

On Slepian–Wolf Theorem with Interaction

Alexander Kozachinskiy^{1,2}(✉)

¹ Lomonosov Moscow State University, Moscow, Russia
kozlach@mail.ru

² National Research University Higher School of Economics, Moscow, Russia

Abstract. In this paper we study interactive “one-shot” analogues of the classical Slepian–Wolf theorem. Alice receives a value of a random variable X , Bob receives a value of another random variable Y that is jointly distributed with X . Alice’s goal is to transmit X to Bob (with some error probability ε). Instead of one-way transmission we allow them to interact. They may also use shared randomness.

We show, that for every natural r Alice can transmit X to Bob using $(1 + \frac{1}{r})H(X|Y) + r + O(\log_2(\frac{1}{\varepsilon}))$ bits on average in $\frac{2H(X|Y)}{r} + 2$ rounds on average. Setting $r = \lceil \sqrt{H(X|Y)} \rceil$ and using a result of [2] we conclude that every one-round protocol π with information complexity I can be compressed to a (many-round) protocol with expected communication about $I + 2\sqrt{I}$ bits. This improves a result by Braverman and Rao [3], where they had $I + 5\sqrt{I}$. Further, we show (by setting $r = \lceil H(X|Y) \rceil$) how to solve this problem (transmitting X) using $2H(X|Y) + O(\log_2(\frac{1}{\varepsilon}))$ bits and 4 rounds on average. This improves a result of [4], where they had $4H(X|Y) + O(\log 1/\varepsilon)$ bits and 10 rounds on average.

In the end of the paper we discuss how many bits Alice and Bob may need to communicate on average besides $H(X|Y)$. The main question is whether the upper bounds mentioned above are tight. We provide an example of (X, Y) , such that transmission of X from Alice to Bob with error probability ε requires $H(X|Y) + \Omega(\log_2(\frac{1}{\varepsilon}))$ bits on average.

Keywords: Slepian–Wolf theorem · Communication complexity · Information complexity

1 Introduction

Assume that Alice receives a value of a random variable X and she wants to transmit that value to Bob. It is well-known [8] that Alice can do it using one message over the binary alphabet of expected length less than $H(X) + 1$. Assume now that there are n independent random variables X_1, \dots, X_n distributed as X , and Alice wants to transmit all X_1, \dots, X_n to Bob. Another classical result from [8] states, that Alice can do it using one message of *fixed* length, namely $\approx nH(X)$, with a small probability of error.

A. Kozachinskiy—Supported by the RFBR grant 16-01-00362 and by the grant of the President of Russian Federation (MK-7312.2016.1).

One of the possible ways to generalize this problem is to provide Bob with a value of another random variable Y which is jointly distributed with X . That is, to let Bob know some partial information about X for free. This problem is the subject of the classical Slepian-Wolf Theorem [9] which asserts that if there are n independent pairs $(X_1, Y_1), \dots, (X_n, Y_n)$, each pair distributed exactly as (X, Y) , then Alice can transmit all X_1, \dots, X_n to Bob, who knows Y_1, \dots, Y_n , using one message of fixed length, namely $\approx nH(X|Y)$, with a small probability of error¹ However, it turns out that a one-shot analogue of this theorem is impossible, if only one-way communication is allowed.

The situation is quite different, if we allow Alice and Bob to *interact*, that is, to send messages in both directions. In [7] Orlitsky studied this problem for the average-case communication when no error is allowed. He showed that if pair (X, Y) is uniformly distributed on it's support, then Alice may transmit X to Bob using at most

$$H(X|Y) + 3 \log_2(H(X|Y) + 1) + 17$$

bits on average and 4 rounds. For the pairs (X, Y) whose support is a Cartesian product Orlitsky showed that error-less transmission of X from Alice to Bob requires $H(X)$ bits on average.

From a result of Braverman and Rao [3], it follows that for arbitrary (X, Y) it is sufficient to communicate at most

$$H(X|Y) + 5\sqrt{H(X|Y)} + O\left(\log_2\left(\frac{1}{\varepsilon}\right)\right)$$

bits on average (here ε stands for the error probability).

We show that for every positive ε and natural r there is a public-coin protocol transmitting X from Alice to Bob with error probability at most ε (for each pair of inputs) using at most

$$\left(1 + \frac{1}{r}\right) H(X|Y) + r + O\left(\log_2 \frac{1}{\varepsilon}\right)$$

bits on average in at most

$$\frac{2H(X|Y)}{r} + 2$$

rounds on average. Furthermore, there is a private-coin protocol with the same properties plus extra $O(\log \log \text{supp}(X, Y))$ bits of communication. Our protocol is inspired by protocol from [1]. The idea of the protocol is essentially the same, we only apply some technical trick to reduce communication.

