Lower Bounds for Interactive Function Computation via Wyner Common Information

A Project Report

submitted by

SHIJIN RAJAKRISHNAN

in partial fulfilment of the requirements
for the award of the degree of

BACHELOR OF TECHNOLOGY

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY MADRAS.

June 2016

THESIS CERTIFICATE

This is to certify that the thesis titled **Lower Bounds for Interactive Function Computation via Wyner Common Information**, submitted by **Shijin Rajakrishnan**, to the Indian Institute of Technology, Madras, for the award of the degree of **Bachelor of Technology**, is a bona fide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Andrew Thangaraj
Research Guide
Professor
Dept. of Electrical Engineering
IIT-Madras, 600 036

Dr. Vinod Prabhakaran
Research Guide
Reader
School of Technology and Computer Science
TIFR, 400 005

Place: Chennai

Date: June 2016

ABSTRACT

KEYWORDS: Communication Complexity, Information Complexity, Wyner Common Information

The question of how much communication is required between collaborating parties to compute a function of their data is of fundamental importance in the fields of theoretical computer science and information theory. In this work, the focus is on coming up with lower bounds on this. The information cost of a protocol is the amount of information the protocol reveals to Alice and Bob about each others inputs, and the information complexity of a function is the infimum of information costs over all valid protocols. For the amortized case, it is known that the optimal rate for the computation is equal to the information complexity. Exactly computing this information complexity is not straight forward however. In this work we lower bound information complexity for independent inputs in terms of the Wyner common information of a certain pair of random variables. We show a structural property for the optimal auxiliary random variable of Wyner common information and exploit this to exactly compute the Wyner common information in certain cases. The lower bound obtained through this technique is shown to be tight for a non-trivial example - equality (EQ) for the ternary alphabet. We also give an example to show that the lower bound may, in general, not be tight.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

r.v.	random variable
s.t.	such that
i.i.d	independently and identically distributed
IC	Information Complexity
Wyn	Wyner
CI	Common Information

NOTATION

X	Input to Alice
Y	Input to Bob
μ	Joint distribution of Alice and Bob's inputs
$\mathcal{X}$	Cardinality of the r.v. X
$\mathcal{Y}$	Cardinality of the r.v. Y
IC	Information Complexity
f	Function to be computed by the interacting parties
Z	Random variable denoting output of the function computed
Π	Transcript of a protocol

CHAPTER 1

Introduction and Background

Communication Complexity aims to characterize the complexity of function computation by estimating the amount of messages that must be exchanged between communicating parties to jointly compute the function, whose inputs are broken up amongst the players so that each player only has partial knowledge of the input. The notion of Communication Complexity was introduced in [14], where the problem of the minimum amount of communication required between two parties, Alice and Bob, who wanted to compute a function $f(x, y)$ that depended on their inputs x and y, respectively, was studied. Over 35 years later, the simplistic model of communication established by Yao has proven to be quite powerful and the development of the field has seen the production of an immense number of interesting upper and lower bounds on the complexity of several diverse communication problems, and in addition, these bounds transfer easily to other computation models. This problem is relevant in several real world applications, such as in VLSI design(where one wants to minimize the amount of energy used by decreasing the amount of electic signals required between the different components), in the study of data structures, in proving space-time tradeoffs for Turing machines, in proving circuit size and depth lower bounds, and in the optimization of computer networks. A more detailed survey of the topic is presented in [7].

The general communication model established by Yao admits any number of finite players, however, the field is split on looking at two-party communication bounds and multi-party communication bounds, and in this thesis, we confine ourselves to the former model. Most of the results in this domain are based on two general purpose methods for proving lower bounds on communication: rectangle based methods and information based methods. Early progress was made using the former, which leads to a combinatorial method of tackling the problem [7], while more recent advances in the area have centered around the notion of information complexity, which measures the amount of information learned by the parties about each other's inputs from the protocol's transcript, rather than a direct estimate of the number of bits required in the

communication between the parties to correctly compute the function. More specifically, if the inputs, X and Y of the two parties comes from a joint distribution μ, then the information cost of a protocol Π (for computing a function f) whose transcript is denoted by M is defined as

$$I(X; M|Y) + I(Y; M|X).$$

Information complexity is the infimum of information costs of valid protocols, i.e., protocols which allow the parties to compute within the desired error performance, and is denoted by $IC_{XY}(Z)$ for the computation of a function $Z = f(X, Y)$.

This quantity has a close connection to the problem of interactive source coding and interactive function computation studied in information theory literature. In particular, works by Kaspi [6] and Ma and Ishwar [8] show that information complexity for zero-error is precisely the rate of communication required to compute with asymptotically vanishing error when the parties are allowed to code over long blocks of independent, identically distributed inputs. While, in general, computing information complexity is not straightforward, it is known exactly for some interesting examples [8] and an algorithm, albeit with run-time exponential in the alphabet size, for approximating it has been proposed [3].

In [9], with the goal of better understanding information complexity, a monotonicity property of interactive protocols was leveraged to obtain lower bounds on the information complexity. The monotonicity property is that of the "tension region" of the views of the two users. Tension region of a pair of random variables was introduced in [10] as a measure of dependence which cannot be captured using a common random variable. The question of how well correlation *can* be captured by a random variable may be formulated in terms of "common information." Two different notions of common information were developed in the 70's, $CI_{\mathsf{GK}}(A; B)$ by Gács-Körner [5], and $CI_{\mathsf{Wyn}}(A; B)$ by Wyner [13].

$$CI_{\mathsf{GK}}(A; B) = \max_{\substack{p_{Q|A,B}: \\ Q-A-B \\ Q-B-A}} I(Q; A, B) \tag{1.1}$$

$$CI_{\mathsf{Wyn}}(A; B) = \min_{\substack{p_{Q|A,B}: \\ A-Q-B}} I(Q; A, B) \tag{1.2}$$

One can define corresponding notions of tension as the gap between mutual information (which accounts for all the correlation, but may not correspond to a common random variable) and common information. More precisely, one can define the non-negative tension quantities $T_{\mathsf{GK}}(A;B) = I(A;B) - CI_{\mathsf{GK}}(A;B)$ and $T_{\mathsf{Wyn}}(A;B) = CI_{\mathsf{Wyn}}(A;B) - I(A;B)$. These notions of tension were identified in [10] as special cases of a unified 3-dimensional notion of *tension region*.

The tension region of a pair of random variables was defined in [10] as the following upward closed region.

Definition 1. *For a pair of random variables A, B, their* tension region $\mathfrak{T}(A;B)$ *is defined as*

$$\mathfrak{T}(A;B) = \{(r_1, r_2, r_3) : \exists Q \textit{ jointly distr. with } A, B$$
$$\textit{s.t. } I(B;Q|A) \leq r_1, I(A;Q|B) \leq r_2, I(A;B|Q) \leq r_3\}.$$

As shown in [10], without loss of generality, we may assume a cardinality bound $|\mathcal{Q}| \leq |\mathcal{A}||\mathcal{B}| + 2$ on the alphabet $\mathcal{Q}$ in the above definition, where $\mathcal{A}$ and $\mathcal{B}$ are the alphabets of A and B, respectively.

In [10], an operational meaning was also obtained for tension region in terms of a generalization of the common information problem of Gács and Körner. Tension region has proved useful in deriving converse results for secure computation. Specifically, it was used to strictly improve upon an upper bound of Ahlswede and Csiszár [1] on the oblivious transfer capacity of channels [11].

Suppose X,Y are the inputs and A,B the outputs of the parties under a protocol. Let M denote the transcript of the protocol. Let $V_A = (X, A, M)$ and $V_B = (Y, B, M)$ denote the views of the parties at the end of the protocol. The key monotonicity property we use is:

Proposition 1 (Theorem 5.4 of [10])**.**

$$\mathfrak{T}(V_A; V_B) \supseteq \mathfrak{T}(X;Y).$$

A consequence of this is the following result:

Theorem 1. *For all X, Y, Z,*

$$IC_{XY}(Z) \geq T_{\mathsf{Wyn}}(XZ; YZ) - T_{\mathsf{Wyn}}(X; Y) + I(X; Z|Y) + I(Y; Z|X).$$

See [9] for a more general result which implies the above lower bound. For the case of independent inputs the $T_{\mathsf{Wyn}}(X; Y)$ term goes to zero. We will give a proof of Theorem 1 for the case of independent inputs in Appendix A. While, the above bound is not always tight[2], we present a non-trivial example where the bound turns out to give a tight result. It is worth noting that the technique of [8] does not easily yield this result.

