

NOTES ON USE OF GENERALIZED ENTROPIES IN COUNTING

ALEXEY E. RASTEGIN

ABSTRACT. We address an idea of applying generalized entropies in counting problems. First, we consider some entropic properties that are essential for such purposes. Using the α -entropies of Tsallis–Havrda–Charvát type, we derive several results connected with Shearer’s lemma. In particular, we derive upper bounds on the maximum possible cardinality of a family of k -subsets, when no pairwise intersections of these subsets may coincide. Further, we revisit the Minc conjecture. Our approach leads to a family of one-parameter extensions of Brégman’s theorem. A utility of the obtained bounds is explicitly exemplified.

1. INTRODUCTION

The concept of entropy is fundamental in both statistical physics and information theory. It plays a certain role in applying information-theoretic ideas to combinatorial problems [17]. Many results of such a kind were reviewed by Radhakrishnan [21] and Galvin [10]. An entropy approach is often used in studies of colorings of graphs [11, 12]. Applications of the entropy as a combinatorial tool are typically based on the Shannon entropy and its conditional form. Meantime, other entropic functions have found to be useful in various questions [2]. The Rényi entropy [25] and the Tsallis–Havrda–Charvát (THC) entropy [14, 29] are especially important extensions of the Shannon entropy. In principle, such entropic functions may have combinatorial or computational applications. For instances, they both have been used in global thresholding approach to image processing [27].

The main goal of this study is to address entropy-based approach to counting problems with use of the Tsallis–Havrda–Charvát entropies. The paper is organized as follows. In Section 2, we recall properties of the THC entropies and prove a useful statement. In Section 3, we obtain THC-entropy versions of some combinatorial results related to the so-called Shearer lemma. In particular, we consider an upper estimate for the maximum possible cardinality of a family of k -subsets of the given set, when subsets obey certain restrictions. In Section 4, we derive one-parameter family of upper bounds on permanents of square $(0, 1)$ -matrices. This family is an extension of the Brégman theorem. We describe an example of utility of the presented extension.

2. DEFINITIONS AND PROPERTIES OF THE THC α -ENTROPIES

In this section, we briefly recall definitions of the Tsallis–Havrda–Charvát entropies and related conditional entropies. Required properties of these entropic

2000 *Mathematics Subject Classification.* 05A20; 05D40; 15A15; 94A17.
The author is grateful to anonymous referees for valuable comments.

functionals are discussed as well. Let discrete random variable X take values on the finite set Ω_X . The non-extensive entropy of strictly positive degree $\alpha \neq 1$ is defined by [29]

$$(2.1) \quad H_\alpha(X) := \frac{1}{1-\alpha} \left(\sum_{x \in \Omega_X} p(x)^\alpha - 1 \right).$$

With the factor $(2^{1-\alpha} - 1)^{-1}$ instead of $(1-\alpha)^{-1}$, this entropic form was considered by Havrda and Charvát [14]. In non-extensive statistical mechanics, the entropy (2.1) is known as the Tsallis entropy. It is instructive to rewrite (2.1) as

$$(2.2) \quad H_\alpha(X) = - \sum_{x \in \Omega_X} p(x)^\alpha \ln_\alpha p(x) = \sum_{x \in \Omega_X} p(x) \ln_\alpha \left(\frac{1}{p(x)} \right).$$

Assuming $\xi > 0$, the so-called α -logarithm is defined as

$$(2.3) \quad \ln_\alpha(\xi) = \begin{cases} \frac{\xi^{1-\alpha} - 1}{1-\alpha}, & \text{for } \alpha > 0, \alpha \neq 1, \\ \ln \xi, & \text{for } \alpha = 1. \end{cases}$$

In the limit $\alpha \rightarrow 1$, the entropy (2.1) gives the standard Shannon entropy

$$(2.4) \quad H_1(X) = - \sum_{x \in \Omega_X} p(x) \ln p(x).$$

For all real $q \in [0, 1]$, we write the binary THC entropy

$$(2.5) \quad h_\alpha(q) := -q^\alpha \ln_\alpha(q) - (1-q)^\alpha \ln_\alpha(1-q).$$

For $q \in (0, 1)$, this function is concave and obeys $h_\alpha(q) = h_\alpha(1-q)$. The THC entropies succeed some natural properties of the Shannon entropy. The maximal value of (2.1) is equal to $\ln_\alpha |\Omega_X|$ and reached with the uniform distribution. For $\alpha \geq 1$, the joint THC entropy of two random variables obeys [9]

$$(2.6) \quad H_\alpha(X, Y) \leq H_\alpha(X) + H_\alpha(Y).$$

In applications of information-theoretic methods, the notion of conditional entropy is widely used [6]. Let us put the particular functional

$$H_1(X|y) = - \sum_x p(x|y) \ln p(x|y),$$

in which the sum is taken over $x \in \Omega_X$. The entropy of X conditional on knowing Y is defined as [6]

$$(2.7) \quad H_1(X|Y) := \sum_y p(y) H_1(X|y) = - \sum_x \sum_y p(x, y) \ln p(x|y),$$

where $p(x|y) = p(x, y)/p(y)$. When the range of summation is clear from the context, we will omit symbols such as Ω_X and Ω_Y .

In the literature, two kinds of the conditional THC entropy have been discussed [9]. These forms are respectively inspired by the two expressions given in (2.2). The first form is defined as [9]

$$(2.8) \quad H_\alpha(X|Y) := \sum_y p(y)^\alpha H_\alpha(X|y),$$

where

$$(2.9) \quad H_\alpha(X|y) := \frac{1}{1-\alpha} \left(\sum_x p(x|y)^\alpha - 1 \right) = - \sum_x p(x|y)^\alpha \ln_\alpha p(x|y)$$

and strictly positive $\alpha \neq 1$. The conditional entropy (2.8) is, up to a factor, the quantity originally introduced by Daróczy [7]. For any $\alpha > 0$, this conditional entropy obeys the chain rule written as [7]

$$(2.10) \quad H_\alpha(X, Y) = H_\alpha(X|Y) + H_\alpha(Y) .$$

Due to nonnegativity of $H_\alpha(X|Y)$, for all $\alpha > 0$ we also have

$$H_\alpha(X, Y) \geq H_\alpha(Y) .$$

The chain rule (2.10) can be extended to more than two variables. Up to reordering of random variables, this result is expressed as [9]

