Kybernetika

Rostislav Matveev; Jacobus W. Portegies

Tropical probability theory and an application to the entropic cone

Kybernetika, Vol. 56 (2020), No. 6, 1133–1153

Persistent URL: http://dml.cz/dmlcz/148503

KYBERNETIKA — VOLUME *56* (2020), NUMBER 6, PAGES 1133–1153

TROPICAL PROBABILITY THEORY AND AN APPLICATION TO THE ENTROPIC CONE

Rostislav Matveev and Jacobus W. Portegies

In a series of articles, we have been developing a theory of *tropical diagrams of probability spaces*, expecting it to be useful for information optimization problems in information theory and artificial intelligence. In this article, we give a summary of our work so far and apply the theory to derive a dimension-reduction statement about the shape of the entropic cone.

Keywords: tropical probability, entropic cone, non-Shannon inequality

Classification: 94A17, 94A24

1. INTRODUCTION

With the aim of developing a systematic approach to an important class of problems in information theory and artificial intelligence, we started in [16] the development of a theory of *tropical diagrams of probability spaces*. One of our intended applications is to characterize, or at least derive important properties of, the *entropic cone*: an important open problem in information theory, to which Fero Matúš made invaluable contributions.

In this article, we give a summary of our work on tropical diagrams so far and apply the technology to derive a statement about the entropic cone.

We briefly recall the definition of the entropic cone. Given a collection of k random variables $\mathsf{X}_1, \dots, \mathsf{X}_k$ and a subset $I \subset \{1, \dots, k\}$, we can record the entropy of the joint random variable X_I. This way, we get a function from $\mathbf{\Lambda}_k$ to $\mathbb{R}$, where $\mathbf{\Lambda}_k$ denotes the set of nonempty subsets of $\{1, \dots, k\}$. We interpret the function as an element of the vector space $\mathbb{R}^{\mathbf{\Lambda}_k}$ and call it the entropy vector of the random variables $\mathsf{X}_1, \dots, \mathsf{X}_k$. In general, we say that a vector in $\mathbb{R}^{\mathbf{\Lambda}_k}$ is entropically representable if it is the entropy vector of some collection of random variables $\mathsf{X}_1, \dots, \mathsf{X}_k$.

The entropic cone is the closure of the set of entropically representable vectors. Entropies of random variables, conditional entropies, mutual information and conditional mutual information are all nonnegative. These conditions are called the Shannon inequalities. For $k \leq 3$, the Shannon inequalities completely describe the entropic cone, but for $k \geq 4$ the situation is much more complicated. Zhang and Yeung showed that the entropic cone and the submodular cone (i. e. the cone cut out by Shannon inequalities) are different [25], by identifying a non-Shannon inequality satisfied by all entropically

DOI: 10.14736/kyb-2020-6-1133

representable vectors. Subsequently, more non-Shannon inequalities were discovered, e. g. [5, 15].

In [12] Matúš discovered several infinite families of linear inequalities satisfied by the entropic cone and used it in a clever way to show that the cone is *not polyhedral*. Other infinite families of *information inequalities* were found in [6] as well as many (more than 200) sporadic inequalities.

In the case of four random variables, the entropic cone is a closed convex cone in $\mathbb{R}^{15}$. Using techniques developed in [16, 17, 18, 19], we show how the dimension of the problem of determining the entropic cone could be reduced from 15 to 11.

During our work on the development of tropical probability we were greatly influenced by the article of Gromov [7] and by numerous discussions with Fero Matúš as well as by his published work, such as [10, 12, 14].

The idea of tropicalization of probability theory is roughly the following. We define the entropic distance $\mathbf{k}(\mathsf{X},\mathsf{Y})$ between two tuples of random variables $\mathsf{X} = (\mathsf{X}_1,\dots,\mathsf{X}_n)$ and $\mathsf{Y} = (\mathsf{Y}_1,\dots,\mathsf{Y}_n)$ as the infimum of the expression

$$\sum_{J\subset\{1,\dots,n\}} \big(\,\mathsf{Ent}(\mathsf{X}_I|\mathsf{Y}_I) + \mathsf{Ent}(\mathsf{Y}_I|\mathsf{X}_I)\big)$$

over all choices of joint distributions $(\mathsf{X}_1,\dots,\mathsf{X}_n,\mathsf{Y}_1,\dots,\mathsf{Y}_n)$.

This is indeed a pseudo-distance, that vanishes only on pairs of isomorphic tuples. Further, we define the asymptotic distance

$$\boldsymbol{\kappa}(\mathsf{X},\mathsf{Y}) := \lim_{k\to\infty}\frac{1}{k}\,\mathbf{k}(\mathsf{X}^k,\mathsf{Y}^k)$$

where X^k stand for the k-fold product of i.i.d. copies of X. The asymptotic distance may vanish on non-isomorphic tuples, and we call such tuples of random variables *asymptotically equivalent*. While two random variables are asymptotically equivalent if and only if they have the same entropy, for $n \geq 2$ the space of asymptotic equivalence classes is infinite-dimensional.

Asymptotically equivalent tuples have the same entropy profiles. However, asymptotic equivalence is a much finer relation than equality of entropy profiles. Asymptotically equivalent tuples have the same solutions to entropy optimization problems, such as Wyner Common Information, [22], or information decomposition quantities introduced in [3].

Two key results from [16] and [17] will be used in the present article. The first one is the tropical version of the Asymptotic Equipartition Property, generalizing the theorem of Chan and Yeung from [4]. While the theorem in [4] asserts that for any tuple of random variables its entropy profile up to scaling can be approximated by the entropy profiles of homogeneous tuples (logarithms of indexes of subgroups and their intersection in some group), we show that any tuple can be approximated by a sequence of homogeneous tuples up to asymptotic equivalence, in particular, preserving the solutions of entropy optimization problems.

In [17], we also generalize the Ahlswede–Körner Theorem, see [1, 2], to the tropical setting. While the original Ahlswede–Körner theorem asserts the equality of entropy profiles of the results of a certain construction, we provide an analogous construction that preserves asymptotic equivalence.

In both generalizations our results are ultimately equalities in infinite-dimensional spaces versus finite collections of equalities of real numbers in the original statements.

The same types of generalizations are possible with the Slepian–Wolf Lemma [20] and Matúš' convolution construction [11]. These will be addressed in future articles.

2. TROPICAL DIAGRAMS

The language of random variables was introduced by Fréchet, Kolmogorov and others, so that joint distributions are automatically defined. For our purposes, this is not a convenient setup, as we often need to vary the joint distributions. That's why we use a different language of diagrams of probability spaces, which we introduce below. A more detailed discussion and proofs of the statements below can be found in [16, 19] and [17].

2.1. Probability spaces

For the purposes of this article, a *probability space* is a set with a probability measure on it which is supported on a finite subset. A *reduction* from one probability space to another is an equivalence class of measure-preserving maps, where two maps are considered equivalent if they coincide on a set of full measure. Note that the target space of a random variable taking values in a finite set is a probability space according to this definition.

