

Article

Optimizing Distributions for Associated Entropic Vectors via Generative Convolutional Neural Networks

Shuhao Zhang ¹, Nan Liu ², Wei Kang ^{1,*} and Haim Permuter ³

¹ School of Information Science and Engineering, Southeast University, Nanjing 211189, China; shzhang@seu.edu.cn

² National Mobile Communications Research Laboratory, Southeast University, Nanjing 211189, China; nanliu@seu.edu.cn

³ Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beersheba 8410501, Israel; haimp@bgu.ac.il

* Correspondence: wkang@seu.edu.cn

Abstract: The complete characterization of the almost-entropic region yields rate regions for network coding problems. However, this characterization is difficult and open. In this paper, we propose a novel algorithm to determine whether an arbitrary vector in the entropy space is entropic or not, by parameterizing and generating probability mass functions by neural networks. Given a target vector, the algorithm minimizes the normalized distance between the target vector and the generated entropic vector by training the neural network. The algorithm reveals the entropic nature of the target vector, and obtains the underlying distribution, accordingly. The proposed algorithm was further implemented with convolutional neural networks, which naturally fit the structure of joint probability mass functions, and accelerate the algorithm with GPUs. Empirical results demonstrate improved normalized distances and convergence performances compared with prior works. We also conducted optimizations of the Ingleton score and Ingleton violation index, where a new lower bound of the Ingleton violation index was obtained. An inner bound of the almost-entropic region with four random variables was constructed with the proposed method, presenting the current best inner bound measured by the volume ratio. The potential of a computer-aided approach to construct achievable schemes for network coding problems using the proposed method is discussed.



Citation: Zhang, S.; Liu, N.; Kang, W.; Permuter, H. Optimizing Distributions for Associated Entropic Vectors via Generative Convolutional Neural Networks. *Entropy* **2024**, *26*, 711. <https://doi.org/10.3390/e26080711>

Academic Editor: Boris Ryabko

Received: 15 June 2024

Revised: 12 August 2024

Accepted: 20 August 2024

Published: 21 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: entropic vectors; entropic region; neural networks; convolutional neural networks; Ingleton score; Ingleton violation index; inner bounds; network coding

1. Introduction

Given n discrete random variables, for a fixed joint distribution, all their $2^n - 1$ (joint) entropies define an entropic vector in the entropy space $\mathbb{R}^{2^n - 1}$. By varying over all possible joint distributions, the set of all entropic vectors defines the entropic region. The entropic region plays a fundamental role in information theory and network coding. The closure, known as the almost-entropic region, yields rate regions for multi-source coded networks [1,2]. However, the complete characterization of the entropic region encounters challenges even when $n = 3$ [3–7], where only the closure, i.e., the almost-entropic region, is fully characterized [3]. When $n = 4$, the characterization of the almost-entropic region also becomes extremely difficult and remains open [8].

In the pursuit of the characterization of the entropic region, one crucial problem is to *determine whether an arbitrary vector in the entropy space is entropic or not*. If all entropic vectors in the entropy space are verified, the entropic region is fully characterized. Additionally, given a point in the rate region of a coded network, the existence of achievable codes relies on the underlying entropic vector [9].

This problem has been tackled from different perspectives in the literature. Information inequalities, which fully characterize all almost-entropic vectors, are shown to be infinitely

many [10] and extremely hard to characterize [8,10–17]. The construction of entropic vectors, however, is more feasible and studied through probability distributions, e.g., [6,7,16,18–28], groups, e.g., [29–35], and matroids, e.g., [36–40]. Regarding the construction of entropic vectors from probability distributions, remarkable research has attempted to numerically obtain probability mass functions (PMFs) for entropic vectors [16,19–27], particularly when the alphabet sizes of the involved random variables are bounded. In this way, the entropic nature of a given vector in the entropy space is immediately verified following the definition if the underlying joint PMF is obtained.

Among these approaches, some are interested in quasi-uniform PMFs [19,27] or PMFs supported on small atoms [22–25], while more general PMFs are considered in [16,20,21,26]. For example, an algorithm is given in [20] to verify binary entropic vectors by recursively constructing PMFs. In [16], the PMFs are numerically obtained with Newton's method. In [21], the study of the almost-entropic region when $n = 4$ is reduced to a three-dimensional tetrahedron and visualized by various maximization procedures with different methods and strategies [21] (Section VII). Recently, a random search algorithm has been introduced [26] to find the nearest entropic vector to a given target vector. This algorithm iteratively tries a few random perturbations on PMFs, striving to decrease the normalized distance between an entropic vector and the target vector.

However, the limitations of the above methods cannot be overlooked. The method in [20] is only feasible for binary entropic vectors. In [16], and [21], formal algorithms and performances of described methods are not systematically discussed in detail. Although [26] presents more analytical and empirical results, the limitation of the randomized algorithm persists, and the convergence performances are not empirically guaranteed. Furthermore, a critical issue in above methods is that only PMFs with relatively small alphabet sizes can be handled. This limitation is vital, since the verification of an underlying entropic vector in the rate region of a coded network requires large alphabets when subpacketizations are necessary for achievable codes.

With these methods, which construct entropic vectors from PMFs, some interesting optimization problems over the entropic region when $n = 4$ can be experimented. It is known that the Ingleton inequality [41] completely characterizes the part of the almost-entropic region achieved by linearly representable entropic vectors [13]; thus, it reduces the study to the entropic vectors violating it. To measure the degree of violation from the Ingleton inequality for an entropic vector, the *Ingleton score* [16,42] and *Ingleton violation index* [19] are proposed from different angles. Although remarkable efforts have been made to minimize the Ingleton score [16,21,26,35,42] and maximize the Ingleton violation index [19,26], their optimal values remain open. In [16], the Ingleton score is numerically optimized and conjectured to be -0.089373 , known as the four-atom conjecture [16] (Conjecture 1). In [21], a technique is proposed to transform an entropic vector into another entropic vector, and the transformed vector is optimized to the Ingleton score -0.09243 , which refutes the four-atom conjecture by [16]. By experimenting with groups [35], the best-known Ingleton score is currently -0.0925 . The Ingleton violation index is optimized in [19] and, subsequently, in [43], and its best known value is currently 0.0281316 , according to [26].

As the complete characterization of the almost-entropic region for $n = 4$ is difficult, inner bounds can be constructed by taking the convex hull of certain acquired entropic vectors, as previously investigated in [16,22,26]. The quality of such an inner bound is measured by the volume ratio, i.e., the percentage of inner-bound polytope volume to the volume of the outer-bound polytope. In [16], several entropic vectors are optimized to form an inner bound with a volume ratio of 53.4815%. Using distributions supported on small atoms, [22] finds entropic vectors that yield an inner bound with the volume ratio 57.8%. In [26], with their proposed grid method, the current best-known volume ratio is 62.4774%, while the largest volume ratio, which requires the complete characterization of the almost-entropic region when $n = 4$, is still open.

In this paper, we develop a novel architecture for optimizing PMFs for associated entropic vectors via convolution neural networks (CNNs). Recently, applications of neural networks

(NNs) can be found in information theory [44–46] and control theory [47,48]. Our motivation arises from the fruitful research on applying NNs to problems in information theory [44–46]. More specifically, in [45], a unique mapping that generates conditional PMFs is defined and approximated using NNs. In [46], a model which generates a capacity-achieving input PMF for a given channel in discrete input spaces is proposed. Consequently, we are especially interested in approximating another mapping, which generates joint PMFs for multiple random variables. Sharing similar ideas with [26], the problem of verifying an entropic vector can be performed by minimizing the normalized distance through NN training. Furthermore, we are motivated to modify NNs due to the special structures and complexities of joint PMFs. Additionally, the developed method can be applied immediately to the optimizations of the Ingleton score and Ingleton violation index, and the construction of inner bounds for the almost-entropic region.

The major contributions of this paper are summarized as follows.

- A novel algorithm is proposed to optimize distributions for entropic vectors with NN training. For each target vector, an NN is trained such that the output PMF produces an entropic vector as close to target as possible. In practice, we implement the algorithm with CNNs, which accelerate and enable the algorithm to generate PMFs with large alphabets.
- The effectiveness of our proposed method is verified by empirical results. More specifically, smaller normalized distances and improved convergence performances are achieved by our proposed method, compared to [26]. In addition, with derived theoretical guarantees, by exploiting the proposed algorithm, the state-of-the-art Ingleton score is reconfirmed, and a new tighter lower bound of the Ingleton violation index is obtained. Furthermore, by utilizing the proposed algorithm, we develop another algorithm to construct a new inner bound of the almost-entropic region ($n = 4$), yielding the current best inner bound measured by the volume ratio.

This paper is organized as follows. Section 2 introduces preliminaries, notations, and the problem statement. The proposed method and algorithm with derived theoretical guarantees are presented in Section 3, and the implementation with CNNs is demonstrated in Section 4. In Section 5, empirical results exploiting the proposed method for several problems when $n = 4$ are presented. Section 6 summarizes the paper in general, and discusses the potential of the proposed method to construct achievable schemes for network coding problems.

2. Preliminaries and Problem Statement

In this section, we provide the preliminaries, notations, and the problem statement of this paper.

2.1. Convex Cones and Convex Polytopes

Given a set $C \subset \mathbb{R}^d$, C is a *pointed cone* if $\mathbf{a} \in C$ implies that $t\mathbf{a} \in C$ for all real $t \geq 0$, and $R_{\mathbf{a}} = \{t\mathbf{a}, t \in \mathbb{R}^+\} \subset C$ is a *ray* of C . Pointed cones are unbounded except $\{\mathbf{0}\}$, where $\mathbf{0}$ is the origin of Euclidean space. In the rest of the paper, all cones are assumed to be pointed and unbounded. For all real $t \geq 0$, $R_{\mathbf{b}} = \{t\mathbf{b}\}$ of C is an *extreme ray* if \mathbf{b} cannot be expressed as the positive linear combination of any $\mathbf{a}_1, \mathbf{a}_2 \in C$. A cone C is convex if C is a convex set.

Given a convex set $C \subset \mathbb{R}^d$, C is a *convex polyhedron* if it is the intersection of finitely many halfspaces (i.e., linear inequalities). A convex cone C is a special polyhedron when the halfspaces that define C contain $\mathbf{0}$ simultaneously, and, in this case, the convex cone C is called *polyhedral*. A convex polyhedron is a *convex polytope* if it is bounded. In the rest of the paper, all polytopes are assumed to be convex. For a convex cone (or convex polytope), if we specify one extreme ray (or vertex) as the *top* and others as the *bases*, then the convex cone (or convex polytope) is often called the *pyramid*.

For a convex polyhedron C , there are two equivalent types of representations, i.e., the *H-representation* and the *V-representation* [49] (Chapter 1). The H-representation is the set of all halfspaces defining C , and the V-representation is the set of all vertices (if C is a convex

polytope) or extreme rays (if C is a polyhedral convex cone) defining C . The transformation between the H-representation and the V-representation for a given convex cone or convex polytope can be numerically performed with the implementation [50–52] of the double description method [53]. More details about convex cones and convex polytopes can be found in [49].

2.2. Entropic Vectors and Entropic Region

Let $n \geq 2$, consider a discrete random vector $\mathbf{X} \triangleq (X_1, X_2, \dots, X_n)$ with a finite index set $\mathcal{N}_n = \{1, 2, \dots, n\}$, and \mathbf{X} takes values in the finite alphabet $\mathcal{X} \triangleq \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_n$ of size $|\mathcal{X}| = m = \prod_{i=1}^n m_i$. The realization of \mathbf{X} is denoted as \mathbf{x} . Let $p(\mathbf{x})$ denote the probability of \mathbf{x} such that $p(\mathbf{x}) = \Pr\{\mathbf{X} = \mathbf{x}\}$. The joint probability mass function (PMF) of \mathbf{X} is denoted as $\mathbf{p} \triangleq [p(\mathbf{x}), \mathbf{x} \in \mathcal{X}]^T \in \Delta_{\mathcal{X}}^m$, where $\Delta_{\mathcal{X}}^m$ is the m -dimensional probability simplex defined as

$$\Delta_{\mathcal{X}}^m \triangleq \left\{ \mathbf{p} \in \mathbb{R}^m \mid \sum p(\mathbf{x}) = 1, p(\mathbf{x}) \geq 0, \forall \mathbf{x} \in \mathcal{X} \right\}. \tag{1}$$

A vector in \mathbb{R}^m is a *valid PMF* if it belongs to the set $\Delta_{\mathcal{X}}^m$.