Example 1 (ternary EQ)**.** *Let X, Y be independent and uniformly distributed over $\{0, 1, 2\}$. The goal is to compute the indicator for the event $(X = Y)$. Theorem 1 gives a lower bound of $H_2\left(\frac{2}{3}\right) + \log_2(3)$ which can be shown to be tight.*

The equality (EQ) function, which determines whether two parties have the same inputs, has been studied extensively. To the best of our knowledge, the only lower bound on information complexity available is the trivial $IC_{XY}(Z) \geq I(X; Z|Y) + I(Y; Z|X)$. The best available upper bound is 4.5 for k-ary EQ computation, for any probability distribution over the inputs [2]. In this paper, we obtain both lower bounds and upper bounds on the information complexity of the EQ function for uniformly distributed inputs. To evaluate our lower bound of Theorem 1, we need to compute Wyner common information (or an equivalent quantity given in (2.4)). Note that computing Wyner common information is, in general, not straightforward [12]. Using standard techniques based on Carathéodory's theorem, an upper bound of $|\mathcal{Q}| \leq |\mathcal{A}| \times |\mathcal{B}| + 2$ on the auxiliary random variable Q of (1.2) is available. We show that it is enough to consider a potentially smaller cardinality for $\mathcal{Q}$ which depends on the number of maximal cliques of the bipartite characteristic graph of $p_{A,B}$ – this is the bipartite graph on $\mathcal{A} \times \mathcal{B}$ such that there is an edge between $a \in \mathcal{A}$ and $b \in \mathcal{B}$ if $p_{A,B}(a, b) > 0$ – such that conditioned on each element of $q \in \mathcal{Q}$, the characteristic graph of $p_{A,B|Q=q}$ is a distinct clique (Theorem 3). This then allows us to compute Wyner common information exactly for certain examples of interest (chapter 3). In particular, the resulting lower bound turns out to be tight for the ternary EQ example above. We also give a randomized protocol

[2] An example where this bound turns out not to be tight is that of computing the AND of two independent uniform bits X, Y, for which information complexity is known [8].

for the 4-ary EQ problem which performs better than deterministic protocols in terms of its information cost, but here our lower bound does not meet the upper bound given by the protocol.

Before delving into the material, we first state a couple of definitions used throughout the text.

Definition 2. *Consider a pair of random variables A and B, jointly distributed according to μ, with alphabets $\mathcal{A}$ and $\mathcal{B}$ respectively. The* characteristic graph *between A and B is the bipartite graph $G = (\mathcal{A} \cup \mathcal{B}, E)$, where the set of edges E is defined as $E = \{(a, b) \mid a \in \mathcal{A}, b \in \mathcal{B}, \mu(a, b) > 0\}$.*

Definition 3. *Given k numbers, $\{i_1, ..., i_k\}$, we define $H_k(i_1, ..., i_k) = \sum_{j=1}^{k} i_j \log\left(\frac{1}{i_j}\right)$*

Definition 4. *The function $\mathbb{1}$ takes as input a closed boolean formula φ and is defined as*

$$\mathbb{1}[\varphi] = \begin{cases} 1 & \text{if } \varphi \text{ is true} \\ 0 & \text{if } \varphi \text{ is false} \end{cases}$$

And finally, we look at the *Log-sum inequality*, which states that given nonnegative numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$,

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq a \log \frac{a}{b}$$

where $a = \sum_{i=1}^{n} a_i$ and $b = \sum_{i=1}^{n} b_i$.

CHAPTER 2

Interactive Function Computation

2.1 Introduction and Problem definition

In the two party interactive function computation model, there are two players - Alice and Bob, with unlimited computational power and access to private random strings - who want to compute a function that depends on both of their inputs. Alice is given an input X and Bob is given an input Y, drawn from a joint distribution μ over $X \times Y$, and they want to compute a function $f : \mathcal{X} \times \mathcal{Y} \mapsto \mathcal{Z}$, where $\mathcal{X}$ is the alphabet over which the random variable X is defined, and likewise $\mathcal{Y}$ is the alphabet of the r.v. Y. The range of the function is $\mathcal{Z}$, and given a distribution on X and Y, we can naturally define a random variable $Z = f(X, Y)$ whose alphabet is $\mathcal{Z}$ and whose distribution depends only on the function f and the distribution μ. Now given inputs $(x, y) \in \mathcal{X} \times \mathcal{Y}$, the objective of Alice and Bob is to each compute the value of $f(x, y)$ exactly. Naturally, for functions that non-trivially depend on both the inputs, this engenders the need for communication between the players. The model of the communication between the players is a 2-way, zero error channel connecting the two players, in which the communication takes place sequentially - first Alice uses the channel to communicate information to Bob, following which Bob uses the channel, and then Alice uses it again and so on until the end of the communication when both have enough information to correctly compute the value of the function f.

The players proceed according to a *protocol* Π, which describes a sequence of steps each player takes based on the input he/she has received and the messages received on the communication link upto that point. Let the messages sent on the link when the players follow a protocol Π be $M_{\Pi,0}, M_{\Pi,1}, ..., M_{\Pi,t}$, where t is the length of the protocol(number of steps), and as defined by the communication model above, the messages indexed by even numbers are sent by Alice and the odd ones by Bob. Define the concatenation of all the messages sent upto the i$^{\text{th}}$ turn, the *transcript*,

$M_\Pi^i = (M_{\Pi,0}, M_{\Pi,1}, ..., M_{\Pi,i-1})$. Also define $M_\Pi = M_\Pi^t$. Note that all the individual messages and hence the transcript upto any stage are random variables that depend on μ and the protocol. Now a protocol Π for computing a function f is said to be *valid* if at the end of the protocol, both Alice and Bob have correctly computed the function value. In other words, $H(Z|X, M_\Pi) = 0$, and $H(Z|Y, M_\Pi) = 0$.

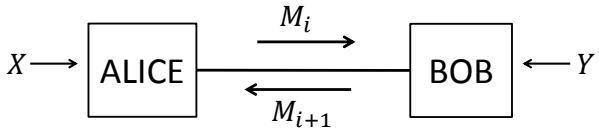

Figure 2.1: The model for the two-party computation

Now Alice initially does not know Bob's input, except for the (possibly) partial information she might have due to her own input X and the distribution μ, and thus conditioned on her input, the message she sends, $M_{\Pi,0}$ cannot reveal/depend on any new information about Y. However, $M_{\Pi,2}$ might contain some new information about Y, as the message sent by Bob, $M_{\Pi,1}$ might reveal something about his input. However this is the only source of Alice's knowledge of Bob's input and hence conditioned on the transcript so far, Alice's next message cannot depend on any new information about Bob that cannot be gotten from X and the transcript thus far. In other words,

$$\forall \text{ even } i, \qquad M_{\Pi,i} - X, M_\Pi^{i-1} - Y \tag{2.1}$$

Using a similar argument about Bob's source of knowledge of Alice's input, it is easy to see that

$$\forall \text{ odd } i, \qquad M_{\Pi,i} - Y, M_\Pi^{i-1} - X \tag{2.2}$$

The entropy of the final transcript $H(M_\Pi)$ is a lower bound for the average number of bits needed for the protocol Π. Further, the *Information Complexity* of the function computed by Π is a lower bound on $H(M_\Pi)$. Before proving this, we look at an important lemma that is useful for it.

