Some New Results on Index Coding When The Number of Data Is Less Than The Number of Receivers

Jungho So, Sangwoon Kwak and Youngchul Sung†
Dept. of Electrical Engineering
KAIST
Daejeon, Korea 305-701
Email: {jhso@, sw.kwak@ and ysung@ee.}kaist.ac.kr

***Abstract*—In this paper, index coding problems in which the number (m) of receivers is larger than that (n) of data are considered. Unlike the case that the two numbers are same ($n = m$), index coding problems with $n \leq m$ are more general and hard to handle. To circumvent this difficulty, problems with $n < m$ are approached via corresponding problems with $n = m$. It is shown that in certain cases, the symmetric capacity and code construction for index coding problems with $n < m$ can be obtained from the existing symmetric capacity result and codes for index coding problems with $n = m$. Such cases include cases with $n < m \leq 5$.**

I. Introduction

Recently, the source coding problem with a broadcast channel with multiple receivers that have side information has drawn much attention from the research community since the problem, named Informed Source Coding On Demand (ISCOD), was first introduced by Birk and Kol [1]. In ISCOD, a transmitter wants to deliver data to receivers by using a broadcast channel with minimum required transmission time, under the assumption that each receiver requires a part of the data and has some side information about the data, and the transmitter knows what side information each receiver has. The problem is often referred to as *index coding*. Index coding has many applications in distributed systems such as distributed storage, satellite communications, network coding, video on demand, cellular networks, etc. [2]

The considered index coding problem in this paper is formally defined as follows.

Definition 1: A transmitter sends a data set $X = \{x_1, x_2, \cdots, x_n : x_i \in \{0,1\}, \forall i\}$ *to* m *receivers* $r_1, r_2, \cdots, r_m$ *by using a broadcast channel. Here, each receiver* r_i *requires* $X[f(i)] := \{x_j | j \in f(i)\}$*, where* $f(i) \subset [n] := \{1, \cdots, n\}$ *is the index set of the required data for receiver* i*, and has side information* $X[N(i)] := \{x_j | j \in N(i)\}$*, where* $N(i) \subset [n]$ *is the index set for the side information for receiver* i*. The function* $E(X, \{N(\cdot)\}) : (0,1)^n \rightarrow (0,1)^l$ *is an index code of length* l *if* $E(X, \{N(\cdot)\})$ *is decodable in the sense that every receiver* r_i *satisfies the following condition:*

$$H(X[f(i)]|E(X, \{N(\cdot)\}), X[N(i)]) = 0 \text{ at } r_i,$$

where $H(\cdot)$ *is the entropy function.* $E(X)$ *shall be used for* $E(X, \{N(\cdot)\})$ *for simplicity.*

Mostly, the index coding problem has been considered in the case in which $n = m$ and $f(1), \cdots, f(m)$ is a partition of $[n]$ with $|f(i)| = 1,\ \forall\ i$ [1], [3]–[5]. (We refer to this case as $\mathcal{C}_1$.) However, one needs to handle general cases with $n \neq m$ to accommodate various system setup. An index coding problem in the case of $n > m$ and $f(i) \cap f(j) = \emptyset$ for $i \neq j$ can trivially be converted to a corresponding problem in the case of $n = m$ and $|f(i)| = 1,\ \forall\ i$ by adding virtual receivers and properly assigning side information for the virtual receivers [1]. However, the case of $n < m$ is not as simple as the case of $n > m$. (We refer to the case of $n < m$ and $|f(i)| = 1\ ,\forall i$ as $\mathcal{C}_2$. Hereafter, we shall use $\mathcal{P}_i$ $(i = 1, 2)$ to represent a particular index coding problem belonging to the case $\mathcal{C}_i$.) Recently, Lubetzky and Stav considered the case $\mathcal{C}_2$ and obtained a lower bound on the minimum length of index coding for $\mathcal{P}_2$ by considering a $\mathcal{P}_1$ properly constructed from the original index coding problem $\mathcal{P}_2$ [6].

In this paper, we investigate the general case $\mathcal{C}_2$ further, and obtain some new results regarding the general case $\mathcal{C}_2$.

