Asymptotical Isoperimetric Problem

Rudolf Ahlswede and Zhen Zhang

Summary

Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets. Let $\rho : \mathcal{X} \times \mathcal{Y} \to [0, +\infty)$ be a distortion measure. The distortion measure defined on the power sets $\rho_n : \mathcal{X}^n \times \mathcal{Y}^n \to [0, +\infty)$ is given by the formula

$$\rho_n(x^n, y^n) = \frac{1}{n} \sum_{i=1}^{n} \rho(x_i, y_i) \tag{1}$$

where $x^n = (x_1, \ldots, x_n)$ and $y^n = (y_1, \ldots, y_n)$ ($\forall i$, $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$) are elements of the power sets $\mathcal{X}^n, \mathcal{Y}^n$, respectively.

Let $\mathcal{A} \subset \mathcal{X}^n$. The d-neighbor of $\mathcal{A}$ is defined by

$$\Gamma^d(A) = \{y^n \in \mathcal{Y}^n : \exists x^n \in \mathcal{A} \text{ s.t. } \rho_n(x^n, y^n) \leq d\}. \tag{2}$$

Define

$$G_n(M, d) = \min\{|\Gamma^d(A)| : |\mathcal{A}| \geq M\}. \tag{3}$$

The isoperimetric problem is to determine this function and to find subsets $\mathcal{A}$ of $\mathcal{X}^n$ for which the minimum in the definition of $G_n(M, d)$ is achieved. This is a well known problem in Extremal Combinatorial Theory. It seems to be very difficult. So far the solutions are known only in some very special cases like for the binary Hamming case.

In this paper, we are interested in asymptotic solutions of the isoperimetric problem which is formulated as follows. Let $R > 0$ and $\delta > 0$. Define

$$\gamma(R, \delta) = \inf_r \frac{1}{n} \log_2(G_n(2^{nR}, \delta n)). \tag{4}$$

The problem is to determine this function for given sets $\mathcal{X}, \mathcal{Y}$ and given distortion measure ρ. Here is our solution in terms of auxiliary random variables and conditional entropy H.

Theorem 1. *For given sets $\mathcal{X}, \mathcal{Y}$ and given distortion measure $\rho : \mathcal{X} \times \mathcal{Y} \to [0, +\infty)$, let $\mathcal{U}$ be a finite set of cardinality N. Let X, Y, U be three random variables jointly distributed over $\mathcal{X} \times \mathcal{Y} \times \mathcal{U}$ with probability mass function P_{XYZ}. Define*

$$\xi_N(R, \delta) = \inf_{(XU): H(X|U) \geq R} \sup_{(XY): E[\rho(X,Y)] \leq \delta} H(Y|U). \tag{5}$$

Then

$$\xi_N(R, \delta) \geq \gamma(R, \delta) \geq \xi_N(R, \delta) - O\left(\frac{ln\ N}{N^{\frac{1}{|\mathcal{X}|}}}\right). \tag{6}$$

The key tool for the proof of this result is the so-called Inherently Typical Subset Lemma, which was developed in [3]. We briefly explain the lemma as follows: For each integer $m > 0$, let $\mathcal{P}_m(\mathcal{X})$ denote the set of all m-types on $\mathcal{X}$, that is

$$\mathcal{P}_m(\mathcal{X}) = \left\{ P \in \mathcal{P}(\mathcal{X}) : P(x) \in \left\{ 0, \frac{1}{m}, \frac{2}{m}, \ldots, \frac{m-1}{m}, 1 \right\} \forall\, x \in \mathcal{X} \right\}, \tag{7}$$

with $\mathcal{P}(\mathcal{X})$ as set of all probability distributions on $\mathcal{X}$.

Let $\mathcal{U}_m = \{u_1, \ldots, u_{|\mathcal{P}_m(x)|}\}$ be an arbitrary set. Since $|\mathcal{U}_m| = |\mathcal{P}_m(\mathcal{X})|$, we can associate with each $P \in \mathcal{P}_m(\mathcal{X})$ an element $u \in \mathcal{U}_m$ so that elements of $\mathcal{U}_m$ associated with distinct m-types are distinct. If $u \in \mathcal{U}_m$ is associated with $P \in \mathcal{P}_m(\mathcal{X})$, for convenience, we shall write P as $P(\cdot|u)$. In terms of this notation, we have

$$\mathcal{P}_m(\mathcal{X}) = \{P(\cdot|u) : u \in \mathcal{U}_m\}. \tag{8}$$

Let A be any subset of $\mathcal{X}^n$. For any $0 \le i \le n-1$, define

$$A_i = \{x^i \in \mathcal{X}^i : x^i \text{ is a prefix of some element of } A\}. \tag{9}$$

Here, we make use of the convention that $A_0 = \{\Lambda\}$, where Λ is the empty string. Assume the integer m is greater than or equal to $2^{16|\mathcal{X}|^2}$.