This improves the result of Braverman and Rao, since setting $r = \lceil \sqrt{H(X|Y)} \rceil$ above we obtain the protocol with expected communication at

¹ This paper is focused only on the non-symmetric version of this problem. In more general version Alice and Bob send messages to the 3rd party Charlie, who must reconstruct both random variables.

most $H(X|Y) + 2\sqrt{H(X|Y)} + O(\log_2(\frac{1}{\epsilon}))$. In [4], it is established a one-shot interactive analogue of the Slepian-Wolf theorem for the bounded-round communication. They showed that Alice may transmit X to Bob using at most $O(H(X|Y) + 1)$ bits and $O(1)$ rounds on average. More specifically, their protocol transmits at most $4H(X|Y) + \log_2(1/\epsilon) + O(1)$ bits on average in 10 rounds on average. Setting $r = \lceil H(X|Y) \rceil$ above we improve this result. Indeed, we obtain the protocol with the expected length at most $2H(X|Y) + O(\log_2(\frac{1}{\epsilon}))$ and the expected number of rounds at most 4.

Actually, in [3] a more general result was established. It was shown there that every one-round protocol π with information complexity I can be compressed to the (many-round) protocol with expected length at most

$$\approx I + 5\sqrt{I}. \tag{1}$$

Using the result from [2], we improve Equ. 1. Namely, we show that every one-round protocol π with information complexity I can be compressed to the (many-round) protocol with expected communication length at most

$$\approx I + 2\sqrt{I}.$$

Are there random variables X, Y for which the upper bound of the form $H(X|Y) + O(\sqrt{H(X|Y)})$ is tight? We make a step towards answering this question: we provide an example of random variables X, Y such that every public-coin communication protocol which transmits X from Alice to Bob with error probability ϵ (with respect to the input distribution and the protocol’s randomness) must communicate at least $H(X|Y) + \Omega(\log_2(\frac{1}{\epsilon}))$ bits on average.

2 Definitions

2.1 Information Theory

Let X, Y be two joint distributed random variables, taking values in the finite sets, respectively, \mathcal{X} and \mathcal{Y} .

Definition 1. *Shannon Entropy of X is defined by the formula*

$$H(X) = \sum_{x \in \mathcal{X}} \Pr[X = x] \log_2 \left(\frac{1}{\Pr[X = x]} \right).$$

Definition 2. *Conditional Shannon entropy of X with respect to Y is defined by the formula:*

$$H(X|Y) = \sum_{y \in \mathcal{Y}} H(X|Y = y) \Pr[Y = y],$$

where $X|Y = y$ denotes a distribution of X , conditioned on the event $\{Y = y\}$.

If X is uniformly distributed in \mathcal{X} then obviously $H(X) = \log_2(|\mathcal{X}|)$. We will also use the fact that the formula for conditional entropy may be re-written as

$$H(X|Y) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \Pr[X = x, Y = y] \log_2 \left(\frac{1}{\Pr[X = x|Y = y]} \right).$$

Generalization of the Shannon entropy is Renyi entropy.

Definition 3. *Renyi entropy of X is defined by the formula*

$$H_2(X) = -\log_2 \left(\sum_{x \in \mathcal{X}} \Pr[X = x]^2 \right).$$

Concavity of log implies that $H(X) \geq H_2(X)$.

The mutual information of two random variables X and Y , conditioned on another random variable Z , can be defined as:

$$I(X : Y|Z) = H(X|Z) - H(X|Y, Z).$$

For the further introduction in information theory see, for example [11].

2.2 Communication Protocols

Assume that we are given jointly distributed random variables X and Y , taking values in finite sets \mathcal{X} and \mathcal{Y} . Let R_A, R_B be a random variables, taking values in finite sets \mathcal{R}_A and \mathcal{R}_B , such that $(X, Y), R_A, R_B$ are mutually independent.

Definition 4. *A private-coin communication protocol is a rooted binary tree, in which each non-leaf vertex is associated either with Alice or with Bob. For each non-leaf vertex v associated with Alice there is a function $f_v : \mathcal{X} \times \mathcal{R}_A \rightarrow \{0, 1\}$ and for each non-leaf vertex u associated with Bob there is a function $g_u : \mathcal{Y} \times \mathcal{R}_B \rightarrow \{0, 1\}$. For each non-leaf vertex one of an out-going edges is labeled by 0 and other is labeled by 1. Finally, for each leaf l there is a function $\phi_l : \mathcal{Y} \times \mathcal{R}_B \rightarrow \mathcal{O}$, where \mathcal{O} denotes the set of all possible Bob's outputs.*

A computation according to a protocol runs as follows. Alice is given $x \in \mathcal{X}$, Bob is given $y \in \mathcal{Y}$. Assume that R_A takes a value r_a and R_B takes a value r_b . Alice and Bob start at the root of the tree. If they are in the non-leaf vertex v associated with Alice, then Alice sends $f_v(x, r_a)$ to Bob and they go by the edge labeled by $f_v(x, r_a)$. If they are in a non-leaf vertex associated with Bob then Bob sends $g_v(y, r_b)$ to Alice and they go by the edge labeled by $g_v(y, r_b)$. When they reach a leaf l Bob outputs the result $\phi_l(y, r_b)$.

A protocol is called *deterministic* if f_v, g_u and ϕ_l do not depend on the values of R_A, R_B .

A *randomized communication protocol* is a distribution over private-coin protocols with the same \mathcal{X} for Alice and the same \mathcal{Y} for Bob. The random variable with this distribution (*public randomness*) is denoted below by R . Before the

execution starts, Alice and Bob sample R to choose the private-coin protocol to be executed.

