Bounding the Entropic Region via Information Geometry

Yunshu Liu John MacLaren Walsh
Dept. of ECE, Drexel University, Philadelphia, PA 19104, USA
yunshu.liu@drexel.edu jwalsh@coe.drexel.edu

***Abstract*—This paper suggests that information geometry may form a natural framework to deal with the unknown part of the boundary of entropic region. An application of information geometry shows that distributions associated with Shannon facets can be associated, in the right coordinates, with affine collections of distributions. This observation allows an information geometric reinterpretation of the Shannon-type inequalities as arising from a Pythagorean style relationship. The set of distributions which violate Ingleton's inequality, and hence are linked with the part of the entropic region which is yet undetermined, is shown also to have a surprising affine information geometric structure in a special case involving four random variables and a certain support. These facts provide strong evidence for the link between information geometry and characterizing the boundary of the entropic region.**

I. Introduction

The region of entropic vectors is a convex cone that is known to be at the core of many yet undetermined fundamental limits for problems in data compression, network coding, and multimedia transmission. This set has recently shown to be non-polyhedral, but its boundaries remain unknown. In §II, we give some background on the region of entropic vectors, which include discussion about inner and outer bounds for it, and the gap between Shannon outer bound and Ingleton inner bound for 4 random variables. In §III we point out that existing inner and outer bounds, which based on linear constructions, are unable to parameterize the boundary of the region of entropic vectors because non-linear dependence structures are necessary to parameterize the unknown part of it. We argue that information geometry may be a tool that could parameterize this unknown part of the boundary where the best known bounds do not meet. This claim is partially substantiated by providing in §III-B an information geometric interpretation of Shannon facets of the region of entropic vectors and the Shannon type inequalities, as well as by providing in §III-C an information geometric characterization of a special type of distributions at four random variables that violate Ingleton Inequalities. The paper concludes in §IV with a number of interesting directions for future work.

II. Bounding the Region of Entropic Vectors

Consider N discrete random variables $\boldsymbol{X} = \boldsymbol{X}_{\mathcal{N}} = (X_1, \dots, X_N)$, $\mathcal{N} = \{1, \dots, N\}$ with joint probability mass function $p_{\boldsymbol{X}}(\boldsymbol{x})$. To every non-empty subset of these random variables $\boldsymbol{X}_{\mathcal{A}} := (X_n \mid n \in \mathcal{A})$, $\mathcal{A} \subset \{1, \dots, N\}$, there is associated a Shannon entropy $H(\boldsymbol{X}_{\mathcal{A}})$ calculated from the marginal distribution $p_{\boldsymbol{X}_{\mathcal{A}}}(\boldsymbol{x}_{\mathcal{A}}) = \sum_{\boldsymbol{x}_{\mathcal{N}\setminus\mathcal{A}}} p_{\boldsymbol{X}_{\mathcal{N}}}(\boldsymbol{x})$ via

$$H(\boldsymbol{X}_{\mathcal{A}}) = \sum_{\boldsymbol{x}_{\mathcal{A}}} -p_{\boldsymbol{X}_{\mathcal{A}}}(\boldsymbol{x}_{\mathcal{A}}) \log_2 p_{\boldsymbol{X}_{\mathcal{A}}}(\boldsymbol{x}_{\mathcal{A}}) \quad (1)$$

One can stack these entropies of different non-empty subsets into a $2^N - 1$ dimensional vector $\boldsymbol{h} = (H(\boldsymbol{X}_{\mathcal{A}})|\mathcal{A} \subseteq \mathcal{N})$, which can be clearly viewed as $\boldsymbol{h}(p_{\boldsymbol{X}})$, a function of the joint distribution $p_{\boldsymbol{X}}$. A vector $\boldsymbol{h}_? \in \mathbb{R}^{2^N-1}$ is said to be *entropic* if there is some joint distribution $p_{\boldsymbol{X}}$ such that $\boldsymbol{h}_? = \boldsymbol{h}(p_{\boldsymbol{X}})$. The region of entropic vectors is then the image of the set $\mathcal{D} = \{p_{\boldsymbol{X}} | p_{\boldsymbol{X}}(\boldsymbol{x}) \geq 0, \ \sum_{\boldsymbol{x}} p_{\boldsymbol{X}}(\boldsymbol{x}) = 1\}$ of valid joint probability mass functions under the function $\boldsymbol{h}(\cdot)$, and is denoted by

$$\Gamma_N^* = \boldsymbol{h}(\mathcal{D}) \subsetneq \mathbb{R}^{2^N-1} \quad (2)$$

It is known that the closure of this set $\bar{\Gamma}_N^*$ is a convex cone [1], but surprisingly little else is known about this cone for $N \geq 4$. Understanding the "shape" and boundaries of the set $\bar{\Gamma}_N^*$ is the subject of this paper. The fundamental importance of $\bar{\Gamma}_N^*$ lies in several contexts in signal processing, compression, network coding and information theory [1].

A. Outer Bounds and Inner Bounds

Viewed as a function $h_{\mathcal{A}} = H(\boldsymbol{X}_{\mathcal{A}})$ of the selected subset, with the convention that $h_{\emptyset} = 0$, entropy is *sub-modular* [1], [2], meaning that

$$h_{\mathcal{A}} + h_{\mathcal{B}} \geq h_{\mathcal{A}\cap\mathcal{B}} + h_{\mathcal{A}\cup\mathcal{B}} \quad \forall \mathcal{A}, \mathcal{B} \subseteq \mathcal{N}, \quad (3)$$

and is also *non-decreasing* and *non-negative*, meaning that

$$h_{\mathcal{A}} \geq h_{\mathcal{B}} \geq 0 \quad \forall \mathcal{B} \subseteq \mathcal{A} \subseteq \mathcal{N}. \quad (4)$$

Viewed as requirements for arbitrary set functions (not necessarily entropy) the inequalities (3) and (4) are known as the *polymatroidal axioms* [1], [2], and a function obeying them is called the *rank function* of a *polymatroid*. If a set function r that obeys the polymatroidal axioms (3) and (4) additionally obeys

$$r_{\mathcal{A}} \leq |\mathcal{A}|, \quad r_{\mathcal{A}} \in \mathbb{Z} \quad \forall \mathcal{A} \subseteq \mathcal{N} \quad (5)$$

then it is the rank function of a *matroid* on the ground set $\mathcal{N}$.

