On Designing Probabilistic Supports to Map the Entropy Region

John MacLaren Walsh, Ph.D. & Alexander Erick Trofimoff
Dept. of Electrical & Computer Engineering
Drexel University
Philadelphia, PA, USA 19104
jmw96@drexel.edu & aet57@drexel.edu

***Abstract*—The boundary of the entropy region has been shown to determine fundamental inequalities and limits in key problems in network coding, streaming, distributed storage, and coded caching. The unknown part of this boundary requires nonlinear constructions, which can, in turn, be parameterized by the support of their underlying probability distributions. Recognizing that the terms in entropy are submodular enables the design of such supports to maximally push out towards this boundary.**

***Index Terms*—entropy region; Ingleton violation**

I. Introduction

From an abstract mathematical perspective, as each of the key quantities in information theory, entropy, mutual information, and conditional mutual information, can be expressed as linear combinations of subset entropies, every linear information inequality, or inequality among different instances of these quantities, specifies a half-space in which the entropy region $\bar{\Gamma}^*_N$ [1], [2] must live, and vice-versa. This in turn implies that characterizing the boundary of $\bar{\Gamma}^*_N$ is equivalent to determining all fundamental inequalities in information theory[1], [3], [2]. Other abstract equivalences have shown that determining it amounts to determining all inequalities in the sizes of N subgroups and their intersections[4], and all inequalities among Kolmogorov complexities[5], and that the faces of $\bar{\Gamma}^*_N$ encode information about implications among conditional independences [6] which are key to the language of graphical models in machine learning. From a more applied perspective, it has been shown the determining the boundary of $\bar{\Gamma}^*_N$ is equivalent to determining the capacity regions of all networks under network coding[7], [8], which in turn have been shown to be key ingredients in building optimal coding protocols for streaming information with low delay over multipath routed networks[9], [10], limits for secret sharing systems[11], as well as fundamental tradeoffs between the amount of information a large distributed information storage system can store and the amount of network traffic it must consume to repair failed disks[12], [13], [14].

The entropy region can be broadly broken into two parts, the part that can be achieved by time-sharing linear codes, and the part that cannot. Here, linear codes construct the discrete random variables (r.v.s) as vectors created by multiplying vectors of r.v.s uniformly distributed over a finite field by matrices with elements drawn from the same field. For collections of exclusively $N \leq 3$ r.v.s, all entropy vectors are achievable by time-sharing linear codes [2]. For $N = 4$ and $N = 5$ r.v.s, the region of entropic vectors (e.v.s) reachable by time-sharing linear codes has been fully determined ([5] and [15], [16], respectively), and is furthermore polyhedral, while the full region of entropic vectors remains unknown for all $N \geq 4$. As such, for $N \geq 4$, and especially for the case of $N = 4$ and $N = 5$ r.v.s, determining $\bar{\Gamma}^*_N$ amounts to determining those e.v.s exclusively reachable with non-linear codes. Of the highest interest in such an endeavor is determining those extreme rays generating points on the boundary of $\bar{\Gamma}^*_N$ requiring such non-linear code based constructions. Once these extremal e.v.s from non-linear codes have been determined, all of the points in $\bar{\Gamma}^*_N$ can be generated through time-sharing their constructions and the extremal linear code constructions.

For $N = 4$ r.v.s, determining whether a entropic vector can be achieved with the time-sharing of linear codes is equivalent to checking whether it obeys Ingleton's inequality [17], [5], and for $N \geq 5$ violation of Ingleton's inequality is a sufficient condition for an entropic point to require nonlinear constructions. These sorts of conditions identify the e.v.s of interest directly through conditions expressed in their entropies, but a method for clearly working these conditions back to structure in the joint probability mass function is yet unknown. Connections with inequalities for the sizes of subgroups of a common finite group [4] has provided one direction for constructions probably sufficient in the limit [18], but the groups must grow arbitrary large in proofs for these constructions to suffice, and extremality of *scores* in various directions is difficult to couple with group properties other than being non-abelian. On the other hand, in practice numerical optimization [19], [20] to map $\bar{\Gamma}^*_N$ consistently exhibits limited support of a non-quasi-uniform nature, motivating a study of arbitrary supports for non-uniform distributions [21]. A particularly simple way of encoding nonlinear dependence between r.v.s in a joint probability mass function is through its probabilistic support, or the collection of N-dimensional outcomes for the N discrete r.v.s which have strictly positive probabilities. **Main Contribution:** This paper advances a nascent structural theory identifying and organizing which probabilistic supports can yield e.v.s in the part of $\bar{\Gamma}^*_N$ reachable exclusively with non-linear codes. Additionally, in order to push the constructions as close to the boundary of $\bar{\Gamma}^*_N$ as possible, structural theory and methods for ranking supports relative to one another in terms of Ingleton score and other measures of code nonlinearity

This material is based upon work support by the National Science Foundation under Grant No. 1812965 & 1421828.

are investigated. The new theorems provide guidance and explanation for observations that were made exclusively with numerical experiments in [21].

