

Average Entropy Functions

Qi Chen, Chen He, Lingge Jiang, Qingchuan Wang

Dept. of Electronic Eng.

Shanghai Jiao Tong Univ.

Shanghai, China 200240

Email: {cq094, chenhe, lgjiang, r6144}@sjtu.edu.cn

Abstract—THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD. The closure of the set of entropy functions associated with n discrete variables, $\bar{\Gamma}_n^*$, is a convex cone in $(2^n - 1)$ -dimensional space, but its full characterization remains an open problem. In this paper, we map $\bar{\Gamma}_n^*$ to an n -dimensional region $\bar{\Phi}_n^*$ by averaging the joint entropies with the same number of variables, and show that the simpler $\bar{\Phi}_n^*$ can be characterized solely by the Shannon-type information inequalities.

I. INTRODUCTION

Given an n -dimensional discrete random vector $\mathbf{X} = (X_1, \dots, X_n)$, for each non-empty subset α of $\mathcal{N} = \{1, 2, \dots, n\}$ there is a joint entropy $H(X_\alpha)$ with $X_\alpha = (X_i)_{i \in \alpha}$, and the $2^n - 1$ joint entropies form the entropy function $(H(X_\alpha))_{\alpha \subseteq \mathcal{N}, \alpha \neq \emptyset}$ of \mathbf{X} . We can then define $\Gamma_n^* \subseteq \mathbb{R}^{2^n - 1}$ as the set of all possible entropy functions involving n discrete random variables, and $\bar{\Gamma}_n^*$ as its closure. A vector $\mathbf{H} \in \mathbb{R}^{2^n - 1}$ is called entropic if $\mathbf{H} \in \Gamma_n^*$, and almost entropic if $\mathbf{H} \in \bar{\Gamma}_n^*$ [1].

All $\mathbf{H} = (H_\alpha)_{\alpha \subseteq \mathcal{N}, \alpha \neq \emptyset} \in \bar{\Gamma}_n^*$ satisfy the following Shannon-type information inequalities for any subsets α, β of \mathcal{N} (we let $H_\emptyset = 0$ for convenience):

$$H_\alpha \geq 0, \quad (1)$$

$$H_\alpha \leq H_\beta, \quad \alpha \subseteq \beta, \quad (2)$$

$$H_\alpha + H_\beta \geq H_{(\alpha \cup \beta)} + H_{(\alpha \cap \beta)}. \quad (3)$$

However, (1)–(3) are not sufficient conditions for an $\mathbf{H} \in \mathbb{R}^{2^n - 1}$ to be almost entropic when $n \geq 4$ [2]. In other words, denoting by Γ_n the set of vectors in $\mathbb{R}^{2^n - 1}$ satisfying (1)–(3), we have

$$\bar{\Gamma}_n^* \subsetneq \Gamma_n, \quad n \geq 4. \quad (4)$$

A number of non-Shannon-type information inequalities satisfied by the members of $\bar{\Gamma}_n^*$ have subsequently been found in [2]–[4], but the full characterization of $\bar{\Gamma}_n^*$ remains an open problem.

In this paper, we will show that an averaged version of $\bar{\Gamma}_n^*$ can be more easily characterized.

Definition 1: For a vector $\mathbf{H} = (H_\alpha)_{\alpha \subseteq \mathcal{N}, \alpha \neq \emptyset} \in \mathbb{R}^{2^n - 1}$, we define its average as

$$\Psi(\mathbf{H}) \triangleq (h_1, \dots, h_n), \quad (5)$$

where $h_k = \binom{n}{k}^{-1} \sum_{|\alpha|=k} H_\alpha$. If \mathbf{H} is the entropy function of random vector \mathbf{X} , we call $\mathbf{h} = \Psi(\mathbf{H})$ the average entropy function. Ψ then maps Γ_n^* to the set $\Phi_n^* \triangleq \Psi(\Gamma_n^*)$ of all

average entropy functions, $\bar{\Gamma}_n^*$ to the closure $\bar{\Phi}_n^*$, and Γ_n to $\Phi_n \triangleq \Psi(\Gamma_n)$.