A protocol is called *public-coin* if it is a distribution over deterministic protocols.

We distinguish between average-case communication complexity and the worst-case communication complexity. The (worst-case) communication complexity of a protocol π , denoted by $CC(\pi)$, is defined as the maximal possible depth of the leaf Alice and Bob may reach in π .

We say that protocol π communicates d bits on average (or expected length of the protocol is equal to d), if the expected depth of the leaf that Alice and Bob reach during the execution of the protocol π is equal to d , where the expectation is taken over X, Y and the protocol’s randomness.

For the further introduction in Communication Complexity see [5].

3 Slepian-Wolf Theorem with Interaction

Consider the following auxiliary problem. Let A be a finite set. Assume that Alice receives an arbitrary $a \in A$ and Bob receives an arbitrary probability distribution μ on A . Alice wants to communicate a to Bob in about $\log(1/\mu(a))$ bits with small probability of error.

Lemma 1. *Let ε be a positive real and r a positive integer. There exists a public coin randomized communication protocol such that for all a in the support of μ the following hold:*

- *in the end of the communication Bob outputs $b \in A$ which is equal to a with probability at least $1 - \varepsilon$;*
- *the protocol communicates at most*

$$\log_2 \left(\frac{1}{\mu(a)} \right) + \frac{\log_2 \left(\frac{1}{\mu(a)} \right)}{r} + r + \log_2 \left(\frac{1}{\varepsilon} \right) + 2$$

bits, regardless of the randomness.

- *the number of rounds in the protocol does not exceed*

$$\frac{2 \log_2 \left(\frac{1}{\mu(a)} \right)}{r} + 2.$$

Proof. Alice and Bob interpret each portion of $|A|$ consecutive bits from the public randomness source as a table of a random function $h : A \rightarrow \{0, 1\}$. That is, we will think that they have access to a large enough family of mutually independent random functions of the type $A \rightarrow \{0, 1\}$. Those functions will be called *hash functions* and their values *hash values* below.

The first set $k = \lceil \log_2 \left(\frac{1}{\varepsilon} \right) \rceil$. Then for all $i = 0, 1 \dots$ Bob sets:

$$S_i = \{x \in A \mid \mu(x) \in (2^{-i-1}, 2^{-i}]\}.$$

At the beginning Alice sends k hash values of a . Then Alice and Bob work in stages numbered $1, 2, \dots$.

On *Stage t* :

1. Alice sends r new hash values of a to Bob so that the total number of hash values of a available to Bob be $k + rt$.
2. For each $i \in \{r(t - 1), \dots, rt - 1\}$ Bob computes set S'_i , which consists of all elements from S_i , which agree with all Alice's hash values.
3. If there exists $i \in \{r(t - 1), \dots, rt - 1\}$ such that $S'_i \neq \emptyset$, then Bob sends 1 to Alice, outputs any element of S'_i and they terminate. Otherwise Bob sends 0 to Alice and they proceed to Stage $t + 1$.

Let us at first show that the protocol terminates for all a in the support of μ . Assume that Alice has a and Bob has μ . Let $i = \left\lfloor \log_2 \left(\frac{1}{\mu(a)} \right) \right\rfloor$ so that $a \in S_i$. The protocol terminates on Stage t where

$$r(t - 1) \leq i \leq rt - 1 \tag{2}$$

or earlier. Indeed all hash values of a available to Bob on Stage t coincide with hash values of some element of S_i (for instance, with those of a).

Thus Alice sends at most $k + rt$ bits to Bob and Bob sends at most t bits to Alice. The left-hand side of (2) implies that $t \leq \frac{i}{r} + 1$. Therefore Alice's communication is bounded by

$$\begin{aligned} k + rt &\leq k + r \left(\frac{i}{r} + 1 \right) \\ &= \left\lceil \log_2 \left(\frac{1}{\varepsilon} \right) \right\rceil + i + r \\ &\leq \log_2 \left(\frac{1}{\mu(a)} \right) + r + \log_2 \left(\frac{1}{\varepsilon} \right) + 1, \end{aligned}$$

and Bob's communication is bounded by

$$t \leq \frac{i}{r} + 1 \leq \frac{\log_2 \left(\frac{1}{\mu(a)} \right)}{r} + 1.$$

These two bounds imply that the total communication length is at most $\log_2 \left(\frac{1}{\mu(a)} \right) + \frac{\log_2 \left(\frac{1}{\mu(a)} \right)}{r} + r + \log_2 \left(\frac{1}{\varepsilon} \right) + 2$. The number of rounds equals the length of Bob's communication, multiplied by 2. Hence this number is at most $\frac{2 \log_2 \left(\frac{1}{\mu(a)} \right)}{r} + 2$. We conclude that the communication and the number of rounds are as short as required.