II. Encoding Supports with Sets of Set Partitions

A collection of N discrete random variables $\boldsymbol{X} := (X_1, \ldots, X_N)$ taking values in the set $\boldsymbol{\mathcal{X}} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$, is often specified by defining the joint probability mass function (PMF) $p_{\boldsymbol{X}} : \boldsymbol{\mathcal{X}} \to [0,1]$, which assigns to each vector $\boldsymbol{x} = (x_1, \ldots, x_N) \in \boldsymbol{\mathcal{X}}$ its probability $p_{\boldsymbol{X}}(\boldsymbol{x}) := \mathbb{P}[\boldsymbol{X} = \boldsymbol{x}]$. Perhaps the most familiar representation of a probabilistic support is the set of outcomes mapped to strictly positive probabilities under the PMF

$$\boldsymbol{\mathcal{X}}_{>0} = \{\boldsymbol{x} \in \boldsymbol{\mathcal{X}} \,| p_{\boldsymbol{X}}(\boldsymbol{x}) > 0\}. \quad (1)$$

A *k-atom probabilistic support*, or k-atom support for short, is one with $|\boldsymbol{\mathcal{X}}_{>0}| = k$.

The key object under study is the closure of the set of entropy vectors reachable with a given k-atom support

$$\mathcal{H}_N^o(\boldsymbol{\mathcal{X}}_{>0}) = \left\{ \boldsymbol{h}(p_{\boldsymbol{X}}) \;\middle|\; \begin{array}{c} \sum_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}}_{>0}} p_{\boldsymbol{X}}(\boldsymbol{x}) = 1, \\ p_{\boldsymbol{X}}(\boldsymbol{x}) \geq 0 \ \forall \boldsymbol{x} \in \boldsymbol{\mathcal{X}}, \\ p_{\boldsymbol{X}}(\boldsymbol{x}) = 0 \ \forall \boldsymbol{x} \notin \boldsymbol{\mathcal{X}}_{>0} \end{array} \right\}, \quad (2)$$

with the explicit goal of providing a structural characterization of those k-atom supports which can generate e.v.s in the unknown part of $\bar{\Gamma}_N^*$ requiring non-linear codes. In this respect, because $\bar{\Gamma}_N^*$ is invariant to the identification of the labels of the N-r.v.s, but identification of a particular k-atom support can break this symmetry, we consider the set of entropy vectors to be reachable from $\boldsymbol{\mathcal{X}}_{>0}$ to be not only $\mathcal{H}_N^o(\boldsymbol{\mathcal{X}}_{>0})$ but

$$\mathcal{H}_N(\boldsymbol{\mathcal{X}}_{>0}) = \{\pi(\boldsymbol{h}) \,| \boldsymbol{h} \in \mathcal{H}_N^o(\boldsymbol{\xi}), \pi \in \mathbb{S}_N\} \quad (3)$$

where $\pi \in \mathbb{S}_N$, a permutation in the symmetric group of order N, acts on the entropy vector $\boldsymbol{h}$ by permuting the subsets of $\{1, ..., N\}$ that index its dimensions. From the viewpoint of the goal of mapping $\bar{\Gamma}_N^*$, those supports $\boldsymbol{\mathcal{X}}_{>0}$ generating the same sets $\mathcal{H}_N(\boldsymbol{\mathcal{X}}_{>0})$ are functionally equivalent. It is thus of interest to group probabilistic supports into equivalence classes based on those entropy vectors $\mathcal{H}_N(\boldsymbol{\mathcal{X}}_{>0})$ that they can reach.

In this regard, an important observation is that a k-atom support induces a collection of N set partitions which directly determine the entropy vectors it can reach. Indeed, labeling $\boldsymbol{\mathcal{X}}_{>0}$ via $\boldsymbol{\mathcal{X}}_{>0} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ with $\boldsymbol{x}_i = [x_{i,1}, \ldots, x_{i,N}]$, we can associate each random variable X_n with a set partition ξ_n whose blocks are the indices i of those outcomes $\boldsymbol{x}_i \in \boldsymbol{\mathcal{X}}_{>0}$ with $x_{i,n}$ equal to some particular outcome from $\mathcal{X}_n$,

$$\xi_n = \{\{i \in \{1, \ldots, k\} \,| x_{i,n} = x_n\} \,| x_n \in \mathcal{X}_n\}. \quad (4)$$

The meet of these N set partitions $\xi_1, \ldots, \xi_N$ is the set of singletons, i.e.

$$\bigwedge_{n \in \{1,\ldots,N\}} \xi_n = \{\{i\} \,| i \in \{1, \ldots, k\}\}. \quad (5)$$

Together with the vector of non-zero probabilities $\boldsymbol{p} = [p_1, \ldots, p_k]$, $p_i = p_{X_1,\ldots,X_N}(\boldsymbol{x}_i)$, $\boldsymbol{\xi}$ determines

$$H(X_n) = H(\xi_n; \mathbf{p}) := -\sum_{\mathcal{B} \in \xi_n} \left(\sum_{i \in \mathcal{B}} p_i \right) \log \left(\sum_{i \in \mathcal{B}} p_i \right), \quad (6)$$

where we have defined $H(\xi; \boldsymbol{p})$, and,

$$H(\boldsymbol{X}_{\mathcal{A}}) = H\left(\bigwedge_{n \in \mathcal{A}} \xi_n; \mathbf{p} \right), \quad (7)$$

so we can formally express the entropic vector as a function

$$\boldsymbol{h}(\boldsymbol{\xi}; \boldsymbol{p}) = \left[H\left(\bigwedge_{n \in \mathcal{A}} \xi_n; \mathbf{p} \right) | \mathcal{A} \subseteq \{1, \ldots, N\} \right], \quad (8)$$

and the set of entropy vectors reachable with a given equivalence class of probabilistic supports as

