Electronic Colloquium on Computational Complexity, Report No. 62 (2014)

On the role of private coins in unbounded-round Information Complexity

Alexander Kozachinsky*

*Moscow State University, Faculty of Mechanics and Mathematics
kozlach@mail.ru

2014

Abstract

We prove a version of "Reversed Newman Theorem" in context of information complexity: every private-coin communication protocol with information complexity I and communication complexity C can be replaced by public-coin protocol with the same behavior so that it's information complexity does not exceed $O\left(\sqrt{IC}\right)$. This result holds for unbounded-round communication whereas previous results in this area dealt with one-way protocols. As an application it gives an undirect way to prove a best-known compression theorem in Information Complexity.

1 Introduction

Information complexity of communication protocol π, denoted by $IC_\mu(\pi)$, is the amount of information Alice and Bob reveal about their inputs while computing π in a assumption that input are distributed according μ. Information complexity is useful foremost in context of a Direct-Sum problem in Communication complexity. Let us firstly describe the substance of this problem. Fix a small constant ϵ. Suppose that you are given an arbitrary function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and probability distribution μ on the set $\mathcal{X} \times \mathcal{Y}$, (here $\mathcal{X}$ is corresponded to Alice and $\mathcal{Y}$ is corresponded to Bob). Define $D_\epsilon^\mu(f)$ as follows:

$$D_\epsilon^\mu(f) = \inf_\pi CC(\pi),$$

where infimum ranges over all deterministic communication protocols π which output 1 bit $\pi(x,y)$, such that $\mu\left\{(x,y) \,|\, \pi(x,y) \neq f(x,y)\right\} \leq \epsilon$. Imagine then, that you task is to compute n copies of f in parallel. Consider function $f^n : (\mathcal{X} \times \mathcal{Y})^n \to \{0,1\}^n$ and probability distribution μ^n on the set $(\mathcal{X} \times \mathcal{Y})^n$, which are defined as follows:

$$f^n\left((x_1,y_1),\ldots,(x_n,y_n)\right) = \left(f(x_1,y_1),\ldots,f(x_n,y_n)\right),$$

ISSN 1433-8092

$$\mu^n\left((x_1,y_1),\dots,(x_n,y_n)\right) = \mu(x_1,y_1)\times\dots\times\mu(x_n,y_n).$$

Here we define $D_\epsilon^{n,\mu^n}(f^n)$:

$$D_\epsilon^{n,\mu^n}(f^n) = \inf_\pi CC(\pi),$$

where infimum ranges over all deterministic communication protocols π which output n bits $\pi_1(x,y),\dots,\pi_n(x,y)$ such that for every i the following holds: $\mu^n\{(x,y)\,|\,\pi_i(x,y)\neq(f^n(x,y))_i\}\leq\epsilon$.

The Direct-Sum question is the question whether $nD_\epsilon^\mu(f)$ and $D_\epsilon^{n,\mu^n}(f^n)$ can considerably differ or not. It is easy to prove that: $nD_\epsilon^\mu(f)\geq D_\epsilon^{n,\mu^n}(f^n)$. You can just get the optimal protocol in the sense of definition for $D_\epsilon^\mu(f)$ and apply it n times. Unfortunately, the opposite inequality is not that trivial.

What strategy can we design in attempt to prove the opposite inequality? Consider the optimal protocol π for f^n in the sense of definition for $D_\epsilon^{n,\mu^n}(f^n)$, so that $CC(\pi)=D_\epsilon^{n,\mu^n}(f^n)$. Using information-theoretic technique, described in [2], you can convert π into the randomized protocol τ computing f, which satisfy following inequalities: $IC_\mu(\tau)\leq\frac{CC(\pi)}{n}$, $CC(\tau)\leq CC(\pi)$ and:

$$\Pr[\tau(x,y)\neq f(x,y)]\leq\epsilon,$$

where probability in the last inequality is taken from distribution μ and the inner randomness of the protocol. Suppose that you are given some numerical function $\phi(I,C)$. Consider the following statement:

For every protocol α which computes function g over the distribution μ with error probability ϵ there exists protocol α' which computes g over distribution μ with error probability 2ϵ such that $CC(\alpha')=O\left(\phi(IC_\mu(\alpha),CC(\alpha)\right)$

Figure 1.1: Comression statement for ϕ

It is not hard to see that compression statement for ϕ implies following inequality for Direct-Sum Problem:[1]

$$D_{2\epsilon}^\mu(f)=O\left(\phi\left(\frac{D_\epsilon^{n,\mu^n}(f^n)}{n},D_\epsilon^{n,\mu^n}(f^n)\right)\right).$$

In order to reach this result you have just to convert τ into the protocol with the properties stated in compression statement for ϕ and then make it deterministic by fixing an optimal choice of random bits.

The following theorem was proved in [1]:

Theorem 1.1. *Compression statement holds for $\phi(I,C)=\sqrt{IC}\log(C)$.*

[1] It is not evident how to decrease error from 2ϵ to ϵ. Instead you may consider the similar inequality stated for $R_\epsilon(f)$, which can be derived using mini-max argument; since $R_{2\epsilon}(f)=\Omega\left(R_\epsilon(f)\right)$ it is not a problem.

Automatically it implies that $D_{\epsilon}^{n,\mu^n}(f^n) = \Omega\,(\sqrt{n}D_{2\epsilon}^{\mu}(f))$(up to logarithmic factor). Next result, proved in [4], gives an improvement of the previous theorem, but with some restriction:

Theorem 1.2. *Compression statement holds for public-coin protocols with* $\phi(I, C) = I\log(C)$.

Unfortunately protocol τ is reached using public coins as well as private coins. We can try to circumvent this confuse considering the problem of simulation private-coin protocol using public-coin protocols.

As marked in [4], this problem is reverse in some sense to the Newman theorem, which states that every public-coin communication protocol can be effectively simulated by private-coin protocol(for the details look in the [6]). Note that in the case of Communication Complexity public coins are more powerful tool than private coins. With respect to Information Complexity it is not true. For example, suppose that Alice receives binary string of length n and privately flips n coins; then she sends to Bob the bit-wise XOR of her input and coins. Information that Bob receives about Alice's input from the message is equal to zero unless Alice's coins are available to Bob; otherwise, Bob can reestablish Alice's input from the message.

We say that two protocols are *distributional-equivalent* if they are defined on the same input space $\mathcal{X} \times \mathcal{Y}$ and for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$ their transcripts, conditioned on (x, y), are same distributed. For every private-coin protocol π with information complexity I out task is to find public-coin protocol with information complexity close to I, which is distributional equivalent to π. First results in this setting were proved in [4] and [3] for the bounded-round protocols. The following estimate, proven in [3], is tight for one-way communication:

Theorem 1.3. *For every one-way private-coin communication protocol π and every distribution μ there exists public-coin communication protocol τ which is distributional-equivalent to π, satisfying following inequality:*

$$IC_{\mu}(\tau) \leq IC_{\mu}(\pi) + \log(IC_{\mu}(\pi)) + O(1).$$

Constant in right-hand side is crucial when you try to generalize last theorem for unbounded-round case. Our contribution is the estimate, which holds for all protocols:

Theorem 1.4. *There exists universal constant $C > 0$ such that for every private-coin protocol π there exists public-coin protocol τ which is distributional-equivalent to π such that for every distribution μ the following holds:*

$$IC_{\mu}(\tau) \leq C\sqrt{IC_{\mu}(\pi)CC(\pi)}.$$

This theorem gives also new, undirect way to prove compression result, stated in theorem 1.1 ; use our result to remove private coins from the protocol and then apply theorem 1.2.

Technique, that we use to bound information complexity of the protocol after removing private-coins, is not new. The key considiration is a fact that on each step of the protocol total variation between Alice's a priori distribution of the next bit in protocol and Bob's one can be bounded, using Pinsker's inequality, by the term related with information complexity of this step. It is worth noting that proof of the theorem 1.1, which contains in [1], uses this fact to bound a number of mistakes in simulating of an original protocol.