$$(2.11) \quad H_\alpha(X_1, X_2, \dots, X_n) = \sum_{j=1}^n H_\alpha(X_j | X_{j-1}, \dots, X_1) ,$$

where $\alpha > 0$. In the case $\alpha = 1$, we obtain the chain rule with the standard conditional entropy (2.7). This property turns to be very essential in entropic approach to counting. The second form of conditional THC entropy is introduced as [9]

$$(2.12) \quad \tilde{H}_\alpha(X|Y) := \sum_y p(y) H_\alpha(X|y) .$$

Although the quantity (2.12) does not share the chain rule, it has found use in some questions [9, 22]. Its definition is based on the formulation, which seems to be more appropriate in the context of dynamical systems and generalized entropy rates [8, 24, 28]. We also have $\tilde{H}_\alpha(X|Y) \leq H_\alpha(X|Y)$ for $\alpha \in (0, 1)$ and $\tilde{H}_\alpha(X|Y) \geq H_\alpha(X|Y)$ for $\alpha \in (1, \infty)$. For $\alpha = 1$, the α -entropies (2.8) and (2.12) both coincide with (2.7).

Using entropic approach in counting, several properties of the conditional entropy are required. One of these properties is the chain rule. The standard conditional entropy also satisfies

$$(2.13) \quad H_1(X|Y_1, \dots, Y_{n-1}, Y_n) \leq H_1(X|Y_1, \dots, Y_{n-1}) .$$

Thus, conditioning on more can only reduce the conditional entropy. This relation is sometimes required in counting [21]. Another very useful property of the standard conditional entropy is formulated as follows. Let $Y \mapsto f(Y)$ be some function, whose domain covers the support of random variable Y . Then we have [21]

$$(2.14) \quad H_1(X|f(Y)) \geq H_1(X|Y) .$$

We shall now establish analogous properties for the conditional α -entropies.

Proposition 1. Let X and Y_1, \dots, Y_n be discrete random variables, where $n \geq 1$. For $\alpha \geq 1$, the conditional entropy (2.8) satisfies

$$(2.15) \quad H_\alpha(X|Y_1, \dots, Y_{n-1}, Y_n) \leq H_\alpha(X|Y_1, \dots, Y_{n-1}) .$$

For $\alpha > 0$, the conditional entropy (2.12) satisfies

$$(2.16) \quad \tilde{H}_\alpha(X|Y_1, \dots, Y_{n-1}, Y_n) \leq \tilde{H}_\alpha(X|Y_1, \dots, Y_{n-1}) .$$

Let $Y \mapsto f(Y)$ be a function of random variable Y . For $\alpha \geq 1$, the conditional entropy (2.8) satisfies

$$(2.17) \quad H_\alpha(X|f(Y)) \geq H_\alpha(X|Y) .$$

For $\alpha > 0$, the conditional entropy (2.12) satisfies

$$(2.18) \quad \tilde{H}_\alpha(X|f(Y)) \geq \tilde{H}_\alpha(X|Y) .$$

Proof. The results (2.15) and (2.16) were proved in [9, 23] and [24], respectively. Let us proceed to (2.17) and (2.18). Since the standard case is known, we assume $\alpha \neq 1$. To each value u of the function, we assign the subset $\omega_u \subseteq \Omega_Y$ such that

$$\omega_u := \{y : y \in \Omega_Y, f(y) = u\} .$$

Then the probabilities are written as

$$(2.19) \quad p(x, u) = \sum_{y \in \omega_u} p(x, y) , \quad p(u) = \sum_{y \in \omega_u} p(y) .$$

The left-hand side of (2.17) is represented as

$$(2.20) \quad H_\alpha(X|f(Y)) = \sum_{u \in \Omega_{f(Y)}} p(u)^\alpha H_\alpha(X|u) .$$

Replacing $p(u)^\alpha$ with $p(u)$, we obtain the expression for $\tilde{H}_\alpha(X|f(Y))$. For strictly positive $\alpha \neq 1$ and $\xi \geq 0$, we introduce the function

$$\eta_\alpha(\xi) = \frac{\xi^\alpha - \xi}{1 - \alpha} .$$

In terms of this function, we now write

$$p(u)^\alpha H_\alpha(X|u) = \sum_{x \in \Omega_X} p(u)^\alpha \eta_\alpha(p(x|u)) .$$

As $\eta_\alpha''(\xi) \leq 0$ for the considered values of α , the function $\xi \mapsto \eta_\alpha(\xi)$ is concave. For fixed x and u , we put numbers $\lambda_y = p(y)/p(u)$ and $\xi_y = p(x, y)/p(y) = p(x|y)$ such that

$$\sum_{y \in \omega_u} \lambda_y = \frac{p(u)}{p(u)} = 1 , \quad \sum_{y \in \omega_u} \lambda_y \xi_y = \frac{p(x, u)}{p(u)} = p(x|u) ,$$

according to (2.19). By Jensen's inequality, we then obtain

$$(2.21) \quad p(u) \eta_\alpha(p(x|u)) \geq p(u) \sum_{y \in \omega_u} \lambda_y \eta_\alpha(\xi_y) = \sum_{y \in \omega_u} p(y) \eta_\alpha(p(x|y)) .$$

$$(2.22) \quad p(u)^\alpha \eta_\alpha(p(x|u)) \geq p(u)^\alpha \sum_{y \in \omega_u} \lambda_y \eta_\alpha(\xi_y) = \sum_{y \in \omega_u} p(u)^{\alpha-1} p(y) \eta_\alpha(p(x|y)) .$$

Summing (2.21) with respect to $x \in \Omega_X$, for all the considered values of α one gets

$$p(u) H_\alpha(X|u) \geq \sum_{y \in \omega_u} p(y) H_\alpha(X|y) .$$

The latter leads to (2.18) after summing with respect to $u \in \Omega_{f(Y)}$. For all $y \in \omega_u$ and $\alpha > 1$, we have $p(u) \geq p(y)$ and $p(u)^{\alpha-1} p(y) \geq p(y)^\alpha$. Combining this with (2.22) and summing with respect to $x \in \Omega_X$, we obtain

$$p(u)^\alpha H_\alpha(X|u) \geq \sum_{y \in \omega_u} p(y)^\alpha H_\alpha(X|y) .$$

Summing this with respect to $u \in \Omega_{f(Y)}$ completes the proof of (2.17). \square

Note that the standard case $\alpha = 1$ of (2.17) and (2.18) can be proved by repeating the above reasons with the concave function $\xi \mapsto -\xi \ln \xi$. In the mentioned ranges of the parameter, the conditional THC entropies (2.8) and (2.12) enjoy the property with respect to conditioning on more. The result (2.15) has allowed to derive the one-parametric extension of entropic Bell inequalities originally given in [3]. Using

(2.14), one can deduce a property useful in entropic approach to Bregman's theorem [20, 21]. We shall now formulate a similar statement for the α -entropies.