Lemma 1. *For any valid protocol* Π*,*

$$I(X;Y) \geq I(X;Y|M_\Pi)$$

Proof. Consider the i^{th} message in the protocol defined by Π. Suppose Alice is sending the message $M_{\Pi,i}$ on the link, and so i is even.

$$\begin{aligned} I(X;Y|M_\Pi^{i-1}) - I(X;Y|M_\Pi^i) &= H(Y|M_\Pi^{i-1}) - H(Y|X,M_\Pi^{i-1}) \\ &\quad - H(Y|M_\Pi^i) + H(Y|X,M_\Pi^i) \\ &\overset{(a)}{=} H(Y|M_\Pi^{i-1}) - H(Y|M_\Pi^i) \\ &\geq 0 \end{aligned}$$

where (a) follows from the fact that $H(Y|X,M_\Pi^{i-1}) = H(Y|X,M_\Pi^i)$, since $M_{\Pi,i} - X, M_\Pi^{i-1} - Y$ forms a markov chain. We get a similar result for the message sent by Bob as well.
Thus, $I(X;Y|M_\Pi^{i-1}) \geq I(X;Y|M_\Pi^i)$, and by repeatedly using this for the length of the protocol, we get $I(X;Y) \geq I(X;Y|M_\Pi)$. □

Now we use this lemma to prove that the information complexity of a function $Z = f(X,Y)$ is a lowerbound on the entropy, $H(M_\Pi)$, of the transcript observed using a protocol Π computing the function.

Lemma 2. *Consider a function $f : \mathcal{X} \times \mathcal{Y} \mapsto \mathcal{Z}$, and a two player interactive protocol Π which computes the function. If $Z = f(X,Y)$, then*

$$H(M_\Pi) \geq \mathsf{IC}_{XY}(Z)$$

Proof.

$$\begin{aligned} H(M_\Pi) &\geq I(M_\Pi; XY) \\ &= I(X;M_\Pi) + I(Y;M_\Pi|X) \\ &= I(X;Y,M_\Pi) - I(X;Y|M_\Pi) + I(Y;M_\Pi|X) \\ &= I(X;Y) + I(X;M_\Pi|Y) - I(X;Y|M_\Pi) + I(Y;M_\Pi|Y) \\ &\overset{(a)}{\geq} I(X;M_\Pi|Y) + I(Y;M_\Pi|X) \geq \mathsf{IC}_{XY}(Z) \end{aligned}$$

where (a) uses Lemma 1. □

Now, in the *amortized* case, when we consider a block of independent identically distributed inputs of length n and a sequence of schemes, one for each block length

n, the following theorem, proved in [8],[4], gives the minimum rate of communication needed to compute a function with a vanishing probability of block error. The rate R of a scheme is defined as the total number of bits exchanged divided by the block length. A rate R is said to be *achievable* if there is a sequence of schemes whose probability of error goes to 0 as $n \to \infty$. The optimal rate R^* is the infimum of all achievable rates.

Theorem 2. *The optimal amortized rate R^* for computing the function $Z = f(X, Y)$ is*

$$R^* = \inf_M [I(X; M|Y) + I(Y; M|X)] = IC_{XY}(Z) \tag{2.3}$$

where the infimum is over all $M = (M_1, M_2, ...)$ satisfying the Markov chain conditions in Equation 2.1 and Equation 2.2, and $H(Z|Y, M) = H(Z|X, M) = 0$.

2.2 Wyner Common Information as a Lower Bound

Now we look at the main result of the report, which is to bound the Information Complexity of a function using the Wyner Common Information between a pair of variables. Recall the definition of the Wyner Common Information and Wyner Tension between two r.v.s

$$\mathsf{CI}_{\text{Wyn}}(A; B) = \min_{\substack{p_{Q|A,B}: \\ A-Q-B}} I(Q; A, B)$$

$$\mathsf{T}_{\text{Wyn}}(A; B) = \mathsf{CI}_{\text{Wyn}}(A; B) - I(A; B)$$

We can rewrite the Wyner Tension as

$$\begin{aligned}
\mathsf{T}_{\text{Wyn}}(A;B) &= \mathsf{CI}_{\text{Wyn}}(A;B) - I(A;B) \\
&= \min_{\substack{p_{Q|A,B}: \\ A-Q-B}} I(Q;A,B) - I(A;B) \\
&= \min_{\substack{p_{Q|A,B}: \\ A-Q-B}} [I(Q;A,B) - I(A;B)] \\
&= \min_{\substack{p_{Q|A,B}: \\ A-Q-B}} [I(A;Q) + I(B;Q|A) - I(A;B)] \\
&\overset{(a)}{=} \min_{\substack{p_{Q|A,B}: \\ A-Q-B}} [I(A;Q) + I(A;B|Q) + I(B;Q|A) - I(A;B)] \\
&= \min_{\substack{p_{Q|A,B}: \\ A-Q-B}} [I(A;B,Q) + I(B;Q|A) - I(A;B)] \\
&= \min_{\substack{p_{Q|A,B}: \\ A-Q-B}} [I(A;Q|B) + I(B;Q|A)]
\end{aligned}$$

where (a) follows from the fact that $A - Q - B$ is a markov chain and therefore $I(A;B|Q) = 0$.

Now consider the case where X and Y are independent random variables. Then, from Theorem 1(or rather, the version of it for independent inputs, as proved in the

Appendix), we can see that

$$
\begin{aligned}
\mathsf{IC}_{XY}(Z) &\geq \mathsf{T}_{\text{Wyn}}(X,Z;Y,Z) + I(X;Z|Y) + I(Y;Z|X)) \\
&= \inf_{\substack{p_{Q|X,Y,Z}: \\ XZ-Q-YZ}} [I(XZ;Q|YZ) + I(YZ;Q|XZ)] + I(X;Z|Y) + I(Y;Z|X) \\
&= \inf_{\substack{p_{Q|X,Y,Z}: \\ XZ-Q-YZ}} [I(XZ;Q|YZ) + I(YZ;Q|XZ) + I(X;Z|Y) + I(Y;Z|X)] \\
&= \inf_{\substack{p_{Q|X,Y,Z}: \\ XZ-Q-YZ}} [I(X;Q|YZ) + I(Y;Q|XZ) + I(X;Z|Y) + I(Y;Z|X)] \\
&= \inf_{\substack{p_{Q|X,Y,Z}: \\ XZ-Q-YZ}} [I(X;Q,Z|Y) + I(Y;Q,Z|X)] \\
&= H(X|Y) + H(Y|X) + \inf_{\substack{p_{Q|X,Y,Z}: \\ XZ-Q-YZ}} [-H(X|Q,Y,Z) - H(Y|Q,X,Z)] \\
&= H(X|Y) + H(Y|X) + \inf_{\substack{p_{Q|X,Y,Z}: \\ XZ-Q-YZ}} [-H(XZ|Q,YZ) - H(YZ|Q,XZ)] \\
&\overset{(a)}{=} H(X|Y) + H(Y|X) + \inf_{\substack{p_{Q|X,Y,Z}: \\ XZ-Q-YZ}} [-H(X,Z|Q) - H(Y,Z|Q)] \\
&= H(X|Y) + H(Y|X) - \sup_{\substack{p_{Q|X,Y,Z}: \\ XZ-Q-YZ}} [H(X,Z|Q) + H(Y,Z|Q)] \\
&= H(X|Y) + H(Y|X) - \sup_{\substack{p_{Q|U,V}: \\ U-Q-V}} [H(U|Q) + H(V|Q)] \qquad (2.4)
\end{aligned}
$$

where (a) follows from the markov chain $X,Z - Q - Y,Z$ and thus $H(X,Z|Q) = H(X,Z|Q,Y,Z)$. We introduce new random variables, $U = X,Z$ and $V = Y,Z$ as well for ease of notation.

Now the problem reduces to computing the supremum term in the above lower bound, $\sup_{\substack{p_{Q|U,V}: \\ U-Q-V}} [H(U|Q) + H(V|Q)]$, where the auxillary r.v. Q is such that, given Q, the random variables U and V are independent. Suppose the r.v. Q has the alphabet $\mathcal{Q}$, and let q be one element from it. What this means is that if $Q = q$, then the characteristic graph of $U,V|Q = q$ should be a bipartite clique, for U and V to be conditionally independent. Thus we shift our focus to only the possible induced bipartite cliques in the characteristic graph of U and V, and group the elements of $\mathcal{Q}$ according to the clique that they conditionally induce.

Since the alphabet of X,Y, and Z are finite, U and V have a finite set of vertices in the characteristic graph and thus the number of possible bipartite cliques is also finite. Hence we end up with a finite number of groups, each group characterized by the same

graph structure they induce on the characteristic graph of U and V, but having potentially different distributions on the edges.