- First, we show that for a given $\mathcal{P}_2$, $\ell^\star(\mathcal{P}_2) = \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ in the case that each datum x_i is required by less than three receivers if linear coding is optimal for $\mathcal{P}_1(G_{cl}(\mathcal{P}_2))$. Here, $G_{cl}(\mathcal{P}_2)$ is a directed graph properly constructed from $\mathcal{P}_2$ as in [6], and $\mathcal{P}_1(G)$ is the $\mathcal{C}_1$-index coding problem equivalent to a directed graph G. (See [1] for the equivalence between G and $\mathcal{P}_1(G)$.) $\ell^\star(\cdot)$ is the minimum index code length of the corresponding problem. In this case, a known algorithm for $\mathcal{P}_1(G_{cl}(\mathcal{P}_2))$ can be used to solve the original $\mathcal{P}_2$. (See Lemma 2 and Remark 1 in Appendix.)
- Second, we further identify certain situations under which $\ell^\star(\mathcal{P}_2) = \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$.
- Third, based on the existing symmetric capacity result for $\mathcal{C}_1$ with $n = m \leq 5$ [4], $\ell^\star(\mathcal{P}_2) \geq \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ [6], and our new result, we compute the symmetric capacity for $\mathcal{C}_2$ with $n < m \leq 5$. In the case to which the first result in the above applies, the above result simply yields

† This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2013R1A1A2A10060852).

This paper has supplementary downloadable material available at http://wisrl.kaist.ac.kr/papers/Num14ISITSoKwakSung.zip provided by the authors. This material is 1.15 MB in size.

978-1-4799-5186-4/14/$31.00 ©2014 IEEE

the desired result. In other cases, we numerically verify that $\ell^\star(\mathcal{P}_2) = \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ when $n < m \leq 5$.

- Finally, we provide an example for which $\ell_L^\star(\mathcal{P}_2) > \ell_L^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$. Note that $\ell^\star(\mathcal{P}_2) \geq \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ by Lubetzky and Stav [6]. See Definition 2 for $\ell_L^\star(\cdot)$.

A. Related Work

In [1], the authors showed that considering the case of a single requested data symbol per a receiver is enough, which means considering the case of $n \leq m$ and $|f(i)| = 1, \ \forall \ i$ is enough, by adding virtual receivers and properly assigning side information for the virtual receivers. In [3], the authors considered the scalar linear index coding for $\mathcal{C}_1$, and showed that in this case, a $\mathcal{P}_1$ can be represented by a directed graph $G(V, E)$ and the optimal index code length of $\mathcal{P}_1$ is given by a graph parameter of $G(V, E)$. In addition, the authors identified certain cases in which linear index coding is optimal, and such cases include directed acyclic graphs, perfect graphs, odd cycles and odd anti-holes, etc. However, in [6], Lubetzky and Stav showed that there exist certain cases in which non-linear index coding strictly outperforms linear index coding. In [4], Arbabjolfaei *et al.* obtained the capacity region for index coding with up to five receivers for $\mathcal{C}_1$. By using this result and setting all user rates the same, one can obtain the symmetric capacity for $\mathcal{C}_1$ with up to five receivers. In [7], the authors considered the case of $\mathcal{C}_2$ and showed that a bipartite graph can be used to represent a $\mathcal{P}_2$.

B. Background

We here provide some definitions and background for later sections.

Definition 2 ($\ell^\star(\mathcal{P}), \ell_L^\star(\mathcal{P})$): Consider an index coding problem $\mathcal{P}$. The minimum length of an index code for $\mathcal{P}$ and the minimum length of a linear index code for $\mathcal{P}$ are denoted by $\ell^\star(\mathcal{P})$ and $\ell_L^\star(\mathcal{P})$, respectively.

First, consider the case $\mathcal{C}_1$. For simplicity, we assume that $f(i) = i$ in this case. A $\mathcal{C}_1$-index coding problem $\mathcal{P}_1$ can be represented by a directed graph $G = (V, E)$, where V is a vertex set and E is an edge set [1]. The directed graph corresponding to $\mathcal{P}_1$ is constructed as follows: [1]

1) The vertex set V is given by $V = [n]$, where vertex i represents r_i and x_i.
2) $(i, j) \in E$ if and only if r_i knows the datum x_j.

In this way, a directed graph G with n nodes is equivalent to an index coding problem $\mathcal{P}_1$ [1]. Here, we will use $\mathcal{P}_1(G)$ (or simply G when no ambiguity) for the $\mathcal{C}_1$-index coding problem equivalent to a directed graph G. It is shown in [3] that $\ell_L^\star(G) = \text{minrk}_2(G)$, where the $\text{minrk}_2(G)$ is defined as the minimum rank of a matrix fitting G. An $n \times n$ binary matrix A fits a graph $G(V, E)$ with $|V| = n$ if $A(i, i) = 1$ for all i; $A(i, j) = 0$ if $(i, j) \notin E$; and $A(i, j) = 0$ or 1 if $(i, j) \in E$. We denote the set of all matrices fitting G by $M(G)$ or $M(\mathcal{P}_1(G))$.

