\ 1-\alpha & \alpha \end{pmatrix} \tag{22}$$

is output symmetric because

$$\mathcal{G}_{\rm o} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\}. \tag{23}$$

It is not input-invariance symmetric because the input distribution $p = [1/2, 0, 1/2]$, rather than the uniform distribution, achieves capacity. The example of the binary erasure channel shows that an input-invariance symmetric channel doesn't have to be output symmetric. For the binary erasure channel

$$\mathcal{G}_{\rm o} = \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right\}, \tag{24}$$

which is not transitive.

B. Wang-Kulkarni-Poor cyclic symmetry

For any bijective function (permutation matrix) $Q : \{1, \cdots, m\} \mapsto \{1, \cdots m\}$. Let Q^i indicate the corresponding i-times self-composition of Q.

Definition 7 *Wang-Kulkarni-Poor (WKP) cyclic symmetry [2]: An n input m output memoryless channel T is WKP cyclic symmetric if there exists a bijective function Q such that $Q^n = I$ and $T(1,j) = T(i, Q^{i-1}(j))$ for $i = 1, \cdots, n$, where $T(i,j)$ is the i^{th} row, j^{th} column entry of T.*

Theorem 6 *A discrete memoryless channel T has WKP cyclic symmetry, only if it is also cyclic-shift symmetric (see Definition 2).*

Proof of Theorem 6: Let t_i for $i \in \{1, \cdots, n\}$ be the $i^{\rm th}$ row of T. $T(1,j) = T(i, Q^{i-1}(j))$ can be rewrote as $t_i = t_1 Q$. Thus, rows of T are permutation of each other and $H(Y|X)$ is invariant to cyclic shifts in the input distribution.

$$T = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} = \begin{bmatrix} t_1 \\ t_1 Q \\ \vdots \\ t_1 Q^{n-1} \end{bmatrix} \tag{25}$$

$$TQ = \begin{bmatrix} t_1 Q \\ t_1 Q^2 \\ \vdots \\ t_1 Q^n \end{bmatrix} = \begin{bmatrix} t_1 Q \\ t_1 Q^2 \\ \vdots \\ t_1 I \end{bmatrix} = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ 1 & & & & 0 \end{bmatrix} T = G_{\rm cyc} T. \tag{26}$$

¿From (26), it is easy to verify that for any input distribution p_1, its cyclic shifts $p_1 G_{\rm cyc}^i$ for $i = 1, \cdots, n-1$ achieve the same $H(Y)$ and the same mutual information as p_1 does.

Therefore, the channel has cyclic-shift symmetry, and also input-invariance symmetry. Q.E.D.

Note, however, that an input-invariance symmetric channel doesn't have to be WKP cyclic symmetric. The example in (8) is input-invariance symmetric but not WKP-cyclic symmetric. This channel is not WKP-cyclic symmetric because G_{cyc} is not a column-preserving row permutation of T.

C. Cover-Thomas symmetry

Cover and Thomas introduced two definitions of symmetry in [4].

Definition 8 *Cover-Thomas (CT) Symmetry [4]: A discrete memoryless channel with transition probability matrix T is CT symmetric if the rows of T are permutations of each other and columns of T are permutations of each other.*

Definition 9 *CT Weak Symmetry [4]: A discrete memoryless channel with transition probability matrix T is CT weakly symmetric if rows of T are permutation of each other and all the column sums are equal.*

It is clear from these definitions that CT weak symmetry includes CT symmetry. These two definitions of symmetry both have the property that uniform distribution achieves capacity [4]. The binary erasure channel has neither CT symmetry nor CT weak symmetry because the column corresponding to the erasure output is not a permutation of the other columns and (typically) does not even have the same column sum as the other columns. Thus the set of CT symmetric channels doesn't contain the set of cyclic-shift symmetric channels or the set of input-invariance symmetric channels.