$$\mathcal{H}_N(\boldsymbol{\mathcal{X}}_{>0}) = \mathcal{H}_N(\boldsymbol{\xi}) := \{\pi(\boldsymbol{h}(\boldsymbol{\xi}; \mathcal{S}_k)) \,| \pi \in \mathbb{S}_N\}, \quad (9)$$

where $\mathcal{S}_k = \left\{ \boldsymbol{p} \middle| \boldsymbol{p} \geq \boldsymbol{0}, \sum_{i=1}^k p_i = 1 \right\}$. Two different supports $\boldsymbol{\mathcal{X}}_{>0}$ can generate identical collections of N set partitions $\boldsymbol{\xi}$, and thus via (9) exactly the same set of entropy vectors. For this reason, we choose to represent a N-variable k-atom probabilistic support not as a set of vectors of outcomes $\boldsymbol{\mathcal{X}}_{>0}$ but instead as the collection of N set partitions $\boldsymbol{\xi}$. Additionally, as we are adding all permutations of the N random variable labels back in (9), which also form the index ordering $\boldsymbol{\xi}$, we can take $\boldsymbol{\xi}$ to be a multi-set, as two different orderings of the indexing of this multiset will yield identical $\mathcal{H}_N(\boldsymbol{\xi})$s. We further restrict consideration to $\boldsymbol{\xi}$ being a *set* of set partitions, as repetition of the same partition is more properly thought of as an N' variable k-atom support with $N' < N$ together with a repetition of identical r.v.s. Finally, since the entropies will be independent of the ordering of the labels $\{1, ..., k\}$ of the elements in (1), from the orbit of the set of permutations of these labels $\mathbb{S}_k$ (the symmetric group of order k), we select the $\boldsymbol{\xi}$ which is minimum under the natural lexicographic ordering.

Definition 1 (Canonical Minimal k-atom support)**:** A canonical minimal k-atom support is a set of N set partitions $\boldsymbol{\xi}$ of the set $\{1, \ldots, k\}$ whose meet is the singletons

$$\bigwedge_{n \in \{1,\ldots,N\}} \xi_n = \{\{i\} | i \in \{1, \ldots, k\}\} \quad (10)$$

and whom is minimal under lexicographic ordering in its orbit under the action of the group $\mathbb{S}_k$.

In addition to grouping together different supports that are equivalent for the purposes of mapping entropy, one of the benefits or representing supports with set partitions is the ordering among entropies can be determined via refinement.

Lemma 1 (Entropy is ordered under refinement.)**:** Let ξ_1 and ξ_2 be two set partitions of the set $\{1, \ldots, k\}$, and suppose the partition ξ_1 refines the partition ξ_2, denoted by $\xi_1 \preceq \xi_2$. Then, for all probability vectors $\boldsymbol{p} = [p_1, \ldots, p_k]^T$, $p_i \geq 0, i \in \{1, \ldots, k\}$, $\sum_{i=1}^k p_i = 1$,

$$H(\xi_1; \mathbf{p}) \geq H(\xi_2; \mathbf{p}) \quad (11)$$

Proof: Let X_1 be a random variable generating the set partition ξ_1 and X_2 be a random variable generating the set partition ξ_2. Since ξ_1 refines ξ_2, their meet $\xi_1 \wedge \xi_2 = \xi_1$. Thus, $H(X_1, X_2) = H(\xi_1 \wedge \xi_2; \boldsymbol{p}) = H(\xi_1; \boldsymbol{p}) = H(X_1)$. Substituting $H(X_1, X_2) = H(\xi_1; \boldsymbol{p})$ and $H(X_2) = H(\xi_2; \boldsymbol{p})$ into the basic entropy inequality $H(X_1, X_2) \geq H(X_2)$ then shows (11). ■

An even more powerful type of inequality linking the entropies achievable with set partitions invokes entropy submodularity via the meet and join operators as shown in the following lemma.

Lemma 2 (Partition Based Entropy Submodularity)**:** Let ξ_1 and ξ_2 be two set partitions of the set $\{1, \ldots, k\}$. For all probability vectors $\boldsymbol{p} = [p_1, \ldots, p_k]^T$, $p_i \geq 0, i \in \{1, \ldots, k\}$, $\sum_{i=1}^k p_i = 1$,

$$H(\xi_1; \mathbf{p}) + H(\xi_2; \mathbf{p}) \geq H(\xi_1 \wedge \xi_2; \mathbf{p}) + H(\xi_1 \vee \xi_2; \mathbf{p}) \quad (12)$$

Proof: This can in fact be viewed as another instance of the submodularity of entropy as a set function of a collection of r.v.s via a clever identification of r.v.s. Indeed, let X_1 be a random variable inducing the set partition ξ_1, X_2 be a random variable inducing the set partition ξ_2, and Z be a random variable inducing the set partition $\xi_1 \vee \xi_2$. Because $\xi_1 \wedge (\xi_1 \vee \xi_2) = \xi_1$, we can think of $H(\xi_1; \mathbf{p})$ as $H(X_1, Z)$, and similarly because $\xi_2 \wedge (\xi_1 \vee \xi_2) = \xi_2$ we can think of $H(\xi_2; \mathbf{p})$ as $H(X_2, Z)$, substituting these and $H(X_1, X_2, Z) = H(\xi_1 \wedge \xi_2; \boldsymbol{p})$ into (12) we transform it into $H(X_1, Z) + H(X_2, Z) \geq H(Z) + H(X_1, X_2, Z)$. This shows that, once we have identified the random variable Z, this inequality can be recognized as a form of entropy submodularity. ■

An interesting interpretation of this theorem is as stating that the entropy of the common information between two r.v.s is upper bounded by their mutual information, since the join partition can be interpreted as the common information computable individually from either the RVs with partitions ξ_1 and ξ_2 in a manner that is agreed upon.