The *tensor product* $X \otimes Y$ of two probability spaces X and Y is the independent product.

2.2. Diagrams of probability spaces

We will consider commutative diagrams of probability spaces and reductions, such as a *two-fan* and a *diamond*, pictured below

$$\begin{array}{ccccc} & & Z & & \\ & \swarrow & & \searrow & \\ X & & & & Y \end{array} \qquad \begin{array}{ccccc} & \swarrow & Z & \searrow & \\ X & & & & Y \\ & \searrow & W & \swarrow & \end{array} . \tag{2.1}$$

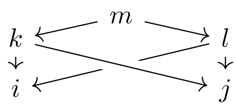
In these diagrams, X, Y, Z and W are probability spaces, and the arrows are reductions. To speak about general diagrams, we will need to specify the arrangement of probability spaces and reductions, i. e. we need to record the underlying combinatorial structure. There are several, equivalent, ways to do so: using a poset category, a partially ordered set (poset), or a directed acyclic graph (DAG) with some additional properties as described below. From our perspective, the language of categories is most convenient for this purpose, but it may not be as familiar as the other two concepts. That is why we will provide a dictionary to convert from one setup to the other. For an easy introduction to categories, functors and natural transformations the reader is refered to the first chapter in the book [9].

2.2.1. Categories, posets and DAGs

A *poset category* is a finite category $\mathbf{G}$ such that for any pair of objects $i, j \in \mathbf{G}$ there is at most one morphism either way. We will require the poset categories used for indexing diagrams to have an additional property, that we describe below after introducing some convenient terminology.

For a morphism $i \to j$ in $\mathbf{G}$, the object i will be called an *ancestor* of j and object j will be called a *descendant* of i.

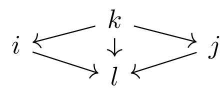
An *indexing category* $\mathbf{G}$ is a finite poset category such that for any two objects $i, j \in \mathbf{G}$ there exists a *minimal common ancestor* $\hat{\imath}$, that is an object $\hat{\imath}$ wich is an ancestor to both i and j and such that any other common ancestor of i and j is also an ancestor of $\hat{\imath}$. For an interested reader an example of a poset category that fails this property is shown below.



Given a poset $(P, \geq)$ such that any subset in P has a supremum (a least common upper bound), one can construct an indexing category $\mathbf{G}$, having as objects the points in the poset, and a unique morphism $i \to j$ for any pair $i \geq j$.

Starting with a DAG, one can construct a poset category by taking the transitive closure of the DAG and considering vertices as objects and arrows as morphisms. The translation of the defining property of indexing categories is straightforward in the DAG language.

A *fan* in a category is a pair of morphisms with the same domain $(i \leftarrow k \to j)$. Such a fan is called *minimal* if whenever it is included in a commutative diagram



the verical arrow $k \to l$ must be an isomorphism.

Indexing categories have the following useful properties, which are elementary to establish. First, for any pair of objects i, j in an indexing category $\mathbf{G}$, there exists a *unique minimal fan* in $\mathbf{G}$ with target objects i and j. Secondly, any indexing category is *initial*, i. e. it has an *initial object* that is an ancestor to any other object in $\mathbf{G}$.

2.2.2. Diagrams

A *diagram of probability spaces* is a functor $\mathcal{X}$ from an indexing category $\mathbf{G} = \{i; \gamma_{ij}\}$ to the category of probability spaces. Essentially, this means that given an indexing category, poset or DAG, we get a $\mathbf{G}$-diagram of probability spaces $\mathcal{X} = \{X_i; \chi_{ij}\}$ by assigning to each object/vertex i a probability space X_i and to each morphism/arrow γ_{ij} a reduction χ_{ij}, requiring that the resulting diagram commutes. We denote the set of all $\mathbf{G}$-diagrams of probability spaces by $\mathbf{Prob}\langle\mathbf{G}\rangle$.

2.2.3. Full diagrams and random variables

Important examples of diagrams are $\mathbf{\Lambda}_n$-diagrams, which we call *full diagrams*, where $\mathbf{\Lambda}_n$ is the poset of non-empty subsets of the set $\{1, \ldots, n\}$ ordered by inclusion. Given an n-tuple of random variables $(\mathsf{X}_1, \ldots, \mathsf{X}_n)$ we can construct a $\mathbf{\Lambda}_n$-diagram

$$\langle \mathsf{X}_1, \ldots, \mathsf{X}_n \rangle := \{X_I; \chi_{IJ}\}$$

by setting X_I equal to the target space of $(X_i : i \in I)$ with the induced distribution and χ_{IJ} equal to the natural projections. On the other hand, starting with a $\mathbf{\Lambda}_n$-diagram we can construct an n-tuple of random variables as reductions from the initial space to n terminal spaces. Diagrams of combinatorial type $\mathbf{\Lambda}_2$ are two-fans, pictured above in (2.1), and $\mathbf{\Lambda}_1$-diagrams are single probability spaces.

2.2.4. Diagrams of diagrams

A *reduction* $\rho : \mathcal{X} \to \mathcal{Y}$ from a $\mathbf{G}$-diagram $\mathcal{X}$ to a $\mathbf{G}$-diagram $\mathcal{Y}$ is a natural transformation (in the sense of category theory) from (the functor) $\mathcal{X}$ to $\mathcal{Y}$. It amounts to specifying a reduction $\rho_i : X_i \to Y_i$ for every i, such that the diagram obtained from $\mathcal{X}$, $\mathcal{Y}$ and the ρ_i's is commutative. Thus, $\mathbf{Prob}\langle \mathbf{G} \rangle$ is itself a category.

Hence, we can also construct diagrams of diagrams. Most important for us are two-fans of $\mathbf{G}$ diagrams,

$$\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y}$$

where $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are $\mathbf{G}$-diagrams, and the arrows are reductions of diagrams. For the space of $\mathbf{H}$-diagrams of $\mathbf{G}$-diagrams we will use the notation $\mathbf{Prob}\langle \mathbf{G} \rangle \langle \mathbf{H} \rangle = \mathbf{Prob}\langle \mathbf{G}, \mathbf{H} \rangle$. Note that an $\mathbf{H}$-diagram of $\mathbf{G}$-diagrams can equivalently be interpreted as a $\mathbf{G}$-diagram of $\mathbf{H}$-diagrams.

2.2.5. Minimal diagrams

A $\mathbf{G}$-diagram $\mathcal{X}$ is called *minimal* if it maps minimal fans in $\mathbf{G}$ to minimal fans in the target category.