Next, consider the case $\mathcal{C}_2$. Handling $\mathcal{C}_2$ is more difficult than handling $\mathcal{C}_1$ for several reasons. For example, a $\mathcal{C}_2$-index coding problem $\mathcal{P}_2$ cannot be represented by a directed graph with equivalence. However, it is shown that it is convenient to express a $\mathcal{P}_2$ as a matrix in a similar way to the case $\mathcal{C}_1$ [6]. Define a matrix set $M(\mathcal{P}_2)$ for $\mathcal{P}_2$ as follows: [6] $A \in M(\mathcal{P}_2)$ if

1) A is a $m \times n$ matrix;
2) $A(i, j) = 1$ for all (i, j) with $j \in f(i)$; and
3) $A(i, j) = 0$ for all (i, j) with $j \notin N(i)$.

It is shown in [6] that $\ell_L^\star(\mathcal{P}_2) = \min_{A \in M(\mathcal{P}_2)} \text{rank}_2(A)$. Furthermore, in [6], an m-vertex directed graph $G_{cl}(\mathcal{P}_2)$ is constructed to capture several properties of $\mathcal{P}_2$. The graph $G_{cl}(\mathcal{P}_2) = (V, E)$ is constructed from $\mathcal{P}_2$ as follows: [6]

1) The set V of m vertices is given by $V = [m]$. Vertex i represents receiver r_i.
2) $(i, j) \in E$ if and only if r_i knows datum $x_{f(j)}$ or r_i and r_j require the same datum.

In $G_{cl}(\mathcal{P}_2)$, a set of receivers that require the same datum forms a clique in the graph $G_{cl}(\mathcal{P}_2)$. Let the clique composed of the receivers that require x_i be denoted by c_i. We will consider that a clique of one node is not a clique (although trivial cliques of size one are considered as cliques in general). This means $|c_i| \geq 2$ in this paper.

Now, consider an example $\mathcal{C}_2$-index coding problem $\mathcal{I}_{EX}$: $X = \{x_1, x_2, x_3\}$, $R = \{r_1, r_2, r_3, r_4, r_5\}$, $f(1) = f(4) = f(5) = 1$, $f(2) = 2$, $f(3) = 3$ and $N(1) = \{2\}$, $N(2) = \{1\}$, $N(3) = \{2\}$, $N(4) = \{3\}$, $N(5) = \emptyset$. The matrix representation for $\mathcal{I}_{EX}$ is given by

$$M(\mathcal{I}_{EX}) = \begin{bmatrix} 1 & * & 0 \\ * & 1 & 0 \\ 0 & * & 1 \\ 1 & 0 & * \\ 1 & 0 & 0 \end{bmatrix}, \tag{1}$$

where $*$ indicates that the corresponding value can be either 0 or 1. Note that, the columns and rows in $M(\mathcal{I}_{EX})$ in (1) represent the data and the receivers, respectively. That is, the columns 1, 2 and 3 represent x_1, x_2 and x_3, respectively, and the rows 1, 2, 3, 4 and 5 represent r_1, r_2, r_3, r_4 and r_5, respectively. Now, consider the $\mathcal{C}_1$-index coding problem $\mathcal{P}_1(G_{cl}(\mathcal{I}_{EX}))$ corresponding to the directed graph $G_{cl}(\mathcal{I}_{EX})$. The matrix representation of $\mathcal{P}_1(G_{cl}(\mathcal{I}_{EX}))$ denoted by $M(\mathcal{P}_1(G_{cl}(\mathcal{I}_{EX})))$ is given by

$$M(\mathcal{P}_1(G_{cl}(\mathcal{I}_{EX}))) = \begin{bmatrix} 1 & * & 0 & * & * \\ * & 1 & 0 & * & * \\ 0 & * & 1 & 0 & 0 \\ * & 0 & * & 1 & * \\ * & 0 & 0 & * & 1 \end{bmatrix}. \tag{2}$$

By the definition of $G_{cl}(\mathcal{I}_{EX})$ and $x_1 = x_{f(1)} = x_{f(4)} = x_{f(5)}$, receivers r_1, r_4 and r_5 form a clique c_1. Thus, $M(\mathcal{P}_1(G_{cl}(\mathcal{I}_{EX})))$ has $*$ in the positions $(1, 4)$, $(1, 5)$, $(4, 1)$, $(4, 5)$, $(5, 1)$ and $(5, 4)$ in (2). Since $N(2) = \{1\}$, receiver r_2 knows $x_1 (= x_{f(1)} = x_{f(4)} = x_{f(5)})$. This implies that $M(\mathcal{P}_1(G_{cl}(\mathcal{I}_{EX})))$ has $*$ in the positions $(2, 1)$, $(2, 4)$ and $(2, 5)$ in (2). Since $N(3) = \{2\}$, receiver r_3 does not know $x_1 (= x_{f(1)} = x_{f(4)} = x_{f(5)})$. This implies that

$M(\mathcal{P}_1(G_{cl}(\mathcal{I}_{EX})))$ has 0 in the positions $(3,1)$, $(3,4)$ and $(3,5)$ and has $*$ in the position $(3,2)$. Similarly, we can incorporate $N(1)$, $N(4)$ and $N(5)$. Note that the columns 1, 4 and 5 corresponding to the clique c_1 of receivers 1, 4 and 5 requiring the same datum x_1 will become the same if 1 is replaced by $*$. This is true even in general cases. (This fact can be used to identify if a given $\mathcal{P}_1$ can be derived from $\mathcal{P}_2$('s) by G_{cl}.) Therefore, in general $\mathcal{C}_2$ with $n < m$ and $|f(i)| = 1$, $\forall i$, at least two receivers require the same datum and thus there exist at least two columns in $M(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ that will become the same if 1 is replaced by $*$.