We consider a subset of components of the random vector \mathbf{X} ; the above quantities can be denoted accordingly. More specifically, we consider the set of random variables $\mathbf{X}_\alpha \triangleq \{X_i, i \in \alpha\}$, $\alpha \subseteq \mathcal{N}_n \setminus \emptyset$, which takes values in the finite alphabet $\mathcal{X}_\alpha = \prod_{i \in \alpha} \mathcal{X}_i$. The realization of \mathbf{X}_α is denoted as \mathbf{x}_α . Let $p_\alpha(\mathbf{x}_\alpha)$ denote the probability of \mathbf{x}_α such that $p_\alpha(\mathbf{x}_\alpha) = \Pr\{\mathbf{X}_\alpha = \mathbf{x}_\alpha\}$. The marginal PMF of \mathbf{X}_α is denoted as $\mathbf{p}_\alpha \triangleq [p_\alpha(\mathbf{x}_\alpha), \mathbf{x}_\alpha \in \mathcal{X}_\alpha]^T$. Then, the Shannon entropy of \mathbf{X}_α is

$$H(\mathbf{X}_\alpha) = - \sum_{\mathbf{x}_\alpha \in \mathcal{X}_\alpha} p_\alpha(\mathbf{x}_\alpha) \log_2 p_\alpha(\mathbf{x}_\alpha), \quad \forall \alpha \subseteq \mathcal{N}_n \setminus \emptyset, \tag{2}$$

where $0 \log_2 0 \triangleq 0$. We define $h_\alpha \triangleq H(\mathbf{X}_\alpha)$, which is the *entropy function*. The vector consisting of entropy functions for all $\alpha \subseteq \mathcal{N}_n \setminus \emptyset$ is the *entropic vector*, as formally defined in the following definition.

Definition 1 (Entropic vectors [54], Chapter 13). *For the random vector \mathbf{X} and index set \mathcal{N}_n , given a joint PMF $\mathbf{p} \in \Delta_{\mathcal{X}}^m$, the entropic vector is defined as*

$$\mathbf{h} \triangleq \left(h_\alpha, \alpha \subseteq \mathcal{N}_n \setminus \emptyset \right)^T, \tag{3}$$

and \mathbf{h} is the associated entropic vector of \mathbf{p} , denoted as $\mathbf{h}^{\mathbf{p}}$.

We note that, by varying α , there are total $2^n - 1$ (joint) entropies for the random vector \mathbf{X} . Thus, each \mathbf{h} can be viewed as a vector in the $(2^n - 1)$ -dimensional Euclidean space, which is defined as the *entropy space*

$$\mathcal{H}_n \triangleq \mathbb{R}^{2^n - 1}, \quad n \geq 2. \tag{4}$$

For example, given a random vector $\mathbf{X} = (X_1, X_2, X_3, X_4)$, let each random variable in \mathbf{X} be uniformly i.i.d.-distributed on alphabet $\{0, 1\}$; then, the associated entropic vector in the entropy space $\mathcal{H}_4 = \mathbb{R}^{15}$ is

$$\begin{aligned} \mathbf{h}^{\mathbf{p}} &= (h_1, h_2, h_3, h_4, h_{12}, h_{13}, h_{14}, h_{23}, h_{24}, h_{34}, h_{123}, h_{124}, h_{134}, h_{234}, h_{1234})^T. \\ &= (1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4)^T. \end{aligned} \tag{5}$$

Allowing infinite alphabets, the region in \mathcal{H}_n consisting of all entropic vectors is the *entropic region* ([54], Chapter 13)

$$\Gamma_n^* \triangleq \left\{ \mathbf{h} \in \mathcal{H}_n \mid \mathbf{h} \text{ is entropic vector} \right\}. \tag{6}$$

The closure of Γ_n^* is the *almost-entropic region* $\bar{\Gamma}_n^*$ ([54], Chapter 15).

Given a set of random variables, *basic inequalities* form the set of inequalities implied by the nonnegativity of all Shannon’s information measures. The nonredundant subset of basic inequalities consists of *elemental inequalities* ([54], Chapter 14), which are defined as the following two types of inequalities for the random vector \mathbf{X}

$$H(X_i | \mathbf{X}_{\mathcal{N}_n \setminus \{i\}}) \geq 0, \quad i \in \mathcal{N}_n, \tag{7}$$

$$I(X_i; X_j | \mathbf{X}_K) \geq 0, \quad i \neq j, K \subset \mathcal{N}_n \setminus \{i, j\}. \tag{8}$$

The inequalities implied by elemental inequalities are *Shannon-type inequalities*. The region in \mathcal{H}_n that consists of vectors satisfying all elemental inequalities is ([54], Chapter 14)

$$\Gamma_n \triangleq \left\{ \mathbf{h} \in \mathcal{H}_n \mid \mathbf{h} \text{ satisfies (7) and (8)} \right\}. \tag{9}$$

Sometimes Γ_n is called the *polymatroidal region* because of the equivalence between the elemental inequalities for vectors in the entropy space and the polymatroidal axioms for polymatroidal rank functions [55].

Although both Γ_n^* and Γ_n are regions within \mathcal{H}_n , the former stands for the associated entropic vectors defined by valid PMFs, while the latter is formed by vectors satisfying all Shannon-type inequalities. Hence, it is reasonable to question the identity between Γ_n^* and Γ_n . It was first discovered by [3,8] that Γ_n is a loose outer bound of Γ_n^* . More specifically, it is known that $\Gamma_2^* = \Gamma_2$, $\Gamma_3^* \neq \Gamma_3$ but $\bar{\Gamma}_3^* = \Gamma_3$, and $\bar{\Gamma}_n^* \subset \Gamma_n$ when $n \geq 4$ [3,8]. These relations reveal that there are inequalities tighter than Shannon-type inequalities as the outer bound of the entropic region, i.e., the inequalities hold for all entropic vectors but cannot be implied by elemental inequalities, and these inequalities are often referred to as the *non-Shannon-type inequalities* ([54], Chapter 15).

We briefly introduce some known structures of Γ_n^* , $\bar{\Gamma}_n^*$ and Γ_n . Both Γ_n and $\bar{\Gamma}_n^*$ are pointed convex cones in the nonnegative orthant of \mathcal{H}_n [54] (Chapter 13). The region Γ_n is polyhedral, while $\bar{\Gamma}_n^*$ is not when $n \geq 4$ [10], i.e., the convex cone Γ_4 ($\bar{\Gamma}_4^*$) is represented with finitely many (infinitely many) extreme rays or finitely many (infinitely many) linear inequalities. For $n \geq 3$, although the almost-entropic region $\bar{\Gamma}_n^*$ is a convex cone, the region Γ_n^* is not convex and $\Gamma_n^* \subset \bar{\Gamma}_n^*$. The complete characterization of $\bar{\Gamma}_n^*$ for $n \geq 4$ is difficult and open. For more details of Γ_n^* , $\bar{\Gamma}_n^*$, and Γ_n , please refer to [54] (Chapter 13–15).

Given the difficulty of characterizing Γ_n^* with infinite alphabets, in order to optimize finite PMFs numerically, it is practical to consider Γ_n^* when the random variables have finite alphabet sizes.

Definition 2 (Alphabet-bounded entropic region [54], Chapter 21). *Given a random vector \mathbf{X} with an index set \mathcal{N}_n , taking values on the finite alphabet \mathcal{X} of size $m = \prod_{i=1}^n m_i$ such that $m < \infty$, the alphabet-bounded entropic region is defined as*

$$\Gamma_{n,\mathcal{X}}^* \triangleq \left\{ \mathbf{h}^{\mathbf{p}} \in \mathcal{H}_n \mid \mathbf{p} \in \Delta_{\mathcal{X}}^m \right\}, \tag{10}$$

where $\mathbf{h}^{\mathbf{p}}$ is the entropic vector associated with the PMF \mathbf{p} , and $\Delta_{\mathcal{X}}^m$ is the probability simplex defined on the finite alphabet \mathcal{X} of size m .

The region $\Gamma_{n,\mathcal{X}}^*$ is the collection of entropic vectors associated with PMFs defined on the finite alphabet \mathcal{X} , and is a compact and closed set.

To characterize the closeness of two given vectors, the most straightforward measurement is the angle between the rays determined by them. Here, we adopt the measure of *normalized distance*, proposed in [26].

Definition 3 (Normalized distance). *Given vectors $\mathbf{a}, \mathbf{b} \in \Gamma_n \setminus \{0\}$, the normalized distance between \mathbf{a} and \mathbf{b} is defined as*

$$d_{\text{norm}}(\mathbf{a}, \mathbf{b}) = \frac{\|\mathbf{a} - \mathbf{b}'\|}{\|\mathbf{b}'\|}, \tag{11}$$

where $\|\cdot\|$ is the l_2 norm, and $\mathbf{b}' \triangleq \arg \inf_{\tilde{\mathbf{b}} \in R_{\mathbf{b}}} \|\mathbf{a} - \tilde{\mathbf{b}}\|$.

We note that the normalized distance is the tangent of the angle between the rays determined by \mathbf{a} and \mathbf{b} . Thus, we also have

$$d_{\text{norm}}(\mathbf{a}, \mathbf{b}) = \frac{\left\| \mathbf{a} - \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{b}\|^2} \mathbf{b} \right\|}{\left\| \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{b}\|^2} \mathbf{b} \right\|}. \tag{12}$$

2.3. Neural Networks

In general, an NN is a computational model, with significant expressive capability for desired functions [56]. Following convention, we now formally define fully connected feedforward multilayer NNs as a family of functions.

Definition 4 (Neural networks [56]). *With $l \in \mathbb{N}$ hidden layers of sizes $d_1, d_2, \dots, d_l \in \mathbb{N}$, fixed input dimension d_0 and output dimension d_{l+1} , where $d_0, d_{l+1} \in \mathbb{N}$, let \circ denote the composition of functions, a fully connected feedforward multilayer NN is defined as the following family of functions*

$$\mathcal{G}_l^{(d_0, d_{l+1})} \triangleq \left\{ g : \mathbb{R}^{d_0} \rightarrow \mathbb{R}^{d_{l+1}} \mid g(\mathbf{x}_0) = f_{l+1} \circ \sigma \circ f_l \circ \dots \circ \sigma \circ f_1(\mathbf{x}_0) \right\}, \tag{13}$$

where, for $j \in \{1, 2, \dots, l+1\}$, $f_j : \mathbb{R}^{d_{j-1}} \rightarrow \mathbb{R}^{d_j}$ is a linear operation

$$f_j(\mathbf{x}_{j-1}) \triangleq \mathbf{W}_j \mathbf{x}_{j-1} + \mathbf{b}_j, \tag{14}$$

with the weight matrices $\mathbf{W}_j \in \mathbb{R}^{d_j \times d_{j-1}}$, the data vectors $\mathbf{x}_{j-1} \in \mathbb{R}^{d_{j-1}}$, and the bias vectors $\mathbf{b}_j \in \mathbb{R}^{d_j}$. The nonlinear activation function $\sigma(\cdot)$ is performed on vectors element-wise, such that, for $j \in \{1, 2, \dots, l\}$,

$$(\mathbf{x}_j)_i = [\sigma(f_j(\mathbf{x}_{j-1}))]_i, \quad i = 1, 2, \dots, d_j. \tag{15}$$

For an arbitrary number of hidden layers and sizes, fully connected feedforward multilayer NNs with the same fixed input and output dimensions are defined as

$$\mathcal{G}^{(d_0, d_{l+1})} \triangleq \bigcup_{l \in \mathbb{N}} \mathcal{G}_l^{(d_0, d_{l+1})}. \tag{16}$$

The set of all possible weights and biases of an NN is the *parameter space*, which is denoted as $\Phi \subset \mathbb{R}^d$ (here, d is defined as the dimension of the parameter space). We sometimes denote an NN function $g \in \mathcal{G}^{(d_0, d_{l+1})}$ with parameters $\phi \in \Phi$ as g_ϕ . Exploiting NNs, desired functions can be *parameterized* with ϕ , and *approximated* by optimizing ϕ with gradient descent methods [57]. Multilayer feedforward NNs are known to be universal approximators for any measurable function as long as the scale of the model is sufficiently large [56].

Activation functions of NNs may vary in forms for different tasks. More specifically, let the input be $\mathbf{y} \in \mathbb{R}^{d_y}$ with elements $y_j, j \in \{1, 2, \dots, d_y\}$; the most common activation functions include the *sigmoid* activation function $\sigma_s(\mathbf{y})_j = \frac{1}{1 + \exp(-y_j)}$ and the *ReLU* activa-

tion function $\sigma_{\mathbb{R}}(\mathbf{y})_j = \max(y_j, 0)$. There are many variations of ReLU, including the *ELU* activation function $\sigma_{\mathbb{E}}(\mathbf{y})_j = \begin{cases} y_j & , y_j \geq 0, \\ e^{y_j} - 1 & , y_j < 0. \end{cases}$ A special activation function is the *softmax* activation function or the *softmax layer*, which is defined as $\sigma_{\text{sm}}(\cdot) : \mathbb{R}^{d_y} \rightarrow \mathbb{R}^{d_y}$, where

$$\sigma_{\text{sm}}(\mathbf{y})_j = \frac{\exp(y_j)}{\sum_{j=1}^{d_y} \exp(y_j)}, \quad j = 1, 2, \dots, d_y. \tag{17}$$

We note that the output vector of the softmax layer has special properties, i.e., $\sigma_{\text{sm}}(\mathbf{y})_j \geq 0$ and $\sum_{j=1}^d \sigma_{\text{sm}}(\mathbf{y})_j = 1$, which make the output vector a valid PMF in (1).