It remains to bound the error probability. An error may occur, if for some t a set S_i considered on Stage t has an element $b \neq a$ which agrees with hash values sent from Alice. At that time Bob has already $k + rt \geq k + i + 1$ hash values, where the inequality follows from (2). The probability that $k + i + 1$ hash values of b

coincide with those of a is 2^{-k-i-1} . Hence by union bound error probability does not exceed

$$\begin{aligned} \sum_{i=0}^{\infty} |S_i| 2^{-k-i-1} &= 2^{-k} \sum_{i=0}^{\infty} |S_i| 2^{-i-1} < 2^{-k} \sum_{i=0}^{\infty} \sum_{x \in S_i} \mu(x) \\ &= 2^{-k} \sum_{x \in A} \mu(x) = 2^{-k} = 2^{-\lceil \log_2(\frac{1}{\varepsilon}) \rceil} \leq \varepsilon. \end{aligned}$$

Theorem 1. *Let X, Y be jointly distributed random variables that take values in the finite sets \mathcal{X} and \mathcal{Y} . Then for every positive ε and positive integer r there exists a public-coin protocol with the following properties.*

- For every pair (x, y) from the support of (X, Y) with probability at least $1 - \varepsilon$ Bob outputs x ;
- The expected length of communication is at most

$$H(X|Y) + \frac{H(X|Y)}{r} + r + \log_2\left(\frac{1}{\varepsilon}\right) + 2.$$

- The expected number of rounds is at most

$$\frac{2H(X|Y)}{r} + 2.$$

Proof. On input x, y , Alice and Bob run protocol of Lemma 1 with $A = \mathcal{X}$, $a = x$ and μ equal to the distribution of X , conditioned on the event $Y = y$. Notice that Alice knows a and Bob knows μ .

Let us show that all the requirements are fulfilled for this protocol. The first requirement immediately follows from the first property of the protocol of Lemma 1.

From the second and the third property of the protocol of Lemma 1 it follows that for input pair x, y out protocol communicates at most:

$$\log_2\left(\frac{1}{\Pr[X = x|Y = y]}\right) + \frac{\log_2\left(\frac{1}{\Pr[X = x|Y = y]}\right)}{r} + r + \log_2\left(\frac{1}{\varepsilon}\right) + 2$$

bits in at most

$$\frac{\log_2\left(\frac{1}{\Pr[X = x|Y = y]}\right)}{r} + 2$$

rounds. Recalling that

$$H(X|Y) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \Pr[X = x, Y = y] \log_2\left(\frac{1}{\Pr[X = x|Y = y]}\right)$$

we see on average the communication and the number of rounds are as short as required.

Theorem 1 provides a trade off between the communication and the number of rounds.

- To obtain a protocol with minimal communication set $r = \lceil \sqrt{H(X|Y)} \rceil$. For such r the protocol communicates at most $H(X|Y) + 2\sqrt{H(X|Y)} + O(\log_2 \frac{1}{\varepsilon})$ bits on average.
- To obtain a protocol with a constant number of rounds on average set, for example, $r = \lceil H(X|Y) \rceil$. For such r the protocol communicates at most $2H(X|Y) + O(\log_2 \frac{1}{\varepsilon})$ bits on average in at most 4 rounds on average.
- In a same manner for every $\delta \in (0, 0.5)$ we can obtain a protocol with the expected communication at most $H(X|Y) + O(H(X|Y)^{0.5+\delta})$ and the expected number of rounds at most $O(H(X|Y)^{0.5-\delta})$.

Remark. One may wonder whether there exists a private-coin communication protocol with the same properties as the protocol of Theorem 1. Newman’s theorem [6] states that every public-coin protocol can be transformed into a private-coin protocol at the expense of increasing the error probability by δ and the worst case communication by $O(\log \log |\mathcal{X} \times \mathcal{Y}| + \log 1/\delta)$ (for any positive δ). Lemma 1 provides an upper bound for the error probability and communication of our protocol for each pair of inputs. Repeating the arguments from the proof of Newman’s theorem, we are able to transform the public-coin protocol of Lemma 1 into a private-coin one with the same trade off between the increase of error probability and the increase of communication length. It follows that for our problem there exists a private-coin communication protocol which errs with probability at most ε and communicates on average as many bits as the public-coin protocol from Theorem 1 plus extra $O(\log \log |\mathcal{X} \times \mathcal{Y}|)$ bits.

4 One-Round Compression

Information complexity of the protocol π with inputs (X, Y) is defined as

$$\begin{aligned}
 IC_\mu(\pi) &= I(X : \Pi|Y, R) + I(Y : \Pi|X, R) \\
 &= I(X : \Pi|Y, R, R_B) + I(Y : \Pi|X, R, R_A) \\
 &= I(X : \Pi, R, R_B|Y) + I(Y : \Pi, R, R_A|X),
 \end{aligned}$$

where R, R_A, R_B denote (shared, Alice’s and Bob’s) randomness, μ stands for the distribution of (X, Y) and Π stands for the concatenation of all bits sent in π (Π is called a *transcript*). The first term is equal to the information which Bob learns about Alice’s input and the second term is equal to the information which Alice learns about Bob’s input. Information complexity is an important concept in the Communication Complexity. For example, information complexity plays the crucial role in the Direct-Sum problem [10].