From the definition (1)–(3) of Γ_n , Φ_n can be given by

$$\Phi_n = \{(h_1, \dots, h_n) \mid h_{k-1} - 2h_k + h_{k+1} \leq 0, \quad k = 1, \dots, n\}, \quad (6)$$

where we let $h_0 = 0$ and $h_{n+1} = h_n$ for convenience. $\bar{\Phi}_n^*$ is obviously a subset of $\bar{\Phi}_n$ since $\bar{\Gamma}_n^* \subseteq \Gamma_n$, but we will show that they are actually equal. In other words, $\bar{\Phi}_n^*$ is characterizable solely with the Shannon-type information inequalities.

Theorem 1: $\bar{\Phi}_n^* = \bar{\Phi}_n$.

This theorem will be proved in the next section.

II. PROOF OF THE THEOREM

It is only necessary to prove that $\bar{\Phi}_n \subseteq \bar{\Phi}_n^*$. We first introduce a one-to-one transform to give $\bar{\Phi}_n$ a simpler form.

Definition 2: For a vector $\mathbf{h} = (h_1, \dots, h_n) \in \mathbb{R}^n$, we define its second-order difference as

$$\Theta(\mathbf{h}) = (g_1, \dots, g_n), \quad (7)$$

where $g_k = h_{k-1} - 2h_k + h_{k+1}$, $k = 1, \dots, n$, with $h_0 = 0$ and $h_{n+1} = h_n$. Θ maps $\bar{\Phi}_n^*$ to $\Lambda_n^* \triangleq \Theta(\bar{\Phi}_n^*)$, $\bar{\Phi}_n$ to $\bar{\Lambda}_n$, and Φ_n to $\Lambda_n \triangleq \Theta(\Phi_n)$.

From (6), we have

$$\Lambda_n = \{(g_1, \dots, g_n) \mid g_k \leq 0, \quad k = 1, \dots, n\}. \quad (8)$$

As Ψ and Θ are both linear maps, and $\bar{\Gamma}_n^*$ is a convex cone [5], $\bar{\Phi}_n^*$ and $\bar{\Lambda}_n^*$ are both convex cones as well. Therefore, to prove that $\bar{\Phi}_n \subseteq \bar{\Phi}_n^*$ or equivalently $\Lambda_n \subseteq \bar{\Lambda}_n^*$, it is sufficient to prove that

$$\mathbf{g}_k \triangleq (\underbrace{0, \dots, 0}_{k-1}, -a, 0, \dots, 0) \in \Lambda_n^* \quad (9)$$

for $k = 1, \dots, n$ and some $a > 0$. In other words, for each k we need to find a random vector \mathbf{X} whose average entropy function is

$$\mathbf{h}_k \triangleq \Theta^{-1}(\mathbf{g}_k) = a \cdot (1, 2, \dots, k, \dots, k). \quad (10)$$

This \mathbf{X} can be constructed from a Reed-Solomon code. Specifically, let q be a power-of-two larger than n , \mathcal{C} be the codeword set of an (n, k) Reed-Solomon code on $\text{GF}(q)$, and $\mathbf{X} = (X_1, \dots, X_n)$ be a random codeword uniformly

distributed over \mathcal{C} , then the entropy function of \mathbf{X} is (10) with $a = \log q$, as shown below.