Since entropy must obey the polymatroidal axioms, the set of all rank functions of polymatroids forms an outer bound for the region of entropic vectors which is often denoted by

$$\Gamma_N = \left\{ \boldsymbol{h} \;\middle|\; \begin{array}{c} \boldsymbol{h} \in \mathbb{R}^{2^N-1} \\ h_{\mathcal{A}} + h_{\mathcal{B}} \geq h_{\mathcal{A}\cap\mathcal{B}} + h_{\mathcal{A}\cup\mathcal{B}} \ \forall \mathcal{A}, \mathcal{B} \subseteq \mathcal{N} \\ h_{\mathcal{P}} \geq h_{\mathcal{Q}} \geq 0 \quad \forall \mathcal{Q} \subseteq \mathcal{P} \subseteq \mathcal{N} \end{array} \right\} \quad (6)$$

Γ_N is a polyhedron, this polyhedral set Γ_N is often known as the Shannon outer bound for $\bar{\Gamma}^*_N$ [1], [2].

While in the low dimensional cases we have $\Gamma_2 = \Gamma^*_2$ and $\Gamma_3 = \bar{\Gamma}^*_3$, for $N \geq 4$, however, $\Gamma_N \neq \bar{\Gamma}^*_N$. Zhang and Yeung first showed this in [2] by proving a new inequality among 4 variables

$$2I(C;D) \leq I(A;B) + I(A;C,D) + 3I(C;D|A) + I(C;D|B)$$

which held for entropies and was not implied by the polymatroidal axioms, which they dubbed a *non-Shannon type* inequality to distinguish it from inequalities implied by Γ_N. For roughly the next decade a few authors produced other new non-Shannon inequalities [3], [4]. In 2007, Matúš [5] showed that $\bar{\Gamma}^*_N$ is not a polyhedron for $N \geq 4$. The proof of this fact was carried out by constructing three sequences of non-Shannon inequalities, including

$$s[I(X_1;X_2|X_3) + I(X_1;X_2|X_4) + I(X_3;X_4) - I(X_1;X_2)] + I(X_2;X_3|X_1) + \frac{s(s+1)}{2}[I(X_1;X_3|X_2) + I(X_1;X_2|X_3)] \geq 0$$

Additionally, Matúš constructed a curve known to be in $\bar{\Gamma}^*_N$ which, together with the infinite sequences of inequalities, was geometrically arranged in a manner that prohibits $\bar{\Gamma}^*_N$ from being a polyhedron. Nonetheless, despite this characterization, the outer bound of even $\bar{\Gamma}^*_4$ is still far away from tight.

Alternately, from the inside, the most common way to generate inner bounds for the region of entropic vectors is to consider special families of distributions for which the entropy function is known to have certain properties. [6] talked about generating inner bounds through representable matroid. [7], [8], [9], [10], [11] focus on calculating inner bounds based on special properties of binary random variables. A third way to generate inner bounds is based on inequalities for subspace arrangements. Define $\mathcal{S}_N$ to be the conic hull of all subspace ranks for N subspaces, it is known that $\mathcal{S}_N$ is an inner bound for $\bar{\Gamma}^*_N$[12]. At present, $\mathcal{S}_N$ is only known for $N \leq 5$. We have $\mathcal{S}_2 = \bar{\Gamma}^*_2 = \Gamma_2$, $\mathcal{S}_3 = \bar{\Gamma}^*_3 = \Gamma_3$ and $\mathcal{S}_4 \subset \bar{\Gamma}^*_4 \subset \Gamma_4$, where $\mathcal{S}_4$ is given by the Shannon type inequalities (i.e. the polymatroidal axioms) together with an additional form of inequality known as *Ingleton*'s inequality [12], [13], [14] which states that for $N = 4$ random variables

$$\begin{aligned} Ingleton_{ij} &= I(X_k;X_l|X_i) + I(X_k;X_l|X_j) \\ &+ I(X_i;X_j) - I(X_k;X_l) \geq 0 \end{aligned} \quad (7)$$

B. The Gap between $\mathcal{S}_4$ and Γ_4

Since the region of entropic vector Γ^*_N is unknown starting from N = 4, we focus on N = 4 in this paper. As we know, $\mathcal{S}_4 \subsetneq \Gamma_4$, where Γ_4 is generated by 28 elemental Shannon type information inequalities[1] and $\mathcal{S}_4$ is generated by the previous 28 Shannon type information inequalities together with six Ingleton's inequalities (7). Matúš in [14] pointed out that Γ_4 is the disjoint union of $\mathcal{S}_4$ and six sets $\{h \in \Gamma_4 | Ingleton_{ij} < 0\}$. The six cones $G^{ij}_4 = \{h \in \Gamma_4 | Ingleton_{ij} \leq 0\}$ are symmetric due to the permutation of inequalities $Ingleton_{ij}$, so it sufficient to study only one of the cones. Furthermore, Matúš give the extreme rays of G^{ij}_4 in Lemma 1 by using the following functions: for $N = \{1,2,3,4\}$, with $I \subseteq N$ and $0 \leq t \leq |N \backslash I|$, define

$$r^I_t(J) = \min\{t, |J \backslash I|\} \ with \ J \subseteq N$$

$$f_{ij}(K) = \begin{cases} 3 & \text{if } K \in \{ik, jk, il, jl, kl\} \\ min\{4, 2|K|\} & \text{otherwise} \end{cases}$$

Lemma 1: (Matúš)[14] The cone $G^{ij}_4 = \{h \in \Gamma_4 | Ingleton_{ij} \leq 0\}$, $i, j \in N$ distinct is the convex hull of 15 extreme rays. They are generated by the 15 linearly independent functions f_{ij}, r^{ijk}_1, r^{ijl}_1, r^{ikl}_1, r^{jkl}_1, $r^{\emptyset}_1$, $r^{\emptyset}_3$, r^i_1, r^j_1, r^{ik}_1, r^{jk}_1, r^{il}_1, r^{jl}_1, r^k_2, r^l_2, where $kl = N - ij$.