Further, we restrict our attention only to maximal bipartite cliques, since a non-maximal clique is just a special case of a maximal clique with some edges taking zero probability.

This is just a characterization of the various elements of $\mathcal{Q}$, but however we haven't yet seen any way to narrow down the search space. To this end, we prove that, in the search for the optimal alphabet, $\mathcal{Q}_{\text{opt}}$, we need only consider one from each of the classes defined above. With this reduction in the alphabet size, we can find the optimal distribution Q as well.

Theorem 3. *For any distribution of random variables U and V, there is a r.v Q_{opt}, s.t $U - Q_{opt} - V$ and that maximizes $H(U|Q) + H(V|Q)$, such that atmost one element q from the alphabet, $\mathcal{Q}_{opt}$, induces any given maximal bipartite cliques on the characteristic graph of U and V.*

Proof. Consider an structure of the induced characteristic graph, say a (k, l)-bipartite clique, as shown in Figure 2.2. Now consider two elements $q_0, q_0' \in \mathcal{Q}$, where Q is a r.v. which satisfies $U - Q - V$, with alphabet $\mathcal{Q}$. The edge distribution on these two cliques, $p_{Q,U,V}(\cdot,\cdot,\cdot)$ are given by the vectors $\mathbf{p}$ and $\mathbf{p}'$, where the elements are indexed as shown in the figure. We now construct another r.v, Q_{new}, with one less element from

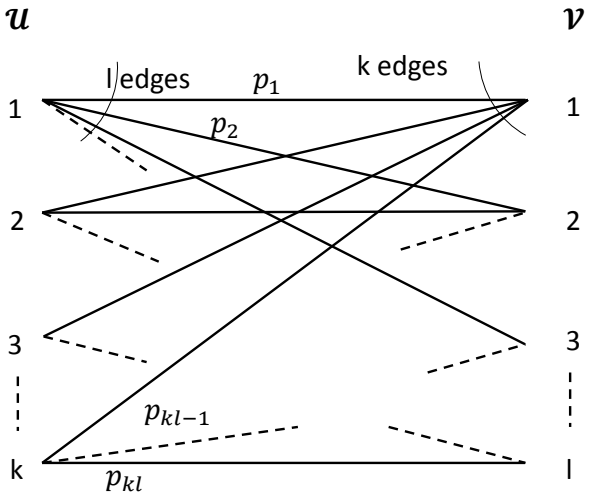

Figure 2.2: Characteristic graph of $U, V|Q = q$ when q is of class q_1

the class which induces this bipartite clique and the same number of elements from the other classes, such that $H(U|Q) + H(V|Q) \leq H(U|Q_{\text{new}}) + H(V|Q_{\text{new}})$. This then

proves the required claim, as given some arbitrary r.v Q, we can construct another r.v Q', satisfying the necessary markov conditions such that its alphabet contains atmost one element inducing any given kind of bipartite clique.

The new r.v. is constructed as follows. The alphabet is $\mathcal{Q}_{\text{new}} = \mathcal{Q} \cup \{q_{\text{new}}\} \setminus \{q_0, q_0'\}$, and the distribution of each of the induced cliques is the same as for Q, on all the elements except when $Q_{\text{new}} = q_{\text{new}}$, in which case, it is the element wise sum $\mathbf{p} + \mathbf{p}'$. This is illustrated in Figure 2.3.

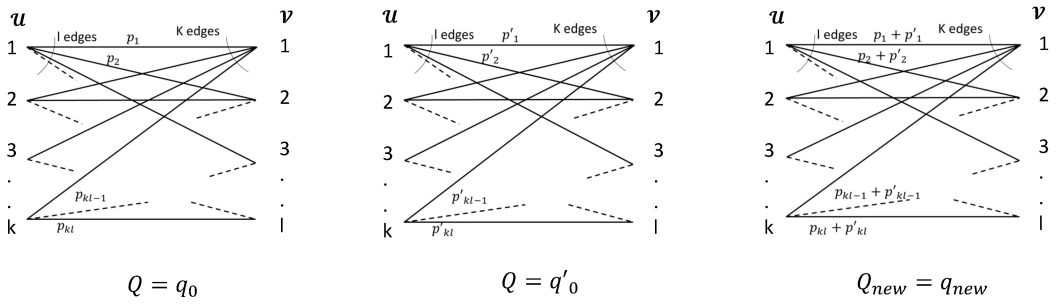

Figure 2.3: Merging of classes

We now prove that $H(U|Q) \leq H(U|Q_{\text{new}})$ and the proof of $H(V|Q) \leq H(V|Q_{\text{new}})$ is similar, thus proving the theorem. First, note that $p_Q(q_0) = \sum_{i=1}^{kl} p_i$, and $p_Q(q_0') = \sum_{i=1}^{kl} p_i'$, and $p_{Q_{\text{new}}}(q_{\text{new}}) = \sum_{i=1}^{kl}(p_i + p_i') = p_Q(q_0) + p_Q(q_0')$. Therefore,

$$\begin{aligned}
&p_Q(q_0)H(U|Q = q_0) + p_Q(q_0')H(U|Q = q_0') - p_{Q_{\text{new}}}(q_{\text{new}})H(U|Q_{\text{new}} = q_{\text{new}}) \\
&\quad = (p_0 + p_0')H\left(\frac{a_1 + a_1'}{p_0 + p_0'}, \frac{a_2 + a_2'}{p_0 + p_0'}, ..., \frac{a_k + a_k'}{p_0 + p_0'}\right) \\
&\qquad - p_0 H\left(\frac{a_1}{p_0}, \frac{a_2}{p_0}, ..., \frac{a_k}{p_0}\right) - p_0' H\left(\frac{a_1'}{p_0'}, \frac{a_2'}{p_0'}, ..., \frac{a_k'}{p_0'}\right)
\end{aligned}$$

where $a_i = p_{(i-1)l+1} + p_{(i-1)l+2} + ... + p_{il}$, likewise $a_i' = p'_{(i-1)l+1} + p'_{(i-1)l+2} + ... + p'_{il}$, and $p_0 = p_Q(q_0)$, $p_0' = p_Q(q_0')$.

Using the *Log-Sum inequality*, we get

$$\sum_{i=1}^{k} \left[(a_i + a_i') \log \frac{a_i + a_i'}{p_0 + p_0'}\right] \leq \sum_{i=1}^{k} \left[a_i \log \frac{a_i}{p_0} + a_i' \log \frac{a_i'}{p_0'}\right]$$

and thus

$$p_Q(q_0)H(U|Q = q_0) + p_Q(q_0')H(U|Q = q_0') \leq p_{Q_{new}}(q_{new})H(U|Q_{new} = q_{new})$$

$$
\begin{aligned}
H(U|Q) &= \sum_{q\in\mathcal{Q}} p_Q(q)H(U|Q=q) \\
&= \sum_{q\in\{q_0,q_0'\}} p_Q(q)H(U|Q=q) + \sum_{q\in\mathcal{Q}\setminus\{q_0,q_0'\}} p_Q(q)H(U|Q=q) \\
&= \sum_{q\in\{q_0,q_0'\}} p_Q(q)H(U|Q=q) + \sum_{q\in\mathcal{Q}_{\text{new}}\setminus\{q_{\text{new}}\}} p_{Q_{\text{new}}}(q)H(U|Q_{\text{new}}=q) \\
&= \sum_{q\in\{q_0,q_0'\}} p_Q(q)H(U|Q=q) - p_{Q_{\text{new}}}(q_{\text{new}})H(U|Q_{\text{new}}=q_{\text{new}}) + H(U|Q_{\text{new}}) \\
&\leq H(U|Q_{\text{new}})
\end{aligned}
$$

□

Having thusly restricted the search space for $\mathcal{Q}_{\text{opt}}$, we look at a few examples of using this bound in the next chapter.

CHAPTER 3

Information Complexity of EQ

We consider the problem, EQ, of testing if the inputs given to Alice and Bob are equal. The inputs, X and Y are assumed to be independently and unifromly distributed.

3.1 Ternary EQ

In the Ternary EQ problem, Alice's input X and Bob's input Y are drawn independently and uniformly at random over a ternary alphabet $\mathcal{X} = \mathcal{Y} = \{\alpha, \beta, \gamma\}$, and the function to be computed is $Z = \mathbb{1}[X = Y]$, the EQ function over a ternary alphabet.