For details on NNs, please refer to [58] (Chapter 20).

2.4. Problem Statement

In this paper, when a target vector $\mathbf{h}_t \in \Gamma_n$ is given, we are interested in finding an entropic vector $\mathbf{h}^{\mathbf{p}}$ in $\Gamma_{n,\mathcal{X}}^*$, such that $d_{\text{norm}}(\mathbf{h}^{\mathbf{p}}, \mathbf{h}_t)$ is relatively small. If \mathbf{h}_t is entropic, we aim to give the underlying PMF to verify and realize \mathbf{h}_t . If \mathbf{h}_t is not entropic, we aim to give an entropic vector close to \mathbf{h}_t and, hopefully, on the boundary of $\Gamma_{n,\mathcal{X}}^*$.

We recall that $d_{\text{norm}}(\mathbf{h}^{\mathbf{p}}, \mathbf{h}_t)$ is the tangent of the angle between the corresponding rays, but the tangent is only strictly increasing when the angle is limited in $[0, \frac{\pi}{2}]$. Restricting the target vector \mathbf{h}_t to be in Γ_n can satisfy this requirement, since Γ_n belongs to the nonnegative orthant of \mathcal{H}_n and $\Gamma_{n,\mathcal{X}}^* \subset \Gamma_n$.

3. Optimizing PMFs for Associated Entropic Vectors via Neural Networks

In this section, we propose the methodology that tackles the problem stated in Section 2.4. More specifically, given a target vector in Γ_n , we propose to train a corresponding NN to identify the entropic vector closest to the target, and, at the same time, provide the corresponding underlying PMF.

In particular, the proposed NN is configured with a final softmax layer to output PMFs, with the input fixed to a specific constant (similar settings can be found in [45,46]). We claim that, for a given target vector, there always exists an NN such that the entropic vector associated with the produced PMF is arbitrarily close to the target in terms of normalized distance. This statement is formalized in the following theorem.

Theorem 1. *For the random vector \mathbf{X} with a fixed and finite alphabet \mathcal{X} , given a target $\mathbf{h}_t \in \Gamma_n$ (entropic or not), we consider the entropic vector $\mathbf{h}^* \in \Gamma_{n,\mathcal{X}}^*$ that yields the minimal normalized distance D^* to \mathbf{h}_t . Then, for $\eta > 0$, there exists a corresponding NN g_{ϕ} with parameters $\phi \in \Phi$ (where $\Phi \subset \mathbb{R}^d$ is the parameter space) which outputs the PMF \mathbf{p}^{ϕ} valid in $\Delta_{\mathcal{X}}^m$ from a specific constant input $a \in \mathbb{R}$, such that*

$$d_{\text{norm}}(\mathbf{h}^{\mathbf{p}^{\phi}}, \mathbf{h}_t) < D^* + \eta, \tag{18}$$

where $\mathbf{h}^{\mathbf{p}^{\phi}}$ is the associated entropic vector of \mathbf{p}^{ϕ} .

The proof of Theorem 1 is straightforward based on the observation that the softmax layer can produce any desired PMF, and both the entropic vector and the normalized distance are continuous functions of the PMF. However, for the completeness of the paper, we provide the proof of Theorem 1 in Appendix B.1. The proof can be described as follows. For a given target vector $\mathbf{h}_t \in \Gamma_n$, let \mathbf{p}^* be one of the underlying PMFs that achieves \mathbf{h}^* , i.e., the entropic vector with minimal normalized distance to \mathbf{h}_t ; then, we can find an NN which outputs \mathbf{p}^{ϕ} from a fixed constant, and \mathbf{p}^{ϕ} approximates \mathbf{p}^* in l_1 norm. By the continuity of the Shannon entropy ([54], Section 2.3), for fixed finite alphabets, we can prove that, as \mathbf{p}^{ϕ} approximates \mathbf{p}^* in l_1 norm, the associated entropic vector $\mathbf{h}^{\mathbf{p}^{\phi}}$ approximates \mathbf{h}^* in normalized distance as well.

Algorithm

In the rest of the section, we give a practical algorithm, i.e., Algorithm 1, to minimize the NN loss function, which is defined as the normalized distance interpreted in (12), i.e.,

$$L(\phi, \mathbf{h}_t, \mathbf{h}^{\mathbf{p}^\phi}) \triangleq \frac{\|\mathbf{h}^{\mathbf{p}^\phi} - \frac{\mathbf{h}^{\mathbf{p}^\phi} \cdot \mathbf{h}_t}{\|\mathbf{h}_t\|^2} \mathbf{h}_t\|}{\|\frac{\mathbf{h}^{\mathbf{p}^\phi} \cdot \mathbf{h}_t}{\|\mathbf{h}_t\|^2} \mathbf{h}_t\|}. \tag{19}$$

Algorithm 1 trains parameters ϕ with a gradient descent method to find an NN that achieves \mathbf{h}^* in Theorem 1, i.e., identify the entropic vector closest to the target and provide the corresponding underlying PMF. The overall architecture of the proposed method is depicted in Figure 1. More specifically, at each iteration, the NN takes a constant as the input, and outputs a joint PMF \mathbf{p}^ϕ . By configuring the last layer of NN as the softmax layer in (17), i.e.,

$$\mathbf{p}_j^\phi = \sigma_{\text{sm}}(\tilde{\mathbf{p}}^\phi)_j = \frac{\exp(\tilde{p}_j^\phi)}{\sum_{j=1}^m \exp(\tilde{p}_j^\phi)}, \quad j = 1, 2, \dots, m, \tag{20}$$

where $\tilde{\mathbf{p}}^\phi \in \mathbb{R}^m$ is the layer input, the resulting PMF \mathbf{p}^ϕ is valid in $\Delta_{\mathcal{X}}^m$. Subsequently, the associated entropic vector $\mathbf{h}^{\mathbf{p}^\phi}$ is computed from \mathbf{p}^ϕ with (2), (3), and (5). The normalized distance between the associated entropic vector and target vector is then evaluated with (19). To optimize NN parameters ϕ , the gradient descent method is performed in turn. The training procedures are iterated for a sufficiently large number N .

Algorithm 1: Optimize PMFs for the associated entropic vector closest to the target vector $\mathbf{h}_t \in \Gamma_{n,\mathcal{X}}$ via NN training

Require: $\mathbf{h}_t, n, m_1, m_2, \dots, m_n$
Ensure: $\mathbf{p}^\phi, \mathbf{h}^{\mathbf{p}^\phi}, d_{\text{norm}}(\mathbf{h}^{\mathbf{p}^\phi}, \mathbf{h}_t)$

- 1 **Initialization**
- 2 | NN parameters ϕ , learning rate γ , a constant a , number of iterations N ;
- 3 **for** $i = 1$ **to** N **do**
- 4 | Calculate the joint PMF $\mathbf{p}^\phi = g_\phi(a)$, such that \mathbf{p}^ϕ is valid according to (1);
- 5 | Obtain all the marginal PMFs \mathbf{p}_α^ϕ ;
- 6 | Compute all (joint) entropies with \mathbf{p}^ϕ and \mathbf{p}_α^ϕ according to (2);
- 7 | Compute the entropic vector $\mathbf{h}^{\mathbf{p}^\phi}$ according to (3) and (5);
- 8 | Evaluate $L(\phi, \mathbf{h}_t, \mathbf{h}^{\mathbf{p}^\phi})$ according to (19);
- 9 | Update NN parameters with gradient descent: $\phi \leftarrow \phi - \gamma \nabla_\phi L(\phi, \mathbf{h}_t, \mathbf{h}^{\mathbf{p}^\phi})$;
- 10 **end**
- 11 $d_{\text{norm}}(\mathbf{h}^{\mathbf{p}^\phi}, \mathbf{h}_t) \leftarrow L(\phi, \mathbf{h}_t, \mathbf{h}^{\mathbf{p}^\phi})$;
- 12 **return** $\mathbf{p}^\phi, \mathbf{h}^{\mathbf{p}^\phi}, d_{\text{norm}}(\mathbf{h}^{\mathbf{p}^\phi}, \mathbf{h}_t)$;

We further implement Algorithm 1 and the NN model in Figure 1 with CNNs, and details are presented in Section 4. With the proposed implementation techniques, the convergence performances of Algorithm 1 for different targets are demonstrated in Section 5.1. Other empirical results on various problems related to the entropic region are demonstrated in Sections 5.2 and 5.3.

Before ending this section, in the following remark, we discuss the connections between our work with existing literature [44–46], which focus on the neural optimization of distributions for problems in information theory as well.

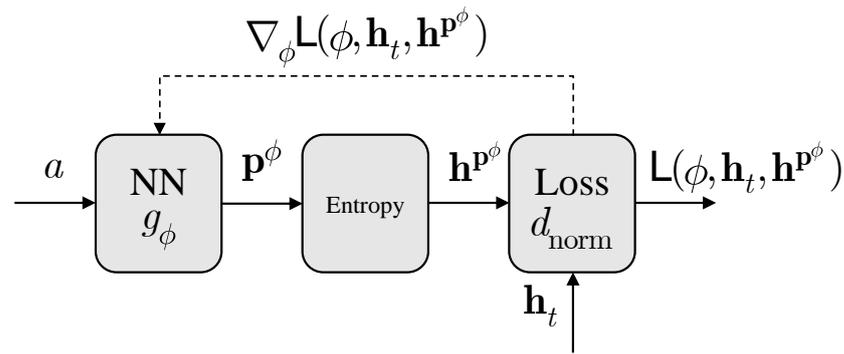


Figure 1. The overall architecture of the proposed method. The dashed arrow indicates the flow of gradients in back-propagation.

Remark: Recently, the optimization of probability distributions utilizing generative NNs in information theory has been studied and applied to plenty of problems in [44–46]. In [44,46], NNs are designed to generate the input distributions of given channels, and optimized to produce a distribution that achieves channel capacity, serving as a part of the joint estimation–optimization architecture for channel capacity. More specifically, over continuous spaces [44], a model named neural distribution transformer (NDT) is proposed; the optimized NDT transfers uniformly distributed samples into data that are distributed according to the desired capacity-achieving distribution. Over discrete alphabets [46], the PMF generator is proposed to optimize the PMF numerically, and capacity-achieving channel input data can be sampled from the optimized PMF. We note that [46] focuses on channels with feedback and memory, and the PMF generator is based on a time-series deep reinforcement learning model. For channels without feedback and memory, the PMF generator in [46] degenerates to a similar architecture as the one proposed in this paper. However, different from sampling data from generated PMFs, the method proposed in this paper directly computes entropic vectors from PMFs, and theoretical guarantees are provided in the context of this task. In [45], NNs are also used to optimize conditional distributions that achieve the rate-distortion function for any given source distribution, in both continuous and discrete spaces. The PMF generator in [45] takes source samples as input, and produces corresponding conditional PMFs. If the source is deterministic in [45], the conditional PMF generator degenerates to the proposed model in this paper as well.

4. Implementation

In this section, we aim to present implementation details of Algorithm 1 and the NN model in Figure 1 with more advanced deep learning techniques. In addition, we demonstrate that the implementation results in significantly larger feasible alphabet sizes for Algorithm 1 than existing methods, which will contribute to constructing achievable schemes for network coding problems, as discussed in Section 6.

4.1. Implementing with Convolutional Neural Networks

We recall that, for a random vector \mathbf{X} , if we assume $|\mathcal{X}_i| = k_i = k, i \in \mathcal{N}_n$ with joint PMFs $\mathbf{p} \in \Delta_{\mathcal{X}}^{k^n}$, then every PMF is a k^n -dimensional vector. In addition, we notice that Algorithm 1 and the model depicted in Figure 1 require the dimension of the output of the NN to match the dimension of the desired PMF. Hence, the size of the last layer of the NN grows polynomially with the alphabet size k of each random variable, and grows exponentially with the number of random variables n . Moreover, the indexing problem to calculate all marginal PMFs in Algorithm 1 is computational expensive as well with large dimensional PMF vectors.

Therefore, we are motivated to implement Algorithm 1 and the model in Figure 1 with CNNs, which transfer the vector data in NNs to tensors. Furthermore, by virtue of high-performance GPUs, the training of CNNs can be notably accelerated.

More specifically, if we focus on the case when $n = 4$, a PMF \mathbf{p} for \mathbf{X} with $|\mathcal{X}_i| = k, i \in \mathcal{N}_n$ is a k^4 -dimensional vector such that $\mathbf{p} \in \Delta_{\mathcal{X}}^{k^4}$. To implement with CNNs, we naturally utilize the 3d-CNNs (available in PyTorch [59]), whose data in each layer is a four-dimensional cube. By modifying the NN in Figure 1 with 3d-CNNs, the model takes a constant tensor \mathbf{A} as input, where $\mathbf{A} \in \mathbb{R}^{1 \times 1 \times 1 \times 1}$, and outputs a PMF tensor $\mathbf{P}^\phi \in [0, 1]^{k \times k \times k \times k}$.