In [6], it is shown that $\ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ is a lower bound of $\ell^\star(\mathcal{P}_2)$, i.e., $\ell^\star(\mathcal{P}_2) \geq \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$. The optimal broadcast rate of an index coding problem $\mathcal{P}$ is not $\ell^\star(\mathcal{P})$, since $\ell^\star$ is the minimum length obtained with the constraint that the length of data is limited as 1. We refer to an index code with the length of data being one as a scalar index code. In general, the vector index code with vector size t is defined as follows:

Definition 3: A transmitter sends a data set $X = \{x_1, x_2, \cdots, x_n : x_i \in \{0,1\}^t, \forall i\}$ to m receivers $r_1, r_2, \cdots, r_m$ by using a broadcast channel. The function $E(X) : (0,1)^{tn} \to (0,1)^l$ is a vector index code of length l with vector size t if $E(X)$ is decodable in the sense that every receiver r_i satisfies the following condition:

$$H(X[f(i)]|E(X), X[N(i)]) = 0 \text{ at } r_i,$$

where $f(i), N(i)$ and $X[N(i)]$ are defined in Definition 1.

By using vector index code, the optimal broadcast rate of an index coding problem is defined as follows:

Definition 4 ($\beta_t(\mathcal{P})$,$\beta(\mathcal{P})$): [8] *Consider an index coding problem $\mathcal{P}$. Let the minimum length of a vector index code with vector size t for $\mathcal{P}$ be $\beta_t(\mathcal{P})$. Then, the optimal broadcast rate of the index coding problem $\mathcal{P}$ is defined as*

$$\beta(\mathcal{P}) := \lim_{t\to\infty} \frac{\beta_t(\mathcal{P})}{t}. \qquad (3)$$

The symmetric capacity of an index coding problem $\mathcal{P}$ is the inverse of optimal broadcast rate, i.e., $\frac{1}{\beta(\mathcal{P})}$.

II. New Results on The Asymmetric Case $\mathcal{C}_2$

In this section, we provide some new results on the asymmetric case $\mathcal{C}_2$. The first result is provided in the following theorem.

Theorem 1: For a $\mathcal{C}_2$-index coding problem $\mathcal{P}_2$, construct a directed graph $G_{cl}(\mathcal{P}_2)$, as mentioned in Section I-B. Then, $\ell^\star(\mathcal{P}_2) = \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ if all c_i's in $G_{cl}(\mathcal{P}_2)$ have size less than or equal to 2 and if linear index coding is optimal for $\mathcal{P}_1(G_{cl}(\mathcal{P}_2))$ constructed from $\mathcal{P}_2$.

Note that Theorem 1 is a refined result of the result $\ell^\star(\mathcal{P}_2) \geq \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ in [6]. To prove Theorem 1, we introduce the following lemma.

Lemma 1: Consider a $\mathcal{C}_2$-index coding problem $\mathcal{I}_1$ with the restriction that each datum x_i is required by only one or two receivers. Without loss of generality, $\mathcal{I}_1$ is defined as follows.

1) Data set $X = \{x_1, \cdots, x_n\}$.
2) Receiver set $R = \{r_1, \cdots, r_m\}$.
3) Required data $f(i) = \begin{cases} i, & \text{for } 1 \leq i \leq n, \\ 1, & \text{for } i = n+1, \\ j_i(\neq 1), & \text{otherwise.} \end{cases}$
4) Side information index set $N = \{N(1), \cdots, N(m)\}$.

Define another index coding problem $\mathcal{I}_2$ from $\mathcal{I}_1$.