We first use Theorem 3 to derive a lower bound for the information complexity of the function.

The characteristic graph of XZ and YZ is shown in Figure 3.1, and it is clear that there are 9 maximal bipartite cliques in this graph(Figure 3.2). From Theorem 3, in searching for the optimal auxillary r.v. Q, we can restrict the cardinality of its alphabet, $\mathcal{Q}$ to only 9, where each element induces one of the bipartite cliques shown.

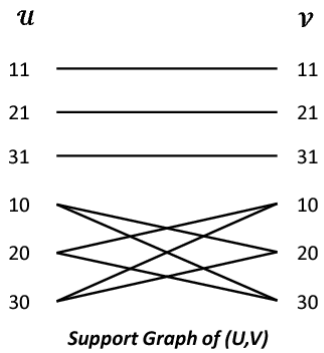

Figure 3.1: Support graph of ternary EQ

Since the input is uniformly distributed, each of the edges in the support graph of XZ and YZ has weight $\frac{1}{9}$, since each is equally likely. This translates to the following

observation on the weights of the induced bipartite cliques:

$$\sum_{q} \Pr_{XZ,YZ,Q}(u,v,q) = \frac{1}{9}, \quad \forall (u,v) \in \text{Supp}(XZ, YZ)$$

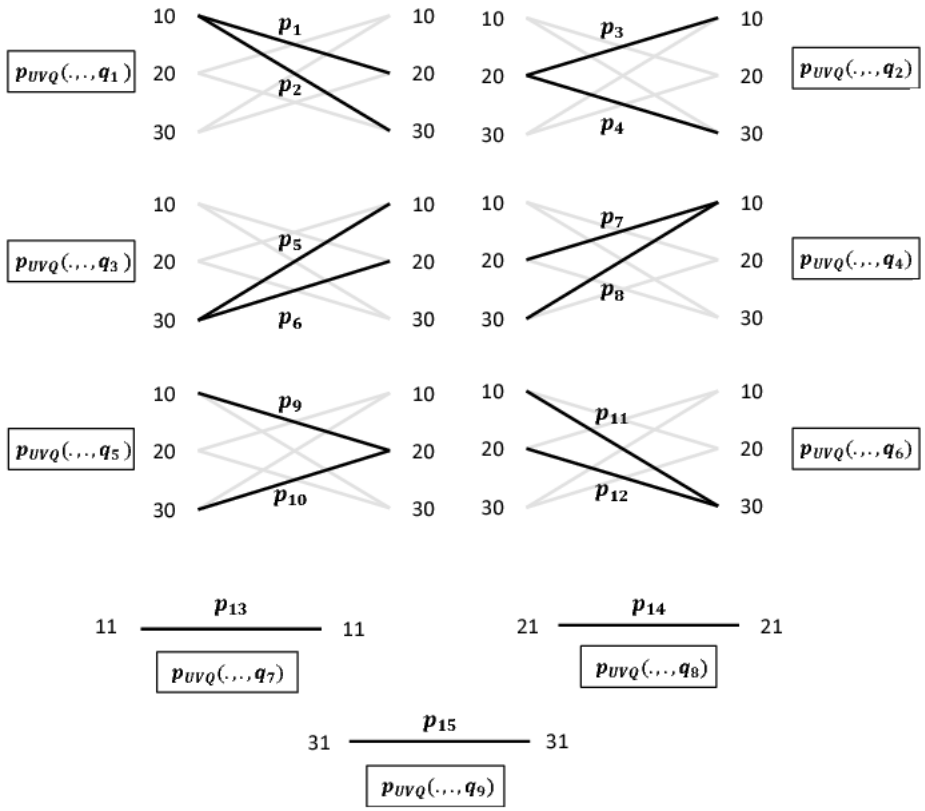

Figure 3.2: Maximal bipartite cliques for ternary EQ

Naming the probability of the edges in the induced cliques as in Figure dos, this leads to the equations

$$\begin{aligned}
p_1 + p_9 &= \frac{1}{9} \\
p_2 + p_7 &= \frac{1}{9} \\
p_3 + p_{11} &= \frac{1}{9} \\
p_4 + p_8 &= \frac{1}{9} \\
p_5 + p_{12} &= \frac{1}{9} \\
p_6 + p_{10} &= \frac{1}{9} \\
p_i &= \frac{1}{9} \quad \forall i \in [13, 15]
\end{aligned}$$

$$\text{Now,} \quad H(U|Q) + H(V|Q) = \sum_{i=1}^{6} p_Q(q_i) H_2\left[\frac{p_{2i-1}}{p_Q(q_i)}\right]$$
$$\leq \sum_{i=1}^{6} p_Q(q_i)$$
$$= \frac{2}{3}$$

where we have used the fact that $H_2(\cdot) \leq 1$ and the set of equations mentioned above. Now from section 2.4, we get $\mathsf{IC}_{XY}(Z) \geq H(X) + H(Y) - \frac{2}{3} = 2\log 3 - \frac{2}{3} = 2.5033$. We use the information complexity as a lower bound for the entropy of the link, $H(M_\Pi)$ for any protocol Π computing the function, for the amortized case, leading to $H(M_\Pi) \geq 2.5033$. We now provide a protcol Π whose link entropy meets this value, thus proving this bound's tightness. The protocol mentioned is only for one repetition - for the amortized case, it has to be repeated over the block of inputs.

Protocol 1: Ternary EQ computation

1. Alice sends her input to Bob
2. Bob computes $Z = \mathbb{1}[X = Y]$ and sends the resultant bit to Alice

---

We can calculate the information cost of this protocol. The entropy of the link during Alice's communication is exactly equal to the entropy of her input r.v, X. And the entropy of the link on Bob's message is equal to the entropy of the output variable, Z. Thus the information cost is $H(X) + H(Z) = \log 3 + H_2(\frac{1}{3}) = 2.5033$. Thus we see that the lower bound developed is tight in this example.

3.2 Two bit EQ

In the Two bit EQ problem, Alice gets two bits as the input $X = (X_0, X_1)$ and Bob gets $Y = (Y_0, Y_1)$ as input where $X_0, X_1, Y_0,$ and Y_1 are bits drawn uniformly and independently at random, and the function to be computed is the Equality function, $Z = \mathbb{1}[(X_0, X_1) = (Y_0, Y_1)]$.

The characteristic graph of XZ and YZ for this problem is shown in Figure 3.3, and the possible bipartite cliques are listed in Figure 3.4. These 18 bipartite cliques are

the only maximal ones possible, and hence from Theorem 3, we need consider only alphabets of size 18, when searching for the optimal $\mathcal{Q}$, such that each element of $\mathcal{Q}$ induces a different bipartite clique in the characteristic graph.