We demonstrate an example of a configuration of the modified model in Table 1. Our empirical results in Section 4.2 show that the model implemented with a CNN is feasible even when k is large, acquiring additional gain from GPU acceleration.

Table 1. An example of a configuration when the NN in Figure 1 is implemented with a CNN.

Layer	Output Dimension	Kernel Size	Padding	Activation Function
Input	$1 \times 1 \times 1 \times 1$	-	-	-
Conv. 3d	$k \times k \times k \times k$	$k \times k \times k$	$k - 1$	ELU
Conv. 3d	$k \times k \times k \times k$	$3 \times 3 \times 3$	1	ELU
Conv. 3d	$k \times k \times k \times k$	$3 \times 3 \times 3$	1	ELU
Flatten	k^4	-	-	softmax
Output	$k \times k \times k \times k$	-	-	-

4.2. Feasibility for Large Alphabets

In the proposed method, despite the fact that the complexity of the output layer in the original NN grows as the alphabet sizes of the random variables increase, after the implementation discussed in Section 4.1, the training of CNNs can be accelerated by high-performance GPUs, resulting in significant gains in the training speed for Algorithm 1. In Figure 2, we present empirical results of the training speed of the proposed method as the alphabet sizes of random variables increase.

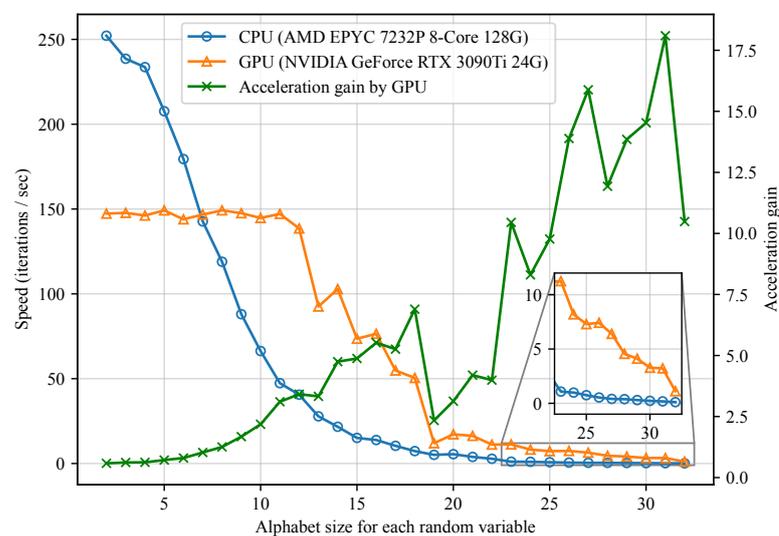


Figure 2. Empirical results on the speed of Algorithm 1 implemented with CNNs. When the alphabet size for every random variable grows from 2 to 32, the iteration speeds and acceleration gains are shown, respectively.

To the best of the authors’ knowledge, when $n = 4$, the feasible alphabet size of the proposed method (i.e., 32 for every random variable) is larger than prior works. For instance, the largest alphabet size considered in [26] is 5. Alphabet sizes of 10 and 11 are mentioned in [16,21], respectively.

5. Empirical Results

In this section, empirical results of the method proposed in Section 3 are presented, focusing on various problems related to Γ_4^* and $\bar{\Gamma}_4^*$. Firstly, we compare our method with the existing random search algorithm [26]. Secondly, we tackle the problems of optimizing the Ingleton score and Ingleton violation index. Lastly, we exploit the proposed method to obtain many entropic vectors that approximate the boundary of Γ_4^* , and use the convex hull of these entropic vectors to form an inner bound of $\bar{\Gamma}_4^*$.

Before presenting these results, we first provide a brief introduction of regions within \mathcal{H}_4 . When $n = 4$, the Ingleton inequality [41] plays an essential role in the characterization of $\bar{\Gamma}_4^*$, and one permutation instance of the Ingleton inequality is defined as follows.

Definition 5 (Ingleton inequality). *For a vector $\mathbf{h} \in \mathcal{H}_4$, one specific permutation instance of the Ingleton inequality [41] is denoted as the inner product between the coefficients \mathbf{I}_{34} and \mathbf{h}*

$$\begin{aligned} \mathbf{I}_{34}\mathbf{h} \triangleq & h_{12} + h_{13} + h_{14} + h_{23} + h_{24} \\ & - h_1 - h_2 - h_{34} - h_{123} - h_{124} \geq 0. \end{aligned} \tag{21}$$

Permuting \mathcal{N}_4 , there exist six instances of the Ingleton inequality in total, which are denoted as $\mathbf{I}\mathbf{h} \geq \mathbf{0}$, $\mathbf{I} \in \mathbb{R}^{6 \times 15}$.

When $n = 4$, the linear rank region is the intersection of \mathcal{H}_4 with $\mathbf{I}\mathbf{h} \geq \mathbf{0}$ (all six instances of Ingleton inequality), i.e.,

$$\Gamma_4^{\text{linear}} \triangleq \{\mathbf{h} \in \mathcal{H}_4 \mid \mathbf{I}\mathbf{h} \geq \mathbf{0}\}. \tag{22}$$

We consider vectors that represent rank functions of linear subspaces (i.e., linearly representable entropic vectors, or entropic vectors that yield rate regions achieved by linear network codes); then, the linear rank region is the closure of the linear hull of such vectors. The region Γ_4^{linear} is a polyhedral convex cone, and is a tight linear inner bound of $\bar{\Gamma}_4^*$ [8,13,60]. However, Γ_n^{linear} is only completely characterized for $n \leq 5$ [15,17]. Nevertheless, when $n = 4$, the relation $\Gamma_4^{\text{linear}} \subset \bar{\Gamma}_4^* \subset \Gamma_4$ holds.

Hence, one is interested in characterizing the part of $\bar{\Gamma}_4^*$ excluding Γ_4^{linear} (i.e., the almost-entropic vectors that cannot be linearly representable or yield rate regions, which cannot be achieved by linear network codes). We recall that $\mathbf{I}\mathbf{h} \geq \mathbf{0}$ represents all six instances of the Ingleton inequality. We define

$$\Lambda_4 \triangleq \{\mathbf{h} \in \Gamma_4 \mid \mathbf{I}\mathbf{h} \geq \mathbf{0} \text{ does not hold}\}, \tag{23}$$

and the entropic part of Λ_4 , i.e.,

$$\Lambda_4^* \triangleq \{\mathbf{h} \in \Gamma_4^* \mid \mathbf{I}\mathbf{h} \geq \mathbf{0} \text{ does not hold}\}. \tag{24}$$

We denote the closure of Λ_4^* as $\bar{\Lambda}_4^*$; then, it is clear that $\Lambda_4^* \subset \bar{\Lambda}_4^* \subset \Lambda_4$. By [60] (Lemma 4), any ray in Λ_4 exactly violates one instance of the Ingleton inequality, i.e., Λ_4 is the union of six symmetric regions, each region corresponds to and violates one instance of the Ingleton inequality, and the six regions are disjointed and share boundary points only. Thus, we may consider Λ_4 and Λ_4^* when only one instance of the Ingleton inequality does not hold. More specifically, when $\mathbf{I}_{34}\mathbf{h} \leq 0$, i.e., the reversed version of (21) holds, we denote Λ_4 and Λ_4^* as Λ_4^{34} and $\Lambda_4^{*,34}$, respectively. Analogously, we can consider different instances of the Ingleton inequality to obtain other five outer regions symmetric to Λ_4^{34} and five regions symmetric to Λ_4^* .

We denote the closure of $\Lambda_4^{*,34}$ as $\bar{\Lambda}_4^{*,34}$; then, in a word, the characterization of $\bar{\Gamma}_4^*$ is reduced to $\bar{\Lambda}_4^{*,34}$.

The only extreme ray of Λ_4^{34} violating $\mathbf{I}_{34}\mathbf{h} \geq 0$ (as well as the Zhang–Yeung inequality [8]) is denoted as $R_{\mathbf{v}}$, where

$$\mathbf{v} = (1, 1, 1, 1, 1.5, 1.5, 1.5, 1.5, 1.5, 2, 2, 2, 2, 2, 2)^T. \quad (25)$$

The region Λ_4^{34} is often called the *pyramid* and is represented by 15 extreme rays, with $R_{\mathbf{v}}$ as the top and the other 14 extreme rays as bases (a complete list of these extreme rays is available in [16] (Section XI)).

In the rest of this section, we focus on the bounded regions as Definition 2 illustrates with $m = k^4$, i.e., each random variable has an equal alphabet size k . The above discussions still hold, as we denote the bounded versions as $\Lambda_{4,\mathcal{X}}^{34}$, $\Lambda_{4,\mathcal{X}}^{*,34}$ and $\bar{\Lambda}_{4,\mathcal{X}}^{*,34}$.

In the following remark, we present the implementation details of Algorithm 1 for experiments in this section.

Remark 1. *In this section, we utilized Algorithm 1 with hyperparameters $n = 4, m_1 = m_2 = m_3 = m_4 = 4$. Overall, the NN model followed from the configuration presented in Table 1 with $k = 4$ and either one or three hidden layers. The NN parameters set ϕ was randomly initialized with the default method of PyTorch [59] (the initial parameters can be fixed by fixing the random seed), the optimizer was selected with Adam optimizer [61] with a learning rate γ tuned in the range of 10^{-2} to 10^{-5} (based on the number of hidden layers), and we selected the constant $a = 1$. The number of iterations N were tuned to be 5×10^4 (however, the algorithm can converge with significantly smaller iterations in most tasks of this section).*

5.1. Comparison with the Random Search Algorithm in [26]

When evaluating the performance of our method, the target was set as \mathbf{v} in (25), which is trivially *not entropic* due to the Zhang–Yeung inequality [8], i.e., $D^* > 0$ in Theorem 1. We aimed to compare the returned normalized distance by Algorithm 1 with that obtained by the random search algorithm in [26], which is the only existing method for verifying entropic vectors with general form of PMFs. Intuitively, if Algorithm 1 is capable of obtaining entropic vectors closer to \mathbf{v} , we are able to approximate finer boundaries of $\Gamma_{4,\mathcal{X}}^*$.

Figure 3 depicts the empirical results of our method. In addition to \mathbf{v} (Target 1), two *entropic* vectors in [13] (Theorem 4, Target 2) and [16] (Conjecture 1, Target 3), which belong to the region $\Lambda_{4,\mathcal{X}}^{*,34}$, are set as targets, and the results are compared to [26] (the numerical data of the results in [26] are available in [62]). The alphabet size of every random variable, i.e., k , is set to 4. In Figure 3, the curves illustrate the convergence performances as the iterations of Algorithm 1 grow, while the shaded regions represent the range of final results obtained by several experiments in [26].

From Figure 3, it is evident that our method outperforms the random search algorithm proposed by [26]. More specifically, for the non-entropic target (Target 1 in Figure 3), the returned normalized distance using our method (0.01895) is smaller than the smallest value obtained by [26] (0.02482). For entropic targets (Targets 2 and 3 in Figure 3), our method returns negligible normalized distance as expected, thus successfully verifying these entropic targets. In addition, the normalized distances obtained by [26] vary across different experiments, and even increase with larger iterations. For example, with Target 1 in Figure 3, the method in [26] returns a normalized distance of 0.02482 with 1226 iterations; however, with 10,091 iterations, a larger normalized distance 0.04283 is obtained [26] (Figure 1).

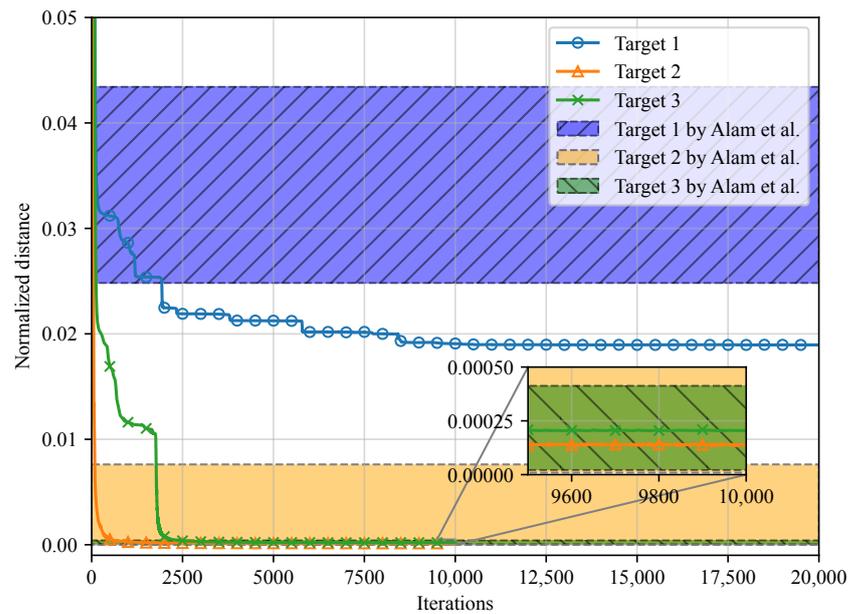


Figure 3. Minimization of normalized distances for different target vectors, compared with the random search algorithm by Alam et al. in [26].