A *minimization* of a two-fan $\hat{\mathcal{Z}} := (\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y})$ of either of probability spaces or diagrams is the minimal fan $\hat{\mathcal{C}} = (\mathcal{A} \leftarrow \mathcal{C} \rightarrow \mathcal{B})$ and a reduction

$$\begin{array}{ccccc} \mathcal{X} & \longleftarrow & \mathcal{Z} & \longrightarrow & \mathcal{Y} \\ \downarrow{\scriptstyle f} & & \downarrow{\scriptstyle h} & & \downarrow{\scriptstyle g} \\ \mathcal{A} & \longleftarrow & \mathcal{C} & \longrightarrow & \mathcal{B} \end{array} \tag{2.2}$$

such that f and g are isomorphisms.

It is shown in [16, Proposition 2.1] that a minimization always exists and is unique up to isomorphism.

We will also refer to a minimal two-fan with $\mathcal{X}$ and $\mathcal{Y}$ as targets, as a coupling between $\mathcal{X}$ and $\mathcal{Y}$.

2.2.6. Tensor product and conditioning

The *tensor product* of two $\mathbf{G}$-diagrams $\mathcal{X} = \{X_i; \chi_{ij}\}$ and $\mathcal{Y} = \{Y_i; \upsilon_{ij}\}$ is $\mathcal{X} \otimes \mathcal{Y} := \{X_i \otimes Y_i; \chi_{ij} \times \upsilon_{ij}\}$.

If $\mathcal{X}$ is a $\mathbf{G}$-diagram, and U is a probability space in $\mathcal{X}$, then the whole diagram $\mathcal{X}$ can be conditioned on an outcome $u \in U$ with positive weight. We denote the conditioned diagram by $\mathcal{X}|u$. A precise definition of this construction is given in [16, Section 2.8].

2.3. The intrinsic and asymptotic entropy distances

For a given a $\mathbf{G}$-diagram $\mathcal{X}$ we may evaluate the entropies of the individual probability spaces, which gives a map

$$\mathsf{Ent}_* : \mathbf{Prob}\langle\mathbf{G}\rangle \to \mathbb{R}^{\mathbf{G}}$$

where the target space $\mathbb{R}^{\mathbf{G}}$ is the vector space of all real-valued functions on the set of objects in $\mathbf{G}$, equipped with the ℓ^1-norm. The entropy is a homomorphism in the sense that $\mathsf{Ent}_*(\mathcal{X} \otimes \mathcal{Y}) = \mathsf{Ent}_*(\mathcal{X}) + \mathsf{Ent}_*(\mathcal{Y})$.

Given a two-fan of $\mathbf{G}$-diagrams $\mathcal{K} = (\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y})$, the *entropy distance* between $\mathcal{X}$ and $\mathcal{Y}$ is defined by

$$\mathrm{kd}(\mathcal{K}) := \| \mathsf{Ent}_*(\mathcal{Z}) - \mathsf{Ent}_*(\mathcal{X}) \|_1 + \| \mathsf{Ent}_*(\mathcal{Z}) - \mathsf{Ent}_*(\mathcal{Y}) \|_1.$$

We use the entropy distance as a measure of deviation of a fan $\mathcal{K}$ from being an isomorphism between $\mathcal{X}$ and $\mathcal{Y}$. Indeed, the entropy distance $\mathrm{kd}(\mathcal{K})$ vanishes if and only if both arrows in $\mathcal{K}$ are isomorphims.

We obtain the *intrinsic entropy distance* $\mathbf{k}(\mathcal{X}, \mathcal{Y})$ between two $\mathbf{G}$-diagrams $\mathcal{X}$ and $\mathcal{Y}$ by taking an infimum over all couplings between $\mathcal{X}$ and $\mathcal{Y}$

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) := \inf\{\mathrm{kd}(\mathcal{K}) : \ \mathcal{K} \text{ coupling between } \mathcal{X} \text{ and } \mathcal{Y}\}.$$

For probability spaces, the intrinsic entropy distance was introduced in [8, 21].

We also define the *asymptotic entropy distance* $\boldsymbol{\kappa}(\mathcal{X}, \mathcal{Y})$ by

$$\boldsymbol{\kappa}(\mathcal{X}, \mathcal{Y}) := \lim_{n \to \infty} \frac{1}{n} \mathbf{k}(\mathcal{X}^n, \mathcal{Y}^n)$$

where $\mathcal{X}^n$ denotes the n-fold independent product of $\mathcal{X}$.

Both $\mathbf{k}$ and $\boldsymbol{\kappa}$ are pseudo-distance functions in that they satisfy all axioms of a distance function, except that they may vanish on pairs of non-identical points, see [16] and [19] for the proofs.

2.4. Tropical diagrams

Tropical objects, as for instance encountered in algebraic geometry, are, roughly speaking, divergent sequences of classical objects (e. g. algebraic varieties), renormalized by viewing them on a log scale with increasing base.

The space of tropical diagrams of probability spaces is defined along similar lines: it consists of certain divergent sequences of diagrams and is endowed with an asymptotic entropy distance, thus achieving a similar renormalization.

Our description below is extremely brief. For details and proofs, we refer the reader to [19].

2.4.1. Quasi-linear sequences

We define the *linear sequence* generated by a $\mathbf{G}$-diagram $\mathcal{X}$ as the sequence $\overrightarrow{\mathcal{X}} := (\mathcal{X}^n : n \in \mathbb{N}_0)$ and we define the distance between two such sequences by

$$\boldsymbol{\kappa}\Big(\overrightarrow{\mathcal{X}}, \overrightarrow{\mathcal{Y}}\Big) := \lim_{n\to\infty} \frac{1}{n} \mathbf{k}\Big(\mathcal{X}^n, \mathcal{Y}^n\Big).$$

Tropical diagrams of probability spaces will be sequences that are almost linear, so that it allows us to define algebraic operations on them, and establish completeness of the space of all tropical diagrams.

We call a sequence $[\mathcal{X}] := (\mathcal{X}(n) : n \in \mathbb{N}_0)$ *quasi-linear* if for every $m, n \in \mathbb{N}$,

$$\boldsymbol{\kappa}\Big(\mathcal{X}(m+n), \mathcal{X}(m) \otimes \mathcal{X}(n)\Big) \le C(m+n)^{3/4}. \tag{2.3}$$

The constant 3/4 above is an artificial choice. As explained in [19] any exponent in the interval $(1/2, 1)$ leads to an equivalent definition of the tropical diagrams. We also define the distance between two quasi-linear sequences by

$$\boldsymbol{\kappa}\Big([\mathcal{X}], [\mathcal{Y}]\Big) := \lim_{n\to\infty} \frac{1}{n} \boldsymbol{\kappa}\Big(\mathcal{X}(n), \mathcal{Y}(n)\Big)$$

and denote by $\mathbf{Prob}[\mathbf{G}]$ the (pseudo-)metric space of all quasi-linear sequences endowed with $\boldsymbol{\kappa}$. Two quasi-linear sequences will be called asymptotically equivalent if they are zero distance apart. Equivalence classes of quasi-linear sequences will be called *tropical diagrams of probability spaces*. In our discussions we will sometimes be sloppy, and make no distinction between equivalence classes and their representatives (quasi-linear sequences). This is harmless, as operations we consider are all $\boldsymbol{\kappa}$-continuous and preserve asymptotic equivalence.