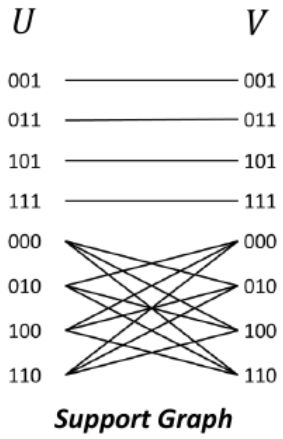

Figure 3.3: Support graph for 2bit EQ

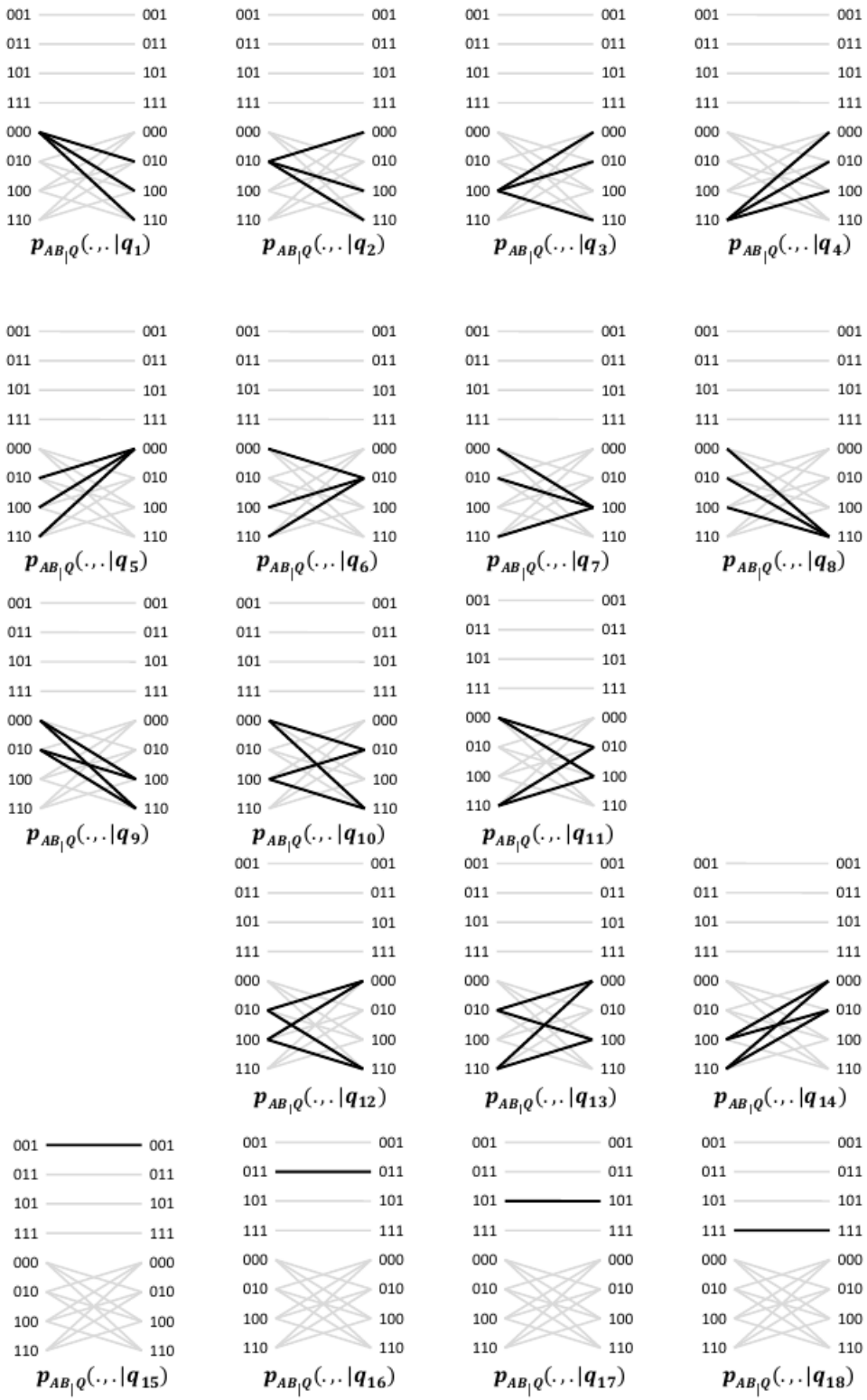

Figure 3.4: Maximal bipartite cliques of 2bit EQ

We have the following equalities:

$$
\begin{aligned}
&p_{Q|UV}(q_1|000,010) + p_{Q|UV}(q_6|000,010) + p_{Q|UV}(q_{10}|000,010) \\
&\qquad + p_{Q|UV}(q_{11}|000,010) = 1 \\
&p_{Q|UV}(q_1|000,100) + p_{Q|UV}(q_7|000,100) + p_{Q|UV}(q_9|000,100) \\
&\qquad + p_{Q|UV}(q_{11}|000,100) = 1 \\
&p_{Q|UV}(q_1|000,110) + p_{Q|UV}(q_8|000,110) + p_{Q|UV}(q_9|000,110) \\
&\qquad + p_{Q|UV}(q_{10}|000,110) = 1 \\
&p_{Q|UV}(q_2|010,000) + p_{Q|UV}(q_5|010,000) + p_{Q|UV}(q_{12}|010,000) \\
&\qquad + p_{Q|UV}(q_{13}|010,000) = 1 \\
&p_{Q|UV}(q_2|010,100) + p_{Q|UV}(q_7|010,100) + p_{Q|UV}(q_9|010,100) \\
&\qquad + p_{Q|UV}(q_{13}|010,100) = 1 \\
&p_{Q|UV}(q_2|010,110) + p_{Q|UV}(q_8|010,110) + p_{Q|UV}(q_9|010,110) \\
&\qquad + p_{Q|UV}(q_{12}|010,110) = 1 \\
&p_{Q|UV}(q_3|100,000) + p_{Q|UV}(q_5|100,000) + p_{Q|UV}(q_{12}|100,000) \\
&\qquad + p_{Q|UV}(q_{14}|100,000) = 1 \\
&p_{Q|UV}(q_3|100,010) + p_{Q|UV}(q_6|100,010) + p_{Q|UV}(q_{10}|100,010) \\
&\qquad + p_{Q|UV}(q_{14}|100,010) = 1 \\
&p_{Q|UV}(q_3|100,110) + p_{Q|UV}(q_8|100,110) + p_{Q|UV}(q_{10}|100,110) \\
&\qquad + p_{Q|UV}(q_{12}|100,110) = 1 \\
&p_{Q|UV}(q_4|110,000) + p_{Q|UV}(q_5|110,000) + p_{Q|UV}(q_{13}|110,000) \\
&\qquad + p_{Q|UV}(q_{14}|110,000) = 1 \\
&p_{Q|UV}(q_4|110,010) + p_{Q|UV}(q_6|110,010) + p_{Q|UV}(q_{11}|110,010) \\
&\qquad + p_{Q|UV}(q_{14}|110,010) = 1 \\
&p_{Q|UV}(q_4|110,100) + p_{Q|UV}(q_7|110,100) + p_{Q|UV}(q_{11}|110,100) \\
&\qquad + p_{Q|UV}(q_{13}|110,100) = 1
\end{aligned}
\tag{3.1}
$$

Now we evaluate the values of $H(U|QV)$ and $H(V|QU)$.

Note that $H(V|Q=q_i,U)=0$ for $i=5,\ldots,8$ and $i=15,\ldots,18$. Therefore, we have

$$
\begin{aligned}
H(V|QU) = & \sum_{\substack{(q,a)\in(q_1,000),(q_2,010),(q_3,100),(q_4,110)\\ (q_9,000),(q_9,010),(q_{10},000),(q_{10},100)\\ (q_{11},000)(q_{11},110),(q_{12},010),(q_{12},100)\\ (q_{13},010),(q_{13},110),(q_{14},100),(q_{14},110)}} p(q,u)H(V|Q=q,U=u) \\
= & \frac{1}{16}\times\{[p(q_1|000,010)+p(q_1|000,100)+p(q_1|000,110)]\cdot H_3(\cdot,\cdot) \\
& +[p(q_2|010,000)+p(q_2|010,100)+p(q_2|010,110)]\cdot H_3(\cdot,\cdot) \\
& +[p(q_3|100,000)+p(q_3|100,010)+p(q_3|100,110)]\cdot H_3(\cdot,\cdot) \\
& +[p(q_4|110,000)+p(q_4|110,010)+p(q_4|110,100)]\cdot H_3(\cdot,\cdot) \\
& +[p(q_9|000,100)+p(q_9|000,110)]\cdot H_2(\cdot) \\
& +[p(q_9|010,100)+p(q_9|010,110)]\cdot H_2(\cdot) \\
& +[p(q_{10}|000,010)+p(q_{10}|000,110)]\cdot H_2(\cdot) \\
& +[p(q_{10}|100,010)+p(q_{10}|100,110)]\cdot H_2(\cdot) \\
& +[p(q_{11}|000,010)+p(q_{11}|000,100)]\cdot H_2(\cdot) \\
& +[p(q_{11}|110,010)+p(q_{11}|110,100)]\cdot H_2(\cdot) \\
& +[p(q_{12}|010,000)+p(q_{12}|010,110)]\cdot H_2(\cdot) \\
& +[p(q_{12}|100,000)+p(q_{12}|100,110)]\cdot H_2(\cdot) \\
& +[p(q_{13}|010,000)+p(q_{13}|010,100)]\cdot H_2(\cdot) \\
& +[p(q_{13}|110,000)+p(q_{13}|110,100)]\cdot H_2(\cdot) \\
& +[p(q_{14}|100,000)+p(q_{14}|100,010)]\cdot H_2(\cdot) \\
& +[p(q_{14}|110,000)+p(q_{14}|110,010)]\cdot H_2(\cdot)\}
\end{aligned}
$$