As identified early, the region of entropic vector $\bar{\Gamma}^*_4$ is known as long as we know the structure of six cones G^{ij}_4. Because of symmetry, we only need focus on one of the six cones, in the rest of the paper, we will focus on understanding G^{34}_4.

III. Perspectives on How to Use Information Geometry to Understand Entropy Geometry

As identified in the previous sections, $\bar{\Gamma}^*_N$ is a convex cone that is known to be non-polyhedral, which suggests that differential geometric tools may be necessary to characterize it. Information geometry, on the other hand, is a discipline in statistics which endows the manifold of probability distributions with a special "dually flat" differential geometric structure created by selecting a Riemannian metric based on the Fisher information and a family of affine connections called the α-connections. Building on these observations, here we wish to sketch out a couple of ideas in an attempt to show that characterizing the boundary of $\bar{\Gamma}^*_N$ could be thought of as a problem in information geometry, and that information geometry may in fact be a convenient framework to utilize to calculate the boundaries of $\bar{\Gamma}^*_N$.

A. Review of Some Relevant Ideas from Information Geometry

Information geometry endows a manifold of probability distributions $p(x;\boldsymbol{\xi})$, parameterized by a vector of real numbers $\boldsymbol{\xi} = [\xi_i]$, with a Riemannian metric, or inner product between tangent vectors, given by the Fisher information:

$$g_{i,j}(\xi) \triangleq \mathbb{E}_{\xi}[\frac{\partial \log p(x;\boldsymbol{\xi})}{\partial \xi_i} \frac{\partial \log p(x;\boldsymbol{\xi})}{\partial \xi_j}] \quad (8)$$

This allows us to calculate an inner product between two tangent vectors $c = \sum_i c_i \partial_{\xi_i}$ and $d = \sum_i d_i \partial_{\xi_i}$ at a particular point ξ in the manifold of probability distributions, as

$$< c, d >_{\xi} = \sum_{i,j} c_i d_j g_{i,j}(\xi) \quad (9)$$

Selecting this Riemannian metric (indeed, we could have selected others), and some additional structure, allows the differential geometric structure to be related to familiar properties of exponential families and mixture families of probability distributions. While the resulting theory is very elegant, it is also somewhat complex, and hence we must omit the general details, referring the interested reader to [15], and only introduce a small subset of the concepts.

In particular, let's focus our attention on the manifold $\mathcal{P}(\boldsymbol{\mathcal{X}})$ of probability distributions for a random vector $\boldsymbol{X} = [X_1, \ldots, X_N]$ taking values on the Cartesian product $\boldsymbol{\mathcal{X}} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_N$, where $\mathcal{X}_n$ is a finite set with values denoted by $v_{i,n}, i \in \{1, \ldots, |\mathcal{X}_n|\}$. To a particular probability mass function $p_{\boldsymbol{X}}$ we can associate a length $\prod_{n=1}^N |\mathcal{X}_n| - 1$

vector, called the m-coordinates or $\boldsymbol{\eta}$-coordinates, by listing the probabilities of all but one of the outcomes in $\boldsymbol{\mathcal{X}}$ into a vector. Such an $\boldsymbol{\eta}$ can be determined uniquely from the probability mass function $p_{\boldsymbol{X}}$, and owing to the fact that the probability mass function must sum to one, the omitted probability can be calculated, and hence the probability mass function can be determined from η.

Alternatively we could parameterize the probability mass function for such a joint distribution with a vector $\boldsymbol{\theta}$, whose $\prod_{n=1}^{N}|\mathcal{X}_n|-1$ elements take the form

$$\boldsymbol{\theta}=\left[\log\left(\frac{p_{\mathbf{X}}(v_{i_1,1},\ldots,v_{i_N,N})}{p_{\mathbf{X}}(v_{1,1},\ldots,v_{1,N})}\right)\middle|\begin{array}{c} i_k\in\{1,2,\ldots,|\mathcal{X}_k|\},\\ k\in\{1,\ldots,N\},\\ \prod_k i_k\neq 1.\end{array}\right]$$

These coordinates provide an alternate unique way of specifying the joint probability mass function $p_{\mathbf{X}}$, called the e-coordinates or $\boldsymbol{\theta}$ coordinates.

A subfamily of these probability mass functions associated with those $\boldsymbol{\eta}$ coordinates that take the form

$$\boldsymbol{\eta}=\mathbf{A}\boldsymbol{p}+\mathbf{b} \tag{10}$$

for some $\boldsymbol{p}$ for any particular fixed $\mathbf{A}$ and $\mathbf{b}$, that is, that lie in an affine submanifold of the $\boldsymbol{\eta}$ coordinates, are said to form a m-*autoparallel* submanifold of probability mass functions. This is not a definition, but rather a consequence of a theorem involving a great deal of additional structure which we must omit here [15].

Similarly, a subfamily of these probability mass functions associated with those $\boldsymbol{\theta}$ coordinates that take the form

$$\boldsymbol{\theta}=\mathbf{A}\boldsymbol{\lambda}+\mathbf{b} \tag{11}$$

for some $\boldsymbol{\lambda}$ for any particular fixed $\mathbf{A}$ and $\mathbf{b}$, that is, that lie in an affine submanifold of the $\boldsymbol{\theta}$ coordinates, are said to form a e-*autoparallel* submanifold of probability mass functions.