The sum of two sequences is defined as element-wise tensor product, and multiplication by a scalar $\lambda \ge 0$ is defined by

$$\lambda \cdot [\mathcal{X}] := (\mathcal{X}(\lfloor \lambda \cdot n \rfloor) : n \in \mathbb{N}_0).$$

The addition and scalar multiplication satisfy the usual associative, commutative and distributive laws up to asymptotic equivalence. Therefore, the space $\mathbf{Prob}[\mathbf{G}]$ has the structure of a convex cone.

The asymptotic distance $\boldsymbol{\kappa}$ is 1-homogeneous

$$\boldsymbol{\kappa}\Big(\lambda \cdot [\mathcal{X}], \lambda \cdot [\mathcal{Y}]\Big) = \lambda\, \boldsymbol{\kappa}\Big([\mathcal{X}], [\mathcal{Y}]\Big)$$

and translation-invariant,

$$\boldsymbol{\kappa}\Big([\mathcal{X}] + [\mathcal{Z}], [\mathcal{Y}] + [\mathcal{Z}]\Big) = \boldsymbol{\kappa}\Big([\mathcal{X}], [\mathcal{Y}]\Big).$$

We show in [19] that the space $\mathbf{Prob}[\mathbf{G}]$ is complete. Together with the algebraic structure, it implies that $\mathbf{Prob}[\mathbf{G}]$ is a closed convex cone in some (generally infinite-dimensional) Banach space B. We call elements in the dual space B^* *entropic quantities*. The entropy functional defined by

$$\mathsf{Ent}_*([\mathcal{X}]) := \lim_{n\to\infty} \frac{1}{n} \mathsf{Ent}_*(\mathcal{X}(n)) \tag{2.4}$$

yields an example of such dual elements.

Less trivial examples are the Wyner Common Information, [22],

$$C(X \leftarrow Z \rightarrow Y) := \inf_W \{I(Z:W) : X \perp\!\!\!\perp Y | W\}$$

or information decomposition quantities defined in [3].

2.5. Homogeneous diagrams and asymptotic equipartition property

We call a diagram of probability spaces $\mathcal{X}$ *homogeneous* if its automorphism group $\mathrm{Aut}(\mathcal{X})$ acts transitively on every space in $\mathcal{X}$.

Examples of homogeneous diagrams can be constructed in the following way. A **G**-*diagram of groups* is a pair consisting of an ambient finite group G and a **G**-diagram of subgroups of G, $(H_i : i \in \mathbf{G})$, where the arrows are inclusions. Starting with a **G**-diagram of groups, we construct a **G**-diagram of probability spaces $\mathcal{X} := \{X_i; \chi_{ij}\}$ by setting $X_i := G/H_i$ with the uniform measure and defining the reduction χ_{ij} to be the natural projection $\chi_{ij} : G/H_i \rightarrow G/H_j$ whenever $H_i \subset H_j$. In fact, every homogeneous diagram arises in this way [16, Section 2.7.1]. We call a homogeneous diagram *Abelian* if it can be constructed in this way with Abelian G.

Starting with a diagram of groups $\{G; H_i : i \in \mathbf{G}\}$ the resulting homogeneous diagram will be minimal if and only if for any $i, j \in \mathbf{G}$ there exists $k \in \mathbf{G}$, such that $H_k = H_i \cap H_j$. If a diagram of groups satisfies this property, we will call it *minimal* as well. When **G** is a full indexing category $\mathbf{G} = \mathbf{\Lambda}_n$, the description of minimal diagrams of groups is especially simple: One needs to specify only the terminal groups, others being obtained by appropriate intersections. We will write $(G;\ H_1, \ldots, H_n)$ for such minimal $\mathbf{\Lambda}_n$-diagram of groups.

The space of quasi-linear sequences of homogeneous **G**-diagrams will be denoted by $\mathbf{Prob}[\mathbf{G}]_{\mathsf{h}}$, the subspace of sequences of Abelian **G**-diagrams by $\mathbf{Prob}[\mathbf{G}]_{\mathsf{Ab}}$, and the space quasi-linear sequences of minimal **G**-diagrams by $\mathbf{Prob}[\mathbf{G}]_{\mathsf{m}}$.

The Asymptotic Equipartition Property for diagrams of probability spaces that we have shown in [16, Theorem 6.1] essentially says that any linear sequences of diagrams of probability spaces is asymptotically equivalent to a quasi-linear sequence of homogeneous diagrams. Together with the density of linear sequences in $\mathbf{Prob}[\mathbf{G}]$, [19, Theorem 5.2], it implies the following theorem.

Theorem 2.1. For any indexing category **G** the spaces $\mathbf{Prob}[\mathbf{G}]_{\mathsf{h}}$ and $\mathbf{Prob}[\mathbf{G}]_{\mathsf{h,m}}$ are dense in $\mathbf{Prob}[\mathbf{G}]$ and $\mathbf{Prob}[\mathbf{G}]_{\mathsf{m}}$, respectively.

This theorem generalizes results of Chan and Yeung in [4], where they show that given a diagram of probability spaces there is a sequence of homogeneous diagrams such that rescaled entropies of spaces in the homogeneous diagrams converge to entropies of the given one. Theorem 2.1, however, is stronger. It implies, in particular, that any entropic quantity such as, for example, the optimum value in some entropy optimization problem, will be approximately the same in the given diagram of probability spaces and in the approximating homogeneous diagram, to any level of precision. Theorem 2.1 gives an equality in an infinite dimensional space as opposed to finite dimensional equalities in the main theorem in [4].

Intuitively, according to the theorem above, one may think of a tropical diagram as a homogeneous diagram of very large probability spaces. Thus, whenever one wants to evaluate a continuous linear functional on a diagram $\mathcal{X}$, one may assume that it is homogeneous and consists of arbitrarily large spaces. We take advantage of this point of view in the next section.

As a trivial, but enlightening, example, consider a two-fan $(X \leftarrow Z \rightarrow Y)$. By Theorem 2.1, a high power $(X^n \leftarrow Z^n \rightarrow Y^n)$ can be approximated by a homogeneous fan $H_X \leftarrow H_Z \rightarrow H_Y$. The entropies of X, Z and Y (and therefore also the mutual information $\mathsf{Ent}(X) + \mathsf{Ent}(Y) - \mathsf{Ent}(Z)$ between X and Y) can be established by just counting in the homogeneous fan: $\mathsf{Ent}(X) \approx \frac{1}{n} \log |H_X|$ where $|H_X|$ denotes the cardinality of H_X. However, the three entropies do not determine the asymptotic equivalence class of the fan, there are many more (in fact, infinitely many) independent entropic quantities. But all can be determined from the homogeneous approximation.

2.6. Tropical conditioning

2.6.1. Conditioning

One of the advantages of homogeneous diagrams is that if a homogeneous diagram $\mathcal{X}$ contains a space U, then the isomorphism class of the conditioned diagram $\mathcal{X}|u$ does not depend on the choice of an atom $u \in U$.