Let j_1, \dots, j_n be distinct indices in $1, \dots, n$. According to the properties of Reed-Solomon codes, given any $x_{j_1}^*, \dots, x_{j_k}^* \in \text{GF}(q)$, there exists a unique $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{C}$ with $x_{j_l} = x_{j_l}^*$, $l = 1, \dots, k$. For any $x_{j_1}^* \in \text{GF}(q)$, there are thus q^{k-1} codewords $\mathbf{x} \in \mathcal{C}$ with $x_{j_1} = x_{j_1}^*$, one for each value combination on $k-1$ other positions, and since \mathbf{X} is equal to each codeword with probability q^{-k} , we have $p(X_{j_1} = x_{j_1}^*) = q^{-1}$, so $H(X_{j_1}) = \log q$. Similarly, $H(X_{j_1}, X_{j_2}) = 2 \log q$, \dots , $H(X_{j_1}, \dots, X_{j_k}) = k \log q$. For $l = k+1, \dots, n$, given x_{j_1}, \dots, x_{j_l} , there is either one matching codeword in \mathcal{C} or none, therefore $p(X_{j_1} = x_{j_1}, \dots, X_{j_l} = x_{j_l})$ is q^{-k} on its support, and $H(X_{j_1}, \dots, X_{j_l}) = k \log q$. Consequently, the average entropy function of \mathbf{X} is (10) with $a = \log q$ as desired, and for each l , all $\binom{n}{l}$ l -variable joint entropies of \mathbf{X} that are being averaged actually have the same value. ■

III. DISCUSSION

Determination of $\bar{\Gamma}_n^*$ is important due to its close connection to the capacity region of general multi-source multi-sink wired networks [6], [7], but this seems to be a difficult problem, and even if a full characterization is found, computational difficulties due to $\bar{\Gamma}_n^*$'s high dimensionality and complex structure might reduce its usefulness in practice [8]. What we have shown is that the region $\bar{\Phi}_n^*$ obtained by averaging the k -variable joint entropies has a much simpler structure: it is not affected by the non-Shannon information inequalities, and the linear Reed-Solomon codes used in the proof suggest that the suboptimality of linear network coding is also hidden by this averaging. On one hand, this means that further work on the characterization of $\bar{\Gamma}_n^*$ must focus on the variation among the k -variable entropies, not just their averages. On the other hand, many practically interesting networks have a somewhat symmetric structure, possibly in a statistical sense, and an appropriately averaged version of $\bar{\Gamma}_n^*$ (not necessarily as simplistic as $\bar{\Phi}_n^*$) might provide a tractable method for the determination of their capacity regions.

Average entropy functions are also closely related to the MAP EXIT functions discussed in e.g. [9] for large n .

ACKNOWLEDGMENT

This paper was supported by National Natural Science Foundation of China Grants No. 60772100 and No. 60872017.

REFERENCES

- [1] R. W. Yeung, "A framework for linear information inequalities," *IEEE Trans. Inf. Theory*, vol. 43, pp. 1924–1934, Nov. 1997.
- [2] Z. Zhang and R. W. Yeung, "On characterization of entropy function via information inequalities," *IEEE Trans. Inf. Theory*, vol. 44, pp. 1440–1452, Nov. 1998.
- [3] X. Yan, R. Yeung and Z. Zhang, "A class of non-Shannon type information inequalities and their applications," *IEEE Int. Symp. Inf. Theory*, Washington, DC, June 2001.
- [4] R. Doughter, C. Freiling and K. Zeger, "Six new non-Shannon information inequalities," *IEEE Int. Symp. Inf. Theory*, Seattle, WA, June 2006.

- [5] Z. Zhang and R. W. Yeung, "A non-Shannon type conditional inequality of information quantities," *IEEE Trans. Inf. Theory*, vol. 43, pp. 1982–1986, Nov. 1997.
- [6] X. Yan, R. Yeung and Z. Zhang, "The capacity region for multi-source multi-sink network coding," *IEEE Int. Symp. Inf. Theory*, Nice, France, June 2007.
- [7] T. Chan and A. Grant, "Dualities between entropy functions and network codes," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4470–4487, Oct. 2008.
- [8] F. Matúš, "Infinitely many information inequalities," *IEEE Int. Symp. Inf. Theory*, Nice, France, June 2007.
- [9] C. Measson, A. Montanari, and R. Urbanke, "Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding," Jun. 2005, arXiv:cs.IT/0506083.