An e-autoparallel submanifold (resp. m-autoparallel submanifold) that is one dimensional, in that its $\boldsymbol{\lambda}$ (resp. $\boldsymbol{p}$) parameter vector is in fact a scalar, is called a e-*geodesic* (resp. m-*geodesic*).

On this manifold of probability mass functions for random variables taking values in the set $\boldsymbol{\mathcal{X}}$, we can also define the Kullback Leibler divergence, or relative entropy, measured in bits, according to

$$D(p_{\boldsymbol{X}}||q_{\boldsymbol{X}})=\sum_{\boldsymbol{x}\in\boldsymbol{\mathcal{X}}}p_{\boldsymbol{X}}(\boldsymbol{x})\log_2\left(\frac{p_{\boldsymbol{X}}(\boldsymbol{x})}{q_{\boldsymbol{X}}(\boldsymbol{x})}\right) \tag{12}$$

Note that in this context $D(p||q)\geq 0$ with equality iff. $p=q$, hence this function is a bit like a distance, however it does *not* in general satisfy symmetry or triangle inequality.

Let $\mathcal{E}$ be a particular e-autoparallel submanifold, and consider a probability distribution $p_{\boldsymbol{X}}$ not necessarily in this submanifold. The problem of finding the point $\overrightarrow{\Pi}_{\mathcal{E}}(p_{\boldsymbol{X}})$ in $\mathcal{E}$ closest in Kullback Leibler divergence to $p_{\boldsymbol{X}}$ defined by

$$\overrightarrow{\Pi}_{\mathcal{E}}(p_{\boldsymbol{X}})\triangleq\arg\min_{q_{\boldsymbol{X}}\in\mathcal{E}}D(p_{\boldsymbol{X}}||q_{\boldsymbol{X}}) \tag{13}$$

is well posed, and is characterized in the following two ways(here $\overrightarrow{\Pi}_{\mathcal{E}}$ with a right arrow means we are minimizing over the second arguments $q_{\boldsymbol{X}}$). The tangent vector of the m-geodesic connecting $p_{\boldsymbol{X}}$ to $\overrightarrow{\Pi}_{\mathcal{E}}(p_{\boldsymbol{X}})$ is orthogonal, in the sense of achieving Riemannian metric value 0, at $\overrightarrow{\Pi}_{\mathcal{E}}(p_{\boldsymbol{X}})$ to the tangent vector of the e-geodesic connecting $\overrightarrow{\Pi}_{\mathcal{E}}(p_{\boldsymbol{X}})$ and any other point in $\mathcal{E}$. Additionally, for any other point $q_{\boldsymbol{X}}\in\mathcal{E}$, we have the Pythagorean like relation

$$D(p_{\boldsymbol{X}}||q_{\boldsymbol{X}})=D(p_{\boldsymbol{X}}||\overrightarrow{\Pi}_{\mathcal{E}}(p_{\boldsymbol{X}}))+D(\overrightarrow{\Pi}_{\mathcal{E}}(p_{\boldsymbol{X}})||q_{\boldsymbol{X}}). \tag{14}$$

B. *Information Geometric Structure of the Shannon Exposed Faces of the Region of Entropic Vectors*

As identified in §II-A, $\Gamma_2=\Gamma_2^*$ and $\Gamma_3=\bar{\Gamma}_3^*$, implying that Γ_2^* and $\bar{\Gamma}_3^*$ are fully characterized by Shannon type information inequalities. For $N=4$, even though $\Gamma_4\neq\bar{\Gamma}_4^*$ and the region is not a polyhedral cone, there are still many exposed faces of $\bar{\Gamma}_N^*$ defined by attaining equality in a particular Shannon type information inequality of the form (3) or (4). Such exposed faces of $\bar{\Gamma}_N^*$ could be referred to as the "Shannon facets" of entropy, and in this section we aim to characterize the distributions associated with these Shannon facets via information geometry.

Let $\mathcal{E}_{\mathcal{A}}^{\perp}$ be the submanifold of probability distributions for which $\boldsymbol{X}_{\mathcal{A}}$ and $\boldsymbol{X}_{\mathcal{A}^c}=\boldsymbol{X}_{\mathcal{N}\setminus\mathcal{A}}$ are independent

$$\mathcal{E}_{\mathcal{A}}^{\perp}=\{p_{\boldsymbol{X}}(\cdot)\,|\,p_{\boldsymbol{X}}(\boldsymbol{x})=p_{\boldsymbol{X}_{\mathcal{A}}}(\boldsymbol{x}_{\mathcal{A}})p_{\boldsymbol{X}_{\mathcal{A}^c}}(\boldsymbol{x}_{\mathcal{A}^c})\quad\forall\boldsymbol{x}\} \tag{15}$$

then we define $\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}$, $\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}$ and $\mathcal{M}_{\mathcal{A}}$ as follows

$$\begin{aligned}
\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}&=\{p_{\boldsymbol{X}}(\cdot)\,|\,p_{\boldsymbol{X}}=p_{\boldsymbol{X}_{(\mathcal{A}\cup\mathcal{B})}}p_{\boldsymbol{X}_{(\mathcal{A}\cup\mathcal{B})^c}}\}\\
\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}&=\left\{p_{\boldsymbol{X}}(\cdot)\,\middle|\,p_{\boldsymbol{X}}=p_{\boldsymbol{X}_{\mathcal{A}}}p_{\boldsymbol{X}_{\mathcal{B}\setminus\mathcal{A}}|\boldsymbol{X}_{\mathcal{A}\cap\mathcal{B}}}p_{\boldsymbol{X}_{(\mathcal{A}\cup\mathcal{B})^c}}\right\}\\
\mathcal{M}_{\mathcal{A}}&=\{p_{\boldsymbol{X}}(\cdot)\,|\,p_{\boldsymbol{X}}=p_{\boldsymbol{X}_{\mathcal{A}}}\cdot\delta_{\mathcal{A}^c}\}
\end{aligned}$$