We will consider the special case when π is *one-round*. In this case Alice sends one message Π to Bob, then Bob outputs the result (based on his input, his randomness, and Alice’s message) and the protocol terminates. Since Alice learns nothing, information complexity can be re-written as

$$I = IC_\mu(\pi) = I(X : \Pi|Y, R).$$

Our goal is to simulate a given one-round protocol π with another protocol τ which has the same input space (X, Y) and whose expected communication complexity is close to I . The new protocol τ may be many-round. The quality of simulation will be measured by the statistical distance. Statistical distance between random variables A and B , both taking values in the set V , equals

$$\delta(A, B) = \max_{U \subset V} |\Pr[A \in U] - \Pr[B \in U]|.$$

One of the main results of [3] is the following theorem.

Theorem 2. *For every one-round protocol π and for every probability distribution μ there is a public-coin protocol τ with expected length (with respect to μ and the randomness of τ) at most $I + 5\sqrt{I} + O(\log_2 \frac{1}{\varepsilon})$ such that for each pair of inputs (x, y) after termination of τ Bob outputs a random variable Π' with $\delta((\Pi|X = x, Y = y), (\Pi'|X = x, Y = y)) \leq \varepsilon$.*

We will show that Theorem 1 and together with the main result of [2] imply that we can replace $5\sqrt{I}$ by about $2\sqrt{I}$ in this theorem. More specifically,

Theorem 3. *For every one-round protocol π and for every probability distribution μ there is a public-coin protocol τ with expected length (with respect to μ and the randomness of τ) at most*

$$I + \log_2(I + O(1)) + 2\sqrt{I + \log_2(I + O(1))} + O\left(\log_2 \frac{1}{\varepsilon}\right)$$

such that for each pair of inputs (x, y) in the protocol τ Bob outputs Π' with $\delta((\Pi|X = x, Y = y), (\Pi'|X = x, Y = y)) \leq \varepsilon$

We want to transmit Alice’s message Π to Bob (who knows Y and his randomness R) in many rounds so that the expected communication length is small. By Theorem 1 this task can be solved with error ε in expected communication

$$H(\Pi|Y, R) + 2\sqrt{H(\Pi|Y, R)} + O\left(\log_2 \frac{1}{\varepsilon}\right). \tag{3}$$

Assume first that the original protocol π uses only public randomness. Then

$$I = I(X : \Pi|Y, R) = H(\Pi|Y, R) - H(\Pi|X, Y, R) = H(\Pi|Y, R).$$

Indeed, $H(\Pi|X, Y, R) = 0$, since Π is defined by X, R . Thus (3) becomes

$$I + 2\sqrt{I} + O\left(\log_2 \frac{1}{\varepsilon}\right)$$

and we are done.

In general case, when the original protocol uses private randomness, I can be much smaller than $H(\Pi|Y, R)$. Fortunately, by the following theorem from [2] we can remove private coins from the protocol with only a slight increase in information complexity.

Theorem 4. *For every one-round protocol π and for every probability distribution μ there is a one-round public-coin protocol π' with information complexity $IC_\mu(\pi') \leq I + \log_2(I + O(1))$ such that for each pairs of inputs (x, y) in the protocol π' Bob outputs Π' for which $\Pi'|X = x, Y = y$ and $\Pi|X = x, Y = y$ are identically distributed.*

Combining this theorem with our main result (Theorem 1), we obtain Theorem 3.

5 A Lower Bounds for the Average-Case Communication

Let (X, Y) be a pair of jointly distributed random variables. Assume that π is a deterministic protocol to transmit X from Alice to Bob who knows Y . Let $\pi(X, Y)$ stand for the result output by the protocol π for input pair (X, Y) . We assume that for at least $1 - \varepsilon$ input pairs this result is correct:

$$\Pr[\pi(X, Y) \neq X] \leq \varepsilon.$$

It is not hard to see that in this case the expected communication length cannot be much less than $H(X|Y)$ bits on average. Moreover, this applies for communication from Alice to Bob only.

Proposition 1. *For every deterministic protocol as above the expected communication from Alice to Bob is at least $H(X|Y) - \varepsilon \log_2 |\mathcal{X}| - 1$.*

The proof of this proposition is omitted due to space constraints.

There are random variables for which this lower bound is tight. For instance, let Y be empty and let X take the value $x \in \{0, 1\}^n$ with probability $\varepsilon/2^n$ (for all such x) and let $X = (\text{the empty string})$ with the remaining probability $1 - \varepsilon$. Then the trivial protocol with no communication solves the job with error probability ε and $H(X|Y) \approx \varepsilon \log_2 |\mathcal{X}|$.

In this section we consider the following question: are there a random variables (X, Y) , for which for every public-coin communication protocol the expected communication is significantly larger than $H(X|Y)$, say close to the upper bound $H(X|Y) + 2\sqrt{H(X|Y)} + \log_2(\frac{1}{\varepsilon})$ of Theorem 1?

Orlitsky [7] showed that if no error is allowed and the support of (X, Y) is a Cartesian product, then every deterministic protocol must communicate $H(X)$ bits on average.

Proposition 2. *Let (X, Y) be a pair of jointly distributed random variables whose support is a Cartesian product. Assume that π is a deterministic protocol, which transmits X from Alice to Bob who knows Y and*

$$\Pr[\pi(X, Y) \neq X] = 0.$$

Then the expected length of π is at least $H(X)$.