Corollary 2. Let the support Ω_Y of random variable Y be partitioned into m mutually disjoint sets ω_j as

$$\Omega_Y = \bigcup_{j=1}^m \omega_j .$$

Let $\varpi_j \subseteq \Omega_X$ be defined as

$$\varpi_j := \{x : x \in \Omega_X, y \in \omega_j, p(x|y) \neq 0\} .$$

If $\varpi_j \neq \varpi_k$ for all $j \neq k$, then

$$(2.23) \quad H_\alpha(X|Y) \leq \sum_{j=1}^m \Pr[Y \in \omega_j]^\alpha \ln_\alpha |\varpi_j| \quad (1 \leq \alpha < \infty) ,$$

$$(2.24) \quad \tilde{H}_\alpha(X|Y) \leq \sum_{j=1}^m \Pr[Y \in \omega_j] \ln_\alpha |\varpi_j| \quad (0 < \alpha < \infty) .$$

Proof. Let us take the function $Y \mapsto f_\omega(Y)$ such that $f_\omega(y) = \varpi_j$ for each $y \in \omega_j$. It then follows from (2.17) and (2.18) that

$$(2.25) \quad H_\alpha(X|Y) \leq H_\alpha(X|f_\omega(Y)) = \sum_{j=1}^m \Pr[Y \in \omega_j]^\alpha H_\alpha(X|\varpi_j) \quad (1 \leq \alpha < \infty) ,$$

$$(2.26) \quad \tilde{H}_\alpha(X|Y) \leq \tilde{H}_\alpha(X|f_\omega(Y)) = \sum_{j=1}^m \Pr[Y \in \omega_j] H_\alpha(X|\varpi_j) \quad (0 < \alpha < \infty) .$$

The quantity $H_\alpha(X|\varpi_j)$ is represented as the sum

$$(2.27) \quad H_\alpha(X|\varpi_j) = \sum_{x \in \varpi_j} \eta_\alpha(p(x|\omega_j)) .$$

The sum of $p(x|\omega_j)$ over $x \in \varpi_j$ is equal to 1, whence the term (2.27) does not exceed $\ln_\alpha |\varpi_j|$. Combining this fact with (2.25) and (2.26) completes the proof. \square

Using Corollary 2, we will obtain upper bounds on conditional α -entropies in some combinatorial problems. To do so, we have to estimate not only cardinalities $|\varpi_j|$, but also probabilities $\Pr[Y \in \omega_j]$. From this viewpoint, the inequality (2.24) seems to be more appropriate.

3. SHEARER'S LEMMA AND INTERSECTIONS OF k -ELEMENT SETS

In this section, we will examine some questions connected with the Shearer lemma [5]. The properties of the THC entropies lead to a lot of inequalities with interesting combinatorial applications. We first note the following.

Proposition 3. Let $X = (X_1, \dots, X_n)$ be a random variable taking values in the set $S = S_1 \times \dots \times S_n$, where each coordinate X_j is a random variable taking values in S_j . For all $\alpha \geq 1$, we have

$$(3.1) \quad H_\alpha(X) \leq \sum_{j=1}^n H_\alpha(X_j) .$$

Proof. The claim (3.1) immediately follows by induction from (2.6). \square

The result (3.1) is a straightforward extension of proposition 15.7.2 of the book [1]. Hence, we can obtain several corollaries. The first of them is posed as follows.

Corollary 4. Let \mathcal{F} be a family of subsets of the set $\{1, \dots, n\}$, and let q_j denote the fraction of members of \mathcal{F} that contain j . For all $\alpha \geq 1$, we have

$$(3.2) \quad \ln_\alpha |\mathcal{F}| \leq \sum_{j=1}^n h_\alpha(q_j) .$$

Proof. To each set $F \in \mathcal{F}$, we assign its characteristic vector $v(F)$, which is a binary n -tuple. Let $X = (X_1, \dots, X_n)$ be the random variable taking values in $\{0, 1\}^n$ such that

$$(3.3) \quad \Pr[X = v(F)] = |\mathcal{F}|^{-1} \quad \forall F \in \mathcal{F} ,$$

whence $H_\alpha(X) = \ln_\alpha |\mathcal{F}|$. The random variable X_j takes values in $\{0, 1\}$ and is j -th value in the characteristic vector. By definition of q_j , the entropy of X_j is equal to $h_\alpha(q_j)$. Combining this with (3.1) completes the proof. \square

The result (3.2) provides an upper estimate for the maximum possible cardinality of a family of subsets. It is an α -entropy version of the basic lemma proved in [16]. The authors of [16] used tools of information theory for studying a family of k -subsets, which satisfy some restrictions. We will further apply (3.2) to a specific family of k -element subsets of the set $\{1, \dots, n\}$. Suppose that a family $\mathcal{G} = \{G_1, \dots, G_m\}$ of m subsets of the set $\{1, \dots, n\}$ obey the implication

$$(3.4) \quad \text{“}\{i, j\} \neq \{s, t\}\text{”} \implies \text{“}G_i \cap G_j \neq G_s \cap G_t\text{”} .$$

That is, no pairwise intersections of the k -subsets may coincide. We aim to estimate cardinality of this family from above. Let us begin with an auxiliary result.

Lemma 5. For $\alpha \in [1, 3.67]$, the function $\lambda \mapsto h_\alpha(\lambda^2)/\lambda$ is concave for $\lambda \in [0, 1/\sqrt{2}]$.