In contrast, the results obtained by our method are consistent within multiple experiments, empirically presenting better convergence performances than [26].

5.2. Optimizing Ingleton Score and Ingleton Violation Index

When optimizing entropic vectors for $n = 4$, it is of great interest to optimize the Ingleton score and Ingleton violation index. The Ingleton score was first proposed in [42] (Conjecture 4.1) and later rigorously defined in [16] (Definition 3).

Definition 6 (Ingleton score [16] (Definition 3)). Given a random vector $\mathbf{X} = (X_1, X_2, X_3, X_4)$, and an entropic vector $\mathbf{h} \in \Gamma_4^*$, the Ingleton score induced by (21) is defined as

$$\mathbb{I}_{34}(\mathbf{h}) \triangleq \frac{\mathbf{I}_{34}\mathbf{h}}{h_{1234}}. \tag{26}$$

The Ingleton score is defined as $\mathbb{I} \triangleq \mathbb{I}_{34}$ due to permutation symmetry, and the optimization problem of \mathbb{I} is

$$\begin{aligned} &\inf_{\mathbf{h}} \mathbb{I}(\mathbf{h}), \\ &\text{s.t. } \mathbf{h} \in \Gamma_4^*. \end{aligned} \tag{27}$$

Similarly, the Ingleton violation index [19] is defined as follows.

Definition 7 (Ingleton violation index [19]). Given a random vector $\mathbf{X} = (X_1, X_2, X_3, X_4)$, and an entropic vector $\mathbf{h} \in \Gamma_4^*$, the Ingleton violation index induced by (21) is defined as

$$\iota_{34}(\mathbf{h}) \triangleq -\frac{\mathbf{I}_{34}\mathbf{h}}{\|\mathbf{h}\|}. \tag{28}$$

Due to permutation symmetry, the Ingleton violation index is defined as $\iota \triangleq \iota_{34}$, and the optimization problem of ι is

$$\begin{aligned} &\sup_{\mathbf{h}} \iota(\mathbf{h}), \\ &\text{s.t. } \mathbf{h} \in \Gamma_4^*. \end{aligned} \tag{29}$$

One is interested in minimizing \mathbb{I} or maximizing ι over Γ_4^* , because their optimal values reveal the largest violations from (21) among entropic vectors in $\Lambda_4^{*,34}$, and, thus, are crucial characterizations of Γ_4^* and $\bar{\Gamma}_4^*$.

There is a rich history of optimizing \mathbb{I} [16,21,26,35,42] and ι [19,26]. More specifically, for $\mathbf{h} \in \bar{\Gamma}_4^*$, [21] proposes a decomposition and two linear transformations on \mathbf{h} . If we denote the composition of these techniques on \mathbf{h} as $\mathbf{g} \triangleq T(\mathbf{h})$, then it is shown that $\mathbf{g} \in \bar{\Gamma}_4^*$, i.e., \mathbf{g} is still almost entropic for any almost-entropic \mathbf{h} . In addition, the Ingleton score optimized with \mathbf{g} in [21] is smaller than the one optimized with \mathbf{h} , and refutes the four-atom conjecture in [16] (please refer to [21] (Section VIII) for details). Since $T(\cdot)$ plays an important role in optimizing the Ingleton score and Ingleton violation index in this paper, for both completeness and clarity, we state the details of the techniques that compose $T(\cdot)$ in Appendix A for interested readers.

Thus, we are motivated to exploit our method to obtain a better Ingleton score and Ingleton violation index, when the objective L in Algorithm 1 is replaced by $\mathbb{I}(\mathbf{h})$, $\iota(\mathbf{h})$, and $\mathbb{I}(\mathbf{g})$, $\iota(\mathbf{g})$. Although Theorem 1 does not directly imply that the proposed method can be exploited to optimize the Ingleton score and Ingleton violation index, we claim that there always exists an NN that optimizes the Ingleton score and Ingleton violation index to the optimal values to any desired of accuracy as indicated by Theorems 2 and 3. The proofs of Theorems 2 and 3 are provided in Appendices B.2 and B.3, respectively.

Theorem 2. For the random vector $\mathbf{X} = (X_1, X_2, X_3, X_4)$ with a fixed and finite alphabet \mathcal{X} , we recall the Ingleton score for entropic vectors defined in Definition 6. We consider the entropic vector $\mathbf{h}^* \in \Gamma_{4,\mathcal{X}}^*$, such that \mathbf{h}^* yields the optimal Ingleton score \mathbb{I}^* . Then, there exists an NN with parameters $\phi \in \Phi$, such that $\mathbf{h}^{\mathbf{p}^\phi}$ is the associated entropic vector of $\mathbf{p}^\phi = g_\phi(a)$ valid in $\Delta_{\mathcal{X}}^m$, and $\mathbf{h}^{\mathbf{p}^\phi}$ yields the corresponding Ingleton score \mathbb{I}^ϕ , which is consistent with \mathbb{I}^* within any desired degree of accuracy $\kappa > 0$, i.e.,

$$|\mathbb{I}^\phi - \mathbb{I}^*| < \kappa, \tag{30}$$

where $|\cdot|$ is the absolute value for scalars.

Theorem 3. For the random vector $\mathbf{X} = (X_1, X_2, X_3, X_4)$ with a fixed and finite alphabet \mathcal{X} , we recall the Ingleton violation index for entropic vectors defined in Definition 7. We consider the entropic vector $\mathbf{h}^* \in \Gamma_{4,\mathcal{X}}^*$, such that \mathbf{h}^* yields the optimal Ingleton violation index ι^* . Then, there exists an NN with parameters $\phi \in \Phi$, such that $\mathbf{h}^{\mathbf{p}^\phi}$ is the associated entropic vector of $\mathbf{p}^\phi = g_\phi(a)$ valid in $\Delta_{\mathcal{X}}^m$, and $\mathbf{h}^{\mathbf{p}^\phi}$ yields the corresponding Ingleton violation index ι^ϕ , which is consistent with ι^* within any desired degree of accuracy $\mu > 0$, i.e.,

$$|\iota^\phi - \iota^*| < \mu, \tag{31}$$

where $|\cdot|$ is the absolute value for scalars.

The optimized numerical results are compared with the state-of-the-art values in Table 2, and the alphabet size of every random variable is set to 4. In Table 2, our results are rounded to ten decimal digits, and the last four digits are significant since the perturbations for small probability values may lead to large absolute errors of entropy functions and score results.

Table 2. Optimized Ingleton scores and Ingleton violation indices.

Objective	Best-Known Results	Our Results
$\inf_{\mathbf{h}} \mathbb{I}(\mathbf{h})$	−0.089373 [16]	−0.0893733002
$\inf_{\mathbf{h}} \mathbb{I}(\mathbf{g})$	−0.0925000777 [35]	−0.0925001031
$\sup_{\mathbf{h}} \iota(\mathbf{h})$	0.028131604 [26]	0.0281316527
$\sup_{\mathbf{h}} \iota(\mathbf{g})$	N/A	0.0288304141

As seen in Table 2, firstly, for the Ingleton score \mathbb{I} , the upper bound -0.0925001031 , which is optimized with \mathbf{g} , is returned (with negligible improvement compared to -0.0925000777 due to numerical stability). After the first discovery of this value [34,35], the method in [26] obtained a close value of -0.092499 as well, and, thus, we reconfirm this upper bound for the third time. Secondly, for the Ingleton violation index ι , we particularly optimize it with \mathbf{g} as well, and a new lower bound 0.0288304 is obtained, which beats the current best value of 0.0281316 . We observe that ι actually measures the sine of the angle between the hyperplane $\mathbf{I}_{34}\mathbf{h} = 0$ and \mathbf{h} ; thus, the obtained new lower bound finds an entropic vector with a larger violation angle from the Ingleton hyperplane, bringing us closer to the optimal ι and the complete characterization of Γ_4^* and $\bar{\Gamma}_4^*$.

The corresponding convergence performances for results in Table 2 are presented in Figure 4. From Figure 4, one can verify the obtained results listed in Table 2, and observe that the proposed method successfully converges to the best known Ingleton score when optimizing $\mathbb{I}(\mathbf{g})$, and converges to the Ingleton violation index, which exceeds the best-known results when optimizing $\iota(\mathbf{g})$.

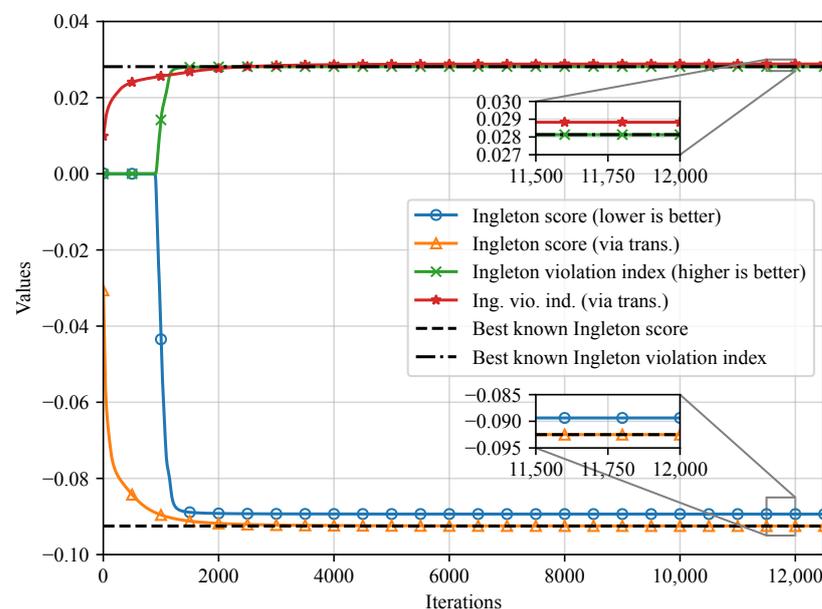


Figure 4. The convergence results on the optimization of the Ingleton score and the Ingleton violation index, where “via trans.” refers to what we optimized with a series of transformations $\mathbf{g} = T(\mathbf{h})$ proposed by [21].

Remark 2. The pursuit of the Ingleton score -0.0925 appears to be difficult with no alternative strategies in the literature. In [26], the search area is constrained using a strategy called hyperplane score reduction with the minimization of normalized distance with relatively large iterations. Additionally, [35] reports that most experiments yield the value -0.09103635 , while returning -0.0925000777 very occasionally, even with millions of searches with massive computation resources in parallel. Similarly, with our method, the direct minimization of $\mathbb{I}(\mathbf{g})$ yields the value -0.09103635 . However, by selecting the entropic point from [13] as the initial point of Algorithm 1 (by setting it as the target and minimizing the normalized distance first), we reproduced the value -0.0925 in a single experiment, with less than 10,000 iterations (which approximately take 6 s).

5.3. Inner Bounding the Almost-Entropic Region

The complete characterization of $\bar{\Gamma}_4^*$ is difficult and open. However, as we find many entropic vectors, a linear inner bound is immediately obtained. For example, for an entropic vector $\mathbf{h} \in \Lambda_{4,\mathcal{X}}^{*,34}$, if we regard \mathbf{h} as the top vertex, then the convex hull of the top \mathbf{h} and other 14 base vertices (refer to [16] (Section XI) for a complete list of these extreme rays) yields an inner bound for $\bar{\Lambda}_{4,\mathcal{X}}^{*,34}$ and $\bar{\Gamma}_4^*$.

In this paper, following [16] (Section XI), the quality of the inner bound is evaluated by the volume ratio. Let $\tilde{\Lambda}_{4,\mathcal{X}}^{34}$ denote the numerical inner bound obtained, and $\text{vol}(\cdot)$ denote the polytope volume computation operation; the volume ratio is defined as

$$R \triangleq \frac{\text{vol}(\tilde{\Lambda}_{4,\mathcal{X}}^{34})}{\text{vol}(\Lambda_{4,\mathcal{X}}^{34})} \times 100\%. \tag{32}$$

Intuitively, \mathbf{h} needs to be closer to the boundary of $\Lambda_{4,\mathcal{X}}^{*,34}$ to obtain a larger volume ratio R .

Because Algorithm 1 easily produces entropic vectors with large \mathbb{I} and better ι , we are inspired to exploit it to develop a new algorithm, i.e., Algorithm 2, to construct an inner bound with a larger volume ratio, by optimizing entropic vectors that violate other specific hyperplanes.