where $\delta_{\mathcal{A}^c}=\begin{cases}1 & \text{if } X_i=v_{1,i}\ \forall i\in\mathcal{A}^c\\ 0 & \text{otherwise}\end{cases}$. Note $\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}$ is a submanifold of $\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}$ such that the random variables $\boldsymbol{X}_{\mathcal{A}\cup\mathcal{B}}$, in addition to being independent from $\boldsymbol{X}_{(\mathcal{A}\cup\mathcal{B})^c}$ form the Markov chain $\boldsymbol{X}_{\mathcal{A}\setminus\mathcal{B}}\leftrightarrow\boldsymbol{X}_{\mathcal{A}\cap\mathcal{B}}\leftrightarrow\boldsymbol{X}_{\mathcal{B}\setminus\mathcal{A}}$. These sets of distributions will be useful because $I(\boldsymbol{X}_{\mathcal{A}\cup\mathcal{B}};\boldsymbol{X}_{(\mathcal{A}\cup\mathcal{B})^c})=h_{\mathcal{A}\cup\mathcal{B}}+h_{(\mathcal{A}\cup\mathcal{B})^c}-h_{\mathcal{N}}=0$ for every distribution in $\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}$ and $\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}$, $I(\boldsymbol{X}_{\mathcal{A}\setminus\mathcal{B}};\boldsymbol{X}_{\mathcal{B}\setminus\mathcal{A}}|\boldsymbol{X}_{\mathcal{A}\cap\mathcal{B}})=h_{\mathcal{A}}+h_{\mathcal{B}}-h_{(\mathcal{A}\cap\mathcal{B})}-h_{(\mathcal{A}\cup\mathcal{B})}=0$ for every distribution in $\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}$, and $h_{\mathcal{N}}-h_{\mathcal{A}}=0$ for every distribution in $\mathcal{M}_{\mathcal{A}}$.

Proposition 1: $\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}\subset\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}\subset\mathcal{P}(\boldsymbol{\mathcal{X}})$, $\mathcal{E}_{\mathcal{A},\mathcal{B}}^{\leftrightarrow,\perp}$ is an e-autoparallel submanifold of $\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}$ and $\mathcal{E}_{\mathcal{A}\cup\mathcal{B}}^{\perp}$ is an e-autoparallel submanifold of $\mathcal{P}(\boldsymbol{\mathcal{X}})$.

Proposition 2: Let $\mathcal{A},\mathcal{B}\subseteq\{1,\ldots,N\}$ be given, where $\mathcal{A}\subseteq\mathcal{B}$, then $\mathcal{M}_{\mathcal{A}}\subseteq\mathcal{M}_{\mathcal{B}}\subseteq\mathcal{P}(\boldsymbol{\mathcal{X}})$, $\mathcal{M}_{\mathcal{A}}$ is a m-autoparallel submanifold of $\mathcal{M}_{\mathcal{B}}$ and $\mathcal{M}_{\mathcal{B}}$ is a m-autoparallel submanifold of $\mathcal{P}(\boldsymbol{\mathcal{X}})$.

The proofs of Proposition 1 and Proposition 2 are based on the properity of exponential family and conditional exponential family, we omitted them due to space limit. These two Propositions have shown that Shannon facets are associated with affine subsets of the family of probability distribution, when it is regarded in an appropriate parameterization. In fact, as we shall presently show, all Shannon Type information

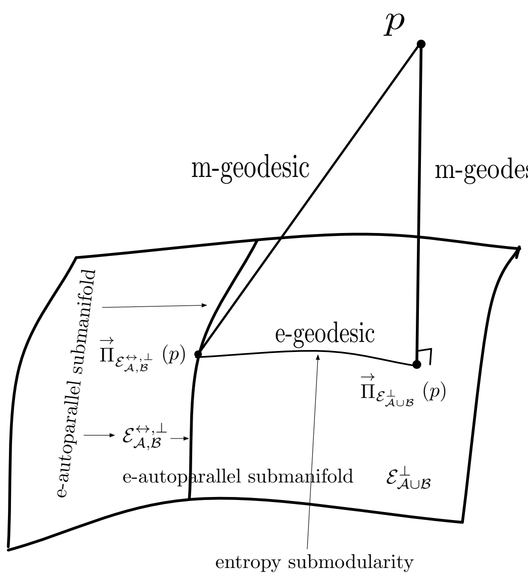

Fig. 1. Submodularity of the entropy function is equivalent to the non-negativity of a divergence $D(\overrightarrow{\pi}_{\mathcal{E}^{\perp}_{\mathcal{A}\cup\mathcal{B}}}(p)||\overrightarrow{\pi}_{\mathcal{E}^{\leftrightarrow,\perp}_{\mathcal{A},\mathcal{B}}}(p))$ between two information projections, one projection ($\overrightarrow{\pi}_{\mathcal{E}^{\leftrightarrow,\perp}_{\mathcal{A},\mathcal{B}}}(p)$) is to a set that is a submanifold the other projection's ($\overrightarrow{\pi}_{\mathcal{E}^{\perp}_{\mathcal{A}\cup\mathcal{B}}}(p)$) set. A Pythagorean style relation shows that for such an arrangement $D(p||\overrightarrow{\pi}_{\mathcal{E}^{\leftrightarrow,\perp}_{\mathcal{A},\mathcal{B}}}(p)) = D(p||\overrightarrow{\pi}_{\mathcal{E}^{\perp}_{\mathcal{A}\cup\mathcal{B}}}(p)) + D(\overrightarrow{\pi}_{\mathcal{E}^{\perp}_{\mathcal{A}\cup\mathcal{B}}}(p)||\overrightarrow{\pi}_{\mathcal{E}^{\leftrightarrow,\perp}_{\mathcal{A},\mathcal{B}}}(p))$.

inequalities correspond to the positivity of divergences of projections to these submanifolds. The nested nature of these submanifolds, and the associated Pythagorean relation, is one way to view the submodularity of entropy, as we shall now explain with Figure 1 and theorem 1.