2 Preliminaries

We denote logarithms in base 2 by log and natural logarithms by ln.

2.1 Information theory

We use standart notion of a Shannon Entropy; if X is a random variable, taking values in the set $\mathcal{X}$, then:

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

Conditional entropy can be defined as follows:

$$H(X|Y) = H(X, Y) - H(Y).$$

Also it can be defined as expectation value $H(X|Y) = E_{Y=y} H(X|Y = y)$, where $X|Y = y$ denote a random variable which distribution is equal to distribution of X, condition on the event $Y = y$. Mutual information between two random variables defined as follows:

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

Mutual information is symmetric: $I(X : Y) = I(Y : X)$; it follows from the well-known fact that $H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$. Conditional mutual information is defined in the same way:

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

Entropy and the mutual information satisfy the chain rule:

Proposition 2.1 (Chain Rule)**.**

$$H(X_1 \dots X_n) = H(X_1) + \sum_{i=2}^{n} H(X_i|X_1 \dots X_{i-1}),$$

$$I(X_1 \dots X_n : Y) = I(X_1 : Y) + \sum_{i=2}^{n} I(X_i : Y|X_1 \dots X_{i-1}).$$

Chain rule holds also for conditional entropy and conditional mutual information.

Let P, Q denote two probability distributions on the set W. We consider two quantities that measure dissimilarity between P and Q: *total variation*:

$$\delta(P,Q) = \sup_{A \subset W} |P\{A\} - Q\{A\}|,$$

and the *information divergence*

$$D(P||Q) = \sum_{w \in W} P(w) \log\left(\frac{P(w)}{Q(w)}\right).$$

We will use the following well-known inequality:

Proposition 2.2 (Pinkser's inequality)**.**

$$\delta(P,Q) \leq \sqrt{\frac{D(P||Q)}{2}}.$$

When α is a number we use letter H also to denote the following function:

$$H(\alpha) = \alpha \log\left(\frac{1}{\alpha}\right) + (1-\alpha)\log\left(\frac{1}{1-\alpha}\right)$$

$H(\alpha)$ is equal to entropy of a random variable ξ with two possible values $\{w_1, w_2\}$ such that $\Pr[\xi = w_1] = \alpha$. We will use the following fact:

Fact 2.1. *If* $\alpha \leq \frac{1}{2}$, *then* $H(\alpha) \leq 2\alpha \log\left(\frac{1}{\alpha}\right)$

Proof. It is sufficient to show that $(1-\alpha)\log\left(\frac{1}{1-\alpha}\right) \leq \alpha \log\left(\frac{1}{\alpha}\right)$ when $\alpha \leq \frac{1}{2}$. Consider function $f(\alpha) = \alpha \log\left(\frac{1}{\alpha}\right) - (1-\alpha)\log\left(\frac{1}{1-\alpha}\right)$. We have:

$$f'(\alpha) = \frac{1}{\ln(2)}\left(\ln\left(\frac{1}{\alpha(1-\alpha)}\right) - 2\right).$$

It means that f grows on $[0, \alpha_0]$, (α_0 is a left root of equation $\alpha(1-\alpha) = \frac{1}{e^2}$); respectively f decrease on $[\alpha_0, \frac{1}{2}]$. Since $f(0) = f(\frac{1}{2}) = 0$ it means that $f(\alpha) \geq 0$ for any $\alpha \in [0, \frac{1}{2}]$. □