Algorithm 2: Construct inner bounds for the almost-entropic region with Algorithm 1

Require: Base entropic vectors \mathcal{B} , top entropic vectors \mathcal{T}
Ensure : V-representation of the inner bound $\tilde{\Lambda}_{4,\mathcal{X}}^{34}$, volume ratio R

- 1 **Initialization**
- 2 Number of iterations N , V-representation set $\mathcal{V} \leftarrow \emptyset$, H-representation set $\mathcal{H} \leftarrow \emptyset$;
- 3 $\mathcal{V} \leftarrow \mathcal{B} \cup \mathcal{T}$;
- 4 **for** $i = 1$ **to** N **do**
- 5 $\mathcal{H} \leftarrow$ the H-representation of \mathcal{V} obtained using the implementation of the double description method [51];
- 6 $\mathcal{V}^\phi \leftarrow \emptyset$;
- 7 **foreach** $\mathbf{c} \in \mathcal{H}$ **do**
- 8 Minimize the objective $\frac{\mathbf{c}\mathbf{h}}{h_{1234}}$ using Algorithm 1, obtain corresponding $\mathbf{h}^{\mathbf{p}^\phi}$;
- 9 **if** $\frac{\mathbf{c}\mathbf{h}^{\mathbf{p}^\phi}}{h_{1234}^{\mathbf{p}^\phi}} < 0$ **and** $\frac{\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi}}{h_{1234}^{\mathbf{p}^\phi}} < 0$ **then**
- 10 $\mathcal{V}^\phi \leftarrow \mathcal{V}^\phi \cup \{\mathbf{h}^{\mathbf{p}^\phi}\}$;
- 11 **end**
- 12 **end**
- 13 $\mathcal{V} \leftarrow \mathcal{V} \cup \mathcal{V}^\phi$;
- 14 $\mathcal{V} \leftarrow$ the V-representation of the convex hull of \mathcal{V} ;
- 15 **end**
- 16 $\text{vol}(\tilde{\Lambda}_{4,\mathcal{X}}^{34}) \leftarrow$ the polytope volume of \mathcal{V} ;
- 17 $R \leftarrow \frac{\text{vol}(\tilde{\Lambda}_{4,\mathcal{X}}^{34})}{\text{vol}(\Lambda_{4,\mathcal{X}}^{34})} \times 100\%$;
- 18 **return** \mathcal{V}, R ;

Algorithm 2 can be interpreted as follows. First, starting from the 14 base entropic vectors of $\Lambda_{4,\mathcal{X}}^{34}$, i.e., \mathcal{B} , we construct an initial inner bound with (an) optimized entropic vector(s), i.e., \mathcal{T} , the initial inner bound can be represented by the set of vertices $\mathcal{V} = \mathcal{B} \cup \mathcal{T}$. By transferring the V-representation of the initial inner bound with H-representation, we obtain many hyperplanes with the set of coefficients \mathcal{H} . Then, for each of these hyperplanes, we minimize their violation score similar to the Ingletton score using Algorithm 2, i.e., we find entropic vectors \mathcal{V}^ϕ with large violations from the initial inner bound (violating $\mathbf{I}_{34}\mathbf{h} \geq 0$ as well). Now, we have extended the vertices of the initial inner bound from \mathcal{V} to $\mathcal{V} \cup \mathcal{V}^\phi$, and expanded the initial inner bound to an inner bound with a larger volume ratio. This procedure can be repeated for inner bounds with larger volume ratios.

Remark 3. Since $\|\mathbf{c}\|_2 < \infty$, by Theorem 2 and its proof in Appendix B.2, the theoretical guarantee for Algorithm 2, i.e., minimizing the hyperplane score $\frac{\mathbf{ch}}{h_{1234}}$ exploiting Algorithm 1, immediately follows.

We present the volume ratio of the inner bound obtained with Algorithm 2 in Table 3, comparing it with all existing results. For reference, the volume ratio of the outer bound obtained with the non-Shannon-type inequalities in [16] (Section VII) (not tight) was 96.4682%, and the volume of the trivial inner bound obtained with $\mathbf{I}_{34}\mathbf{h} \geq 0$ was 0. From Table 3, one can see that the volume ratio of the inner bound obtained by our method is larger than existing results. Comparing with the state-of-the-art result [26], the ratio 62.4774% is improved to 72.0437%, leading to the current best inner bound of $\bar{\Gamma}_4^*$ measured by volume ratio. Thus, our proposed method takes another step forward in the complete characterization of $\bar{\Gamma}_4^*$.

Table 3. Volume ratios of inner bounds of $\bar{\Lambda}_{4,\mathcal{X}}^{*,34}$.

Methods	Volume Ratio R (%)
Newton’s method [16]	53.4815
Non-isomorphic supports [22]	57.8
Grid approach [26]	62.4774
Ours ¹	66.1340 72.0437

¹ For details of the results, refer to Remarks 4 and 5.

Remark 4. Here, we give details on how we obtained the two results listed in Table 3. For the 66.1340% result, the set \mathcal{T} consisted of three entropic vectors that yielded the Ingleton score -0.0925001031 , the Ingleton violation indices 0.028131653 and 0.0288304141 in Table 2, respectively, and the entropic vector that yielded the Ingleton score -0.09103635 , as discussed in Remark 2. Furthermore, after $N = 1$ iteration of Algorithm 2, we obtained 152 vertices. After the convex hull operation, there were 121 entropic vectors, which gave us an inner bound with a volume ratio of 66.1340%. For the 72.0437% result, the set \mathcal{T} consisted of one entropic vector that yielded the Ingleton score -0.0925001031 in Table 2. Furthermore, after $N = 2$ iterations of Algorithm 2, we obtained 2585 vertices, and this inner bound yielded an estimated (as illustrated in Remark 5) volume ratio of 72.0437%.

Remark 5. There are several algorithms available for polytope volume computation, as listed in [63]; we chose “Qhull” [64,65] and “lrs” [66,67], which only require the V-representation of a polytope, while providing the convex hull operation as well. However, all existing numerical methods become computationally intractable with high dimension and a large number of vertices (details of their computation complexities can be found in [65,67]). In our case, “Qhull” was able to compute the result 66.1340% from 152 entropic vectors in \mathbb{R}^{15} , but became intractable when computing another result where the number of entropic vectors was 2585. Nevertheless, “lrs” provides a function called “Estimate”, which allowed us to estimate the computational time and volume result, without diving into the complete computation process. In this manner, it is estimated that the volume ratio computed from 2585 entropic vectors is 72.0437%. The estimation process took several days, and it is suggested that the complete computation will take more than one year.

6. Conclusions and Discussion

This paper introduced a novel architecture to parameterize and optimize PMFs to verify entropic vectors with NNs. Given a target vector, an algorithm is proposed to identify the entropic vector closest to it via NN training. Empirical results demonstrate smaller normalized distances, and improved convergence performances. Optimized Ingleton scores are presented, and a new lower bound on the Ingleton violation index was obtained. A state-of-the-art inner bound of $\bar{\Gamma}_4^*$ was constructed, which was measured by the volume ratio. However, there exist computation burdens in the proposed algorithms. Although

Algorithm 1 achieves a larger feasible alphabet size than previous works, its efficiency is still constrained by the alphabet size, as Figure 2 shows. Algorithm 2, which requires auxiliary algorithms to calculate the polytope volume, is limited by the auxiliary algorithms due to high dimensionality and the large number of entropic vectors, as Remark 5 discusses.

Future work includes developing a computer-aided approach to construct achievable schemes for network coding problems using the proposed method. The Linear Programming bound [68], which yields Shannon outer bounds for rate regions of network coding problems, is not tight in general due to $\Gamma_n \neq \bar{\Gamma}_n^*$. However, with Algorithm 1, the corner points obtained by the Linear Programming bound can be verified to be entropic or not. Furthermore, if a corner point is verified to be entropic, the returned PMF indicates the coding scheme. Similar arguments have been raised in [26] as well, where the network instances are simple, and the largest alphabet size is small. Nevertheless, due to smaller and more consistent normalized distances depicted in Figure 3 and the feasibility for large alphabets discussed in Section 4.2, we believe that our proposed method can be applied to network coding problems with larger alphabet sizes, and/or more random variables.

Author Contributions: Conceptualization, S.Z., N.L. and W.K.; methodology, S.Z., N.L., W.K. and H.P.; formal analysis, S.Z., N.L., W.K. and H.P.; writing—original draft preparation, S.Z.; writing—review and editing, S.Z., N.L., W.K. and H.P.; supervision, N.L. and W.K.; Funding acquisition, N.L., W.K. and H.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by the National Natural Science Foundation of China under Grants 62361146853, 62071115, and 62371129, and the Research Fund of the National Mobile Communications Research Laboratory, Southeast University (No. 2024A05).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to thank John MacLaren Walsh and Satyajit Thakor for their suggestions on polytope volume computation.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyzes, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. The Decomposition and Linear Transformations Proposed in [21]

In [21], a decomposition and two linear transformations on a given entropic vector are proposed to obtain another entropic vector that yields a better Ingleton score than the four-atom conjecture in [16]. Here, we give an introduction to these techniques.

More specifically, for any vector $\mathbf{h} \in \Gamma_n$, it is decomposed into two parts, i.e., the modular part \mathbf{h}^{mod} and the tight part \mathbf{h}^{ti} , where the element of \mathbf{h}^{mod} is

$$h_\alpha^{\text{mod}} \triangleq \sum_{i \in \alpha} (h_{\mathcal{N}_n} - h_{\mathcal{N}_n \setminus \{i\}}), \quad \forall \alpha \subseteq \mathcal{N}_n \setminus \emptyset, \tag{A1}$$

and the element of \mathbf{h}^{ti} is

$$h_\alpha^{\text{ti}} \triangleq h_\alpha - \sum_{i \in \alpha} (h_{\mathcal{N}_n} - h_{\mathcal{N}_n \setminus \{i\}}), \quad \forall \alpha \subseteq \mathcal{N}_n \setminus \emptyset, \tag{A2}$$

such that $\mathbf{h} = \mathbf{h}^{\text{mod}} + \mathbf{h}^{\text{ti}}$. It is shown that, for any $\mathbf{h} \in \bar{\Gamma}_n^*$, there is $\mathbf{h}^{\text{ti}} \in \bar{\Gamma}_n^*$ [21] (Section III).

Moreover, two linear transformations on vectors in \mathcal{H}_4 are proposed. We first introduce notations to define these mappings. For any $\mathbf{h} \in \Gamma_n$, it satisfies the submodularity of polymatroidal axioms, i.e., for any $\alpha, \beta \subseteq \mathcal{N}_n \setminus \emptyset$, there is $h_\alpha + h_\beta - h_{\alpha \cup \beta} - h_{\alpha \cap \beta} \geq 0$. We further denote the submodularity as the inner product between the coefficient vector $\Delta_{\alpha, \beta}$ and the vector \mathbf{h} , i.e., $\Delta_{\alpha, \beta} \mathbf{h} \geq 0$. For $\beta \subseteq \mathcal{N}_n$ and $0 \leq s \leq |\mathcal{N}_n \setminus \beta|, s \in \mathbb{N}$, we define the

vector \mathbf{r}_s^β with the element $(r_s^\beta)_\alpha = \min\{s, |\alpha \setminus \beta|\}, \forall \alpha \subseteq \mathcal{N}_n \setminus \emptyset$. Now, for any $\mathbf{h} \in \mathcal{H}_4$, the two linear transformations proposed in [21] (Section VI) can be defined as

$$A_{3,4}\mathbf{h} \triangleq \mathbf{h} + \left(\mathbf{r}_1^{\{3\}} - \mathbf{r}_1^\emptyset\right) \left(\Delta_{3,4}\mathbf{h}\right), \tag{A3}$$

and

$$B_{34,1}\mathbf{h} \triangleq \mathbf{h} + \left(\mathbf{r}_2^{\{1\}} - \mathbf{r}_3^\emptyset\right) \left(\Delta_{134,234}\mathbf{h}\right). \tag{A4}$$

Given a vector $\mathbf{h} \in \bar{\Gamma}_4^*$, we define $T(\mathbf{h}) \triangleq A_{3,4}B_{34,1}\mathbf{h}^{\text{ti}}$, where \mathbf{h}^{ti} is the tight part of \mathbf{h} as (A2) defines, and $A_{3,4}B_{34,1}$ is the composition of the two linear transformations as (A3) and (A4) define. Then, if we denote $\mathbf{g} \triangleq T(\mathbf{h})$, it is shown that $\mathbf{g} \in \bar{\Gamma}_4^*$, i.e., \mathbf{g} is still almost entropic for any almost-entropic \mathbf{h} [21] (Section VI).

Please refer to [21] for further details of these techniques.

Appendix B. Proofs of Theorem 1–3

Appendix B.1. Proof of Theorem 1

We first define an NN-based mapping which approximates the joint PMF of the random vector \mathbf{X} (similar definitions can be found in [45,46]). We define the continuous mapping $g_\phi : \mathcal{U} \rightarrow \Delta_{\mathcal{X}}^m$. The set $\mathcal{U} \subset \mathbb{R}$ is defined with $|\mathcal{U}| \triangleq 1$ such that $a \in \mathcal{U}, a < \infty$, and $\Delta_{\mathcal{X}}^m$ is the probability simplex of \mathbf{X} . The output of the mapping is denoted as $\mathbf{p}^\phi \triangleq g_\phi(a)$, where $\mathbf{p}^\phi \in \Delta_{\mathcal{X}}^m$.