Theorem 1: The submodularity (3) of the entropy function can be viewed as a consequence of Pythagorean style information projection relationships depicted in Figure 1. In particular, submodularity is equivalent to the inequality

$$D\left(\overrightarrow{\Pi}_{\mathcal{E}^{\perp}_{\mathcal{A}\cup\mathcal{B}}}(p_{\boldsymbol{X}})||\overrightarrow{\Pi}_{\mathcal{E}^{\leftrightarrow,\perp}_{\mathcal{A},\mathcal{B}}}(p_{\boldsymbol{X}})\right) \geq 0 \tag{16}$$

since

$$D\left(\overrightarrow{\Pi}_{\mathcal{E}^{\perp}_{\mathcal{A}\cup\mathcal{B}}}(p_{\boldsymbol{X}})||\overrightarrow{\Pi}_{\mathcal{E}^{\leftrightarrow,\perp}_{\mathcal{A},\mathcal{B}}}(p_{\boldsymbol{X}})\right) = H(\boldsymbol{X}_{\mathcal{A}}) + H(\boldsymbol{X}_{\mathcal{B}}) - H(\boldsymbol{X}_{\mathcal{A}\cap\mathcal{B}}) - H(\boldsymbol{X}_{\mathcal{A}\cup\mathcal{B}})$$

Proof: The projections are

$$\overrightarrow{\Pi}_{\mathcal{E}^{\perp}_{\mathcal{A}\cup\mathcal{B}}}(p_{\boldsymbol{X}}) = p_{\boldsymbol{X}_{\mathcal{A}\cup\mathcal{B}}} p_{\boldsymbol{X}_{(\mathcal{A}\cup\mathcal{B})^c}}, \text{ and} \tag{17}$$

$$\overrightarrow{\Pi}_{\mathcal{E}^{\leftrightarrow,\perp}_{\mathcal{A},\mathcal{B}}}(p_{\boldsymbol{X}}) = p_{\boldsymbol{X}_{\mathcal{A}\setminus\mathcal{B}}|\boldsymbol{X}_{\mathcal{A}\cap\mathcal{B}}} p_{\boldsymbol{X}_{\mathcal{B}}} p_{\boldsymbol{X}_{(\mathcal{A}\cup\mathcal{B})^c}} \tag{18}$$

since for $\overrightarrow{\Pi}_{\mathcal{E}^{\perp}_{\mathcal{A}\cup\mathcal{B}}}(p_{\boldsymbol{X}})$ given by (17) for every $q_{\boldsymbol{X}} \in \mathcal{E}^{\perp}_{\mathcal{A}\cup\mathcal{B}}$ we can reorganize the divergence as

$$D(p_{\boldsymbol{X}}||q_{\boldsymbol{X}}) = D(p_{\boldsymbol{X}}||\overrightarrow{\Pi}_{\mathcal{E}^{\perp}_{\mathcal{A}\cup\mathcal{B}}}(p_{\boldsymbol{X}})) + D(\overrightarrow{\Pi}_{\mathcal{E}^{\perp}_{\mathcal{A}\cup\mathcal{B}}}(p_{\boldsymbol{X}})||q_{\boldsymbol{X}_{(\mathcal{A}\cup\mathcal{B})}} q_{\boldsymbol{X}_{(\mathcal{A}\cup\mathcal{B})^c}})$$

and for $\overrightarrow{\Pi}_{\mathcal{E}^{\leftrightarrow,\perp}_{\mathcal{A},\mathcal{B}}}(p_{\boldsymbol{X}})$ given by (18) for every $q_{\boldsymbol{X}} \in \mathcal{E}^{\leftrightarrow,\perp}_{\mathcal{A},\mathcal{B}}$ we can reorganize the divergence as

$$D(p_{\boldsymbol{X}}||q_{\boldsymbol{X}}) = D(p_{\boldsymbol{X}}||\overrightarrow{\Pi}_{\mathcal{E}^{\leftrightarrow,\perp}_{\mathcal{A},\mathcal{B}}}(p_{\boldsymbol{X}})) + D(\overrightarrow{\Pi}_{\mathcal{E}^{\leftrightarrow,\perp}_{\mathcal{A},\mathcal{B}}}(p_{\boldsymbol{X}})||q_{\boldsymbol{X}_{\mathcal{A}\setminus\mathcal{B}}|\boldsymbol{X}_{\mathcal{A}\cap\mathcal{B}}} q_{\boldsymbol{X}_{\mathcal{B}}} q_{\boldsymbol{X}_{(\mathcal{A}\cup\mathcal{B})^c}})$$

The remainder of the theorem is proved by substituting (17) and (18) in the equation for the divergence. ■

C. *Information Geometric Structure of Ingleton-Violating Entropic Vectors & their Distributions*

As identified in §III-A, for a given random vector $\boldsymbol{X} = [X_1, \dots, X_N]$ and given cardinality $\boldsymbol{\mathcal{X}} = \mathcal{X}_1 \times \mathcal{X}_2 \times \dots \times \mathcal{X}_N$, the set $\mathcal{D}$ of joint probability mass functions for $\boldsymbol{X}$ is a $K = \prod_{n=1}^{N} |\mathcal{X}_n| - 1$ dimensional manifold, with m-coordinates $\boldsymbol{\eta}$. Now suppose a particular $K - M$ elements of $\boldsymbol{\eta}$ are all zero with $M < K$. This is a M dimensional face on the boundary of $\mathcal{D}$, and we will label a distribution in it as P_{Matom}.