This result can be easily generalized to the case when π is public-coin.

The main result of this section states that there are random variables (X, Y) such that transmission of X from Alice to Bob with error probability ε requires $H(X|Y) + \Omega\left(\log_2\left(\frac{1}{\varepsilon}\right)\right)$ bits on average.

The random variables X, Y are specified by two parameters, $\delta \in (0, 1/2)$ and $n \in \mathbb{N}$. Both random variables take values in $\{0, 1, \dots, n\}$ and are distributed as follows: Y is distributed uniformly in $\{0, 1, \dots, n\}$ and $X = Y$ with probability $1 - \delta$ and X is uniformly distributed in $\{0, 1, \dots, n\} \setminus \{X\}$ with the remaining probability δ . That is,

$$\Pr[X = i, Y = j] = \frac{(1 - \delta)\delta_{ij} + \frac{\delta}{n}(1 - \delta_{ij})}{n + 1},$$

where δ_{ij} stands for the Kronecker’s delta. Notice that X is uniformly distributed on $\{0, 1, \dots, n\}$ as well. A straightforward calculation reveals that

$$\Pr[X = i|Y = j] = \frac{\Pr[X = i, Y = j]}{\Pr[Y = j]} = (1 - \delta - \frac{\delta}{n})\delta_{ij} + \frac{\delta}{n}$$

and

$$H(X|Y) = (1 - \delta) \log_2\left(\frac{1}{1 - \delta}\right) + \delta \log_2\left(\frac{n}{\delta}\right) = \delta \log_2 n + O(1).$$

We will think of δ as a constant, say $1/4$. For one-way protocol we are able to show that communication length must be close to $\log n$, which is about $1/\delta$ times larger than $H(X|Y)$:

Proposition 3. *Assume that π is a one-way deterministic protocol, which transmits X from Alice to Bob who knows Y and*

$$\Pr[\pi(X, Y) \neq X] \leq \varepsilon.$$

Then the expected length of π is at least $(1 - \frac{\varepsilon}{\delta}) \log_2(n + 1) - 2$.

This result explains why the one-way one-shot analogue of the Slepian–Wolf theorem is not possible.

Proof. Let S be the number of leafs in π . For each $j \in \{0, 1, \dots, n\}$

$$\#\{i \in \{0, 1, \dots, n\} \mid \pi(i, j) = i\} \leq S.$$

Hence the error probability ε is at least $(n + 1 - S)\frac{\delta}{n}$. This implies that

$$S \geq n\left(1 - \frac{\varepsilon}{\delta}\right) + 1 \geq (n + 1)\left(1 - \frac{\varepsilon}{\delta}\right).$$

Let $\Pi(X)$ denote the leaf Alice and Bob reach in π (since the protocol is one-way, the leaf depends only on X). The expected length of $\Pi(X)$ is at least $H(\Pi)$ (identify each leaf with the binary string, written on the path from the

root to this leaf in the protocol tree; the set of all these strings is prefix-free). Let l_1, l_2, \dots, l_S be the list of all leaves in the support of the random variable $\Pi(X)$. As X is distributed uniformly, we have

$$\Pr[\Pi = l_i] \geq \frac{1}{n+1}$$

for all i . The statement follows from

Lemma 2. *Assume that $p_1, \dots, p_k, q_1, \dots, q_k \in (0, 1)$ satisfy*

$$\sum_{i=1}^k p_i = 1,$$

$$\forall i \in \{1, \dots, k\} \quad p_i \geq q_i.$$

Then

$$\sum_{i=1}^k p_i \log_2 \frac{1}{p_i} \geq \sum_{i=1}^k q_i \log_2 \frac{1}{q_i} - 2.$$

The proof of this technical lemma is omitted due to space constraints. The lemma implies that

$$\begin{aligned} H(\Pi) &= \sum_{i=1}^S \Pr[\Pi = l_i] \log_2 \left(\frac{1}{\Pr[\Pi = l_i]} \right) \\ &\geq \frac{S}{n+1} \log_2(n+1) - 2 \geq \left(1 - \frac{\varepsilon}{\delta}\right) \log_2(n+1) - 2. \end{aligned}$$

The next theorem states that for any fixed δ every two-way public-coin protocol with error probability ε must communicate about $H(X|Y) + (1 - \delta) \log_2(1/\varepsilon)$ bits on average.

Theorem 5. *Assume that π is a public-coin communication protocol which transmits X from Alice to Bob who knows Y and*

$$\Pr[X' \neq X] \leq \varepsilon,$$

where X' denotes the Bob's output and the probability is taken with respect to input distribution and public randomness of π . Then the expected length of π is at least

$$(1 - \delta - \delta/n) \log_2 \left(\frac{\delta}{\varepsilon + \delta/n} \right) + (\delta - 2\varepsilon) \log_2(n+1) - 2\delta.$$