Proof. We left out the case $\alpha = 1$, for which the concavity was reported in [16] for all $\lambda \in [0, 1]$. During the proof, we will use the following generalization of Bernoulli's inequality (see, e.g., section 2.4 of the book [19]). For $-1 < x \neq 0$, one has

$$(3.5) \quad (1+x)^r > 1+rx \quad (r \notin [0, 1]) ,$$

$$(3.6) \quad (1+x)^r < 1+rx \quad (0 < r < 1) .$$

For $\alpha > 1$, we can write the expression

$$(3.7) \quad (\alpha-1) \frac{h_\alpha(\lambda^2)}{\lambda} = -\lambda^{2\alpha-1} + \frac{1-(1-\lambda^2)^\alpha}{\lambda} .$$

The term $-\lambda^{2\alpha-1}$ is concave with respect to λ for all $\alpha > 1$. We will show concavity of the second term in the right-hand side of (3.7). Let us use the second derivative test. For arbitrary function $\xi \mapsto f(\xi)$, one has a general expression

$$(3.8) \quad \frac{d^2}{d\lambda^2} \frac{f(\lambda^2)}{\lambda} = \frac{2}{\lambda^3} \left(2\xi^2 f''(\xi) - \xi f'(\xi) + f(\xi) \right) ,$$

where $\xi = \lambda^2$. Substituting $f_\alpha(\xi) = 1 - (1-\xi)^\alpha$ finally gives

$$(3.9) \quad 2\xi^2 f''_\alpha(\xi) - \xi f'_\alpha(\xi) + f_\alpha(\xi) = 1 + (1-\xi)^{\alpha-2} (-1 + c_1 \xi + c_2 \xi^2) .$$

The coefficients in (3.9) are calculated as

$$(3.10) \quad c_1 = 2 - \alpha, \quad c_2 = -1 + 3\alpha - 2\alpha^2 = \frac{1}{8} - 2 \left(\alpha - \frac{3}{4} \right)^2.$$

We will show that the quantity (3.9) is not positive for $\alpha \in [1, 3.67]$ and $\lambda \in [0, 1/\sqrt{2}]$. Let us consider separately the cases $\alpha \in [1, 2]$ and $\alpha \in [2, 3.67]$.

Taking the intervals $\alpha \in [1, 2]$ and $\xi \in [0, 1]$, we have

$$(3.11) \quad (1 - \xi)^{2-\alpha} \leq 1 - (2 - \alpha)\xi = 1 - c_1\xi \quad (1 \leq \alpha \leq 2).$$

This formula is based on (3.6) with $x = -\xi$ and $r = 2 - \alpha$. Due to (3.11), we rewrite (3.9) in the form

$$(3.12) \quad (1 - \xi)^{\alpha-2} \left((1 - \xi)^{2-\alpha} - 1 + c_1\xi + c_2\xi^2 \right) \leq (1 - \xi)^{\alpha-2} c_2\xi^2 \leq 0,$$

where $c_2 \leq 0$ for $\alpha \in [1, 2]$ by (3.10). Here, the concavity takes place for all $\lambda \in [0, 1]$.

The case $\alpha \geq 2$ is more complicated to analysis. Here, we introduce the positive parameter $\beta = \alpha - 2$. The condition of negativity of (3.9) is then rewritten as

$$(1 - \xi)^\beta (1 + \beta\xi + \gamma\xi^2) - 1 =: F_\beta(\xi) \geq 0,$$

where $\gamma = -c_2 = 3 + 5\beta + 2\beta^2$. This inequality is to be proved for $\xi = \lambda^2 \leq 1/2$.

We begin with the case $\beta \in [0, 1]$. Using the polynomial $p_\beta(\xi) = 1 + \beta\xi + \gamma\xi^2$, we write the derivative

$$\frac{dF_\beta}{d\xi} = (1 - \xi)^{\beta-1} \left((1 - \xi)p'_\beta(\xi) - \beta p_\beta(\xi) \right).$$

Doing simple calculations, we easily obtain

$$(3.13) \quad (1 - \xi)p'_\beta(\xi) - \beta p_\beta(\xi) = \xi \left((2\gamma - \beta - \beta^2) - \gamma(2 + \beta)\xi \right).$$

As $\xi \geq 0$, the derivative $dF_\beta/d\xi$ is not negative, whenever

$$(3.14) \quad \xi \leq \frac{2\gamma - \beta - \beta^2}{\gamma(2 + \beta)} = \frac{6 + 9\beta + 3\beta^2}{(3 + 5\beta + 2\beta^2)(2 + \beta)}.$$

For $\beta \in [0, 1]$, the right-hand side of (3.14) monotonically decreases with β from 1 at $\beta = 0$ up to 0.6 at $\beta = 1$. The condition (3.14) is clearly satisfied for all $\xi \in [0, 1/2]$. Here, the function $F_\beta(\xi)$ does not decrease. Combining this with $F_\beta(0) = 0$, we finally get $F_\beta(\xi) \geq 0$ for all $\beta \in [0, 1]$ and $\xi \in [0, 1/2]$.

For $\beta \geq 1$, we apply $(1 - \xi)^\beta \geq 1 - \beta\xi$ due to (3.5). Thus, the quantity of interest obeys

$$F_\beta(\xi) \geq (1 - \beta\xi)(1 + \beta\xi + \gamma\xi^2) - 1 \geq (\gamma - \beta^2)\xi^2 - \beta\gamma\xi^3.$$

The latter is not negative, whenever $\gamma - \beta^2 \geq \beta\gamma\xi$. Due to $\xi \leq 1/2$, we can focus on the inequality $2(\gamma - \beta^2) - \beta\gamma \geq 0$, or

$$(3.15) \quad 6 + 7\beta - 3\beta^2 - 2\beta^3 \geq 0.$$

Inspecting roots of the polynomial, the condition (3.15) holds for all $\beta \in [0, 1.67]$, though we use it only for $\beta \in [1, 1.67]$. The latter completes the proof for $\alpha \in [3, 3.67]$. \square

We have shown concavity of the function $\lambda \mapsto h_\alpha(\lambda^2)/\lambda$ for $\alpha \in [1, 3.67]$ and $\lambda \in [0, 1/\sqrt{2}]$. When $\alpha \in [1, 2]$, the concavity actually holds for $\lambda \in [0, 1]$. We now formulate the desired estimate as follows.