For $\boldsymbol{X} = [X_4 X_3 X_2 X_1]$, it is easy to prove $\boldsymbol{h}(P_{2atom}) \subseteq \mathcal{S}_4$ for any P_{2atom} and any cardinality $\boldsymbol{\mathcal{X}}$. Our numerical experiments also suggest that $\boldsymbol{h}(P_{3atom}) \subseteq \mathcal{S}_4$, however, this is not true for P_{4atom}, that is to say, $\boldsymbol{h}(P_{4atom}) \cap G_4^{ij} \neq \emptyset$ for some P_{4atom}. Here we focus on the the 4 atom support (0000)(0110)(1010)(1111) that have distributions violate $Ingleton_{34}$. A distribution with this support has been used in [5] by Matúš to show that $\bar{\Gamma}_N^*$ is not a polyhedron for $N \geq 4$, and it is also the 4 atom support utilized with another distribution in [16] associated with the 4 atom conjecture. Next we will use Information Geometry to help understand the 4 atom distributions that violate Ingleton's inequalities.

Denote $\mathcal{D}_{4atom} \subset \mathcal{D}$ the manifold of all probability mass functions for four binary random variables with the support (0000)(0110)(1010)(1111). Parameterize these distributions with the parameters $\eta_1 = \mathbb{P}(\boldsymbol{X} = 0000) = \alpha$, $\eta_2 = \mathbb{P}(\boldsymbol{X} = 0110) = \beta - \alpha$, $\eta_3 = \mathbb{P}(\boldsymbol{X} = 1010) = \gamma - \alpha$, $\eta_4 = \mathbb{P}(\boldsymbol{X} = 1111) = 1 - \sum_{i=1}^{3} \eta_i = 1 + \alpha - \beta - \gamma$, yielding the m-coordinates of $\mathcal{D}_{4atom}$. The associated e-coordinates can be calculated as $\theta_i = \log_2 \frac{\boldsymbol{\eta}_i}{\boldsymbol{\eta}_4}$ for $i = 1, 2, 3$. Now we consider the submanifold $\mathcal{D}_{u1} = \{p_x \in \mathcal{D}_{4atom} | I(X_3; X_4) = 0\}$, by Proposition 1, $\mathcal{D}_{u1}$ is a e-autoparallel submanifold of $\mathcal{D}_{4atom}$. In fact, an equivalent definition is $\mathcal{D}_{u1} = \{p_x \in \mathcal{D}_{4atom} | -\theta_1 + \theta_2 + \theta_3 = 0\}$. Numerical calculation illustrated in the Figure 2 lead to the following proposition, which we have verified numerically.

Proposition 3: Let $\mathcal{D}_{m1} = \{p_x \in \mathcal{D}_{4atom} | Ingleton_{34} = 0\}$, then $\mathcal{D}_{m1}$ is a e-autoparallel submanifold of $\mathcal{D}_{4atom}$ and is parallel with $\mathcal{D}_{u1}$ in e-coordinate, an equivalent definition of $\mathcal{D}_{m1}$ is $\mathcal{D}_{m1} = \{p_x \in \mathcal{D}_{4atom} | -\theta_1 + \theta_2 + \theta_3 = \log_2(\frac{0.5-\alpha_0}{\alpha_0})^2\}$, where α_0 is the solution of $-\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha) = \frac{1+2\alpha}{2}$ in $0 < \alpha < \frac{1}{2}$.

In fact, using this equivalent definition, we can also determine all the distributions in $\mathcal{D}_{4atom}$ that violate $Ingleton_{34} \geqslant 0$ as the submanifold $\mathcal{D}_{Vio} = \{p_x \in \mathcal{D}_{4atom} | -\theta_1 + \theta_2 + \theta_3 < \log_2(\frac{0.5-\alpha_0}{\alpha_0})^2\}$. Because we are dealing with $\mathcal{D}_{4atom}$, a 3 dimensional manifold, we can use a plot to visualize our results in Fig. 2. In Fig. 2, besides $\mathcal{D}_{u1}$ and $\mathcal{D}_{m1}$, we also

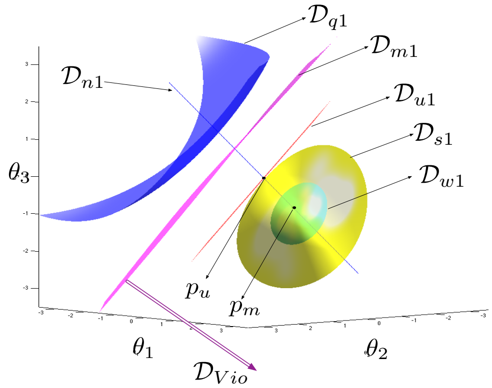

Fig. 2. Manifold $\mathcal{D}_{4atom}$ in θ coordinate

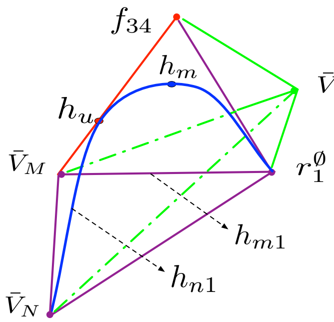

Fig. 3. G_4^{34}: one of the six gaps between $\mathcal{S}_4$ and Γ_4

plot the following submanifolds and points:

$$\begin{aligned}
\mathcal{D}_{q1} &= \{p_x \in \mathcal{D}_{4atom} \mid Ingleton_{34} = 0.1\} \\
\mathcal{D}_{s1} &= \{p_x \in \mathcal{D}_{4atom} \mid Ingleton_{34} = -0.126\} \\
\mathcal{D}_{w1} &= \{p_x \in \mathcal{D}_{4atom} \mid Ingleton_{34} = -0.16\} \\
\mathcal{D}_{n1} &= \{p_x \in \mathcal{D}_{4atom} \mid \beta = \gamma = 0.5\} \\
p_u &= \{p_x \in \mathcal{D}_{4atom} \mid \alpha = 0.25, \beta = \gamma = 0.5\} \\
p_m &= \{p_x \in \mathcal{D}_{4atom} \mid \alpha \approx 0.33, \beta = \gamma = 0.5\}
\end{aligned}$$

As we can see from Fig. 2, $\mathcal{D}_{m1}$ and $\mathcal{D}_{u1}$ are e-autoparallel and parallel to each other. As $Ingleton_{34}$ goes from 0 to negative values, the hyperplane becomes a non-symmetric ellipsoid, and as $Ingleton_{34}$ becomes smaller and smaller, the ellipsoid shrinks, finally shrinking to a single point p_m at $Ingleton_{34} \approx -0.1699$, the point associated with the four atom conjecture in [16]. Also for each e-autoparallel submanifold $\mathcal{D}_{\forall e} \subset \mathcal{D}_{Vio}$ that is parallel to $\mathcal{D}_{m1}$ in e-coordinate, the minimum Ingleton of $\mathcal{D}_{\forall e}$ is achieved at point $\mathcal{D}_{\forall e} \cap \mathcal{D}_{n1}$, where $\mathcal{D}_{n1}$ is the *e-geodesic* in which marginal distribution of X_3 and X_4 are uniform, i.e. $\beta = \gamma = 0.5$.

