# A new proof Ahlswede — Gács — Körner theorem on common information

Nikolai Vereshchagin
Moscow State University
Email: ver@mccme.ru

September 9, 2002

Let $\xi, \eta$ be random variables having a joint distribution, with values in finite alphabets $S$ and $T$, respectively. Consider the bipartite graph $G$ with left nodes from $S$ and right nodes from $T$, where we put an edge between nodes $s$ and $t$ if $\Pr[\xi = s, \eta = t] > 0$. The pair $(\xi, \eta)$ is called *irreducible* if this graph is connected.

Let $\alpha, \beta$ be the results of $n$ independent trials of $\xi, \eta$, respectively. We want to encode $\alpha, \beta$ so that its code consists of three parts $u, v, w$. The parts $u, v$ encode $\alpha$ and the parts $w, v$ encode $\beta$. We are interested in how large should be the length of $u, v, w$ in order to make this possible. More specifically define $R_n^\varepsilon$ (the rate region) as the set of all triples $\langle k, l, m\rangle$ such that there are encoding functions

$$u : S \times T \to \{0,1\}^k, \quad v : S \times T \to \{0,1\}^l, \quad w : S \times T \to \{0,1\}^m$$

and decoding functions

$$f\colon \{0,1\}^{k+l} \to S, \quad g\colon \{0,1\}^{m+l} \to T$$

such that with probability at least $\varepsilon$ we have

$$f(u(\alpha,\beta), v(\alpha,\beta)) = \alpha, \quad g(w(\alpha,\beta), v(\alpha,\beta)) = \beta.$$

**Theorem 1.** *For any $\varepsilon > 0$ and for any irreducible $\xi, \eta$ there is $c$ such that for any $n$ any triple $\langle k, l, m\rangle$ in the rate region $R_n^\varepsilon$ satisfies the inequality*

$$\min\{H(\alpha)-k, H(\beta)-m\}+\max\{H(\alpha)-k, H(\beta)-m\}/c\log n < l+c\sqrt{n}+c. \quad (1)$$

Essentially this theorem appeared in a stronger form in [1] and later in [3]. However the technique used in our proof is different from that of [1] and [3]. Namely, we use Markov chains (first used in this context in [2]). We hope that our proof is more clear than the proof in [2] (of a weaker assertion).

Note that the theorem implies that $\alpha, \beta$ have no common information in the sense of Gács and Körner [2]. Indeed, assume that $k + l$ is close to $H(\alpha)$ and $m + l$ is close to $H(\beta)$, say $k + l \le H(\alpha) + O(\sqrt{n})$ and $m + l \le H(\beta) + O(\sqrt{n})$. Then the inequality (1) implies $l \le O(\sqrt{n}\log n)$.

*Proof.* The proof is based on the following property of the distribution of $\langle \alpha, \beta \rangle$.

**Lemma.** For any $A \subset S^n$ and any $B \subset T^n$ we have

$$\Pr[\alpha \in A, \beta \in B] \le (\Pr[\alpha \in A] + \Pr[\beta \in B])(3 \min\{\Pr[\alpha \in A], \Pr[\beta \in B]\})^{1/N}$$

for some $N \le c \log n$ where $c$ depends only on $\varepsilon, \xi, \eta$.

Assume the lemma. Call a string $a \in S^n$ *typical* if for any letter $s \in S$ the difference between the fraction of $s$ in $a$ and $\Pr[\xi = s]$ is less than $d\sqrt{n}$. The value $d$ (depending on $\varepsilon, \xi, \eta$) is chosen so that $\langle \alpha, \beta \rangle$ is typical with probability at least $1 - \varepsilon/4$. For any typical $a$ the probability that $\alpha = a$ is $2^{-H(\alpha)+O(\sqrt{n})}$ (the constant in $O(\sqrt{n})$ depends on $\varepsilon, \xi, \eta$). Define typical strings in $T^n$ in the same way.

Fix a string $\bar{v}$ of length $l$ and let $A_{\bar{v}}$ be the set of all $f(\bar{u}, \bar{v})$ for $\bar{u} \in \{0,1\}^k$. In a similar way let $B_{\bar{v}}$ be the set of all $g(\bar{w}, \bar{v})$ for $\bar{w} \in \{0,1\}^m$. Assume that the triple $\langle k, l, m \rangle$ belongs to the region $R_n^\varepsilon$. Then with probability at least $\varepsilon$ there is a string $\bar{v}$ of length $l$ such that $\alpha \in A_{\bar{v}}$ and $\beta \in B_{\bar{v}}$ (let $\bar{v} = v(\alpha, \beta)$). Therefore, with probability at least $\varepsilon/2$ there is $\bar{v}$ such that $\alpha \in A_{\bar{v}}$, $\beta \in B_{\bar{v}}$ and both $\alpha, \beta$ are typical.

On the other hand the sets $A_{\bar{v}}$ and $B_{\bar{v}}$ have at most $2^k$ and $2^m$ elements respectively. Hence

$$\Pr[\alpha \in A_{\bar{v}} \text{ and } \alpha \text{ is typical}] \le 2^{k-H(\alpha)+O(\sqrt{n})},$$
$$\Pr[\beta \in B_{\bar{v}} \text{ and } \beta \text{ is typical}] \le 2^{m-H(\beta)+O(\sqrt{n})}.$$

By the lemma the probability that $\alpha \in A_{\bar{v}}$, $\beta \in B_{\bar{v}}$ and both $\alpha$ and $\beta$ are typical is at most

$$2^{\max\{k-H(\alpha), m-H(\beta)\} + \min\{k-H(\alpha), m-H(\beta)\}/N + O(\sqrt{n})}.$$

Multiplying this bound by the number of $v$'s we should get a number at least $\varepsilon/2$:

$$\varepsilon/2 \le 2^{l + \max\{k-H(\alpha), m-H(\beta)\} + \min\{k-H(\alpha), m-H(\beta)\}/N + O(\sqrt{n})}.$$

Taking the logarithm we obtain the inequality (1).