2.2 Communication Protocols

In order to prove our result we have to give a formal definition of private-coin communication protocol. Let $\mathcal{Z}$ be the set of the possible outputs and let $\delta : \{0,1\}^* \to \{A, B\} \cup \mathcal{Z}$ be the function which decides who's turn to communicate unless it's time to produce output. This function determines three sets $\mathcal{A} = \{s \in \{0,1\}^* \,|\, \delta(s) = A\}$, $\mathcal{B} = \{s \in \{0,1\}^* \,|\, \delta(s) = B\}$, $\mathcal{O} = \{s \in \{0,1\}^* \,|\, \delta(s) \in \mathcal{Z}\}$. Finally let $p : \mathcal{X} \times \mathcal{A} \to [0,1]$ and $q : \mathcal{Y} \times \mathcal{B} \to [0,1]$ be the

1. Alice receives $x \in \mathcal{X}$, Bob receives $y \in \mathcal{Y}$; they add some bits to the string s, starting with empty string $s = \lambda$;
2. If $s \in \mathcal{A}$, Alice uses her private randomness to produce one bit b with probability for b to be 0 equal to $p(x, s)$; then Alice and Bob add b to s;
3. If $s \in \mathcal{B}$, Bob acts similarly to Alice;
4. If $s \in \mathcal{O}$, Alice and Bob output $\delta(s)$ and terminate.

Figure 2.2: Private-coin protocol run-time

functions which instruct Alice and Bob how to communicate. In the figure 2.2 we describe how private-coin communication protocol proceed.

We say that protocol is deterministic, if values of functions p and q lie in $\{0, 1\}$. We define public-coin protocol as a random variable R taking values in set of deterministic protocols. Concatenation of all bits Alice and Bob send to each other is called *transcript* of the protocol π; the maximum length of the transcript in protocol π is called communication complexity of the protocol π, denoted by $CC(\pi)$.

For the formal definition of communication complexity of functions and for classic results in this area, see the book [5].

2.3 Information Complexity

Suppose that you are given a communication protocol π and suppose that it's input space $\mathcal{X} \times \mathcal{Y}$ is distributed according μ. Transcript of the protocol π from this point becomes a random variable which distribution depends on μ and inner randomness of the protocol. Denote this random variable by Π. We define information complexity of the protocol π as follows:

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

The following fact proved in [2]

Proposition 2.3. *$IC_\mu(\pi) \leq CC(\pi)$.*

If π is public-coin, $IC_\mu(\pi)$ can be represented in a shorter form:

Proposition 2.4. *If π is a public-coin protocol, then:*

$$IC_\mu(\pi) = H(\Pi|R, Y) + H(\Pi|R, X).$$

Proof. We have:

$$\begin{aligned} I(X : \Pi, R|Y) &= H(\Pi, R|Y) - H(\Pi, R|X, Y) \\ &= H(\Pi|R, Y) + H(R|Y) - H(\Pi|R, X, Y) - H(R|X, Y). \end{aligned}$$

$H(\Pi|R,X,Y)=0$, since Π is determined by R,X,Y; $H(R|Y)=H(R|X,Y)=H(R)$ since R and X,Y are independent. Hence $I(X:\Pi,R|Y)=H(\Pi|R,Y)$. Similarly we proof that $I(Y:\Pi,R|X)=H(\Pi|R,X)$. □

3 Simulation for one-bit protocols

In this section we prove theorem 1.4 for one-way protocols of depth 1. It means that Alice receives her input and sends just 1 bit to Bob, using random bits, after what protocol terminates.

Proposition 3.1. *There exists a universal constant $C>0$, such that for every one-bit private-coin protocol π there exists one-bit public-coin protocol τ which is distributional-equivalent to π, such that for every distribution μ the following holds:*

$$IC_\mu(\tau)\le C\sqrt{IC_\mu(\pi)}.$$

Proof. Suppose that we are given private-coin protocol π of depth 1; it means that there is some set $\mathcal{X}$ of Alice's inputs and there is a function $p:\mathcal{X}\to[0,1]$ such that on input $x\in\mathcal{X}$ Alice send 0 to Bob with probability $p(x)$. Also we are given probability distribution μ on the set $\mathcal{X}$ which defines random variable X.