1) Data set $X' = \{x_1, \cdots, x_n, x_{n+1}\}$.
2) Receiver set $R' = \{r'_1, \cdots, r'_m\}$.
3) Required data $f'(i) = \begin{cases} i, & \text{for } 1 \leq i \leq n, \\ n+1, & \text{for } i = n+1, \\ j_i(\neq 1), & \text{otherwise.} \end{cases}$
4) Side information index set $N' = \{N'(1), \cdots, N'(m)\}$,

where $N'(i)$ is defined as

$$N'(i) := \begin{cases} N(i) \cup \{n+1\}, & \text{if } 1 \in N(i) \text{ or } f'(i) = 1, \\ N(i) \cup \{1\}, & \text{if } f'(i) = n+1, \\ N(i), & \text{otherwise.} \end{cases}$$

(See Fig. 1 for an example.) Then, $\ell^\star(\mathcal{I}_1) = \ell^\star(\mathcal{I}_2)$ if there exists an optimal index code $E(X')$ for $\mathcal{I}_2$, that satisfies

$$H(x_j|E(X'), X'[N(i)]) = 0 \text{ or } 1, \quad \forall i, j \in [n]. \qquad (4)$$

Proof: See Appendix. ■

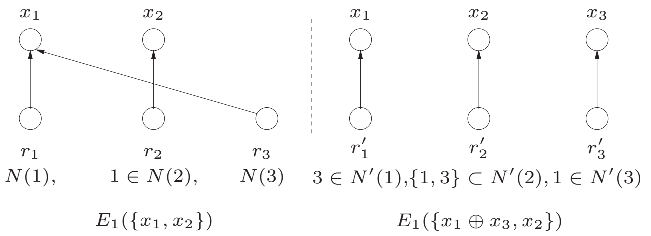

Fig. 1. An example of $\mathcal{I}_1$ and $\mathcal{I}_2$ for $n = 2$ and $m = 3$

Now, we prove proof of Theorem 1.

Proof of Theorem 1: First, reorder the indices of data and receivers of $\mathcal{P}_2$ to obtain $\mathcal{I}_1$. (This does not change the problem.) Then, obtain $\mathcal{I}_2$ according to the construction described in Lemma 1. If $\mathcal{I}_2$ is not a $\mathcal{C}_1$-index coding problem, obtain $\mathcal{I}_3$ from $\mathcal{I}_2$ again according to the construction described in Lemma 1 by setting $\mathcal{I}_2$ in the first construction as $\mathcal{I}_1$ in the second construction. Iterate this procedure until we have $\mathcal{I}_{m-n+1}$. One can easily see that the last $\mathcal{I}_{m-n+1}$ is the index-reordered version of $\mathcal{P}_1(G_{cl}(\mathcal{P}_2))$. By the linear optimality assumption in Theorem 1, the assumption (4) of Lemma 1 is valid to apply Lemma 1 to $\mathcal{I}_{m-n}$ and $\mathcal{I}_{m-n+1}$. By Lemma 1 applied to $\mathcal{I}_{m-n}$ and $\mathcal{I}_{m-n+1} = \mathcal{P}_1(G_{cl}(\mathcal{P}_2))$, $\ell^\star(\mathcal{I}_{m-n}) = \ell^\star(\mathcal{I}_{m-n+1})$ and by Lemma 2, there exist a linear optimal code for $\mathcal{I}_{m-n}$ with length $\ell^\star(\mathcal{I}_{m-n}) = \ell^\star(\mathcal{I}_{m-n+1})$ constructed from the assumed linear optimal code for $\mathcal{I}_{m-n+1}$ according to Remark 1. Now consider $\mathcal{I}_{m-n-1}$ and $\mathcal{I}_{m-n}$ and apply Lemma 1 again. Iterate this procedure until we have $\mathcal{P}_2 = \mathcal{I}_1$ and $\mathcal{I}_2$. Then, we have $\ell^\star(\mathcal{P}_2) = \ell^\star(\mathcal{I}_1) = \ell^\star(\mathcal{I}_2) = \cdots = \ell^\star(\mathcal{I}_{m-n+1}) = \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$. Furthermore, an optimal linear code $\mathcal{P}_2$ is constructed from an optimal linear code for $\mathcal{P}_1(G_{cl}(\mathcal{P}_2))$ through this process. ■