Proposition 6. Let $\mathcal{G} = \{G_1, \dots, G_m\}$ be a family containing m k -subsets of the set $\{1, \dots, n\}$, and let \mathcal{G} satisfy the property (3.4). Let λ_j denote a proportion of those members of \mathcal{G} that contain j . If the precondition

$$(3.16) \quad \lambda_j \leq \frac{1}{\sqrt{2}}$$

holds for all $j \in \{1, \dots, n\}$, then for all $\alpha \in [1, 3.67]$,

$$(3.17) \quad \ln_\alpha \binom{m}{2} \leq k \frac{h_\alpha(\lambda^2)}{\lambda}, \quad \lambda := \sum_{j=1}^n \frac{\lambda_j}{k} \lambda_j.$$

Proof. Let us consider pairwise intersections of members of \mathcal{G} . Each $j \in \{1, \dots, n\}$ will appear in a proportion

$$q_j^* = \binom{m}{2}^{-1} \binom{\lambda_j m}{2} = \lambda_j^2 - \frac{\lambda_j(1 - \lambda_j)}{m - 1}.$$

In the ratio, the denominator gives the number of all pairwise intersections; the numerator is the number of those pairwise intersections that contain j . The binary entropy (2.5) is concave for $q \in (0, 1)$ and reaches its maximum at the point $q = 1/2$. Hence, it does not decrease on $(0, 1/2)$. Then the precondition (3.16) provides

$$h_\alpha(q_j^*) \leq h_\alpha(\lambda_j^2).$$

Combining this with Corollary 4, for $\alpha \geq 1$ we obtain

$$(3.18) \quad \ln_\alpha \binom{m}{2} \leq \sum_{j=1}^n h_\alpha(\lambda_j^2) = k \sum_{j=1}^n \frac{\lambda_j}{k} \frac{h_\alpha(\lambda_j^2)}{\lambda_j}.$$

We further use $\sum_{j=1}^n \lambda_j/k = 1$ and concavity of the function $\lambda \mapsto h_\alpha(\lambda^2)/\lambda$. Combining (3.18) with the Jensen inequality completes the proof. \square

We have obtained an implicit upper bound on $m = |\mathcal{G}|$ in terms of $k = |G_j|$ and the average proportion λ of sets containing a particular element. Our result is a parametric extension of one of the statements proved in [16]. It also differs in the following two respects. First, the precondition (3.16) is now imposed. On the other hand, the formula (3.17) is more explicit in the sense that no unknown asymptotically small terms appear. For the prescribed value of $\lambda \in [0, 1/\sqrt{2}]$, we could optimize a bound with respect to the parameter α . The authors of [16] also consider a family of k -sets, in which the intersection of no two is contained in a third. Such estimates are connected with one of questions raised by Erdős.

The statement of Proposition 3 allows a certain extension. In the case of Shannon entropies, extension of such a kind has been proved by Shearer [5]. It is often referred to as the Shearer lemma [12, 15, 21]. Its generalization in terms of the THC entropies is posed as follows.

Proposition 7. Let $X = (X_1, \dots, X_n)$ be a random variable taking values in the set $S = S_1 \times \dots \times S_n$, where each coordinate X_j is a random variable taking values in S_j . For a subset I of $\{1, \dots, n\}$, let $X(I)$ denote the random variable $(X_j)_{j \in I}$. Suppose that \mathcal{G} is a family of subsets of $\{1, \dots, n\}$ and each $j \in \{1, \dots, n\}$ belongs to at least k members of \mathcal{G} . For $\alpha \geq 1$, we then have

$$(3.19) \quad k H_\alpha(X) \leq \sum_{G \in \mathcal{G}} H_\alpha(X(G)).$$

Proof. Following [21], we will apply the chain rule. Using (2.11), for $\alpha \geq 1$ we have

$$\begin{aligned}
 H_\alpha(X) &= \sum_{j=1}^n H_\alpha(X_j \mid (X_k : k < j)) , \\
 H_\alpha(X(G)) &= \sum_{j \in G} H_\alpha(X_j \mid (X_k : k \in G, k < j)) \\
 (3.20) \qquad &\geq \sum_{j \in G} H_\alpha(X_j \mid (X_k : k < j)) .
 \end{aligned}$$

The step (3.20) follows from (2.15), since any string $(X_k : k < j)$ contains more elements than $(X_k : k \in G, k < j)$. Summing (3.20) with respect to all $G \in \mathcal{G}$ gives

$$(3.21) \qquad \sum_{G \in \mathcal{G}} H_\alpha(X(G)) \geq k \sum_{j=1}^n H_\alpha(X_j \mid (X_k : k < j)) ,$$

because each $j \in \{1, \dots, n\}$ belongs to at least k members of \mathcal{G} . □

The statement of Proposition 7 is a THC-entropy extension of the Shearer lemma. A related geometric picture was described in [21]. Interesting geometric applications are also discussed in [1]. An immediate consequence of (3.19) is posed as follows.

Corollary 8. Let N be a finite set, and let \mathcal{F} be a family of subsets of N . Let $\mathcal{G} = \{G_1, \dots, G_m\}$ be a family of subsets of N such that each element of N appears in at least k members of \mathcal{G} . For each $1 \leq j \leq m$, we define $\mathcal{F}_j := \{F \cap G_j : F \in \mathcal{F}\}$. For $\alpha \geq 1$, we then have

$$(3.22) \qquad k \ln_\alpha |\mathcal{F}| \leq \sum_{j=1}^m \ln_\alpha |\mathcal{F}_j| .$$

For $\alpha = 1$, the formula (3.22) is reduced to a result originally proved in [5]. Some applications of the latter were also described in [5]. Of course, applications of such a kind can further be considered on the base of (3.22). In some cases, a family of one-parameter relations may give a stronger bound. An explicit example of this situation is the case of upper bounds on permanents of square $(0, 1)$ -matrices.

4. UPPER BOUNDS ON PERMANENTS OF $(0, 1)$ -MATRICES

In this section, we will derive a family of one-parameter upper bounds on the permanent of a square $(0, 1)$ -matrix. The well-known upper bound on permanents has been conjectured by Minc [18] and later proved by Brégman [4]. Brégman's proof is based on the duality theorem of convex programming and properties of doubly stochastic matrices. A short elementary proof of this result was given by Schrijver [26]. Schrijver also mentioned an upper bound for permanents of arbitrary nonnegative matrices. A similar proof with randomization is explained in [1]. Developing an approach with randomization, Radhakrishnan presented an entropy-based proof [20]. Our aim is to study the question with use of the THC entropies. First, we recall preliminary facts. Let $\mathbf{A} = [[a(i, j)]]$ be a nonnegative $n \times n$ -matrix, and let S_n denote the set of all permutations on $\{1, \dots, n\}$. The permanent of \mathbf{A} is defined

as

$$(4.1) \quad \text{per}(\mathbf{A}) := \sum_{\sigma \in S_n} \prod_{i=1}^n a(i, \sigma(i)) .$$

We further consider matrices with elements $a(i, j) \in \{0, 1\}$. By $S \subseteq S_n$, we mean the set of permutations σ such that $a(i, \sigma(i)) = 1$ for all $i \in \{1, \dots, n\}$. It is obvious that $\text{per}(\mathbf{A}) = |S|$. It is assumed that the matrix contain no rows of only zeros, since otherwise its permanent is certainly zero. We claim the following.