On the other hand, the set of input-invariance symmetric channels doesn't contain the set of CT symmetric channels or the set of CT weakly symmetric channels. For example, consider the type of transition probability matrices given by the Latin square [5], [8]

$$\begin{bmatrix} a & b & c & d & e \\ b & a & d & e & c \\ c & e & a & b & d \\ d & c & e & a & b \\ e & d & b & c & a \end{bmatrix}.$$

It is clear that a channel with this type of transition probability matrix has CT symmetry and CT weak symmetry. However, this class of channels can hardly have input-invariance symmetry because the only column-preserving row permutation for this transition probability matrix is the identity matrix I when a, b, c, d, e are distinct. Particularly, when $a = 1/31, b = 2/31, c = 4/31, d = 8/31, e = 16/31$, for the input distribution $p = [1/2, 1/2, 0, 0, 0]$, the only permutation of p achieves the same mutual information between X and Y as p does is p itself. Thus, this channel is not input-invariance symmetric and it shows that the set of input-invariance symmetric channels doesn't contain the set of CT symmetric channels or the set of CT weak symmetric channels.

D. Gallager symmetry and Chen-Yang symmetry

Now we turn to another class of symmetric channels defined by Gallager [3] and its extended class of symmetric channels defined by Chen and Yang [5].

Definition 10 *Gallager (G) Symmetry [3]: A discrete memoryless channel is defined to be Gallager's symmetric if the set of outputs can be partitioned into subsets in such a way that for each subset the matrix of transition probabilities (using inputs as rows and outputs of the subset as columns) has the property that each row is a permutation of each other row and each column (if more than 1) is a permutation of each other column.*

Definition 11 *Chen-Yang (CY) Symmetry [5]: A discrete memoryless channel is defined to be CY-symmetric if the set of outputs can be partitioned into subsets in such a way that for each subset the matrix of transition probabilities (using inputs as rows and outputs of the subset as columns) has the property that each row is a permutation of each other row and the columns sums are equal.*

It is clear from these definitions that G symmetry includes CT symmetry and that CY-symmetry includes G symmetry and CT weak symmetry (and of course CT symmetry). Thus, the example which shows that input-invariance symmetry doesn't include CT symmetry also shows that input-invariance symmetry doesn't include Gallager's symmetry or GCT-symmetry.

On the other hand, a channel satisfying the matrix conditions for input-invariance symmetry in Theorem 3 is also G symmetric. We conjecture that both G symmetry or CY symmetry include input-invariance symmetry, but since Theorem three is sufficient (but has not been shown to be necessary) this conjecture remains unproven.

The uniform distribution achieves capacity of both G-symmetric channels [3] and CY-Symmetric channels [5]. Therefore, all classes of symmetric channels discussed in this paper except output symmetric channels have the property that uniform distribution achieves capacity. Among these class of symmetric channels, we conjecture that the class of CY-symmetric channels is the largest and includes all others. (The only remaining question in this regard is whether CY symmetry includes all cases of input-invariance symmetry.) In contrast, cyclic-shift symmetry is simple to understand and to prove and covers all the fundamental channels.

V. Conclusion

This paper introduced a definition of symmetry in discrete memoryless channels requiring that the mutual information remain invariant under certain permutations of the input distribution. This definition does not produce a larger class of symmetric channels. The largest class appears to be the CY-symmetric channels defined by Chen and Yang [5]. However, the information-invariance approach requires only a simple convexity argument while still covering all the basic channels.

ACKNOWLEDGMENT

We are grateful to Jun Chen for his assistance. This work was supported by Boeing and by ST Microelectronics and the University of California through the UC Discovery Program.

REFERENCES

[1] H. S. Witsenhausen and A. D. Wyner. A conditional entropy bound for a pair of discrete random variables. *IEEE Trans. Inform. Theory*, IT-21:493–501, September 1975.

[2] C.-C. Wang, S. R. Kulkarni, and H. V. Poor. Finite-dimensional bounds on Z_m and binary LDPC codes with belief propagation decoders. *IEEE Trans. Inform. Theory*, IT-53:56–81, January 2007.

[3] R. G. Gallager. *Information Theory and Reliable Communication*. New York: Wiley, 1963.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. New York: Wiley, 1991.

[5] J. Chen, D.-K. He, and E.-H. Yang. On the codebook-level duality between Slepian-Wolf coding and channel coding. *IEEE Information Theory and Applications Workshop*, pages 84–93, Jan. 29-Feb. 2 2007.

[6] N. Liu and S. Ulukus. The capacity region of a class of discrete degraded interference channels. *IEEE Information Theory and Applications Workshop*, Jan. 29-Feb. 2 2007.

[7] R. D. Wesel. *Information Theory Course Notes at UCLA* . 2000.

[8] J. H. van Lint and R. M. Wilson. *A Course in Combinatorics*. Cambridge, U.K.: Cambridge Univ. Press, 2001.