In fact, a key property of linear codes, which generate the r.v.s X_n as vectors with elements drawn from some finite field $GF(q)$, that resulting from multiplying a vector of independent r.v.s uniformly distributed over $GF(q)$ by some matrix, is that equality is obtained in the inequality (12). It was this fact that enabled Ingleton to prove Ingleton's inequality, and Dougherty Freiling and Zeger derived the region of e.v.s reachable with linear codes for 5 r.v.s also by exploiting this fact.

Not only does entropy expressed in a terms of set partitions obey a submodular relationship, but also the terms within a partition's entropy are submodular, as shown next.

Lemma 3 (Submodularity of entropy terms)**:** For any probability vector $\boldsymbol{p} \in \mathcal{S}_k$, the set function

$$f_{\boldsymbol{p}}(\mathcal{A}) := -\left(\sum_{i \in \mathcal{A}} p_i\right) \log \left(\sum_{i \in \mathcal{A}} p_i\right) \quad (13)$$

is submodular, so that for any $\mathcal{A}, \mathcal{B} \subseteq \{1, \ldots, k\}$

$$f_{\boldsymbol{p}}(\mathcal{A}) + f_{\boldsymbol{p}}(\mathcal{B}) - f_{\boldsymbol{p}}(\mathcal{A} \cap \mathcal{B}) - f_{\boldsymbol{p}}(\mathcal{A} \cup \mathcal{B}) \geq 0. \quad (14)$$

Proof: Define a ternary random variable X_1 whose set partition representation is $\xi_1 = \{\mathcal{A}, \mathcal{B} \setminus \mathcal{A}, (\mathcal{A} \cup \mathcal{B})^c\}$ and a second ternary random variable X_2 whose set partition representation is $\xi_2 = \{\mathcal{B}, \mathcal{A} \setminus \mathcal{B}, (\mathcal{A} \cup \mathcal{B})^c\}$. The join partition is $\xi_1 \vee \xi_2 = \{\mathcal{A} \cup \mathcal{B}, (\mathcal{A} \cup \mathcal{B})^c\}$ and the meet partition is $\xi_1 \wedge \xi_2 = \{\mathcal{A} \cap \mathcal{B}, \mathcal{A} \setminus \mathcal{B}, \mathcal{B} \setminus \mathcal{A}, (\mathcal{A} \cup \mathcal{B})^c\}$. Plugging these partitions into (12), the partition-based submodularity of entropy, and canceling repeated identical terms $f((\mathcal{A} \cup \mathcal{B})^c), f(\mathcal{A} \setminus \mathcal{B}), f(\mathcal{B} \setminus \mathcal{A})$ between positive and negative terms, one obtains the inequality (14). ■

This submodularity of these block components of entropy provide a very powerful tool for determining whether the collection of entropy vectors reachable with a given probabilistic support live in a particular half-space.

Theorem 1: The entropy vectors reached by the probabilistic support $\boldsymbol{\xi}$ will all live in the halfspace $\{\boldsymbol{h} | \boldsymbol{c}^T \boldsymbol{h} \geq 0\}$, so that $\boldsymbol{c}^T \mathcal{H}_N(\boldsymbol{\xi}) \geq 0$, if for all $\pi \in \mathbb{S}_N$,

$$\sum_{\mathcal{B} \subseteq \{1, \ldots, k\}} d_{\mathcal{B}, \pi} H(\boldsymbol{Y}_{\mathcal{B}}) \geq 0 \quad (15)$$

is a balanced Shannon type information inequality in k-variables $Y_1, \ldots, Y_k$, where

$$d_{\mathcal{B}, \pi} := \begin{cases} \sum_{\mathcal{A} \in \mathcal{F}(\boldsymbol{\xi}; \mathcal{B}, \pi)} c_{\mathcal{A}} & \mathcal{B} \in \mathcal{F}(\boldsymbol{\xi}) \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

with $\mathcal{F}(\boldsymbol{\xi}) := \bigcup_{\mathcal{A} \subseteq \{1, \ldots, N\}} \bigwedge_{n \in \mathcal{A}} \xi_n$ and

$$\mathcal{F}(\boldsymbol{\xi}; \mathcal{B}, \pi) := \left\{ \mathcal{A} \subseteq \{1, \ldots, N\} \,\middle|\, \mathcal{B} \in \bigwedge_{n \in \pi(\mathcal{A})} \xi_n \right\}. \quad (17)$$

Proof: Grouping together terms by common blocks, we can rewrite for any entropy vector $\pi(\boldsymbol{h}(\boldsymbol{\xi}; \boldsymbol{p}))$ in $\mathcal{H}_N(\boldsymbol{\xi})$

$$\begin{aligned} \sum_{\mathcal{A} \subseteq \{1, \ldots, N\}} c_{\mathcal{A}} H(\boldsymbol{X}_{\pi(\mathcal{A})}) &= \sum_{\mathcal{A} \subseteq \{1, \ldots, N\}} c_{\mathcal{A}} H\left(\bigwedge_{n \in \pi(\mathcal{A})} \xi_n; \mathbf{p}\right) \\ &= \sum_{\mathcal{A} \subseteq \{1, \ldots, k\}} c_{\mathcal{A}} \sum_{\mathcal{B} \in \bigwedge_{n \in \pi(\mathcal{A})} \xi_n} f_{\boldsymbol{p}}(\mathcal{B}) \\ &= \sum_{\mathcal{B} \in \mathcal{F}(\boldsymbol{\xi})} d_{\mathcal{B}, \pi} f_{\boldsymbol{p}}(\mathcal{B}) \quad (18) \end{aligned}$$