Proposition 9. Let \mathbf{A} be a $n \times n$ (0, 1)-matrix with $\text{per}(\mathbf{A}) \neq 0$, and let $r_i \neq 0$ be a number of ones in i -th row ($i = 1, \dots, n$). For all $\alpha \geq 1$, the permanent of \mathbf{A} obeys the inequality

$$(4.2) \quad \ln_\alpha(\text{per}(\mathbf{A})) \leq \sum_{i=1}^n \frac{1}{r_i} \sum_{j=1}^{r_i} \ln_\alpha(j) .$$

Proof. Let σ be a random permutation chosen uniformly from S . We then have the value $H_\alpha(\sigma) = \ln_\alpha |S|$, which coincides with the left-hand side of (4.2). We will show that, for $\alpha \geq 1$, the entropy $H_\alpha(\sigma)$ does not exceed the right-hand side of (4.2). Let us choose a random permutation $\tau \in S_n$ uniformly. Using the chain rule (2.11), for each permutation τ we can write

$$(4.3) \quad \begin{aligned} H_\alpha(\sigma) &= H_\alpha(\sigma(\tau(1))) + H_\alpha(\sigma(\tau(2)) \mid \sigma(\tau(1))) + \dots \\ &\quad \dots + H_\alpha(\sigma(\tau(n)) \mid \sigma(\tau(1)), \dots, \sigma(\tau(n-1))) \end{aligned}$$

$$(4.4) \quad \begin{aligned} &\leq \tilde{H}_\alpha(\sigma(\tau(1))) + \tilde{H}_\alpha(\sigma(\tau(2)) \mid \sigma(\tau(1))) + \dots \\ &\quad \dots + \tilde{H}_\alpha(\sigma(\tau(n)) \mid \sigma(\tau(1)), \dots, \sigma(\tau(n-1))) . \end{aligned}$$

Here, the second inequality holds for $\alpha \geq 1$. To the given permutation τ and index $i \in \{1, \dots, n\}$, we assign the integer $k(\tau, i) \in \{1, \dots, n\}$ such that

$$k(\tau, i) := \tau^{-1}(i) , \quad \sigma(\tau(k)) = \sigma(i) .$$

Summing (4.4) over all $\tau \in S_n$, we further obtain

$$(4.5) \quad \begin{aligned} |S_n| H_\alpha(\sigma) &\leq \sum_{\tau \in S_n} \left\{ \tilde{H}_\alpha(\sigma(\tau(1))) + \tilde{H}_\alpha(\sigma(\tau(2)) \mid \sigma(\tau(1))) + \dots \right. \\ &\quad \left. \dots + \tilde{H}_\alpha(\sigma(\tau(n)) \mid \sigma(\tau(1)), \dots, \sigma(\tau(n-1))) \right\} \\ &= \sum_{\tau \in S_n} \sum_{i=1}^n \tilde{H}_\alpha(\sigma(i) \mid \sigma(\tau(1)), \dots, \sigma(\tau(k-1))) . \end{aligned}$$

At the last step, we gather the contributions of different $\sigma(i)$ separately. For the given $\sigma \in S$, $\tau \in S_n$, and $i \in \{1, \dots, n\}$, we define $R_i(\sigma, \tau)$ to be the set of those column indices that differ from $\sigma(\tau(1)), \dots, \sigma(\tau(k-1))$ and give 1's in i -th row [20]. By definition of r_i , we have $|R_i(\sigma, \tau)| \leq r_i$. Using (2.24), we then rewrite (4.5) as

$$|S_n| H_\alpha(\sigma) \leq \sum_{i=1}^n \sum_{\tau \in S_n} \sum_{j=1}^{r_i} \Pr_\sigma[|R_i(\sigma, \tau)| = j] \ln_\alpha(j) .$$

Dividing this relation by $|S_n|$ and taking into account the uniform distribution of $\tau \in S_n$, we immediately obtain

$$(4.6) \quad H_\alpha(\sigma) \leq \sum_{i=1}^n \sum_{j=1}^{r_i} \Pr_{\sigma, \tau} \left[|R_i(\sigma, \tau)| = j \right] \ln_\alpha(j) .$$

We now recall a principal observation of [20] that

$$\Pr_{\sigma, \tau} \left[|R_i(\sigma, \tau)| = j \right] = \frac{1}{r_i} .$$

Combining this with (4.6) completes the proof. □

The statement of Theorem 9 leads to an one-parameter family of upper bounds on permanents. In the limit $\alpha \rightarrow 1^+$, the relation (4.2) leads to the previous result [1]

$$(4.7) \quad \text{per}(\mathbf{A}) \leq \prod_{i=1}^n (r_i!)^{1/r_i} .$$

It was conjectured in [18] and then proved in several ways [4, 20, 26]. This result can naturally be reformulated as an upper bound on the number of perfect matchings in a bipartite graph [10, 21].