The lower bound in this theorem is quite complicated and comes from its proof. To understand this bound assume that δ is a constant, say $\delta = 1/4$,

and $\frac{1}{n} \leq \varepsilon \leq \frac{1}{\log_2 n}$. Then $H(X|Y) = (1/4) \log_2 n + O(1)$ and the lower bound becomes

$$\left(1 - \frac{1}{4} - \frac{1}{4n}\right) \log_2 \left(\frac{\frac{1}{4}}{\varepsilon + \frac{1}{4n}}\right) + (1/4 - 2\varepsilon) \log_2(n + 1) - \frac{1}{2}$$

Condition $\frac{1}{n} \leq \varepsilon$ implies that the first term is equal to

$$(3/4) \log_2 \left(\frac{1}{\varepsilon}\right) - O(1).$$

Condition $\varepsilon \leq \frac{1}{\log_2 n}$ implies that the second term is equal to

$$(1/4) \log_2 n - O(1).$$

Therefore under these conditions the lower bound becomes

$$(1/4) \log_2 n + (3/4) \log_2 \left(\frac{1}{\varepsilon}\right) - O(1) = H(X|Y) + (3/4) \log_2 \left(\frac{1}{\varepsilon}\right) - O(1).$$

Proof. Let us start with the case when π is deterministic. Let $\Pi = \Pi(X, Y)$ denote the leaf Alice and Bob reach in the protocol π for input pair (X, Y) . As we have seen, the expected length of communication is at least the entropy $H(\Pi(X, Y))$. Let l_1, \dots, l_S denote all the leaves in the support of the random variable $\Pi(X, Y)$. The set $\{(x, y) \mid \Pi(x, y) = l_i\}$ is a combinatorial rectangle $R_i \subset \{0, 1, \dots, n\} \times \{0, 1, \dots, n\}$. Imagine $\{0, 1, \dots, n\} \times \{0, 1, \dots, n\}$ as a table in which Alice owns columns and Bob owns rows. Let h_i be the height of R_i and w_i be the width of R_i . Let d_i stand for the number of diagonal elements in R_i (pairs of the form (j, j)). By definition of (X, Y) we have

$$\Pr[\Pi(X, Y) = l_i] = \frac{(1 - \delta)d_i}{n + 1} + \frac{\delta(h_i w_i - d_i)}{n(n + 1)}. \tag{4}$$

The numbers $\{\Pr[\Pi(X, Y) = l_i]\}_{i=1}^S$ define a probability distribution over the set $\{1, 2, \dots, S\}$ and its entropy equals $H(\Pi(X, Y))$. Equation (4) represents this distribution as a weighted sum of the following distributions: $\left\{\frac{d_i}{n+1}\right\}_{i=1}^S$ and $\left\{\frac{h_i w_i}{(n+1)^2}\right\}_{i=1}^S$. That is, Eq. (4) implies that

$$\{\Pr[\Pi = l_i]\}_{i=1}^S = (1 - \delta - \delta/n) \left\{\frac{d_i}{n + 1}\right\}_{i=1}^S + (\delta + \delta/n) \left\{\frac{h_i w_i}{(n + 1)^2}\right\}_{i=1}^S.$$

Since entropy is concave, we have

$$\begin{aligned} H(\Pi) &= H(\{\Pr[\Pi = l_i]\}_{i=1}^S) \\ &\geq (1 - \delta - \delta/n) H\left(\left\{\frac{d_i}{n + 1}\right\}_{i=1}^S\right) + (\delta + \delta/n) H\left(\left\{\frac{h_i w_i}{(n + 1)^2}\right\}_{i=1}^S\right) \end{aligned} \tag{5}$$

The lower bound of the theorem follows from lower bounds of the entropies of these distributions.

A lower bound for $H\left(\left\{\frac{d_i}{n+1}\right\}_{i=1}^S\right)$. In each row of R_i there is at most 1 element (x, y) , for which $\pi(x, y) = x$. The rectangle R_i consists of d_i diagonal elements and hence there are at least $d_i^2 - d_i$ elements (x, y) in R_i for which $\pi(x, y) \neq x$. Summing over all i we get

$$\varepsilon \geq \sum_{i=1}^S \frac{\delta(d_i^2 - d_i)}{n(n+1)}$$

and thus

$$\sum_{i=1}^S \left(\frac{d_i}{n+1}\right)^2 \leq \frac{\varepsilon + \delta/n}{\delta}.$$

Since Renyi entropy is a lower bound for the Shannon entropy, we have

$$H\left(\left\{\frac{d_i}{n+1}\right\}_{i=1}^S\right) \geq \log_2 \left(\frac{1}{\sum_{i=1}^S \left(\frac{d_i}{n+1}\right)^2} \right) \geq \log_2 \left(\frac{\delta}{\varepsilon + \delta/n} \right).$$

A lower bound for $H\left(\left\{\frac{h_i w_i}{(n+1)^2}\right\}_{i=1}^S\right)$. In R_i , there are at most h_i good pairs (for which π works correctly). At most d_i of them has probability $\frac{1-\delta}{n+1}$. Hence

$$\Pr[\Pi = l_i, \pi(X, Y) = X] \leq \frac{(1-\delta)d_i}{n+1} + \frac{\delta(h_i - d_i)}{n(n+1)}$$

and

$$\begin{aligned} 1 - \varepsilon &\leq \Pr[\pi(X, Y) = X] = \sum_{i=1}^S \Pr[\Pi = l_i, \pi(X, Y) = X] \\ &\leq \sum_{i=1}^S \left(\frac{(1-\delta)d_i}{n+1} + \frac{\delta(h_i - d_i)}{n(n+1)} \right) = 1 - \delta - \delta/n + \frac{\delta}{n(n+1)} \sum_{i=1}^S h_i. \end{aligned}$$