In [4], the capacity region of any index coding problem with $n = m \leq 5$ is obtained. By setting the rate for each receiver the same, their capacity region result yields the symmetric capacity, which is the inverse of the optimal broadcast rate β defined in Definition 4. The obtained symmetric capacity result in the case of $n = m \leq 5$ in [4] can be used to obtain the symmetric capacity in the case of $n < m \leq 5$, based on Theorem 1 and the result $\ell^\star(\mathcal{P}_2) \geq \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ by Lubetzky and Stav [6]. There exist 9846 non-isomorphic $\mathcal{C}_1$-index coding problems with $n = m \leq 5$ [4]. (The symmetric capacity of each of the 9846 problems is known due to [4].) By numerical study, we verified that for 9818 problems out of the total 9846 problems the symmetric capacity is achieved by simple scalar linear coding, whereas for the remaining 28 problems simple scalar linear coding is not sufficient. Among the total 9846 $\mathcal{C}_1$-problems, 3018 problems are $\mathcal{P}_1(G_{cl}(\mathcal{P}_2))$ for some $\mathcal{P}_2$ (we can identify this based on the properties of $G_{cl}(\mathcal{P}_2)$ described in Section I-B), that is, each of them is a $\mathcal{C}_1$-problem corresponding to a directed graph $G_{cl}(\mathcal{P}_2)$ generated from some $\mathcal{P}_2$ as described in Section I-B. Among the 28 $\mathcal{C}_1$-problems for which simple scalar linear coding is not sufficient, only 6 problems are $\mathcal{P}_1(G_{cl}(\mathcal{P}_2))$ for some $\mathcal{P}_2$. We verified that the symmetric capacity of each of these 6 $\mathcal{C}_1$-problems (known by [4]) is achieved by vector linear coding with length $t = 2$. For each of the 3018 $\mathcal{C}_1$-problems to which one or more $\mathcal{C}_2$-problem(s) map(s) through the G_{cl} operation, we traced back all the corresponding $\mathcal{C}_2$-problem(s) $\mathcal{P}_2$('s). In this way, we covered all $\mathcal{C}_2$-problems $\{\mathcal{P}_2\}$ with $n < m \leq 5$. (This paper has supplementary downloadable material available at http://wisrl.kaist.ac.kr/papers/Num14ISITSoKwakSung.zip provided by the authors.)

Case 1: If $G_{cl}(\mathcal{P}_2)$ have only size 2 cliques with $n = m \leq 5$, we have $\beta(\mathcal{P}_2) = \beta(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ by Theorem 1 since for any $\mathcal{P}_1$ with $n = m \leq 5$ linear index coding is sufficient.

Case 2: If $G_{cl}(\mathcal{P}_2)$ has c_i whose size is more than 2, $M(G_{cl}(\mathcal{P}_2))$ should have 3 or more columns that are the same after replacing 1 with $*$ (see Section I-B). In this case, we verified numerically that there exists a scalar linear index code for $\mathcal{P}_2$ that achieves $\beta(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$. (Again, $\beta(\mathcal{P}_1)$ is known by [4].)

Case 3: For $\mathcal{P}_2$ corresponding (by G_{cl}) to each of the 6 $\mathcal{C}_1$-problems for which vector linear coding with length 2 is optimal, we observed that each datum is required by one or two receivers. Hence, Theorem 1 applies and we have $\beta(\mathcal{P}_2) = \beta(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$.

Hence, every $\mathcal{C}_2$-index coding problem $\mathcal{P}_2$ with $n < m \leq 5$ satisfies $\beta(\mathcal{P}_2) = \beta(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$. Since $\ell^\star(\mathcal{P}_2) \geq \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ by Lubetzky and Stav [6], the symmetric capacity of $\mathcal{P}_2$ is $\beta(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$, when $n < m \leq 5$.

III. Counter Example

Up to now, we have seen that $\ell^\star(\mathcal{P}_2) = \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ in many cases. One might conjecture $\ell^\star(\mathcal{P}_2) = \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ in general, going beyond $\ell^\star(\mathcal{P}_2) \geq \ell^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ [6]. Although we could not come up with a counter example for this, we have a counter example with $\ell_L^\star(\mathcal{P}_2) > \ell_L^\star(\mathcal{P}_1(G_{cl}(\mathcal{P}_2)))$ in the case of linear index coding, which is provided in the below.

Consider a $\mathcal{C}_2$-index coding problem $\mathcal{I}_c$ with $n = 4 < m = 6$:

1) Data set $X = \{x_1, x_2, x_3, x_4\}$.
2) Receiver set $R = \{r_1, r_2, \cdots, r_6\}$.
3) Required data $f(1) = f(2) = f(3) = 1$, $f(4) = 2$, $f(5) = 3$ and $f(6) = 1$.
4) Side information $N(1) = \{2,3\}$, $N(2) = \{2,4\}$, $N(3) = \{3,4\}$, $N(4) = \{1,4\}$, $N(5) = \{1,2\}$ and $N(6) = \{1,3\}$.

The matrix representation of $\mathcal{I}_c$ and $\mathcal{P}_1(G_{cl}(\mathcal{I}_c))$ are given respectively by

$$M(\mathcal{I}_c) = \begin{bmatrix} 1 & * & * & 0 \\ 1 & * & 0 & * \\ 1 & 0 & * & * \\ * & 1 & 0 & * \\ * & * & 1 & 0 \\ * & 0 & * & 1 \end{bmatrix}, \; M(G_{cl}(\mathcal{I}_c)) = \begin{bmatrix} 1 & * & * & * & * & 0 \\ * & 1 & * & * & 0 & * \\ * & * & 1 & 0 & * & * \\ * & * & * & 1 & 0 & * \\ * & * & * & * & 1 & 0 \\ * & * & * & 0 & * & 1 \end{bmatrix}, \tag{5}$$

where $\min_{A \in M(\mathcal{I}_c)} \text{rank}_2(A) = 3$ and $\min_{A \in M(G_{cl}(\mathcal{I}_c))} \text{rank}_2(A) = 2$. A matrix achieving $\min_{A \in M(G_{cl}(\mathcal{I}_c))} \text{rank}_2(A)$ is given by

$$\underset{A \in M(G_{cl}(\mathcal{I}_c))}{\arg\min} \text{rank}_2(A) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}. \tag{6}$$

Therefore, one can see that $\ell_L^\star(\mathcal{I}_c) > \ell_L^\star(\mathcal{P}_1(G_{cl}(\mathcal{I}_c)))$.