Lemma 3 establishes that $f_{\boldsymbol{p}}(\cdot)$ is a submodular function for every $\boldsymbol{p} \in \mathcal{S}_k$. A balanced Shannon-type inequality is one which can be expressed as a sum of submodularity inequalities. As such, if (15) is a balanced Shannon-type inequality, (18) will be ≥ 0 for all $\boldsymbol{p} \in \mathcal{S}_k$. ■

Checking whether or not an information expression is a balanced Shannon type information inequality, in turn, can be completed by running a linear program. Thus, Thm. 1 enables us to determine and prove whether or not a given support gives only entropy vectors restricted to a certain halfspace, for instance obeying Ingleton's inequality, by running a linear program. Additionally, Thm. 1 also enables us to provide a firm mathematical proof of this fact after running the linear program, removing the possibility for numerical errors leading to incorrect conclusions.

For the purposes of mapping $\bar{\Gamma}^*_N$, Thm. 1's primary use is as a tool to weed out supports that can not be helpful in growing the inner bound, for instance initially for 4-variables, by determining that they can never violate Ingleton's inequality. Any support that survives this weeding out process will have some $\pi \in \mathbb{S}_N$ such that $\mathcal{H}^\pi_N(\boldsymbol{\xi}) := \pi(\mathcal{H}^o_N(\boldsymbol{\xi}))$ can not be proven, via $f_{\boldsymbol{p}}(\cdot)$ submodularity alone, to live in the half space $\{\boldsymbol{h}|\boldsymbol{c}^T\boldsymbol{h} \geq 0\}$ associated with one of the inner bounding inequalities. We say that the pair $(\boldsymbol{\xi}, \pi)$ then form a $\boldsymbol{c}$*-violation candidate*. Attention then turns to comparing the amount that these violation candidates push out in the direction $\boldsymbol{c}$, and submodularity proves to be useful in this regard as well. The following definition makes the notion of "pushing further out in the direction $\boldsymbol{c}$" precise as the definition of $\boldsymbol{c}$-domination.

Definition 2 (Domination of Violation Candidates)**:** The $\boldsymbol{c}$-violation candidate $(\boldsymbol{\xi}, \pi)$ *dominates* the $\boldsymbol{c}$-violation candidate $(\boldsymbol{\xi}', \pi')$ if

$$\sup_{\boldsymbol{h}\in\mathcal{H}^\pi_N(\boldsymbol{\xi})} -\frac{\boldsymbol{c}^T\boldsymbol{h}}{h_{\{1,..,N\}}} \geq \sup_{\boldsymbol{h}'\in\mathcal{H}^{\pi'}_N(\boldsymbol{\xi}')} -\frac{\boldsymbol{c}^T\boldsymbol{h}'}{h'_{\{1,..,N\}}} \quad (19)$$

Submodularity is also a powerful tool for proving $\boldsymbol{c}$-domination as pointed out in the following theorem.

Theorem 2: The $\boldsymbol{c}$-violation candidate $(\boldsymbol{\xi}, \pi)$ *dominates* the $\boldsymbol{c}$-violation candidate $(\boldsymbol{\xi}', \pi')$ if for some $\sigma \in \mathbb{S}_k$

$$\sum_{\mathcal{B}\subseteq\{1,\ldots,k\}} q_{\mathcal{B},\sigma} H(\boldsymbol{Y}_{\mathcal{B}}) \geq 0 \quad (20)$$

is a balanced Shannon-type information inequality in k-variables, where

$$q_{\mathcal{B},\sigma} = d'_{\mathcal{B},\pi',\sigma} - d_{\mathcal{B},\pi} \quad (21)$$

with $d_{\mathcal{B},\pi}$ defined according to (16) applied to $\boldsymbol{\xi}$ and

$$d'_{\mathcal{B},\pi',\sigma} := \begin{cases} \sum_{\mathcal{A}\in\mathcal{F}(\boldsymbol{\xi}';\sigma(\mathcal{B}),\pi')} c_{\mathcal{A}} & \sigma(\mathcal{B}) \in \mathcal{F}(\boldsymbol{\xi}') \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

Proof: Follows from evaluating $\boldsymbol{c}^T\boldsymbol{h}(\boldsymbol{\xi};\boldsymbol{p}) - \boldsymbol{c}^T\pi(\boldsymbol{h}(\boldsymbol{\xi}';\boldsymbol{p}'))$ and identifying $\boldsymbol{p}' = \sigma(\boldsymbol{p})$, yielding identical joint entropies $h_{\{1,\ldots,N\}}$. ∎

III. Theoretical Conditions for Ingleton Violating Supports

One of the most important applications of Thm. 1 is to rule out those supports which always obey Ingleton's inequality, $\text{Ingleton}_{ij} = h_{ij} + h_{ik} + h_{i\ell} + h_{jk} + h_{j\ell} - h_i - h_j - h_{k\ell} - h_{ijk} - h_{ij\ell}$. The following example shows how.