Since any tropical diagram is asymptotically equivalent to a homogeneous tropical diagram, we can use this independence of u to define an operation of conditioning of a tropical **G**-diagram $[\mathcal{X}]$ on a space $[U]$ in it, obtaining another tropical **G**-diagram denoted by $[\mathcal{X}|U]$. In the tropical setting the diagram $[\mathcal{X}|U]$ depends (Lipschitz-)continuously on $[\mathcal{X}]$ and $[U]$. This subject is discussed in more details in [17].

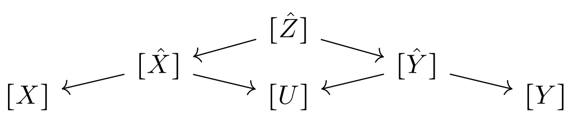
2.6.2. Entropy and mutual information

Now that we defined $[\mathcal{X}|U]$ as a tropical diagram, its entropy $\mathsf{Ent}_*([\mathcal{X}|U])$ is defined by the limit in (2.4). At the same time, it equals the limit

$$\lim_{n\to\infty} \frac{1}{n} \mathsf{Ent}_*(\mathcal{X}(n)|U(n)) := \lim_{n\to\infty} \frac{1}{n} \int_{u \in U(n)} \mathsf{Ent}_*(\mathcal{X}(n)|u) \, \mathrm{d}p(u).$$

In [19] it is shown that the space of single tropical probability spaces, $\mathbf{Prob}[\mathbf{\Lambda}_1]$, is isomorphic to $\mathbb{R}_{\geq 0}$, with the isomorphism given by the entropy. Thus, a tropical probability space is completely determined by its entropy, and we will simply write

- $[X]$ for $\mathsf{Ent}([X])$.
- $[X : Y] := [X] + [Y] - [Z]$ for the mutual information between X and Y in the minimal two-fan $[X] \leftarrow [Z] \rightarrow [Y]$
- $[X : Y|U] := [\hat{X}] + [\hat{Y}] - [\hat{Z}] - [U]$ for the conditional mutual information between X and Y, where $[\hat{X}]$, $[\hat{Y}]$, $[\hat{Z}]$ and $[U]$ are the spaces in the minimal diagram



3. TROPICAL AHLSWEDE–KÖRNER THEOREM

In this section we describe two operations on tropical diagrams, *arrow contraction and expansion*. Given a tropical diagram $[\mathcal{Z}]$, arrow contraction is a modification of the diagram in such a way that a certain arrow becomes an isomorphism, while keeping control of what happens to some other parts of $[\mathcal{Z}]$. Arrow expansion is an inverse operation. We will apply these techniques in the next section to derive a dimension reduction result for the entropic cone.

The full construction is quite involved. Here we will only describe a corollary necessary for our purposes, and refer the reader to [17] for the full results and details.

3.1. Admissible and reduced sub-fans

Suppose $\mathcal{Z}$ is a **G**-diagram and X is an element in it. By the *ideal* generated by X we mean the sub-diagram $\lceil X \rceil$ of $\mathcal{Z}$, that consists of the target spaces of all morphisms starting in X and all (available in $\mathcal{Z}$) arrows between them. We will sometimes refer to spaces in $\lceil X \rceil$ as the *descendants* of X. The ideal generated by a space X included in some diagram will be denoted $\lceil X \rceil$.

If $[\mathcal{Z}]$ is a diagram of tropical probability spaces with the tropical space $[X]$ in it, in order to unclutter notations we will write

$$\lceil X \rceil := \lceil [X] \rceil .$$

An *admissible sub-fan* $(X \leftarrow Z \rightarrow U)$ in a diagram $\mathcal{Z}$ is a *minimal* sub-fan such that the space U is terminal, i. e. it is not the domain of definition of any (non-identity) morphism in $\mathcal{Z}$. An admissible fan will be called reduced if $Z \rightarrow X$ is an isomorphism.

A diagram with an admissible fan is illustrated schematically in Figure 1. Two more concrete examples are shown in Figures 2 and 3.

If an arrow $Y \rightarrow X$ in a diagram $\mathcal{Z}$ is an isomorphism, then we can identify the spaces X and Y, thus changing the combinatorial structure of $\mathcal{Z}$. We call such change *arrow collapse*. Examples of the process of collapsing an arrow can be seen in Figures 1, 2 and 3.

3.2. Tropical Ahlswede–Körner Theorem

Suppose $[\mathcal{Z}]$ and $[\mathcal{Z}']$ are two tropical **G**-diagrams, containing admissible fans $([X] \leftarrow [Z] \rightarrow [U])$ and $([X'] \leftarrow [Z'] \rightarrow [U'])$, respectively, which correspond to each other under the combinatorial isomorphism between $[\mathcal{Z}]$ and $[\mathcal{Z}']$. Suppose the fan $([X'] \leftarrow [Z'] \rightarrow [U'])$ is reduced. Denote $[\mathcal{X}] := \lceil X \rceil$ and $[\mathcal{X}'] := \lceil X' \rceil$.

We say that $[\mathcal{Z}']$ *is obtained from* $[\mathcal{Z}]$ *by arrow contraction* or, alternatively, $[\mathcal{Z}]$ *is obtained from* $[\mathcal{Z}']$ *by arrow expansion*, if

$$[\mathcal{X}] = [\mathcal{X}'] \tag{3.1}$$

$$[\mathcal{X}|U] = [\mathcal{X}'|U']. \tag{3.2}$$

The other spaces in $[\mathcal{Z}]$ outside of $[\mathcal{X}]$ and $[U]$ may change in an uncontrolled manner. In view of equality (3.1) we will identify diagrams $[\mathcal{X}]$ and $[\mathcal{X}']$.

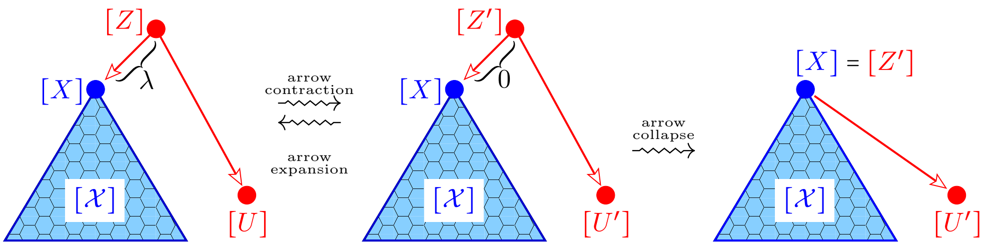


Fig. 1. Arrow contraction/expansion and arrow collapse. Here $[\mathcal{X}] \coloneqq \lceil X \rceil$. In the left diagram the fan $[X] \leftarrow [Z] \rightarrow [U]$ is admissible. In the middle diagram $[X] \leftarrow [Z'] \rightarrow [U']$ is admissible and reduced. The diagrams may contain some other spaces beyond those shown. During contraction/expansion we don't have control over the other parts of the diagram.