Let B denote the Alice's message in protocol π. Let Q denote the distribution of B and let P_x denote the distribution of $B|X=x$. It is easy to to see that:

$$Q=\sum_{x\in\mathcal{X}}\mu(x)P_x.$$

We define public-coin protocol τ as follows:

1. Alice receives value x of a random variable X;
2. Alice and Bob publicly sample R uniformly at random from $[0,1]$;
3. If $R\le p(x)$, then Alice sends $B=0$ to Bob; otherwise, Alice sends $B=1$ to Bob.

Note that τ does not depends on μ. It is clear that Alice's message B, conditioned on $X=x$, is distributed according to P_x. Hence τ is distributional-equivalent to π. Let $B(x,t)$ denote Alice's message in protocol τ conditioned on $X=x$ and $R=t$. The only thing we have to do from now is to estimate Information Complexity of τ. By proposition 2.4 we have $IC_\mu(\tau)=H(B|R)$. By definition we have:

$$H(B|R)=\int_0^1 H(B|R=t)dt.$$

Set $I=IC_\mu(\pi)$. We will use the following fact:

Fact 3.1. *$I = E_{x\leftarrow\mu} D(P_x||Q)$.*

Proof.

$$\begin{aligned}
I &= I(X:B) = H(B) - H(B|X)\\
&= \sum_{b\in\{0,1\}} Q(b)\log\left(\frac{1}{Q(b)}\right) - \sum_{x\in\mathcal{X}}\mu(x)\sum_{b\in\{0,1\}} P_x(b)\log\left(\frac{1}{P_x(b)}\right)\\
&= \sum_{b\in\{0,1\}}\left(\sum_{x\in\mathcal{X}}\mu(x)P_x(b)\right)\log\left(\frac{1}{Q(b)}\right) - \sum_{x\in\mathcal{X}}\mu(x)\sum_{b\in\{0,1\}} P_x(b)\log\left(\frac{1}{P_x(b)}\right)\\
&= \sum_{x\in\mathcal{X}}\mu(x)\sum_{b\in\{0,1\}} P_x(b)\log\left(\frac{P_x(b)}{Q(b)}\right) = E_{x\leftarrow\mu}D(P_x||Q)
\end{aligned}$$

□

By Pinsker's inequality (proposition 2.2) we have: $\delta(P_x, Q) \le \sqrt{D(P_x||Q)/2}$. Using this inequality and fact 3.1 we get:

$$E_{x\leftarrow\mu}\delta^2(P_x, Q) \le I/2. \tag{1}$$

Consider set Ω, which is defined as follows:

$$\Omega = \{t \in [0,1] \,|\, |t - Q(0)| > \sqrt{I}\}.$$

It is clear that $\Pr[R \in [0,1]/\Omega] \le 2\sqrt{I}$. Trivially we conclude that:

$$\int\limits_{[0,1]/\Omega} H(B|R=t)dt \le 2\sqrt{I}. \tag{2}$$

WLOG we assume that $\mathcal{X}$ contains some special element a with the following properties: $\mu(a) = 0$ and $p(a) = Q(0)$. Fix $t \in \Omega$. The following statement holds:

$$H(B|R=t) = H\left(\mu\{x \,|\, B(x,t) \ne B(a,t)\}\right).$$

Let us show that if $B(x,t) \ne B(a,t)$, then $\delta(P_x, Q) \ge |t - Q(0)|$. Suppose that $B(x,t) \ne B(a,t)$; by definition of τ it implies that t lies on the segment between $P_x(0)$ and $Q(0)$; hence $\delta(P_x, Q) \ge |P_x(0) - Q(0)| \ge |t - Q(0)|$.By that and by Markov's inequality applied to 1 we get:

$$\begin{aligned}
\mu\{x \,|\, B(x,t) \ne B(a,t)\} &\le \mu\{x \,|\, \delta(P_x, Q) \ge |t - Q(0)|\}\\
&\le \frac{I}{2(t - Q(0))^2} \le \frac{1}{2}.
\end{aligned}$$