Example 1 (Support Incapable of Violating Ingleton)**:** Consider the support $\boldsymbol{\xi} = \{\xi_1, \xi_2, \xi_3, \xi_4\}$ with $\xi_1 = \{\{1\},\{2\},\{3\},\{4\}\}$, $\xi_2 = \{\{1\},\{2\},\{3,4\}\}$, $\xi_3 = \{\{1\},\{2,3\},\{4\}\}$, and $\xi_4 = \{\{1\},\{2,3,4\}\}$. Theorem 1 can be used to prove that this support is incapable of violating Ingleton's inequality, and is thus restricted to the part of $\bar{\Gamma}^*_N$ reachable with linear codes. Careful evaluation shows

$$\text{Ingleton}_{12}(\boldsymbol{\xi}) = f_{\boldsymbol{p}}(\{2\}) + f_{\boldsymbol{p}}(\{3\}) - f_{\boldsymbol{p}}(\{2,3\}) \quad (23)$$
$$\text{Ingleton}_{13}(\boldsymbol{\xi}) = f_{\boldsymbol{p}}(\{3\}) + f_{\boldsymbol{p}}(\{4\}) - f_{\boldsymbol{p}}(\{3,4\}) \quad (24)$$
$$\text{Ingleton}_{14}(\boldsymbol{\xi}) = f_{\boldsymbol{p}}(\{2,3\}) + f_{\boldsymbol{p}}(\{3,4\}) - f_{\boldsymbol{p}}(\{3\}) - f_{\boldsymbol{p}}(\{2,3,4\}) \quad (25)$$
$$\text{Ingleton}_{23}(\boldsymbol{\xi}) = 0 \quad (26)$$
$$\text{Ingleton}_{24}(\boldsymbol{\xi}) = f_{\boldsymbol{p}}(\{2,3\}) + f_{\boldsymbol{p}}(\{4\}) - f_{\boldsymbol{p}}(\{2,3,4\}) \quad (27)$$
$$\text{Ingleton}_{34}(\boldsymbol{\xi}) = f_{\boldsymbol{p}}(\{2\}) + f_{\boldsymbol{p}}(\{3,4\}) - f_{\boldsymbol{p}}(\{2,3,4\}) \quad (28)$$

Each of these are a form of simple submodular inequality, or equivalently, a balanced Shannon-type inequality, and thus are ≥ 0 no matter what $\boldsymbol{p} \in S_k$ is selected.

Using exact linear programs (e.g. [22]) to provide the submodular constructions, and checking every equivalence class of k-atom supports (see [21] for how to apply [23], [24] to construct these) enables one to obtain Table I col. 0 - 6.

4	75	31	73	29	2	2	1
5	2665	349	2558	313	107	36	2
6	105726	6442	99769	5627	5957	815	14
7	5107735	160365	4763013	136776	344722	23589	53

TABLE I

In terms of the number of atoms in the support (col. 0): The number (henceforth #) of non-isomorphic supports (col. 1), and # of new equivalence classes (forms henceforth) of Ingleton expressions (equation (18)), not reachable with $k' < k$ atoms, that k-atoms supports map to (col. 2), # of supports that submodularity proves can never violate any Ingleton's inequality (col. 3), # of forms submodularity shows are never negative (col. 4), remaining # of supports that can possibly violate Ingleton's inequality (col. 5), # of possibly negative forms (col. 6), and # of Ingleton-dominant forms (col. 7).

IV. Maximizing Ingleton Violation

With those supports that can violate Ingleton (cf. Table I), and thus live in the non-linear part of $\bar{\Gamma}^*_N$, in-hand, one shifts attention to comparing how far their scores can push out in various directions. Thm. 2 was built for this, as the following two examples, and Table I col. 7 demonstrate.

Example 2: Thm. 2 can be used to prove that $\boldsymbol{\xi} = \{\xi_1, \xi_2, \xi_3, \xi_4\}$ with $\xi_1 = \{\{1\},\{2,3,4\}\}$, $\xi_2 = \{\{1,2\},\{3,4\}\}$, $\xi_3 = \{\{1,2,3\},\{4\}\}$, $\xi_4 = \{\{1,3\},\{2,4\}\}$ Ingleton-dominates $\boldsymbol{\xi}'$ with $\xi'_1 = \{\{1\},\{2,3,4\}\}$, $\xi'_2 = \{\{1,2,3\},\{4\}\}$, $\xi'_3 = \{\{1,2,4\},\{3\}\}$, $\xi'_4 = \{\{1,3,4\},\{2\}\}$. For all i,j $\text{Ingleton}_{ij}(\boldsymbol{\xi}')$ is some permutation of

$$-f_{\boldsymbol{p}}(\{1\}) - f_{\boldsymbol{p}}(\{1,2,3\}) - f_{\boldsymbol{p}}(\{1,2,4\}) - f_{\boldsymbol{p}}(\{2\}) - f_{\boldsymbol{p}}(\{3\}) - f_{\boldsymbol{p}}(\{4\}) + f_{\boldsymbol{p}}(\{1,2\}) + f_{\boldsymbol{p}}(\{1,3\}) + f_{\boldsymbol{p}}(\{1,4\}) + f_{\boldsymbol{p}}(\{2,3\}) + f_{\boldsymbol{p}}(\{2,4\})$$