Arrow contraction, expansion and collapse are illustrated in Figures 1, 2 and 3.

Note that equation (3.2) is in general an equality in an infinite-dimensional space. But as a simple consequence we have that for any two spaces $[X_1]$ and $[X_2]$ in $[\mathcal{X}]$ and the corresponding spaces $[X_1']$ and $[X_2']$ in $[\mathcal{X}']$ the following equalities hold:

$$\begin{aligned}
[X_1|U] &= [X_1'|U'] \\
[X_1 : X_2|U] &= [X_1' : X_2'|U'] \\
[X_1 : U] &= [X_1' : U'] \\
[X_1 : U|X_2] &= [X_1' : U'|X_2'] \\
[U'] &= [X : U].
\end{aligned} \tag{3.3}$$

Indeed, the first equality follows directly from equality (3.2). The next three can be proven by expanding the right- and left-hand sides into summands of the form $[A|U]$ and $[B]$ for some $[A]$ and $[B]$ in $[\mathcal{X}]$. The last one follows from the fact that $[U']$ is a descendant of $[X]$ in $[\mathcal{Z}']$ and therefore $[X|U'] = [X] - [U']$.

Next we state the following contraction result from [17] which is a generalization of Ahlswede–Körner Theorem from [1, 2].

Theorem 3.1. (Tropical Ahlswede–Körner Theorem) Let $([X] \leftarrow [Z] \rightarrow [U])$ be an admissible fan in some tropical **G**-diagram $[\mathcal{Z}]$. Then there exists a **G**-diagram $[\mathcal{Z}']$ containing an admissible fan $([X'] \leftarrow [Z'] \rightarrow [U'])$, corresponding to the original admissible fan through the combinatorial isomorphism, such that, with the notations $\mathcal{X} = \lceil X \rceil$ and $\mathcal{X}' = \lceil X' \rceil$, the diagram $[\mathcal{Z}']$ satisfies

i. $[\mathcal{X}'|U'] = [\mathcal{X}|U]$

ii. $[\mathcal{X}'] = [\mathcal{X}]$

iii. $[Z'|X'] = 0$

The evaluation of entropy of an individual space in a tropical diagram is a 1-Lipschitz linear functional, while the operation of conditioning is also a Lipschitz map, [18]. Thus, in the settings of Theorem 3.1 the following inequalities hold: for any two spaces $[X_1]$ and $[X_2]$ in $[\mathcal{X}]$ and corresponding spaces $[X_1']$ and $[X_2']$ in $[\mathcal{X}']$

$$\begin{aligned} &[X_1|U] = [X_1'|U'] \\ &[X_1 : X_2|U] = [X_1' : X_2'|U'] \\ &[X_1 : U] = [X_1' : U'] \\ &[X_1 : U|X_2] = [X_1' : U'|X_2'] \\ &[U'] = [X : U]. \end{aligned} \tag{3.4}$$

The following much simpler theorem from [17] is the reverse of Theorem 3.1.

Theorem 3.2. Given a reduced admissible sub-fan $([X] \leftarrow [Z'] \to [U'])$ in a tropical $\mathbf{G}$-diagram $\mathcal{Z}'$ and a non-negative number $\lambda \geq 0$, there is another $\mathbf{G}$-diagram $[\mathcal{Z}]$, such that $[Z|X] = \lambda$.

3.3. Examples

To illustrate the discussion above, we consider two examples of admissible sub-fans and arrow contraction and expansion.

3.3.1.

As a first example, suppose we are given a tropical two-fan $[\mathcal{Z}] = ([X] \leftarrow [Z] \to [U])$ as in Figure 2. We may ask the following question:

Can the mutual information between $[X]$ and $[U]$ be captured by a tropical space $[V]$? More precisely, is there a diamond extension

$$\begin{array}{ccccc} & \swarrow & [Z] & \searrow & \\ [X] & & & & [U] \\ & \searrow & [V] & \swarrow & \end{array}$$

such that $[V] = [X : U]$ (or equivalently $[X : U|V] = 0$)?

The answer is, in general, no. More precisely, the mutual information is representable by a random variable if and only if Wyner Common Information $C(X, Y)$ is equal to the mutual information $I(X : Y)$. However, by contracting and collapsing the arrow $[Z] \to [X]$ we can still obtain a reduction $([X] \to [V])$, where $[V]$ has the required "size", i. e. its entropy equals the mutual information between $[X]$ and $[U]$. If we want to, we can still keep the spaces $[Z]$ and $[U]$ in the diagram after contraction/collapse. Note, however, that in general there will be no reduction $[U] \to [V]$ commuting with the other reductions.

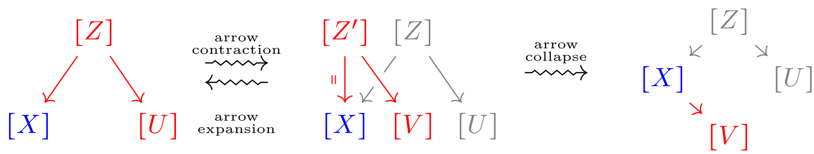


Fig. 2. Contraction/expansion and arrow collapse in a two-fan.

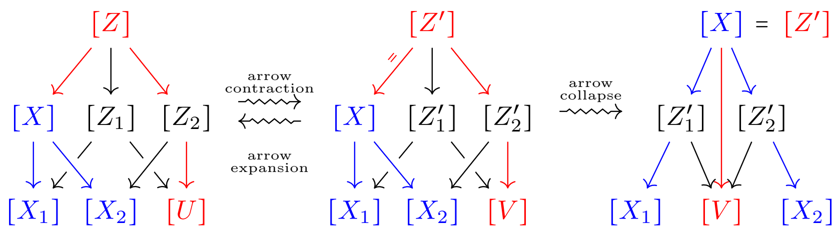


Fig. 3. Arrow contraction and expansion in a $\mathbf{\Lambda}_3$-diagram.

3.3.2.

As a second example, consider the tropical $\mathbf{\Lambda}_3$-diagram

$$[\mathcal{Z}] = \langle [X_1], [X_2], [U] \rangle$$

shown in Figure 3. Such examples can be particularly useful when the space $[U]$ is chosen to satisfy additional properties. For instance, it could be chosen such that the diagrams $[X_1]$ and $[X_2]$ are independent conditioned on $[U]$. We will discuss such extensions elsewhere. The fan $([X] \leftarrow [Z] \rightarrow [U])$ is admissible and the ideal $\lceil X \rceil$ is the fan $([X_1] \leftarrow [X] \rightarrow [X_2])$.

If we contract $[Z] \rightarrow [U]$, we obtain a diagram with a new space $[V]$, that has the same properties relative to $\lceil X \rceil = \langle [X_1], [X_2] \rangle$. The arrow expansion can be seen by reading the picture backwards.