Using an estimate from fact 2.1, we derive:

$$\begin{aligned}
\int\limits_{\Omega} H(B|R=t)dt &\leq \int\limits_{\Omega} H\left(\frac{I}{2(t-Q(0))^2}\right) dt \\
&\leq 2\int\limits_{\Omega} \frac{I}{2(t-Q(0))^2}\log\left(\frac{2(t-Q(0))^2}{I}\right) dt \\
&\leq \sqrt{2}\sqrt{I}\int\limits_{\Omega} \frac{I}{2(t-Q(0))^2}\log\left(\frac{2(t-Q(0))^2}{I}\right) d\frac{\sqrt{2}(t-Q(0))}{\sqrt{I}} \\
&\leq \sqrt{2}\int\limits_{|y|>\sqrt{2}} \frac{\log(y^2)}{y^2}dy\sqrt{I}.
\end{aligned}$$

We showed that

$$\int\limits_{\Omega} H(B|R=t)dt \leq D\sqrt{I}, \tag{3}$$

where $D = \sqrt{2}\int\limits_{|y|>\sqrt{2}} \frac{\log(y^2)}{y^2}dy$ (it is correct to define D this way since integral is convergent). Fixing $C = 2 + D$ we conclude from 2 and 3:

$$\begin{aligned}
IC_{\mu}(\tau) = H(B|R) &= \int\limits_{0}^{1} H(B|R=t)dt \\
&= \int\limits_{[0,1]/\Omega} H(B|R=t)dt + \int\limits_{\Omega} H(B|R=t)dt^{\cdot} \\
&\leq (2+D)\sqrt{I} = C\sqrt{I},
\end{aligned}$$

□

Here we give an example from [3], which shows that our bound for one-bit protocols is tight. Suppose that Alice receives 0 or 1 with the same probability and then sends one bit to Bob, which is equal to her input with probability $\frac{1}{2}+\epsilon$ and differs with probability $\frac{1}{2}-\epsilon$. The proportion of those random bits, on which Alice always sends her input to Bob, is at least $1-2(\frac{1}{2}-\epsilon) = 2\epsilon$. Hence information complexity of every public-coin protocol for this task is at least 2ϵ. At the same time simple calculations show that if random bits are private, then information complexity drops to $\Theta(\epsilon^2)$.

4 Generalizaton for all protocols

In this section we extend the result of the previous section to all protocols.

Proof of theorem 1.4. Suppose that π is arbitrary private-coin communication protocol defined in figure 2.2. Set $N = CC(\pi)$ and let $\Pi = \Pi_1 \dots \Pi_N$ denote transcript of π. Set $\Pi_{<k} = \Pi_1 \dots \Pi_{k-1}$. Also we are given some probability distribution μ on $\mathcal{X} \times \mathcal{Y}$ which defines two random variables X an Y. Here we define public-coin protocol τ:

1. Alice receives value x of random variable X, Bob receives value y of a random variable Y; they add some bits to the string s, starting with empty string $s = \lambda$;

2. Alice and Bob publicly sample $R = (R_1, \dots, R_N)$ uniformly at random from $[0,1]^N$;

3. If $s \in \mathcal{A}$, Alice produces one bit b: if $R_{|s|+1} \leq p(x,s)$ then $b = 0$, otherwise $b = 1$; after that Alice and Bob add b to s;

4. If $s \in \mathcal{B}$, Bob acts similarly to Alice;

5. If $s \in \mathcal{O}$, Alice and Bob output $\delta(s)$ and terminate.

Note that τ does not depend on μ. By definition τ is distributional-equivalent to π. By chain rule(proposition 2.1), applied to protocol π we have:

$$\begin{aligned} IC_\mu(\pi) &= I(X : \Pi | Y) + I(Y : \Pi | X) \\ &= \sum_{i=1}^{N} I(X : \Pi_k | Y, \Pi_{<k}) + I(Y : \Pi_k | X, \Pi_{<k}) \\ &= \sum_{i=1}^{N} I_k, \end{aligned}$$