Meanwhile $\boldsymbol{\xi}$ can be proven to obey all forms of Ingletons inequality except $\text{Ingleton}_{13}(\boldsymbol{\xi})$ which takes the form, up to atom permutation,

$$-f_{\boldsymbol{p}}(\{1\}) - f_{\boldsymbol{p}}(\{1,2,3\}) - f_{\boldsymbol{p}}(\{1,2,4\}) - f_{\boldsymbol{p}}(\{2\}) - f_{\boldsymbol{p}}(\{3,4\}) + f_{\boldsymbol{p}}(\{1,2\}) + f_{\boldsymbol{p}}(\{1,3\}) + f_{\boldsymbol{p}}(\{1,4\}) + f_{\boldsymbol{p}}(\{2,3\}) + f_{\boldsymbol{p}}(\{2,4\})$$

Selecting an appropriate σ we see the difference between the Ingleton value of $\boldsymbol{\xi}'$ and that of $\boldsymbol{\xi}$ is of the form $f_{\boldsymbol{p}}(\{3\}) + f_{\boldsymbol{p}}(\{4\}) - f_{\boldsymbol{p}}(\{3,4\})$ which is positive due to the submodularity of $f_{\boldsymbol{p}}(\cdot)$. Since the Ingleton values of $\boldsymbol{\xi}$ are $\leq$ the Ingleton values of $\boldsymbol{\xi}'$ while they will both have, owing to the joint entropy being associated with the meet partition, identical joint entropies h_{1234}, the Ingleton score reached by $\boldsymbol{\xi}$ is provably lower than the Ingleton score reached by $\boldsymbol{\xi}'$.

In fact Thm. 2 can even be used, with appropriate selection, to prove Ingleton-dominance of a k-atom support over a k'-atom support for $k' > k$, as shown in the following example.

Example 3: Thm. 2 implies that the support $\boldsymbol{\xi}$ identified in Example 2 also Ingleton-dominates over the 5-atom support $\boldsymbol{\xi}''$ with $\xi_1'' = \{\{1\},\{2,3\},\{4,5\}\}$, $\xi_2'' = \{\{1,2,3,4\},\{5\}\}$, $\xi_3'' = \{\{1,2,4,5\},\{3\}\}$, and $\xi_4'' = \{\{1,3,5\},\{2,4\}\}$. To see this, first split the atom 2 by replacing it $2 \mapsto \{2,5\}$ in $\boldsymbol{\xi}$ to get the non-minimal $\boldsymbol{\xi}^o$ with Ingleton value

$$\begin{aligned} -f_{\boldsymbol{p}}(\{1\}) - f_{\boldsymbol{p}}(\{1,2,3,5\}) - f_{\boldsymbol{p}}(\{1,3,4\}) - f_{\boldsymbol{p}}(\{2,5\}) \\ -f_{\boldsymbol{p}}(\{3\}) - f_{\boldsymbol{p}}(\{4\}) + f_{\boldsymbol{p}}(\{1,2,5\}) + f_{\boldsymbol{p}}(\{1,3\}) \\ +f_{\boldsymbol{p}}(\{1,4\}) + f_{\boldsymbol{p}}(\{2,3,5\}) + f_{\boldsymbol{p}}(\{3,4\}) \end{aligned}$$

Next, evaluating $\mathrm{Ingleton}_{23}(\boldsymbol{\xi}'')$ we have

$$\begin{aligned} -f_{\boldsymbol{p}}(\{1\}) - f_{\boldsymbol{p}}(\{1,2,3,4\}) - f_{\boldsymbol{p}}(\{1,2,3,5\}) - f_{\boldsymbol{p}}(\{2\}) \\ -f_{\boldsymbol{p}}(\{3\}) - f_{\boldsymbol{p}}(\{4\}) - f_{\boldsymbol{p}}(\{5\}) + f_{\boldsymbol{p}}(\{1,2\}) + f_{\boldsymbol{p}}(\{1,2,3\}) \\ +f_{\boldsymbol{p}}(\{1,4\}) + f_{\boldsymbol{p}}(\{2,4\}) + f_{\boldsymbol{p}}(\{3,4\}) + f_{\boldsymbol{p}}(\{3,5\}) \end{aligned}$$

So, we see that $\mathrm{Ingleton}_{23}(\boldsymbol{\xi}'') - \mathrm{Ingleton}(\boldsymbol{\xi}^o)$ is of the form

$$\begin{aligned} -f_{\boldsymbol{p}}(\{1,2,3,4\}) - f_{\boldsymbol{p}}(\{1,2,5\}) - f_{\boldsymbol{p}}(\{1,3\}) - f_{\boldsymbol{p}}(\{2\}) \\ -f_{\boldsymbol{p}}(\{2,3,5\}) - f_{\boldsymbol{p}}(\{5\}) + f_{\boldsymbol{p}}(\{1,2\}) + f_{\boldsymbol{p}}(\{1,2,3\}) \\ +f_{\boldsymbol{p}}(\{1,3,4\}) + f_{\boldsymbol{p}}(\{3,5\}) + 2f_{\boldsymbol{p}}(\{2,5\}) \end{aligned} \quad (29)$$