IV. Conclusion

In this paper, we have considered the index coding problem when the number of data is less than the number of receivers, and have shown that in certain cases, the symmetric capacity and code construction for index coding problems in the considered case can be obtained from the existing symmetric capacity result and codes for index coding problems when the number of data is the same as that of receivers. Such cases include cases with up to five receivers for which the symmetric capacity is already known when the number of data is the same as that of receivers.

Appendix: Proof of Lemma 1

The index coding problem $\mathcal{I}_2$ in Lemma 1 is constructed by adding x_{n+1} to $\mathcal{I}_1$ in Lemma 1. We can easily see that

$$\ell^\star(\mathcal{I}_1) \geq \ell^\star(\mathcal{I}_2), \tag{7}$$

since every index code $E_1(x_1, \cdots, x_n)$ for $\mathcal{I}_1$ is a valid index code for $\mathcal{I}_2$ if x_1 is replaced with $x_1 \oplus x_{n+1}$, i.e., $E_1(x_1 \oplus x_{n+1}, x_2, \cdots, x_n)$, where $\oplus$ indicates modulo-2 addition. (r_1 can decode $x_1 \oplus x_{n+1}$ and thus, so can r_1' but r_1' has additional side information x_{n+1}. Similarly, r_{n+1} can decode $x_1 \oplus x_{n+1}$ and thus, so can r_{n+1}' but r_{n+1}' has additional side information x_1. See Fig. 1.) Hence, to prove Lemma 1, we need to show that $\ell^\star(\mathcal{I}_1) \leq \ell^\star(\mathcal{I}_2)$ under the assumption. Since $\ell^\star(\mathcal{I}_2)$ is the minimum $\mathcal{I}_2$ code length, there exists an $\mathcal{I}_2$-code $E^\star(X')$ with length $\ell^\star(\mathcal{I}_2)$ and suppose that $E^\star(X')$

satisfies the assumption (4). Then, by Lemma 2, there exists an $\mathcal{I}_1$-code with length $\ell^\star(\mathcal{I}_2)$. Consequently, we have

$$\ell^\star(\mathcal{I}_1) \leq \ell^\star(\mathcal{I}_2). \tag{8}$$

By (7) and (8), we have $\ell^\star(\mathcal{I}_1) = \ell^\star(\mathcal{I}_2)$. ■

Lemma 2: There exists an $\mathcal{I}_1$-code with length $\bar{\ell}$ if there exists an $\mathcal{I}_2$-code* $E_2(X')$ with length $\bar{\ell}$ satisfying

$$H(x_j | E_2(X'), X'[N(i)]) = 0 \text{ or } 1, \quad \forall i, j. \tag{9}$$

Proof: Let $E_2(X')$ be an $\mathcal{I}_2$-code with length $\bar{\ell}$: $E_2 : (0,1)^{n+1} \to (0,1)^{\bar{\ell}}$, satisfying (9). Under the assumption (9), we only need to consider the following cases:
Case 1: Suppose that $H(x_1|E_2(X'), X'[N(1)]) = H(x_{n+1}| E_2(X'), X'[N(n+1)]) = 0$. Then, receiver r'_1 can decode x_1 without knowing x_{n+1}, since $n+1 \notin N(1)$, and receiver r'_{n+1} can decode x_{n+1} without knowing x_1, since $1 \notin N(n+1)$. Consequently, $E_2(\{x_1, \cdots, x_n, x_1\})$ is a valid index code for $\mathcal{I}_1$ by the construction of $\mathcal{I}_2$ from $\mathcal{I}_1$.
Case 2: Suppose that at least one of the following equations holds:

$$H(x_1|E_2(X'), X'[N(1)]) = 1, \tag{10}$$

$$H(x_{n+1}|E_2(X'), X'[N(n+1)]) = 1. \tag{11}$$

Case 2-1: Assume that (10) holds. Applying the chain rule in two different orders, we have

$$\begin{aligned} &H(x_1, x_{n+1}|E_2(X'), X'[N(1)]) \\ &= H(x_{n+1}|E_2(X'), X'[N(1)]) + H(x_1|E_2(X'), X'[N'(1)]), \quad (12) \\ &= H(x_1|E_2(X'), X'[N(1)]) + H(x_{n+1}|E_2(X'), X'[N(1)], x_1), \quad (13) \end{aligned}$$