where $I_k = I(X : \Pi_k | Y, \Pi_{<k}) + I(Y : \Pi_k | X, \Pi_{<k})$. By proposition 2.4 we have $IC_\mu(\tau) = H(\Pi | R, Y) + H(\Pi | R, X)$. By chain rule we get:

$$\begin{aligned} IC_\mu(\tau) &= H(\Pi | R, Y) + H(\Pi | R, X) \\ &= \sum_{i=1}^{N} H(\Pi_k | R, Y, \Pi_{<k}) + H(\Pi_k | R, X, \Pi_{<k}) \\ &= \sum_{i=1}^{N} I'_k, \end{aligned}$$

where $I'_k = H(\Pi_k | R, Y, \Pi_{<k}) + H(\Pi_k | R, X, \Pi_{<k})$.

Fix arbitrary $k \in \{1, \dots, N\}$, $s \in \{0,1\}^{k-1}$. WLOG it is Alice turn to communicate, that is $\delta(s) = A$. Hence for arbitrary $x \in \mathcal{X}$ we have

$$H(\Pi_k | R, X = x, \Pi_{<k} = s) = 0,$$

$$I(Y : \Pi_k | X = x, \Pi_{<k} = s) = 0.$$

Fix then arbitrary $y \in \mathcal{Y}$. s and y define private-coin protocol π' of depth 1: according to π' Alice acts as though she were running protocol π from the point when message story is equal to s. Similarly public-coin protocol τ' of depth 1 can be defined from protocol τ. Note that protocol τ' with respect to π' is exactly the same to one that was constructed in proof of proposition 3.1. We derive

$$\begin{aligned} H\left(\Pi_k | R_k, Y = y, \Pi_{<k} = s\right) &= IC_\mu(\tau') \\ &\leq C\sqrt{IC_\mu(\pi')} \\ &= C\sqrt{I\left(X : \Pi_k | Y = y, \Pi_{<k} = s\right)}. \end{aligned}$$

After averaging over all $s \in \{0,1\}^{k-1}$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$ by concavity of root we get:

$$\begin{aligned} I'_k &= H\left(\Pi_k | R, Y, \Pi_{<k}\right) + H\left(\Pi_k | R, X, \Pi_{<k}\right) \\ &\leq H\left(\Pi_k | R_k, Y, \Pi_{<k}\right) + H\left(\Pi_k | R_k, X, \Pi_{<k}\right) \\ &\leq C\sqrt{I\left(X : \Pi_k | Y, \Pi_{<k}\right) + I\left(Y : \Pi_k | X, \Pi_{<k}\right)} = C\sqrt{I_k}. \end{aligned}$$

Using Cauchy–Schwarz inequality we conclude

$$\begin{aligned} IC_\mu(\tau) &= I'_1 + \ldots + I'_N \\ &\leq C\left(\sqrt{I_1} + \ldots + \sqrt{I_N}\right) \\ &\leq C\sqrt{(I_1 + \ldots + I_N)N} = C\sqrt{IC_\mu(\pi) CC(\pi)}. \end{aligned}$$

□

References

[1] Barak, B., Braverman, M., Chen, X., and Rao, A. How to compress interactive communication. *SIAM Journal on Computing 42*, 3 (2013), 1327–1363.

[2] Braverman, M. Interactive information complexity. In *Proceedings of the 44th symposium on Theory of Computing* (2012), ACM, pp. 505–524.

[3] Braverman, M., and Garg, A. Public vs private coin in bounded-round information.

[4] Brody, J., Buhrman, H., Koucky, M., Loff, B., Speelman, F., and Vereshchagin, N. Towards a reverse newman's theorem in interactive information complexity. In *Computational Complexity (CCC), 2013 IEEE Conference on* (2013), IEEE, pp. 24–33.

[5] Kushilevitz, E., and Nisan, N. *Communication complexity.* Cambridge university press, 2006.

[6] Newman, I. Private vs. common random bits in communication complexity. *Information processing letters 39*, 2 (1991), 67–71.