Now we will map some submanifolds of $\mathcal{D}_{4atom}$ to the entropic region. From Lemma 1, G_4^{34}, one of the six gaps between $\mathcal{S}_4$ and Γ_4 is characterized by extreme rays $\bar{V}_P = (\bar{V}_M, \bar{V}_R, r_1^{\emptyset}, f_{34})$, where $\bar{V}_M = (r_1^{13}, r_1^{14}, r_1^{23}, r_1^{24}, r_2^1, r_2^2)$ and $\bar{V}_R = (r_1^{123}, r_1^{124}, r_1^{134}, r_1^{234}, r_1^3, r_1^4, r_3^{\emptyset})$. In addition, we define $\bar{V}_N = (r_1^1, r_1^2, r_1^{12})$, and use Fig. 3 to help visualize G_4^{34}.

In Fig. 3, f_{34} is one of the six Ingleton-violating extreme ray of Γ_4, $\bar{V}_M$,$\bar{V}_R$ and $r_1^{\emptyset}$ are all extreme rays of $\mathcal{S}_4$ that make $Ingleton_{34} = 0$. Based on the information geometric characterization, the mapping from $\mathcal{D}_{4atom}$ to Γ_4^* is straight forward: the curve $h_{n1} = \boldsymbol{h}(\mathcal{D}_{n1})$, the point $h_u = \boldsymbol{h}(p_u)$ and the point $h_m = \boldsymbol{h}(p_m)$.

IV. Conclusions and Future Work

In this paper, we aimed to suggest that information geometry may provide a useful framework to parameterize this unknown part of the region of entropic vectors. Along these lines, it was first shown that Shannon facets possess certain information geometric properties and Shannon-type inequalities had a natural information geometric structure. Next, we characterized a special type of 4 atom distributions via Information Geometry, and showed there is a natural information geometric explanation for it. The characterization also suggests it is possible to generate better inner bounds by growing in the number of atoms. Future work includes study the distributions with arbitrarily large cardinality that set Ingleton to zero via Information Geometry and extend our work to more variables.

Acknowledgements

The authors gratefully acknowledge NSF support of this research through grants CCF-1016588 and CCF-1053702.

References

[1] Raymond W. Yeung, *Information Theory and Network Coding*. Springer, 2008.

[2] Zhen Zhang and Raymond W. Yeung, "On Characterization of Entropy Function via Information Inequalities," *IEEE Trans. on Information Theory*, vol. 44, no. 4, Jul. 1998.

[3] K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin, "A new class of non-Shannon-type inequalities for entropies," *Communication in Information and Systems*, vol. 2, no. 2, pp. 147–166, December 2002.

[4] R. Dougherty, C. Freiling, and K. Zeger, "Six new non-Shannon information inequalities," in *IEEE International Symposium on Information Theory (ISIT)*, July 2006, pp. 233–236.

[5] František Matúš, "Infinitely Many Information Inequalities," in *IEEE Int. Symp. Information Theory (ISIT)*, Jun. 2007, pp. 41–44.

[6] Babak Hassibi, Sormeh Shadbakht, Matthew Thill, "On Optimal Design of Network Codes," in *Information Theory and Applications*, UCSD, Feb. 2010, presentation.

[7] J. M. Walsh and S. Weber, "A Recursive Construction of the Set of Binary Entropy Vectors," in *2009 Allerton Conference on Communication, Control, and Computing*.

[8] ——, "A Recursive Construction of the Set of Binary Entropy Vectors and Related Inner Bounds for the Entropy Region," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, Oct. 2011.

[9] ——, "Relationships Among Bounds for the Region of Entropic Vectors in Four Variables," in *2010 Allerton Conference on Communication, Control, and Computing*.

[10] Congduan Li, J. M. Walsh, S. Weber, "A computational approach for determining rate regions and codes using entropic vector bounds," in *2012 Allerton Conference on Communication, Control, and Computing*.

[11] Congduan Li, J. Apte, J. M. Walsh, S. Weber, "A new computational approach for determining rate regions and optimal codes for coded networks," in *IEEE International Symposium on Network Coding*, 2013.

[12] D. Hammer, A. Romashschenko, A. Shen, N. Vereshchagin, "Inequalities for Shannon Entropy and Kolmogorov Complexity," *Journal of Computer and System Sciences*, vol. 60, pp. 442–464, 2000.

[13] A. W. Ingleton, "Representation of Matroids," in *Combinatorial Mathematics and its Applications*, D. J. A. Welsh, Ed. San Diego: Academic Press, 1971, pp. 149–167.

[14] F. Matúš and M. Studený, "Conditional Independences among Four Random Variables I," *Combinatorics, Probability and Computing*, no. 4, pp. 269–278, 1995.

[15] S. Amari and H. Nagaoka, *Methods of Information Geometry*. American Mathematical Society Translations of Mathematical Monographs, 2004, vol. 191.

[16] Randall Dougherty, Chris Freiling, Kenneth Zeger, "Non-Shannon Information Inequalities in Four Random Variables," Apr. 2011, arXiv:1104.3602v1.