and, similarly

$$\begin{aligned}
H(U|QV) =& \frac{1}{16} \times \{[p(q_5|010,000) + p(q_5|100,000) + p(q_5|110,000)] \cdot H_3(\cdot,\cdot) \\
&+ [p(q_6|000,010) + p(q_6|100,010) + p(q_6|110,010)] \cdot H_3(\cdot,\cdot) \\
&+ [p(q_7|000,100) + p(q_7|010,100) + p(q_7|110,100)] \cdot H_3(\cdot,\cdot) \\
&+ [p(q_8|000,110) + p(q_8|010,110) + p(q_8|100,110)] \cdot H_3(\cdot,\cdot) \\
&+ [p(q_9|000,100) + p(q_9|010,100)] \cdot H_2(\cdot) \\
&+ [p(q_9|000,110) + p(q_9|010,110)] \cdot H_2(\cdot) \\
&+ [p(q_{10}|000,010) + p(q_{10}|100,010)] \cdot H_2(\cdot) \\
&+ [p(q_{10}|000,110) + p(q_{10}|100,110)] \cdot H_2(\cdot) \\
&+ [p(q_{11}|000,010) + p(q_{11}|110,010)] \cdot H_2(\cdot) \\
&+ [p(q_{11}|000,100) + p(q_{11}|110,100)] \cdot H_2(\cdot) \\
&+ [p(q_{12}|010,000) + p(q_{12}|100,000)] \cdot H_2(\cdot) \\
&+ [p(q_{12}|010,110) + p(q_{12}|100,110)] \cdot H_2(\cdot) \\
&+ [p(q_{13}|010,000) + p(q_{13}|110,000)] \cdot H_2(\cdot) \\
&+ [p(q_{13}|010,100) + p(q_{13}|110,100)] \cdot H_2(\cdot) \\
&+ [p(q_{14}|100,000) + p(q_{14}|110,000)] \cdot H_2(\cdot) \\
&+ [p(q_{14}|100,010) + p(q_{14}|110,010)] \cdot H_2(\cdot)\}
\end{aligned}$$

Now using $H_2(\cdot) \leq 1$, and $H_3(\cdot,\cdot) \leq 2$, and adding the above 2 equations, we get each of the 12 equations (3.1) twice. Replacing these with the value 1 for each such quadruple, we get the desired supremum value.

$$H(U|Q) + H(V|Q) \leq \frac{1}{16}(24) = 1.5$$

The upper bound is attained when the distribution on the 4 edge classes is uniform, i.e. the probability metric associated with each edge of a 4-edge class is same and equal to $\frac{1}{32}$. This implies that $\sup_{\substack{p_{Q|U,V} \\ U-Q-V}} [H(U|Q) + H(V|Q)] = 1.5$, and hence $\mathsf{IC}_{\mathsf{XY}}(Z) \geq H(X|Y) + H(Y|X) - 1.5 = 4 - 1.5 = 2.5$.

Now, as in the previous section, we explore protocols for computing the 2 bit EQ problem, so as to arrive at an upper bound on $\mathsf{IC}_{\mathsf{XY}}(Z)$ as well.

Consider the following randomized protocol for the problem.

Definitions: Let Alice's input X be uniform in $\mathcal{A} = \{1, 2, 3, 4\}$, and Bob's input Y uniform in $\mathcal{B} = \{1, 2, 3, 4\}$. Define the sets $\mathbf{a} = \{1, 2\}, \mathbf{b} = \{1, 3\}, \mathbf{c} = \{1, 4\}, \mathbf{d} = \{2, 3\}, \mathbf{e} = \{2, 4\}, \mathbf{f} = \{3, 4\}$.

Protocol 2: Two bit EQ computation - Randomized

1. Alice uniformly picks $\mathbf{u} \in$ {a,b,c,d,e,f} such that $X \in \mathbf{u}$, and sends it to Bob.
2. If $Y \in \mathbf{u}$, Bob sends 1. Else he sends 1 or 0 with equal probability.

   If Bob's message is 0, the protocol terminates and $Z = 0$.

   If it is 1, protocol proceeds to step 3.
3. Alice reveals her input.
4. Bob computes Z and sends the result to Alice.

If $X = Y$, which occurs with probability $\frac{1}{4}$, both parties learn 2 bits. If $X \neq Y$, but $Y \in \mathbf{u}$, which happens with probability $\frac{1}{4}$, then Bob sends 1, and thus they proceed to step 3. If $Y \notin \mathbf{u}$, then Bob sends 1 with probability $\frac{1}{2}$. So, given that Bob sends 1, Bob's input $Y \in \mathbf{u}$ with probability $\frac{1}{2}$. Hence if the protocol goes to step 3, Alice's uncertainty about Bob's input is $H_3(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) = 1.5$ at the end of the protocol. If it stops at step 2, Alice and Bob each would have learnt only 1 bit about each other. Therefore, the information cost is, $\frac{1}{4}(4) + \frac{1}{4}(4 - \frac{3}{2}) + \frac{1}{4}(4 - \frac{3}{2}) + \frac{1}{4}(2) = 2.75$

3.3 Multibit EQ

In the case of Multibit EQ, Alice and Bob are each given n bits $X = (X_0, ..., X_k)$ and $Y = (Y_0, ..., Y_k)$, with all bits being iid drawn from $\mathcal{B}(\frac{1}{2})$ and the function to be computed is the EQ function on k bits, i.e, checking if all the k bits of both parties are the same.

Now the maximal bipartite cliques in this new setting for the characteristic graph of $U = XZ$ and $V = YZ$ will be functions of k. For each i there will be ${}^kC_i$ maximal bipartite cliques with i nodes from the U side, with $Z = 0$ and $k - i$ nodes from the V side, with $Z = 0$. So the total number of classes would be $\sum_{i=1}^{k-1} {}^kC_i + k$, where

the final k classes are for the $Z = 1$ case, each containing one edge. For the $Z = 0$ cliques, we refer to a maximal bipartite clique with i nodes from the U set as belonging to a class $\mathcal{L}_i$. Given some edge with $Z = 0$, connecting $U = u$ and $V = v$: (u, v), we can enumerate the number of classes, $\mathcal{L}_i$ that contains the edge. Each edge (u, v), with $u \neq v$, occurs in classes $\mathcal{L}_1$ only once, in classes $\mathcal{L}_2$ ${}^{k-2}C_1$ times, and in general, occurs ${}^{k-2}C_{i-1}$ times in the classes $\mathcal{L}_i$. Now as in the earlier cases, each of the edges has a probability $p_{U,V,Q}(u, v, q)$ associated with it, which leads to a set of constraints:

$$p_{U,V}(u, v) = \sum_q p_{U,V,Q}(u, v, q) = \frac{1}{k^2} \quad \forall (u, v) \tag{3.2}$$

In addition to these constraints the $p_{U,V,Q}(u, v, q)$ should be such that $U - Q - V$. Now,

$$\begin{aligned}
&H(U|Q) + H(V|Q) \\
&= \sum_{q_i} [H(U|Q = q_i)p_Q(q_i) + H(V|Q = q_i)p_Q(q_i)] \\
&\leq \sum_{\mathcal{L}_1} p_Q(q_i) \log(k - 1) + \sum_{\mathcal{L}_2} p_Q(q_i)(1 + \log(k - 2)) \\
&\quad + \cdots + \sum_{\mathcal{L}_{k-1}} p_Q(q_i) \log(k - 1)
\end{aligned} \tag{3.3}$$