We now consider a significance of the one-parameter bound (4.2). It is instructive to consider a concrete example. Let $n \times n$ -matrix \mathbf{A} have elements $a(1, j) = 1$ for all $j = 1, \dots, n$ and $a(i, j) = \delta(i, j)$ for $i = 2, \dots, n$. That is, our matrix is obtained from the identity $n \times n$ -matrix by filling its first row with ones. We then have $\text{per}(\mathbf{A}) = 1$. On the other hand, one gives $r_1 = n$ and $r_2 = \dots = r_n = 1$. Let us compare values of the bounds (4.2) and (4.7). It is easy to apply (4.2) in the case $\alpha = 2$, since $\ln_2(\xi) = 1 - 1/\xi$ due to (2.3). For $\alpha = 2$, the upper bound (4.2) gives

$$(4.8) \quad 1 - \frac{1}{\text{per}(\mathbf{A})} \leq \frac{1}{n} \sum_{j=1}^n \left(1 - \frac{1}{j} \right) = 1 - \frac{\mathcal{H}_n}{n} .$$

By \mathcal{H}_n , we denote the n -th harmonic number [13]. It is well known that the asymptotic expansion of this number for large n is written as [13]

$$\mathcal{H}_n = \ln n + \gamma + O(1/n) ,$$

where γ is the Euler–Mascheroni constant. From (4.8), we immediately obtain

$$(4.9) \quad \text{per}(\mathbf{A}) \leq \frac{n}{\mathcal{H}_n} = \frac{n}{\ln n} \left\{ 1 + O\left(\frac{1}{\ln n} \right) \right\} .$$

Substituting the same collection of numbers r_i into (4.7) gives

$$(4.10) \quad \text{per}(\mathbf{A}) \leq (n!)^{1/n} = \frac{n}{e} \left\{ 1 + O\left(\frac{1}{n} \right) \right\} .$$

At the last step, we used the Stirling approximation. For sufficiently large n , the upper bound (4.9) is significantly stronger than (4.10). On the other hand, both the bounds are very far from the actual value of permanent. Nevertheless, our example has shown a relevance of the result (4.2) proved for $\alpha \geq 1$.

We can further ask for extending bounds with values $\alpha \in (0, 1)$. The corresponding result can be obtained by an immediate extension of Schrijver’s proof [26]. We have the following statement.

Proposition 10. Let \mathbf{A} be a $n \times n$ $(0, 1)$ -matrix with $\text{per}(\mathbf{A}) \neq 0$, and let $r_i \neq 0$ be a number of ones in i -th row ($i = 1, \dots, n$). For $\alpha \in (0, 1)$, the permanent of \mathbf{A} obeys the inequality

$$(4.11) \quad -\ln_\alpha \left(\frac{1}{\text{per}(\mathbf{A})} \right) \leq \sum_{i=1}^n \frac{1}{r_i} \sum_{j=1}^{r_i} \ln_\alpha(j).$$

Proof. For convenience, we introduce the function

$$g_\alpha(\xi) := \frac{\xi^\alpha - \xi}{\alpha - 1} = -\xi \ln_\alpha \left(\frac{1}{\xi} \right),$$

where $\xi > 0$ and $\alpha \in (0, 1)$. For these values of α , the function $\xi \mapsto g_\alpha(\xi)$ is convex. Due to the Jensen inequality, we have

$$(4.12) \quad g_\alpha \left(\frac{1}{r} \sum_{k=1}^r \xi_k \right) \leq \frac{1}{r} \sum_{k=1}^r g_\alpha(\xi_k).$$

We will prove (4.11) by induction on n . For $n = 1$, the result is trivial. Suppose that the claim is already proved for $(n-1) \times (n-1)$ -matrices. The permanent of a $n \times n$ -matrix can be decomposed as

$$(4.13) \quad \text{per}(\mathbf{A}) = \sum_{\substack{k=1 \\ a(i,k)=1}}^n \text{per}(\mathbf{A}(i, k)).$$

Here, the submatrix $\mathbf{A}(i, k)$ is obtained from \mathbf{A} by eliminating the i -th row and the k -th column. Combining (4.12) with (4.13) gives

$$(4.14) \quad \text{per}(\mathbf{A}) (-1) \ln_\alpha \left(\frac{r_i}{\text{per}(\mathbf{A})} \right) = r_i g_\alpha \left(\frac{\text{per}(\mathbf{A})}{r_i} \right) \leq \sum_{\substack{k=1 \\ a(i,k)=1}}^n g_\alpha \left\{ \text{per}(\mathbf{A}(i, k)) \right\}.$$

From the definition of the α -logarithm, we have the identity

$$(4.15) \quad \ln_\alpha(r\xi) = \ln_\alpha(r) + r^{1-\alpha} \ln_\alpha(\xi).$$

Summing (4.14) with respect to $i \in \{1, \dots, n\}$, we therefore obtain

$$(4.16) \quad \text{per}(\mathbf{A}) \left\{ -\sum_{i=1}^n \ln_\alpha(r_i) - n \ln_\alpha \left(\frac{1}{\text{per}(\mathbf{A})} \right) \right\} \leq \sum_{i=1}^n r_i g_\alpha \left(\frac{\text{per}(\mathbf{A})}{r_i} \right)$$

$$(4.17) \quad \leq \sum_{i=1}^n \sum_{\substack{k=1 \\ a(i,k)=1}}^n \text{per}(\mathbf{A}(i, k)) (-1) \ln_\alpha \left(\frac{1}{\text{per}(\mathbf{A}(i, k))} \right)$$

$$(4.18) \quad = \sum_{\sigma \in S} \sum_{i=1}^n (-1) \ln_\alpha \left(\frac{1}{\text{per}\{\mathbf{A}(i, \sigma(i))\}} \right).$$

To prove (4.16), we used (4.15) and the relation $n \leq \sum_{i=1}^n r_i^{1-\alpha}$ satisfied for $\alpha \in (0, 1)$. To justify (4.18), we note the following fact. In the double sum (4.18), the number of terms from any pair (i, k) equals the number of those $\sigma \in S$ for which $\sigma(i) = k$. The latter number is $\text{per}(\mathbf{A}(i, k))$ for $a(i, k) = 1$, and zero otherwise. We

now apply the induction hypothesis to each per $\{\mathbf{A}(i, \sigma(i))\}$ in (4.18). The left-hand side of (4.16) is no greater than

$$(4.19) \quad \sum_{\sigma \in S} \sum_{i=1}^n \left\{ \sum_{\substack{\ell \neq i \\ a(\ell, \sigma(i))=0}} \frac{1}{r_\ell} \sum_{j=1}^{r_\ell} \ln_\alpha(j) + \sum_{\substack{\ell \neq i \\ a(\ell, \sigma(i))=1}} \frac{1}{r_\ell - 1} \sum_{j=1}^{r_\ell - 1} \ln_\alpha(j) \right\}$$

$$(4.19) \quad = \sum_{\sigma \in S} \sum_{\ell=1}^n \left\{ \sum_{\substack{i \neq \ell \\ a(\ell, \sigma(i))=0}} \frac{1}{r_\ell} \sum_{j=1}^{r_\ell} \ln_\alpha(j) + \sum_{\substack{i \neq \ell \\ a(\ell, \sigma(i))=1}} \frac{1}{r_\ell - 1} \sum_{j=1}^{r_\ell - 1} \ln_\alpha(j) \right\}$$

$$(4.20) \quad = \sum_{\sigma \in S} \sum_{\ell=1}^n \left\{ \frac{n - r_\ell}{r_\ell} \sum_{j=1}^{r_\ell} \ln_\alpha(j) + \sum_{j=1}^{r_\ell - 1} \ln_\alpha(j) \right\}.$$