4. ENTROPIC CONE

In this section we define the submodular, entropic and Abelian cones associated to an indexing category $\mathbf{G}$ and prove a dimension reduction result for the entropic cone of four random variables in Theorem 4.1. This result is known, see [13]. In fact, the result in the aforementioned article is stronger then Theorem 4.1. Our methods are somewhat different and we believe, that using them will lead to a deeper understanding of the structure of the entropic cone and other related problems.

4.1. Vector-spaces $\mathbb{R}^{\mathbf{G}}$ and $\mathbb{R} \otimes \mathbf{G}$

Given an indexing category $\mathbf{G}$ we consider two linear spaces associated to it. Recall that the vector space $\mathbb{R}^{\mathbf{G}}$ is the space of functions from $\mathbf{G}$ to $\mathbb{R}$. The second space, dual to the first one, is $\mathbb{R}\otimes\mathbf{G}$ – the vector-space of formal finite linear combinations of objects in $\mathbf{G}$ with real coefficients. These two vector spaces are in natural duality defined by

$$\text{For } f\in\mathbb{R}^{\mathbf{G}} \text{ and } \sum_{i\in\mathbf{G}}\lambda_i\otimes i\in\mathbb{R}\otimes\mathbf{G},\quad \left\langle f,\sum_{i\in\mathbf{G}}\lambda_i\otimes i\right\rangle := \sum_{i\in\mathbf{G}}\lambda_i f(i).$$

The collection of vectors $\{1\otimes i\}_{i\in\mathbf{G}} =: \{[i]\}_{i\in\mathbf{G}}$ forms a basis of the space $\mathbb{R}\otimes\mathbf{G}$, and we denote the dual basis in $\mathbb{R}^{\mathbf{G}}$ by $\{f_i\}_{i\in\mathbf{G}}$. We also consider the following special vectors in $\mathbb{R}\otimes\mathbf{G}$:

- The basis vectors $[i] := 1\otimes i$.
- $[i|j] := [\hat\imath]-[j]$, where $\hat\imath$ is the top object in a minimal fan $i\leftarrow\hat\imath\rightarrow j$ in $\mathbf{G}$.
- $[i:j] := [i]+[j]-[\hat\imath]$, where $\hat\imath$ is top object in a minimal fan $i\leftarrow\hat\imath\rightarrow j$ in $\mathbf{G}$.
- $[i:j|k] := [\hat\imath]+[\hat\jmath]-[k]-[l]$ where objects $\hat\imath$, $\hat\jmath$ and $[l]$ are included in the following minimal diagram in $\mathbf{G}$

$$\begin{array}{ccccccc} & & & l & & & \\ & & \hat\imath \swarrow\!\!\!\nwarrow & & \searrow\!\!\!\nearrow \hat\jmath & & \\ i \leftarrow & & & k & & \rightarrow & j \end{array}$$

$$i \leftarrow \hat\imath \leftarrow l \rightarrow \hat\jmath \rightarrow j,\qquad \hat\imath \rightarrow k \leftarrow \hat\jmath$$

This notations are consistent with the notations for entropy and mutual information, which we introduced in Section 2.6.2, in the sense that for a $\mathbf{G}$-diagram $\mathcal{X}=\{X_i;\chi_{ij}\}$ holds

$$\begin{aligned}\langle \mathsf{Ent}_*\mathcal{X},\,[i]\rangle &= [X_i]\\ \langle \mathsf{Ent}_*\mathcal{X},\,[i|j]\rangle &= [X_i|X_j]\\ \langle \mathsf{Ent}_*\mathcal{X},\,[i:j]\rangle &= [X_i:X_j]\\ \langle \mathsf{Ent}_*\mathcal{X},\,[i:j|k]\rangle &= [X_i:X_j|X_k].\end{aligned}$$

4.2. Submodular, entropic and Abelian cones

Let $\mathbf{G}=\{i;\gamma_{ij}\}$ be an indexing category. We define three closed, convex cones in $\mathbb{R}^{\mathbf{G}}$: the submodular cone $\Gamma_{\mathsf{sm}}(\mathbf{G})$, the entropic cone $\Gamma(\mathbf{G})$ and the Abelian cone $\Gamma_{\mathsf{Ab}}(\mathbf{G})$.

4.2.1. The submodular cone

The submodular cone $\Gamma_{\mathsf{sm}}(\mathbf{G})\subset\mathbb{R}^{\mathbf{G}}$ consists of nonnegative, non-decreasing, submodular functions on the set of objects in the category (points in the poset or vertices in the DAG) $\mathbf{G}$. In essence, these are the functions on $\mathbf{G}$ that satisfy Shannon-like inequalities. More formally it is defined as follows.

The properties nonnegativity, monotonicity and submodularity are defined through linear inequalities. Every linear inequality for $f\in\mathbb{R}^{\mathbf{G}}$ can be written in the form $\langle f,v\rangle\ge 0$ for some $v\in\mathbb{R}\otimes\mathbf{G}$. A function $f\in\mathbb{R}^{\mathbf{G}}$ is called

- *positive*, if $\langle f,\,[i]\rangle\ge 0$ for every object $i\in\mathbf{G}$
- *monotone*, if $\langle f,\,[i|j]\rangle\ge 0$, for every $i,j\in\mathbf{G}$.

- *submodular*, if $\langle f, [i:j]\rangle \geq 0$ and $\langle f, [i:j|k]\rangle \geq 0$ for every $i,j,k \in \mathbf{G}$

The submodular cone is dual to the cone spanned by Shannon-like inequalities

$$\Gamma_{\mathsf{sm}}(\mathbf{G}) := \left\{ f \in \mathbb{R}^{\mathbf{G}} : \langle f, v\rangle \geq 0 \text{ for all } v \in \mathsf{SH} \right\}$$

where $\mathsf{SH} := \{[i], [i|j], [i:j], [i:j|k] : i,j,k \in \mathbf{G}\}$.

4.2.2. The entropic cone

The entropic cone consists of functions on $\mathbf{G}$ that are realizable as entropies of tropical $\mathbf{G}$-diagrams of probability spaces, i. e. it is the image under the entropy map Ent_* of the tropical cone of minimal diagrams indexed by $\mathbf{G}$:

$$\Gamma(\mathbf{G}) := \mathsf{Ent}_*(\mathbf{Prob}[\mathbf{G}]_{\mathsf{m}}).$$

In view of the tropical AEP Theorem 2.1 one can equivalently define

$$\Gamma(\mathbf{G}) := \text{Closure}\left(\mathsf{Ent}_*(\mathbf{Prob}[\mathbf{G}]_{\mathsf{mh}})\right)$$

where by $\mathbf{Prob}[\mathbf{G}]_{\mathsf{mh}}$ we mean the space of minimal, homogeneous, tropical $\mathbf{G}$-diagrams. As we explained in Section 2.2.3, when $\mathbf{G} = \mathbf{\Lambda}_n$, diagrams correspond to n-tuples of random variables. In this case, the entropic cone is equal to the closure of the set of entropically representable vectors, i. e. vectors whose coordinates are entropies of the n random variables and their joints, see [23].