The last inequality implies that

$$\sum_{i=1}^S h_i \geq (1 - \varepsilon/\delta)(n+1)^2.$$

Since $h_i \leq n+1$, we have

$$\begin{aligned} \sum_{i=1}^S \frac{h_i w_i}{(n+1)^2} \log_2 \left(\frac{(n+1)^2}{h_i w_i} \right) &\geq \sum_{i=1}^S \frac{h_i w_i}{(n+1)^2} \log_2 \left(\frac{(n+1)^2}{(n+1)w_i} \right) \\ &= -\log_2(n+1) + \sum_{i=1}^S h_i \frac{w_i}{(n+1)^2} \log_2 \left(\frac{(n+1)^2}{w_i} \right). \end{aligned}$$

Obviously $\frac{w_i}{(n+1)^2} \geq \frac{1}{(n+1)^2}$. By Lemma 2 we get

$$\begin{aligned} \sum_{i=1}^S h_i \frac{w_i}{(n+1)^2} \log_2 \left(\frac{(n+1)^2}{w_i} \right) &\geq \left(\sum_{i=1}^S h_i \right) \frac{1}{(n+1)^2} \log_2 ((n+1)^2) - 2 \\ &\geq (2 - 2\varepsilon/\delta) \log_2(n+1) - 2. \end{aligned}$$

Thus

$$H \left(\left\{ \frac{h_i w_i}{(n+1)^2} \right\}_{i=1}^S \right) \geq (1 - 2\varepsilon/\delta) \log_2(n+1) - 2,$$

and the theorem is proved for deterministic protocols.

Assume now that π is a public-coin protocol with public randomness R and let r be a possible value of R . Let π_r stand for the deterministic communication protocol obtained from π by fixing $R = r$. For any protocol τ let $\|\tau\|$ denote the random variable representing communication length of τ (which may depend on the input and the randomness). Finally, set $\varepsilon_r = \Pr[X' \neq X | R = r]$

Note that π_r transmits X from Alice to Bob with error probability at most ε_r (with respect to input distribution). Since π_r is deterministic, the expected length of π_r is at least:

$$\mathbb{E}\|\pi_r\| \geq (1 - \delta - \delta/n) \log_2 \left(\frac{\delta}{\varepsilon_r + \delta/n} \right) + (\delta - 2\varepsilon_r) \log_2(n+1) - 2\delta.$$

Since $\mathbb{E}_{r \sim R} \varepsilon_r = \varepsilon$ and by concavity of \log :

$$\begin{aligned} \mathbb{E}\|\pi\| &= \mathbb{E}_{r \sim R} \mathbb{E}\|\pi_r\| \\ &\geq \mathbb{E}_{r \sim R} \left[(1 - \delta - \delta/n) \log_2 \left(\frac{\delta}{\varepsilon_r + \delta/n} \right) + (\delta - 2\varepsilon_r) \log_2(n+1) - 2\delta \right] \\ &\geq (1 - \delta - \delta/n) \log_2 \left(\frac{\delta}{\varepsilon + \delta/n} \right) + (\delta - 2\varepsilon) \log_2(n+1) - 2\delta. \end{aligned}$$

References

1. Bauer, B., Moran, S., Yehudayoff, A.: Internal compression of protocols to entropy
2. Braverman, M., Garg, A.: Public vs private coin in bounded-round information. In: Esparza, J., Fraigniaud, P., Husfeldt, T., Koutsoupias, E. (eds.) ICALP 2014. LNCS, vol. 8572, pp. 502–513. Springer, Heidelberg (2014)
3. Braverman, M., Rao, A.: Information equals amortized communication. In: 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science (FOCS), pp. 748–757. IEEE (2011)
4. Brody, J., Buhrman, H., Koucky, M., Loff, B., Speelman, F., Vereshchagin, N.: Towards a reverse newman’s theorem in interactive information complexity. In: 2013 IEEE Conference on Computational Complexity (CCC), pp. 24–33. IEEE (2013)
5. Kushilevitz, E., Nisan, N.: Communication Complexity. Cambridge University Press, Cambridge (2006)

6. Newman, I.: Private vs. common random bits in communication complexity. *Inf. Process. Lett.* **39**(2), 67–71 (1991)
7. Orlitsky, A.: Average-case interactive communication. *IEEE Trans. Inf. Theory* **38**(5), 1534–1547 (1992)
8. Shannon, C.E.: A mathematical theory of communication. *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **5**(1), 3–55 (2001)
9. Slepian, D., Wolf, J.K.: Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory* **19**(4), 471–480 (1973)
10. Weinstein, O.: Information complexity and the quest for interactive compression. *ACM SIGACT News* **46**(2), 41–64 (2015)
11. Yeung, R.W.: *Information Theory and Network Coding*. Springer, New York (2008)