In the step (4.19), we change an order of summation. The step (4.20) is posed as follows. First, the number of i such that $i \neq \ell$ and $a(\ell, \sigma(i)) = 0$ is equal to $(n - r_\ell)$. Second, the number of i such that $i \neq \ell$ and $a(\ell, \sigma(i)) = 1$ is equal to $(r_\ell - 1)$. These observations allow to compute the sums with respect to i and get (4.20). Adding the term $\text{per}(\mathbf{A}) \sum_{1 \leq \ell \leq n} \ln_\alpha(r_\ell)$ to both (4.16) and (4.20), we immediately obtain

$$\text{per}(\mathbf{A}) (-n) \ln_\alpha \left(\frac{1}{\text{per}(\mathbf{A})} \right) \leq \text{per}(\mathbf{A}) \sum_{\ell=1}^n \frac{n}{r_\ell} \sum_{j=1}^{r_\ell} \ln_\alpha(j).$$

The latter completes the proof. \square

In the limit $\alpha \rightarrow 1^-$, the result (4.11) leads to the previous result (4.7). In this regard, it is a proper extension of (4.2) to the parameter range $\alpha \in (0, 1)$. Together, the bounds (4.2) and (4.11) cover all the values $\alpha > 0$.

REFERENCES

- [1] Alon, N., Spencer, J.H.: The Probabilistic Method. John Wiley & Sons, New York (2008)
- [2] Bengtsson, I., Życzkowski, K.: Geometry of Quantum States: An Introduction to Quantum Entanglement. Cambridge University Press, Cambridge (2006)
- [3] Braunstein, S.L., Caves, C.M.: Information-theoretic Bell inequalities. Phys. Rev. Lett. **61**, 662–665 (1988)
- [4] Brégman, L.M.: Some properties of nonnegative matrices and their permanents. Soviet Math. Dokl. **14**, 945–949 (1973)
- [5] Chung, F.R.K., Frankl, P., Graham, R., Shearer, J.B.: Some intersection theorems for ordered sets and graphs. J. Comb. Theory, Ser. A **43**, 23–37 (1986)
- [6] Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, New York (1991)
- [7] Daróczy, Z.: Generalized information functions. Inform. Control **16**, 36–51 (1970)
- [8] Falniowski, F.: Generalized conditional entropy – determinicity of a process and Rokhlin’s formula. Open Sys. Information Dyn. **22**, 1550025 (2015)
- [9] Furuichi, S.: Information-theoretical properties of Tsallis entropies. J. Math. Phys. **47**, 023302 (2006)
- [10] Galvin, D.: Three tutorial lectures on entropy and counting. arXiv:1406.7872 [math.CO] (2014)
- [11] Galvin, D.: Counting colorings of a regular graph. Graphs Combin. **31**, 629–638 (2015)
- [12] Galvin, D., Tetali, P.: On weighted graph homomorphisms. arXiv:1206.3160 [math.CO] (2012)

- [13] Graham, R.L., Knuth, D.E., Patashnik, O.: Concrete Mathematics. Addison-Wesley, New York (1994)
- [14] Havrda, J., Charvát, F.: Quantification methods of classification processes: concept of structural α -entropy. *Kybernetika* **3**, 30–35 (1967)
- [15] Kahn, J.: An entropy approach to the hard-core model on bipartite graphs. *Comb. Prob. Comp.* **10**, 219–237 (2001)
- [16] Kleitman, D.J., Shearer, J., Sturtevant, D.: Intersections of k -element sets. *Combinatorica* **1**, 381–384 (1981)
- [17] Madiman, M., Marcus, A.W., Tetali, P.: Entropy and set cardinality inequalities for partition-determined functions. *Random Struct. Algorithms* **40**, 1–26 (2012)
- [18] Minc, H.: Upper bounds for permanents of $(0, 1)$ -matrices. *Bull. Amer. Math. Soc.* **69**, 789–791 (1963)
- [19] Mitrinović, D.S.: Analytic Inequalities. Springer-Verlag, Berlin (1970)
- [20] Radhakrishnan, J.: An entropy proof of Brégman’s theorem. *J. Comb. Theory, Ser. A* **77**, 161–164 (1997)
- [21] Radhakrishnan, J.: Entropy and counting. In J.C. Misra, ed., *Computational Mathematics, Modelling and Algorithms*, Narosa Publishers, New Delhi, 146–168 (2003)
- [22] Rastegin, A.E.: Convexity inequalities for estimating generalized conditional entropies from below. *Kybernetika* **48**, 242–253 (2012)
- [23] Rastegin, A.E.: Tests for quantum contextuality in terms of q -entropies. *Quantum Inf. Comput.* **14**, 0996–1013 (2014)
- [24] Rastegin, A.E.: Further results on generalized conditional entropies. *RAIRO–Theor. Inf. Appl.* **49**, 67–92 (2015)
- [25] Rényi, A.: On measures of entropy and information. In J. Neyman, ed., *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 547–561 (1961)
- [26] Schrijver, A.: A short proof of Minc’s conjecture. *J. Comb. Theory, Ser. A* **25**, 80–83 (1978)
- [27] Shitong, W., Chung, F.L.: Note on the equivalence relationship between Rényi-entropy based and Tsallis-entropy based image thresholding. *Pattern Recognition Lett.* **26**, 2309–2312 (2005)
- [28] Słomczyński, W., Kwapien, J., Życzkowski, K.: Entropy computing via integration over fractal measures. *Chaos* **10**, 180–188 (2000)
- [29] Tsallis, C.: Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **52**, 479–487 (1988)

DEPARTMENT OF THEORETICAL PHYSICS, IRKUTSK STATE UNIVERSITY, IRKUTSK, GAGARIN BV. 20, 664003, RUSSIA

E-mail address: rast@api.isu.ru; alexrastegin@mail.ru