Definition 1. *$A \subset \mathcal{X}^n$ is called m-inherently typical if there exists a mapping $\phi : \bigcup_{i=0}^{n-1} A_i \to \mathcal{U}_m$ such that the following holds:*

(i) There exists an n-type $Q \in \mathcal{P}_n(\mathcal{X} \times \mathcal{U}_m)$ such that for any $x^n \in A$,

$$P_{x^n u^n}(x, u) = Q(x, u), \quad x \in \mathcal{X}, \quad u \in \mathcal{U}_m \tag{10}$$

where $u^n = (u_1, u_2, \ldots, u_n) \in \mathcal{U}_m^n$ is a sequence defined by $u_i = \phi(x^{i-1})$ for all $i : 1 \le i \le n$, (such a sequence is called a sequence associated with x^n through ϕ) and for any $x \in \mathcal{X}$ and any $u \in \mathcal{U}_m$,

$$P_{x^n u^n}(x, u) = \frac{1}{n} |\{i : (x_i, u_i) = (x, u)\}|. \tag{11}$$

(ii) If $(\hat{X}, \hat{U})$ is a pair of random variables taking values on $\mathcal{X} \times \mathcal{U}_m$ with joint distribution Q, then

$$\frac{1}{n} \log |A| \le H(\hat{X}|\hat{U}) \le \frac{1}{n} \log |A| + \frac{\log^2 m}{m}. \tag{12}$$

Let $A \subset \mathcal{X}^n$ be m-inherently typical. Let ϕ be the corresponding mapping such that (10) and (12) are satisfied. For any random vector $\tilde{X}^n = (\tilde{X}_1, \ldots, \tilde{X}_n)$ taking values on A, we define another random vector $\tilde{U}_m = (\tilde{U}_1, \ldots, \tilde{U}_n)$ by letting $\tilde{U}_i = \phi(\tilde{X}^{i-1})$ for all $i : 1 \le i \le n$. Clearly (10) implies that with probability one, the following holds:

$$P_{\tilde{X}^n \tilde{U}^n}(x, u) = \frac{1}{n} \sum_{1}^{n} Pr\{\tilde{X}_i = x, \tilde{U}_i = u\} \quad x \in \mathcal{X}, \quad u \in \mathcal{U}_m. \tag{13}$$

Note that the left hand side of (13) is the frequency, i.e. the average over time, and the right hand side is the average probability over the ensemble. Intuitively, therefore, (13) just says that with probability one, the average over time is equal to the average over the ensemble. This is where the word "inherently typical" comes from. In typical applications, the random vector $\tilde{X}^n$ is often assumed to be uniformly distributed on A. In this case

$$\frac{1}{n}\log|A| = \frac{1}{n}H(\tilde{X}^n) = \frac{1}{n}\sum_{i=1}^{n} H(\tilde{X}_i|\tilde{X}^{i-1}). \tag{14}$$

Let I be a random variable taking values uniformly on $\{1,\ldots,n\}$ and independent of $\tilde{X}^n$. Let $\tilde{X} = \tilde{X}_I$ and $U = (\tilde{X}^{I-1}, I)$, then

$$\frac{1}{n}\log|A| = H(\tilde{X}|U). \tag{15}$$

If we extend the mapping ϕ in the obvious way so that $\phi(U) = \phi(\tilde{X}^{I-1})$ whenever $U = (\tilde{X}^{I-1}, I)$, then it is not hard to see that $\tilde{X}$ and $\tilde{U}$ have the joint distribution $P_{\tilde{X}\tilde{U}} = Q$ where $\tilde{U} = \phi(U)$. Therefore, (12) just says that

$$H(\tilde{X}|U) \le H(\tilde{X}|\tilde{U}) \le H(\tilde{X}|U) + \frac{\log^2 m}{m}. \tag{16}$$

Lemma 1 (Inherently Typical Subset Lemma). *For any $m \ge 2^{16|\mathcal{X}|^2}$, n satisfying $((m+1)^{5|\mathcal{X}|+4}ln(n+1))/n \le 1$, and any $A \subset \mathcal{X}^n$, there exists an m-inherently typical subset $\tilde{A} \subset A$ such that*

$$\frac{1}{n}\log\frac{|A|}{|\tilde{A}|} \le |\mathcal{X}|(m+1)^{|\mathcal{X}|}\frac{\log(n+1)}{n}. \tag{17}$$

References

[1] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, Inc., N.Y., 1981.

[2] L.H. Harper, Optimal numbering and isoperimetric problems on graphs, Journal Comb. Th., No. 1, 385–394, 1966.

[3] R. Ahlswede, E.-H. Yang and Z. Zhang, Identification via compressed data, IEEE Trans. on Inform. Theory, Vol. 43, No. 1, 48-70, Jan. 1997.