which can be expressed as the sum of the following three submodularity inequalities $f_{\boldsymbol{p}}(\{1,2\}) + f_{\boldsymbol{p}}(\{2,5\}) - f_{\boldsymbol{p}}(\{1,2,5\}) - f_{\boldsymbol{p}}(\{2\}) \geq 0$, $f_{\boldsymbol{p}}(\{2,5\}) + f_{\boldsymbol{p}}(\{3,5\}) - f_{\boldsymbol{p}}(\{2,3,5\}) - f_{\boldsymbol{p}}(\{5\}) \geq 0$, and $f_{\boldsymbol{p}}(\{1,2,3\}) + f_{\boldsymbol{p}}(\{1,3,4\}) - f_{\boldsymbol{p}}(\{1,2,3,4\}) - f_{\boldsymbol{p}}(\{1,3\}) \geq 0$. Thus, $\mathrm{Ingleton}_{23}(\boldsymbol{\xi}'') - \mathrm{Ingleton}(\boldsymbol{\xi}^o) \geq 0$ for all values of $\boldsymbol{p}$. Next, we observe that $\mathrm{Ingleton}(\boldsymbol{\xi}^o) = \mathrm{Ingleton}(\boldsymbol{\xi})$ and $h_{1234}(\boldsymbol{\xi}) = h_{1234}(\boldsymbol{\xi}^o)$, while the refinement inequality (11) shows that $h_{1234}(\boldsymbol{\xi}^o) \leq h_{1234}(\boldsymbol{\xi}'')$. Thus, we see that the Ingleton score reachable with $\boldsymbol{\xi}$ *provably* outpowers those reachable with $\boldsymbol{\xi}''$.

In order to map $\bar{\Gamma}_N^*$, extremizing other cost functions while violating Ingleton is also required. Observe that Thm. 2 is equally suited to this purpose, and is equally amenable to proofs constructed via linear programs.

V. Conclusions

By recognizing submodularity of the components of entropy when restricted to a certain probabilistic support, two theorems were constructed that help determine which supports are best for mapping the boundary of the entropy region.

References

[1] Raymond W. Yeung, "A Framework for Linear Information Inequalities," *IEEE Trans. on Information Theory*, vol. 43, no. 6, Nov. 1997.

[2] ——, *Information Theory and Network Coding*. Springer, 2008.

[3] 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.

[4] T. Chan and R. Yeung, "On a relation between information inequalities and group theory," *IEEE Trans. on Information Theory*, vol. 48, no. 7, pp. 1992 – 1995, Jul. 2002.

[5] 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.

[6] F. Matúš, "Conditional Independences among Four Random Variables III: Final Conclusion," *Combinatorics, Probability and Computing*, vol. 8, no. 3, pp. 269–276, May 1999.

[7] Xijin Yan, Raymond W. Yeung, and Zhen Zhang, "An Implicit Characterization of the Achievable Rate Region for Acyclic Multisource Multisink Network Coding," *IEEE Trans. on Information Theory*, vol. 58, no. 9, pp. 5625–5639, Sep. 2012.

[8] T. Chan and A. Grant, "Dualities between entropy functions and network codes," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4470–4487, October 2008.

[9] J. M. Walsh, S. P. Weber, J. C. de Oliveira, A. Eryilmaz, M. Médard, "Trading Rate for Delay at the Application and Transport Layers (Guest Editorial)," *IEEE J. Sel. A. Comm.*, vol. 29, no. 5, pp. 913–915, May 2011.

[10] J. M. Walsh, S. Weber, and C. wa Maina, "Optimal Rate Delay Tradeoffs and Delay Mitigating Codes for Multipath Routed and Network Coded Networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5491–5510, Dec. 2009.

[11] L. Csirmaz, "The size of share must be large," *J. Cryptology*, vol. 10, no. 4, pp. 223–231, 1997.

[12] A. Dimakis, P. Godfrey, Y. Wu, M. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4539–4551, Sept 2010.

[13] C. Tian, "Characterizing the rate region of the (4,3,3) exact-repair regenerating codes," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 5, pp. 967–975, May 2014.

[14] Fangwei Ye, Kenneth W. Shum, and Raymond W. Yeung, "The rate region for secure distributed storage systems," *IEEE Trans. Inform. Theory*, vol. 63, no. 11, pp. 7038–7051, Nov. 2017.

[15] Randall Dougherty, Chris Freiling, Kenneth Zeger, "Linear rank inequalities on five or more variables," submitted to SIAM J. Discrete Math. arXiv:0910.0284.

[16] Ryan Kinser, "New Inequalities for Subspace Arrangements," *J. of Comb. Theory Ser. A*, vol. 188, no. 1, pp. 152–161, Jan. 2011.

[17] 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.

[18] Wei Mao, Matthew Thill, and Babak Hassibi, "On group network codes: Ingleton-bound violations and independent sources," in *IEEE International Symposium on Information Theory (ISIT)*, Jun. 2010.

[19] F. Matúš and L. Csirmaz, "Entropy region and convolution," Oct. 2013, arXiv:1310.5957v1.

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

[21] Y. Liu and J. M. Walsh, "Non-Isomorphic Distribution Supports for Calculating Entropic Vectors," in *53rd Annual Allerton Conference on Communication, Control, and Computing*, Oct. 2015. [Online]. Available: http://www.ece.drexel.edu/walsh/Liu_Allerton_2015.pdf

[22] David Applegate, William Cook, Sanjeeb Dash, Daniel Espinoza, "Exact Solutions to Linear Programming Problems," *Operations Research Letters*, vol. 35, no. 6, pp. 693–699, Nov. 2007.

[23] A. Betten, M. Braun, H. Fripertinger, A. Kerber, A. Kohnert, and A. Wassermann, *Error-Correcting Linear Codes: Classification by Isometry and Applications*. Springer, 2006.

[24] Bernd Schmalz, "t-Designs zu vorgegebener Automorphismengruppe," *Bayreuther Mathematische Schriften*, no. 41, pp. 1–164, 1992, Dissertation, Universität Bayreuth, Bayreuth.