In the defined mapping g_ϕ , the set \mathcal{U} is compact and g_ϕ is continuous. Thus, we can utilize the universal approximation theorem [56] to guarantee that there exists a parameter set ϕ such that the output of NN $\mathbf{p}^\phi = g_\phi(a)$ is closed to any desired PMF $\mathbf{p}^* = g^*(a)$. We note that the universal approximation theorem holds regardless of whether l_1 or l_2 norm is used. Thus, we have the following statement. For the random vector \mathbf{X} with finite alphabet \mathcal{X} , given a desired joint PMF $\mathbf{p}^* \in \Delta_{\mathcal{X}}^m$ and $\epsilon > 0$, there exists an NN with parameters $\phi \in \Phi$, where $\Phi \subset \mathbb{R}^d$ is the parameter space, and the parameterized mapping $g_\phi \in \mathcal{G}^{(1,m)}$ generates the PMF $\mathbf{p}^\phi = g_\phi(a)$, such that

$$\|\mathbf{p}^\phi - \mathbf{p}^*\|_1 < \epsilon, \tag{A5}$$

where $\|\cdot\|_1$ is the l_1 norm.

Then, we assume that \mathbf{h}^* is associated with PMF \mathbf{p}^* ; we need to prove that, when $\mathbf{p}^\phi \rightarrow \mathbf{p}^*$ in l_1 norm, we have $\mathbf{h}^{\mathbf{p}^\phi} \rightarrow \mathbf{h}^*$ in l_2 norm.

We consider \mathbf{X} with an index set \mathcal{N}_n and a fixed finite alphabet \mathcal{X} ; from [54] (Section 2.3), we have $h_{\mathcal{N}_n}^{\mathbf{p}^\phi} \rightarrow h_{\mathcal{N}_n}^*$ in absolute value, as $\mathbf{p}^\phi \rightarrow \mathbf{p}^*$ in variational distance [54] (Definition 2.23) (i.e., $\|\mathbf{p}^\phi - \mathbf{p}^*\|_1 \rightarrow 0$). More specifically, for any $\eta' > 0$, there exists $\epsilon' > 0$, such that, for all

$$\|\mathbf{p}^\phi - \mathbf{p}^*\|_1 < \epsilon', \tag{A6}$$

where $\|\cdot\|_1$ is the l_1 norm, there is

$$|h_{\mathcal{N}_n}^{\mathbf{p}^\phi} - h_{\mathcal{N}_n}^*| < \eta', \tag{A7}$$

where $|\cdot|$ is the absolute value for scalars.

We note that, for any $\alpha \subseteq \mathcal{N}_n \setminus \emptyset$, there is $\|\mathbf{p}_\alpha^\phi - \mathbf{p}_\alpha^*\|_1 \leq \|\mathbf{p}^\phi - \mathbf{p}^*\|_1 < \epsilon'$ by [26] (Corollary 1), and, thus, we have $|h_\alpha^{\mathbf{p}^\phi} - h_\alpha^*| \leq |h_{\mathcal{N}_n}^{\mathbf{p}^\phi} - h_{\mathcal{N}_n}^*| < \eta'$. Consequently, we have

$$\sum_{\alpha \subseteq \mathcal{N}_n \setminus \emptyset} |h_\alpha^{\mathbf{p}^\phi} - h_\alpha^*| = \|\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{h}^*\|_1 < (2^n - 1) \cdot \eta', \tag{A8}$$

and, by the relationship between l_1 and l_2 norm, we have

$$\|\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{h}^*\| \leq \|\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{h}^*\|_1 < (2^n - 1) \cdot \eta'. \tag{A9}$$

We note that \mathbf{h}^* yields the minimal normalized distance to \mathbf{h}_t , i.e., there exist $\mathbf{h}'_t \in R_{\mathbf{h}_t}$ such that $\|\mathbf{h}^* - \mathbf{h}'_t\| = \|\mathbf{h}'_t\| \cdot D^*$ by Definition 3. Then, by the triangular inequality, we have

$$\begin{aligned} \|\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{h}'_t\| &\leq \|\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{h}^*\| + \|\mathbf{h}^* - \mathbf{h}'_t\| \\ &< (2^n - 1) \cdot \eta' + \|\mathbf{h}'_t\| \cdot D^*. \end{aligned} \tag{A10}$$

Thus, for any desired degree of accuracy $\eta > 0$, if we choose $\eta' = \frac{\|\mathbf{h}'_t\|}{2^n - 1} \eta$ and ϵ to be equal to the corresponding ϵ' , then, there exists an NN, such that

$$\|\mathbf{p}^\phi - \mathbf{p}^*\|_1 < \epsilon, \tag{A11}$$

and

$$\frac{\|\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{h}'_t\|}{\|\mathbf{h}'_t\|} = d_{norm}(\mathbf{h}^{\mathbf{p}^\phi}, \mathbf{h}_t) < D^* + \eta. \tag{A12}$$

Hence, the proof of Theorem 1 is complete. \square

Appendix B.2. Proof of Theorem 2

We recall the proof of Theorem 1; we assume that \mathbf{h}^* is associated with PMF \mathbf{p}^* ; the main idea is to bound (30) as $\|\mathbf{p}^\phi - \mathbf{p}^*\|_1 < \epsilon$. The proof can be divided into two steps.

In the first step, we need to bound the term $|\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{I}_{34}\mathbf{h}^*|$. More specifically, following (A9) of Theorem 1, we have proved that, for any $\eta' > 0$, there exists $\epsilon' > 0$, such that, for all $\|\mathbf{p}^\phi - \mathbf{p}^*\|_1 < \epsilon'$, there is $\|\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{h}^*\| < (2^n - 1) \cdot \eta'$. Similarly, if we choose $\eta' = \frac{\eta}{2^3 - 1}$ and ϵ to be equal to the corresponding ϵ' , then, there exists an NN such that $\|\mathbf{p}^\phi - \mathbf{p}^*\|_1 < \epsilon$ and $\|\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{h}^*\| < \eta$. Then, by the Cauchy-Schwarz inequality, we have

$$|\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{I}_{34}\mathbf{h}^*| = |\mathbf{I}_{34}(\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{h}^*)| \leq \|\mathbf{I}_{34}\| \cdot \|\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{h}^*\| < \|\mathbf{I}_{34}\| \cdot \eta. \tag{A13}$$

In the second step, to complete the rest of proof of (30), we need to bound the term

$\left| \frac{\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi}}{h_{1234}^{\mathbf{p}^\phi}} - \frac{\mathbf{I}_{34}\mathbf{h}^*}{h_{1234}^*} \right|$. More specifically, by the triangular inequality, we derive that

$$\left| \frac{\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi}}{h_{1234}^{\mathbf{p}^\phi}} - \frac{\mathbf{I}_{34}\mathbf{h}^*}{h_{1234}^*} \right| = \left| \frac{\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi} \cdot (h_{1234}^* - h_{1234}^{\mathbf{p}^\phi}) + h_{1234}^{\mathbf{p}^\phi} \cdot (\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{I}_{34}\mathbf{h}^*)}{h_{1234}^{\mathbf{p}^\phi} \cdot h_{1234}^*} \right| \tag{A14}$$

$$\leq \left| \frac{\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi} \cdot (h_{1234}^* - h_{1234}^{\mathbf{p}^\phi})}{h_{1234}^{\mathbf{p}^\phi} \cdot h_{1234}^*} \right| + \left| \frac{h_{1234}^{\mathbf{p}^\phi} \cdot (\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{I}_{34}\mathbf{h}^*)}{h_{1234}^{\mathbf{p}^\phi} \cdot h_{1234}^*} \right|. \tag{A15}$$

We note that, from (A15), following the proof of Theorem 1, in the first term, we already have $|h_{1234}^* - h_{1234}^{\mathbf{p}^\phi}| < \eta'$ as $\|\mathbf{p}^\phi - \mathbf{p}^*\|_1 < \epsilon'$. In the second term, there is $|\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{I}_{34}\mathbf{h}^*| < \|\mathbf{I}_{34}\| \cdot \eta$ by (A13). Thus, $\left| \frac{\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi}}{h_{1234}^{\mathbf{p}^\phi}} - \frac{\mathbf{I}_{34}\mathbf{h}^*}{h_{1234}^*} \right|$ can be further bounded as

$$\left| \frac{\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi}}{h_{1234}^{\mathbf{p}^\phi}} - \frac{\mathbf{I}_{34}\mathbf{h}^*}{h_{1234}^*} \right| < \left| \frac{\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi}}{h_{1234}^{\mathbf{p}^\phi} \cdot h_{1234}^*} \right| \cdot \eta' + \frac{\|\mathbf{I}_{34}\|}{|h_{1234}^*|} \cdot \eta \tag{A16}$$

$$= \frac{\eta}{|h_{1234}^*|} \left(\left| \frac{\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi}}{h_{1234}^{\mathbf{p}^\phi}} \right| \cdot \frac{1}{2^3 - 1} + \|\mathbf{I}_{34}\| \right). \tag{A17}$$

Because for any entropic $\mathbf{h}^{\mathbf{p}^\phi}$ and \mathbf{h}^* , we have $h_{1234}^{\mathbf{p}^\phi} > 0$ and $h_{1234}^* > 0$, while $|\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi}| < \infty$, $h_{1234}^{\mathbf{p}^\phi} < \infty$ and $h_{1234}^* < \infty$ hold for a fixed and finite alphabet, and $\|\mathbf{I}_{34}\|$ is a bounded constant. Thus, for any desired degree of accuracy $\kappa > 0$, we choose ϵ to be equal to (A17). Then, there always exists an NN such that

$$\|\mathbf{p}^\phi - \mathbf{p}^*\|_1 < \epsilon, \tag{A18}$$

and

$$\left| \frac{\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi}}{h_{1234}^{\mathbf{p}^\phi}} - \frac{\mathbf{I}_{34}\mathbf{h}^*}{h_{1234}^*} \right| < \kappa. \tag{A19}$$

With (A14) and (A19), by the permutation symmetry of the Ingleton inequality and Ingleton score, (30) is immediately proved.

Hence, the proof of Theorem 2 is complete. \square

Appendix B.3. Proof of Theorem 3

Similar to the proof of Theorem 2, we recall the proof of Theorem 1; we assume that \mathbf{h}^* is associated with PMF \mathbf{p}^* ; the main idea is to bound (30) as $\|\mathbf{p}^\phi - \mathbf{p}^*\|_1 < \epsilon$. The proof can be divided into two steps.

In the first step, we bound the term $|\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{I}_{34}\mathbf{h}^*|$, which is already proved in the first step of the proof of Theorem 3.

In the second step, similar to the proof of Theorem 2, we need to bound $\left| \frac{\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi}}{\|\mathbf{h}^{\mathbf{p}^\phi}\|} - \frac{\mathbf{I}_{34}\mathbf{h}^*}{\|\mathbf{h}^*\|} \right|$.

We first need to bound the term $\left| \|\mathbf{h}^{\mathbf{p}^\phi}\| - \|\mathbf{h}^*\| \right|$. By the triangular inequality and the proof of Theorem 2, if we choose $\eta' = \frac{\eta}{2^3-1}$ and ϵ to be equal to the corresponding ϵ' , we can derive

$$\|\mathbf{h}^{\mathbf{p}^\phi}\| = \|\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{h}^* + \mathbf{h}^*\| \leq \|\mathbf{h}^{\mathbf{p}^\phi} - \mathbf{h}^*\| + \|\mathbf{h}^*\| < \eta + \|\mathbf{h}^*\|, \tag{A20}$$

i.e., $\|\mathbf{h}^{\mathbf{p}^\phi}\| - \|\mathbf{h}^*\| < \eta$. We can similarly prove that $\|\mathbf{h}^*\| - \|\mathbf{h}^{\mathbf{p}^\phi}\| < \eta$, thus

$$-\eta < \|\mathbf{h}^{\mathbf{p}^\phi}\| - \|\mathbf{h}^*\| < \eta, \tag{A21}$$

and, equivalently,

$$\left| \|\mathbf{h}^{\mathbf{p}^\phi}\| - \|\mathbf{h}^*\| \right| < \eta. \tag{A22}$$

As we have bounded $\left| \|\mathbf{h}^{\mathbf{p}^\phi}\| - \|\mathbf{h}^*\| \right|$, in the rest of the proof, following the same proof strategy as Theorem 3 (i.e., (A14) to (A19)), it is straightforward to prove that, for any desired degree of accuracy $\mu > 0$, by properly choosing μ , there always exists an NN such that

$$\|\mathbf{p}^\phi - \mathbf{p}^*\|_1 < \epsilon, \tag{A23}$$

and

$$\left| \frac{\mathbf{I}_{34}\mathbf{h}^{\mathbf{p}^\phi}}{\|\mathbf{h}^{\mathbf{p}^\phi}\|} - \frac{\mathbf{I}_{34}\mathbf{h}^*}{\|\mathbf{h}^*\|} \right| < \mu. \tag{A24}$$

With (A23) and (A24), by the permutation symmetry of the Ingleton inequality and Ingleton score, (31) is immediately proved.