Case I: k is even: Using the fact that if we have 2 non-negative integers a and b such that $a + b = k$ (a constant), the maximum value of ab is when $a = b = \frac{k}{2}$, we get $(\log(i) + \log(k - i)) \leq (\log(\frac{k}{2}) + \log(\frac{k}{2}))$. Using this in (3.3), we get

$$\begin{aligned}
&H(U|Q) + H(V|Q) \\
&\quad \leq \sum_{\mathcal{L}_1} p_Q(q_i) 2 \log \frac{k}{2} + .. + \sum_{\mathcal{L}_{k-1}} p_Q(q_i) 2 \log \frac{k}{2}
\end{aligned}$$

$$= 2 \log(\frac{k}{2}) \sum_{\mathcal{L}_1, .., \mathcal{L}_{k-1}} p_Q(q_i) = 2(1 - \frac{1}{k}) \log(\frac{k}{2}) \tag{3.4}$$

Consider the distribution $p(u, v, q) = \frac{1}{k^2 \left({}^{k-2}C_{\frac{k-2}{2}} \right)}$ for all the edges in classes $\mathcal{L}_{\frac{k}{2}}$, and $p = 0$ for all the other edges in the $Z = 0$ set(Of course, for all the edges with $Z = 1$, we need $p = \frac{1}{k^2}$ so as to satisfy the constraints in (3.2)). It is easy to verify that this distribution ensures that $U - Q - V$, and hence is a valid Q choice. For this

distribution, the value of $H(U|Q) + H(V|Q)$ is,

$$ {}^{k}C_{\frac{k}{2}} \cdot \frac{\left(\frac{k}{2}\right)\left(\frac{k}{2}\right)}{k^2\left({}^{k-2}C_{\frac{k-2}{2}}\right)} \cdot 2\log\frac{k}{2} = \quad 2\left(1-\frac{1}{k}\right)\log\frac{k}{2} $$

and so, $\sup_{\substack{p_{Q|U,V}:\\ U-Q-V}} H(U|Q) + H(V|Q) = 2\left(1-\frac{1}{k}\right)\log\frac{k}{2}$.

From (section 2.4), we get

$IC_{XY}(Z) \geq 2\log(k) - 2\left(1-\frac{1}{k}\right)\log\frac{k}{2} = 2 + \frac{2}{k}\log\frac{k}{2}$.

Case II: k is odd: Like in the previous case, one can see that

$$ H(U|Q) + H(V|Q) \leq \left[\log(\frac{k-1}{2}) + \log(\frac{k+1}{2})\right]\left(1-\frac{1}{k}\right) $$

Again, we can consider the distribution $p = \frac{1}{k^2\left({}^{k-2}C_{\frac{k-1}{2}}\right)}$ for all the edges in classes $\mathcal{L}_{\frac{k+1}{2}}$, so that

$$ H(U|Q) + H(V|Q) = \left[\log\left(\frac{k-1}{2}\cdot\frac{k+1}{2}\right)\right]\left(1-\frac{1}{k}\right) $$

So $\sup_{\substack{p_{Q|U,V}:\\ U-Q-V}} [H(U|Q) + H(V|Q)] = \left[\log\left(\frac{k^2-1}{4}\right)\right]\left(1-\frac{1}{k}\right)$, and from (2.4), we get $IC_{XY}(Z) \geq 2\log(k) - \left[\log\left(\frac{k-1}{2}\cdot\frac{k+1}{2}\right)\right]\left(1-\frac{1}{k}\right)$.

CHAPTER 4

Conclusion

In this report we demonstrated a method for obtaining lower bounds on information complexity of functions under independent input distributions via computing the Wyner common information of a pair of related random varibales. We showed the tightness of our lower bound for the ternary EQ function. For the 2-bit EQ function, our lower bound works out to 2.5, while we obtained an upper bound of 2.75 by giving a randomized protocol. Following this, we extended the results to the general case of k-ary EQ function, where our lower bound converges to 2 as $k \to \infty$. Repeated use of 2-bit EQ computation protocol gives an upper bound of 3.667 as $k \to \infty$.

APPENDIX A

Proof of Theorem 1

Consider the case when X and Y are independent. From Lemma 1, $I(X;Y|M) \leq I(X;Y) = 0$, and hence $I(X;Y|M) = 0$. Using this, for any valid protocol with transcript M,

$$\begin{aligned} I(XZ;YZ|M) &= I(X;Y|M) + I(Z;Y|MX) \\ &\quad + I(X;Z|MY) + I(Z;Z|XYM) \\ &\leq 0 + H(Z|MX) + H(Z|MY) + H(Z|XYM) \\ &\overset{(a)}{=} 0 \end{aligned}$$

(a) is because all the four terms are 0. Hence $I(XZ;YZ|M) = 0$ and the Markov chain $XZ - M - YZ$.

$$\begin{aligned} \text{Now, } I(X;M|Y) + I(Y;M|X) &\overset{(a)}{=} I(X;MZ|Y) + I(Y;MZ|X) \\ &\overset{(b)}{=} I(X;Z|Y) + I(Y;Z|X) \\ &\quad + I(XZ;M|YZ) + I(YZ;M|XZ) \\ &\overset{(c)}{\geq} I(X;Z|Y) + I(Y;Z|X) + T_{\mathsf{Wyn}}(XZ;YZ) \end{aligned} \tag{A.1}$$

where (a) follows from the fact that $0 \leq I(X;Z|MY) \leq H(Z|MY) = 0$, (b) is true as $I(XZ;M|YZ) = I(X;M|YZ) + H(Z|MYZ) = I(X;M|YZ)$, (c) is a result of the relaxation $XZ - M - YZ$. This implies that the information complexity of the setting $IC_{XY}(Z) \geq I(X;Z|Y) + I(Y;Z|X) + T_{\mathsf{Wyn}}(XZ;YZ)$, thus proving Theorem 1 for independent inputs.

REFERENCES

[1] **Ahlswede, R.** and **I. Csiszár**, On oblivious transfer capacity. *In Information Theory, Combinatorics, and Search Theory*. Springer, 2013, 145–166.

[2] **Braverman, M.** (2015). Interactive information complexity. *SIAM Journal on Computing*, **44**(6), 1698–1739.

[3] **Braverman, M.** and **J. Schneider** (2015). Information complexity is computable. *arXiv preprint arXiv:1502.02971*.

[4] **El Gamal, A.** and **Y.-H. Kim**, *Network information theory*. Cambridge university press, 2011.

[5] **Gács, P.** and **J. Körner** (1973). Common information is far less than mutual information. *Problems of Control and Information Theory*, **2**(2), 149–162.

[6] **Kaspi, A. H.** (1985). Two-way source coding with a fidelity criterion. *Information Theory, IEEE Transactions on*, **31**(6), 735–740.

[7] **Kushilevitz, E.** and **N. Nisan**, *Communication Complexity*. Cambridge University Press, 1997. ISBN 9780521560672. URL `https://books.google.co.in/books?id=yiV6pwAACAAJ`.

[8] **Ma, N.** and **P. Ishwar** (2013). The infinite-message limit of two-terminal interactive source coding. *Information Theory, IEEE Transactions on*, **59**(7), 4071–4094.

[9] **Prabhakaran, M. M.** and **V. M. Prabhakaran** (2014). Tension bounds for information complexity. *arXiv preprint arXiv:1408.6285*.

[10] **Prabhakaran, V. M.** and **M. M. Prabhakaran** (2014). Assisted common information with an application to secure two-party sampling. *Information Theory, IEEE Transactions on*, **60**(6), 3413–3434.

[11] **Rao, K. S.** and **V. M. Prabhakaran**, A new upperbound for the oblivious transfer capacity of discrete memoryless channels. *In Information Theory Workshop (ITW), 2014 IEEE*. IEEE, 2014.

[12] **Witsenhausen, H. S.** (1976). Values and bounds for the common information of two discrete random variables. *SIAM Journal on Applied Mathematics*, **31**(2), 313–333.

[13] **Wyner, A. D.** (1975). The common information of two dependent random variables. *Information Theory, IEEE Transactions on*, **21**(2), 163–179.

[14] **Yao, A. C.-C.**, Some complexity questions related to distributive computing (preliminary report). *In Proceedings of the eleventh annual ACM symposium on Theory of computing*. ACM, 1979.

LIST OF PAPERS BASED ON THESIS

Rajakrishnan, S., Sundara Rajan, S., V. Prabhakaran (2016). Lower bounds for interactive function computation via wyner common information.
arXiv preprint arXiv:1602.02390.