So it remains to prove the lemma. To this end consider elements of the set $S \cup T$ as states of a Markov chain. Let the probability of transition $s \to t$ be equal to $\Pr[\eta = t | \xi = s]$ and symmetrically let the probability of transition $t \to s$ be equal to $\Pr[\xi = s | \eta = t]$. Consider the $n$th power of this chain. (No transitions from $S$ to $T$ and vice versa are possible.) More specifically, the set of states of the product chain is $S^n \cup T^n$ and the probability of transition $\langle s_1, \ldots, s_n \rangle \to \langle t_1, \ldots, t_n \rangle$ is equal to the product of probabilities of transitions $s_1 \to t_1$, $\ldots$, $s_n \to t_n$. The crucial point is that the product chain is very oblivious: starting in any state $a \in S^n$ we get into a state $a' \in S^n$ after $O(\log n)$ steps with probability close to $\Pr[\alpha = a']$. Namely, there is a $c$ depending on $\xi, \eta$ such that for any even $N$, $N \ge c \log n$, the following holds. For any $a, a' \in S^n$ the probability to get from $a$ to $a'$ in $N$ steps is at most $3 \Pr[\alpha = a']$ and for any $b, b' \in T^n$ the probability to get from $b$ to $b'$ in $N$ steps is at most

$3\Pr[\beta = b']$. (The proof of this fact should be well known. Here is a sketch. Let $a = \langle s_1, \ldots, s_n \rangle$ and $a' = \langle s'_1, \ldots, s'_n \rangle$. The probability to get from $a$ to $a'$ in $N$ steps is equal to the product of probabilities to get from $s_i$ to $s'_i$ in $N$ steps (for $i = 1, \ldots, n$). As the pair $\langle \xi, \eta \rangle$ is irreducible, the probability to get from $s_i$ to $s'_i$ in $N$ steps is upper bounded by $\Pr[\xi = s'_i](1 + 2^{-fN})$ where $f$ depends on $\xi, \eta$ only. Let $c = 1/f$ so that $1 + 2^{-fc \log n} = 1 + 1/n$. Then the probability to get from $a$ to $a'$ in $N$ steps is at most $(1 + 1/n)^n < 3$ times bigger than the product of probabilities $\Pr[\xi = s'_i]$, which is equal to $\Pr[\alpha = a']$.)

Now we will use this fast convergence property of the product chain to upper bound $\Pr[\alpha \in A, \beta \in B]$ in terms of $\Pr[\alpha \in A]$, $\Pr[\beta \in B]$.

Assume that $\Pr[\alpha \in A] \leq \Pr[\beta \in B]$ (the other case is entirely similar). Starting with $A' = B' = \emptyset$ choose any node $a \in A \setminus A'$ such that the probability of transition $a \to B \setminus B'$ is at most $\gamma$, for $\gamma = (3\Pr[\alpha \in A])^{1/N}$. Include the chosen node into $A'$. If there is no such $a'$, choose any node $b \in B \setminus B'$ such that the probability of transition $b \to A \setminus A'$ is at most $\gamma$. Repeat this while there is an appropriate $a$ or there is an appropriate $b$. We claim that at the end we have $A' = A$ and $B' = B$. First observe that $A = A'$ if and only if $B = B'$. So assume that $A \setminus A'$ is not empty and $a \in A \setminus A'$. Then with probability more than $\gamma^N$ we get from $a$ in $N$ steps to $A \setminus A'$. On the other hand, by fast convergence property the probability of getting from $a$ to $A \setminus A'$ in $N$ steps is at most $3\Pr[\alpha \in A \setminus A'] \leq 3\Pr[\alpha \in A]$. This contradicts the choice of $\gamma$.

Thus it suffices to proof that at any step we have

$$\Pr[\langle \alpha, \beta \rangle \in A' \times B \cup A \times B'] \leq \gamma(\Pr[\alpha \in A'] + \Pr[\beta \in B']).$$

At the start this is obviously true. When we add a new $a$ in $A'$ then the set $A' \times B \cup A \times B'$ increases by the set $\{a\} \times (B \setminus B')$ and the probability that a pair $\langle \alpha, \beta \rangle$ gets into the latter set is equal to

$$\Pr[\alpha = a] \Pr[\beta \in B \setminus B' | \alpha = a] \leq \gamma \Pr[\alpha = a].$$

Hence adding new node $a$ in $A'$ increases the probability $\Pr[\langle \alpha, \beta \rangle \in A' \times B \cup A \times B']$ by at most $\gamma \Pr[\alpha = a]$. In a similar way we can prove that adding a new node $b$ in $B'$ increases the probability of $\Pr[\langle \alpha, \beta \rangle \in A' \times B \cup A \times B']$ by at most $\gamma \Pr[\beta = b]$. □

# References


[1] R. Ahlswede, J. Körner, On the connection between the entropies of input and output distributions of discrete memoryless channels, *Proceedings of the 5th Brasov Conference on Probability Theory, Brasov, 1974*. Editura Academiei, Bucuresti, 1977, 13–23

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

[3] Yu. Makarychev. Graduate thesis, 2001.