Thus, the proof of Theorem 3 is complete. \square

References

1. Yan, X.; Yeung, R.W.; Zhang, Z. An implicit characterization of the achievable rate region for acyclic multisource multisink network coding. *IEEE Trans. Inf. Theory* **2012**, *58*, 5625–5639.
2. Hassibi, B.; Shadbakht, S. Normalized entropy vectors, network information theory and convex optimization. In Proceedings of the 2007 IEEE Information Theory Workshop on Information Theory for Wireless Networks, Bergen, Norway, 1–6 July 2007; pp. 1–5.
3. Zhang, Z.; Yeung, R.W. A non-Shannon-type conditional inequality of information quantities. *IEEE Trans. Inf. Theory* **1997**, *43*, 1982–1986.
4. Matúš, F. Piecewise linear conditional information inequality. *IEEE Trans. Inf. Theory* **2005**, *52*, 236–238.
5. Chen, Q.; Yeung, R.W. Characterizing the entropy function region via extreme rays. In Proceedings of the 2012 IEEE Information Theory Workshop, Lausanne, Switzerland, 3–7 September 2012; pp. 272–276.
6. Tiwari, H.; Thakor, S. On characterization of entropic vectors at the boundary of almost entropic cones. In Proceedings of the 2019 IEEE Information Theory Workshop (ITW), Visby, Sweden, 25–28 August 2019; pp. 1–5.
7. Thakor, S.; Saleem, D. A Quasi-Uniform Approach to Characterizing the Boundary of the Almost Entropic Region. In Proceedings of the 2022 IEEE Information Theory Workshop (ITW), Mumbai, India, 1–9 November 2022; pp. 541–545.
8. Zhang, Z.; Yeung, R.W. On characterization of entropy function via information inequalities. *IEEE Trans. Inf. Theory* **1998**, *44*, 1440–1452.
9. Chan, T.; Grant, A. Dualities between entropy functions and network codes. *IEEE Trans. Inf. Theory* **2008**, *54*, 4470–4487.
10. Matus, F. Infinitely many information inequalities. In Proceedings of the 2007 IEEE International Symposium on Information Theory, Nice, France, 24–29 June 2007; pp. 41–44.
11. Li, C.T. The undecidability of conditional affine information inequalities and conditional independence implication with a binary constraint. *IEEE Trans. Inf. Theory* **2022**, *68*, 7685–7701.
12. Li, C.T. Undecidability of network coding, conditional information inequalities, and conditional independence implication. *IEEE Trans. Inf. Theory* **2023**, *69*, 3493–3510.
13. Hammer, D.; Romashchenko, A.; Shen, A.; Vereshchagin, N. Inequalities for Shannon entropy and Kolmogorov complexity. *J. Comput. Syst. Sci.* **2000**, *60*, 442–464.
14. Dougherty, R.; Freiling, C.; Zeger, K. Six new non-Shannon information inequalities. In Proceedings of the 2006 IEEE International Symposium on Information Theory, Seattle, WA, USA, 9–14 July 2006; pp. 233–236.
15. Dougherty, R.; Freiling, C.; Zeger, K. Linear rank inequalities on five or more variables. *arXiv* **2009**, arXiv:0910.0284.
16. Dougherty, R.; Freiling, C.; Zeger, K. Non-Shannon information inequalities in four random variables. *arXiv* **2011**, arXiv:1104.3602.
17. Dougherty, R. Computations of linear rank inequalities on six variables. In Proceedings of the 2014 IEEE International Symposium on Information Theory, Honolulu, HI, USA, 29 June–4 July 2014; pp. 2819–2823.
18. Hassibi, B.; Shadbakht, S. On a construction of entropic vectors using lattice-generated distributions. In Proceedings of the 2007 IEEE International Symposium on Information Theory, Nice, France, 24–29 June 2007; pp. 501–505.
19. Shadbakht, S.; Hassibi, B. MCMC methods for entropy optimization and nonlinear network coding. In Proceedings of the 2010 IEEE International Symposium on Information Theory, Austin, TX, USA, 13–18 June 2010; pp. 2383–2387.
20. Walsh, J.M.; Weber, S. A recursive construction of the set of binary entropy vectors and related algorithmic inner bounds for the entropy region. *IEEE Trans. Inf. Theory* **2011**, *57*, 6356–6363.
21. Matúš, F.; Csirmaz, L. Entropy region and convolution. *IEEE Trans. Inf. Theory* **2016**, *62*, 6007–6018.
22. Liu, Y.; Walsh, J.M. Non-isomorphic distribution supports for calculating entropic vectors. In Proceedings of the 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 29 September–2 October 2015; pp. 634–641.
23. Liu, Y.; Walsh, J.M. Mapping the region of entropic vectors with support enumeration & information geometry. *arXiv* **2015**, arXiv:1512.03324.
24. Liu, Y. Extremal Entropy: Information Geometry, Numerical Entropy Mapping, and Machine Learning Application of Associated Conditional Independences. Ph.D. Thesis, Drexel University, Philadelphia, PA, USA, 2016.
25. Walsh, J.M.; Trofimoff, A.E. On designing probabilistic supports to map the entropy region. In Proceedings of the 2019 IEEE Information Theory Workshop (ITW), Visby, Sweden, 25–28 August 2019; pp. 1–5.
26. Alam, S.; Thakor, S.; Abbas, S. Inner bounds for the almost entropic region and network code construction. *IEEE Trans. Commun.* **2020**, *69*, 19–30.
27. Saleem, D.; Thakor, S.; Tiwari, A. Recursive algorithm to verify quasi-uniform entropy vectors and its applications. *IEEE Trans. Commun.* **2020**, *69*, 874–883.
28. Liu, S.; Chen, Q. Entropy functions on two-dimensional faces of polymatroidal region of degree four. In Proceedings of the 2023 IEEE International Symposium on Information Theory (ISIT), Taipei, Taiwan, 25–30 June 2023; pp. 78–83.
29. Chan, H.L.; Yeung, R.W. A combinatorial approach to information inequalities. In Proceedings of the 1999 Information Theory and Networking Workshop (Cat. No. 99EX371), Metsovo, Greece, 27 June–1 July 1999; p. 63.
30. Chan, T.H.; Yeung, R.W. On a relation between information inequalities and group theory. *IEEE Trans. Inf. Theory* **2002**, *48*, 1992–1995.

31. Chan, T.H. Group characterizable entropy functions. In Proceedings of the 2007 IEEE International Symposium on Information Theory, Nice, France, 24–29 June 2007; pp. 506–510.
32. Mao, W.; Thill, M.; Hassibi, B. On Ingleton-violating finite groups. *IEEE Trans. Inf. Theory* **2016**, *63*, 183–200.
33. Boston, N.; Nan, T.T. Large violations of the Ingleton inequality. In Proceedings of the 2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 1–5 October 2012; pp. 1588–1593.
34. Nan, T.T. Entropy Regions and the Four-Atom Conjecture. Ph.D. Thesis, The University of Wisconsin-Madison, Madison, WI, USA, 2015.
35. Boston, N.; Nan, T.T. Violations of the Ingleton inequality and revising the four-atom conjecture. *Kybernetika* **2020**, *56*, 916–933.
36. Dougherty, R.; Freiling, C.; Zeger, K. Networks, matroids, and non-Shannon information inequalities. *IEEE Trans. Inf. Theory* **2007**, *53*, 1949–1969.
37. Li, C.; Weber, S.; Walsh, J.M. Multilevel diversity coding systems: Rate regions, codes, computation, & forbidden minors. *IEEE Trans. Inf. Theory* **2016**, *63*, 230–251.
38. Li, C.; Weber, S.; Walsh, J.M. On multi-source networks: Enumeration, rate region computation, and hierarchy. *IEEE Trans. Inf. Theory* **2017**, *63*, 7283–7303.
39. Chen, Q.; Cheng, M.; Bai, B. Matroidal Entropy Functions: A Quartet of Theories of Information, Matroid, Design, and Coding. *Entropy* **2021**, *23*, 323. [[CrossRef](#)] [[PubMed](#)]
40. Chen, Q.; Cheng, M.; Bai, B. Matroidal entropy functions: Constructions, characterizations and representations. *IEEE Trans. Inf. Theory* **2024**. [[CrossRef](#)]
41. Ingleton, A.W. Representation of matroids. *Comb. Math. Its Appl.* **1971**, *23*, 149–167.
42. Csirmaz, L. The dealer’s random bits in perfect secret sharing schemes. *Stud. Sci. Math. Hung.* **1996**, *32*, 429–438.
43. Shadbakht, S. Entropy Region and Network Information Theory. Ph.D. Thesis, California Institute of Technology, Pasadena, CA, USA, 2011.
44. Tsur, D.; Aharoni, Z.; Goldfeld, Z.; Permuter, H. Neural estimation and optimization of directed information over continuous spaces. *IEEE Trans. Inf. Theory* **2023**, *69*, 4777–4798.
45. Tsur, D.; Huleihel, B.; Permuter, H. Rate distortion via constrained estimated mutual information minimization. In Proceedings of the 2023 IEEE International Symposium on Information Theory (ISIT), Taipei, Taiwan, 25–30 June 2023; pp. 695–700.
46. Tsur, D.; Aharoni, Z.; Goldfeld, Z.; Permuter, H. Data-driven optimization of directed information over discrete alphabets. *IEEE Trans. Inf. Theory* **2023**, *70*, 1652–1670.
47. Wang, X.; Karimi, H.R.; Shen, M.; Liu, D.; Li, L.W.; Shi, J. Neural network-based event-triggered data-driven control of disturbed nonlinear systems with quantized input. *Neural Netw.* **2022**, *156*, 152–159. [[PubMed](#)]
48. Wang, M.; Zhu, S.; Shen, M.; Liu, X.; Wen, S. Fault-Tolerant Synchronization for Memristive Neural Networks with Multiple Actuator Failures. *IEEE Trans. Cybern.* **2024**, 1–10. [[CrossRef](#)]
49. Ziegler, G.M. *Lectures on Polytopes*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 152.
50. Fukuda, K.; Prodon, A. Double description method revisited. In Proceedings of the Franco-Japanese and Franco-Chinese Conference on Combinatorics and Computer Science, Brest, France, 3–5 July 1995; Springer: Berlin/Heidelberg, Germany, 1995; pp. 91–111.
51. Fukuda, K. cdd, cddplus and cddlib. Available online: https://people.inf.ethz.ch/fukudak/cdd_home/ (accessed on 18 May 2024).
52. Troffaes, M. pycddlib. Available online: <https://github.com/mcmtroffaes/pycddlib> (accessed on 18 May 2024).
53. Motzkin, T.S.; Raiffa, H.; Thompson, G.L.; Thrall, R.M. The double description method. *Contrib. Theory Games* **1953**, *2*, 51–73.
54. Yeung, R.W. *Information Theory and Network Coding*; Springer Science & Business Media: New York, NY, USA, 2008.
55. Fujishige, S. Polymatroidal dependence structure of a set of random variables. *Inf. Control* **1978**, *39*, 55–72.
56. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366.
57. Bertsekas, D.P. *Nonlinear Programming*; Athena Scientific: Nashua, NH, USA, 1998.
58. Shalev-Shwartz, S.; Ben-David, S. *Understanding Machine Learning: From Theory to Algorithms*; Cambridge University Press: Cambridge, UK, 2014.
59. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
60. Matúš, F.; Studený, M. Conditional independences among four random variables I. *Comb. Probab. Comput.* **1995**, *4*, 269–278.
61. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
62. Alam, S.; Thakor, S.; Abbas, S. Numerical Data: Inner Bounds for the Almost Entropic Region and Network Code Construction. Available online: <https://github.com/mdsultanalam/NumericalDataInnerBoundsNetworkCodeConstruction> (accessed on 18 May 2024).
63. Büeler, B.; Enge, A.; Fukuda, K. Exact volume computation for polytopes: A practical study. In *Polytopes—Combinatorics and Computation*; Birkhäuser: Basel, Switzerland, 2000; pp. 131–154.
64. Barber, C.B.; Dobkin, D.P.; Huhdanpaa, H. The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw. (TOMS)* **1996**, *22*, 469–483.
65. Barber, C.B.; Dobkin, D.P.; Huhdanpaa, H. Qhull. Available online: <http://www.qhull.org> (accessed on 18 May 2024).

-
66. Avis, D. A Revised Implementation of the Reverse Search Vertex Enumeration Algorithm. In *Polytopes—Combinatorics and Computation*; Birkhäuser: Basel, Switzerland, 2000; pp. 177–198.
 67. Avis, D. lrs. Available online: <http://cgm.cs.mcgill.ca/~avis/C/lrs.html> (accessed on 18 May 2024).
 68. Yeung, R.W. A framework for linear information inequalities. *IEEE Trans. Inf. Theory* **1997**, *43*, 1924–1934.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.