where (12) is because $[X'[N(1)], x_{n+1}] = X'[N'(1)]$. Since $E_2(X')$ is a valid index code for $\mathcal{I}_2$, $H(x_1| E_2(X'), X'[N'(1)]) = 0$, which is the second term in (12). This implies that (12) ≤ 1 and (13) ≤ 1, because the entropy of a binary variable is one at most and (12) = (13). Since $H(x_1| E_2(X'), X'[N(1)]) = 1$ in (13) by assumption, $H(x_{n+1}| E_2(X'), X'[N(1)], x_1) = 0$ because (13) ≤ 1. Therefore, r'_1 can decode x_{n+1} if r'_1 does not know x_{n+1} but knows x_1. This means that r_1 can decode x_{n+1} if r_1 knows x_1, when $E_2(X')$ is used for $\mathcal{I}_1$. Furthermore, by the construction of $\mathcal{I}_2$ from $\mathcal{I}_1$, $N(n+1)$ for r_{n+1} is $N(n+1) = N'(n+1)\backslash\{1\}$. Thus, r_{n+1} can decode x_{n+1} if r_{n+1} knows x_1 additionally, when $E_2(X')$ is used for $\mathcal{I}_1$. Consequently, $E_2(\{0, x_2, \cdots, x_n, x_1\})$ is a valid index code for $\mathcal{I}_1$. Here, constant 0 is known to all receivers beforehand.

Case 2-2: Assume that (11) holds. By applying the chain rule to $H(x_1, x_{n+1}|E_2(X'), X'[N(n+1)])$ and using similar techniques to the above, we can show that $E_2(\{x_1, x_2, \cdots, x_n, 0\})$ is a valid index code for $\mathcal{I}_1$. ■

Remark 1: (*Construction of $\mathcal{I}_1$-codes from $\mathcal{I}_2$-codes when a datum is required by at most two receivers):* Under the assumption (9), we have either (a) $H(x_1|E_2(X'), X'[N(1)]) = H(x_{n+1}|E_2(X'), X'[N(n+1)]) = 0$ or (b) at least one of the following equations holds:

$$\text{(b1)} \quad H(x_1|E_2(X'), X'[N(1)]) = 1, \tag{14}$$

$$\text{(b2)} \quad H(x_{n+1}|E_2(X'), X'[N(n+1)]) = 1. \tag{15}$$

Given an index code E_2 for $\mathcal{I}_2$, check the conditions (a), (b1) and (b2). Use $E_2(\{x_1, \cdots, x_n, x_1\})$, $E_2(\{0, x_2, \cdots, x_n, x_1\})$, or $E_2(\{x_1, x_2, \cdots, x_n, 0\})$ in the cases of (a), (b1), or (b2), respectively. Then, the selected code is a valid code for $\mathcal{I}_1$.

*Decodability at receivers 1 and $n+1$ implies $H(x_1|E_2(X'), X'[N'(1)]) = H(x_{n+1} |E_2(X'), X'[N'(n+1)]) = 0$.

References

[1] Y. Birk and T. Kol, "Coding on demand by an informed source (ISCOD) for efficient broadcast of different supplemental data to caching clients," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2825–2830, 2006.

[2] S. Jafar, "Wireless index coding," in *Global Communications Conference (GLOBECOM), 2012 IEEE*, pp. 2334–2339, 2012.

[3] Z. Bar-Yossef, Y. Birk, T. S. Jayram, and T. Kol, "Index coding with side information," *IEEE Trans. Inform. Theory*, vol. 57, no. 3, pp. 1479–1494, 2011.

[4] F. Arbabjolfaei, B. Bandemer, Y.-H. Kim, E. Sasoglu, and L. Wang, "On the capacity region for index coding," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, pp. 962–966, 2013.

[5] K. Shanmugam, A. G. Dimakis, and M. Langberg, "Local graph coloring and index coding," *ArXiv pre-print cs.IT/1301.5359*, Feb. 2013.

[6] E. Lubetzky and U. Stav, "Nonlinear index coding outperforming the linear optimum," *IEEE Trans. Inform. Theory*, vol. 55, no. 8, pp. 3544–3551, 2009.

[7] A. S. Tehrani, A. G. Dimakis, and M. J. Neely, "Bipartite index coding," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pp. 2246–2250, 2012.

[8] N. Alon, E. Lubetzky, U. Stav, A. Weinstein, and A. Hassidim, "Broadcasting with side information," in *Foundations of Computer Science, 2008. FOCS '08. IEEE 49th Annual IEEE Symposium on*, pp. 823–832, 2008.