4.2.3. The Abelian cone

The Abelian cone consists of entropy vectors of Abelian tropical diagrams

$$\Gamma_{\mathsf{Ab}}(\mathbf{G}) := \mathsf{Ent}_*(\mathbf{Prob}[\mathbf{G}]_{\mathsf{Ab,m}}).$$

The following two inclusions follow from the definitions and the fact that entropy satisfies Shannon inequalities.

$$\Gamma_{\mathsf{sm}}(\mathbf{G}) \supset \Gamma(\mathbf{G}) \supset \Gamma_{\mathsf{Ab}}(\mathbf{G}). \tag{4.1}$$

4.2.4. The cases of $\mathbf{G} = \mathbf{\Lambda}_1$, $\mathbf{\Lambda}_2$, and $\mathbf{\Lambda}_3$

In this cases all three cones coincide. Essentially it means that any tuple of numbers, that satisfy Shannon inequalities can be realized as entropies of Abelian diagrams, see [24, Theorem 2].

4.3. The case $\mathbf{G} = \mathbf{\Lambda}_4$

The Zhang–Yeung non-Shannon information inequality ([25]) shows that the submodular cone $\Gamma_{\mathsf{sm}}(\mathbf{\Lambda}_4)$ is strictly larger than the entropic cone $\Gamma(\mathbf{\Lambda}_4)$. It is also known that $\Gamma(\mathbf{\Lambda}_4)$ is strictly larger than $\Gamma_{\mathsf{Ab}}(\mathbf{\Lambda}_4)$, see for example [12]. Hence, both inclusions in (4.1) are proper.

The cone $\Gamma_{\mathsf{sm}}(\mathbf{\Lambda}_4)$ is polyhedral by definition, and it is known that the cone $\Gamma_{\mathsf{Ab}}(\mathbf{\Lambda}_4)$ is polyhedral as well, see, for example, [6]. In contrast, the entropic cone $\Gamma(\mathbf{\Lambda}_4)$ is not polyhedral, as has been shown by Matúš in [12].

There are many upper and lower bounds for $\Gamma(\mathbf{\Lambda}_4)$. The upper bounds are in the form of linear inequalities, some of them organized in infinite families. A large list can be found in [6]. Lower bounds are in the form of points in the complement $\Gamma(\mathbf{\Lambda}_4) \setminus \Gamma_{\mathsf{Ab}}(\mathbf{\Lambda}_4)$.

Note that there is an action of symmetric group S_4 on $\mathbf{\Lambda}_4$, $\mathbf{Prob}[\mathbf{\Lambda}_4]$, $\Gamma_{\mathsf{sm}}(\mathbf{\Lambda}_4)$, $\Gamma(\mathbf{\Lambda}_4)$ and $\Gamma_{\mathsf{Ab}}(\mathbf{\Lambda}_4)$.

We will adopt Matúš' notations, where an integer (in small bold face) represents the set of its decimal digits (eg $\mathbf{24} \leftrightarrow \{2,4\} \in \mathbf{\Lambda}_4$).

4.3.1. Ingleton inequalities and the Abelian cone $\Gamma_{\mathsf{Ab}}(\mathbf{\Lambda}_4)$

In addition to the Shannon inequalities, Abelian diagrams also satisfy six Ingleton inequalities, corresponding to the Ingleton vector

$$\mathsf{ing}(\mathbf{12};\mathbf{34}) := -[\mathbf{1}:\mathbf{2}] + [\mathbf{1}:\mathbf{2}|\mathbf{3}] + [\mathbf{1}:\mathbf{2}|\mathbf{4}] + [\mathbf{3}:\mathbf{4}] \in \mathbb{R}\otimes\mathbf{\Lambda}_4$$

and five other vectors obtained by permuting the coordinates.

The cone $\Gamma_{\mathsf{Ab}}(\mathbf{\Lambda}_4)$ is a polyhedral cone dual to the cone spanned by SH and six Ingleton vectors. Its structure is well-known: it coincides with the cone called $\mathbf{H}^{\square}$ in [14]. It has 35 extremal rays, grouped into ten S_4-orbits.

4.3.2. The submodular cone $\Gamma_{\mathsf{sm}}(\mathbf{\Lambda}_4)$

We will represent vectors in $\mathbb{R}^{\mathbf{\Lambda}_4}$ by writing their coordinates in the following order

$$\begin{pmatrix} [\mathbf{1234}] \\ [\mathbf{123}],[\mathbf{124}],[\mathbf{134}],[\mathbf{234}] \\ [\mathbf{12}],[\mathbf{13}],[\mathbf{14}],[\mathbf{23}],[\mathbf{24}],[\mathbf{34}] \\ [\mathbf{1}],[\mathbf{2}],[\mathbf{3}],[\mathbf{4}] \end{pmatrix}.$$

The cone $\Gamma_{\mathsf{sm}}(\mathbf{\Lambda}_4)$ has 41 extremal rays, grouped into eleven S_4-orbits: the 35 rays that are extremal for $\Gamma_{\mathsf{Ab}}(\mathbf{\Lambda}_4)$ and six special rays in the S_4-orbit of a ray generated by the vector

$$\mathsf{spc}(\mathbf{12};\mathbf{34}) := \begin{pmatrix} 4 \\ 4,4,4,4 \\ 3,3,3,3,3,4 \\ 2,2,2,2 \end{pmatrix}.$$

Note that $\langle \mathsf{spc}(\mathbf{12};\mathbf{34}), \mathsf{ing}(\mathbf{12};\mathbf{34})\rangle = -1$. It is known that $\mathsf{spc}(\mathbf{12};\mathbf{34})$ and the other special vectors are not in $\Gamma(\mathbf{\Lambda}_4)$; they are neither representable as entropy vectors of some diagram of probability spaces nor can they be approximated by representable vectors. This has been first shown by Z. Zhang and R.W. Yeung in [25]. Later, many other information inequalities obstructing the representability of $\mathsf{spc}(\mathbf{12};\mathbf{34})$ were discovered by several authors. The article [6] contains a comprehensive list and references.

4.3.3. The non-Ingleton cone

The closure of the complement

$$\Gamma_{\mathsf{sm}}(\mathbf{\Lambda}_4) \setminus \Gamma_{\mathsf{Ab}}(\mathbf{\Lambda}_4)$$

is the union of six cones with disjoint interiors, permuted by the action of S_4. The stabilizer D_2 of this action is the dihedral subgroup of S_4 preserving the partition $\mathbf{1234} = \mathbf{12} \cup \mathbf{34}$. It has order four and is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$.

Consider one of these cones, containing $\mathsf{spc}(\mathbf{12};\mathbf{34})$ and denote it by N. We will call it the non-Ingleton cone. The cone N has a 14-dimensional simplex as a base. The vertices $a_1, \dots, a_{15}$ and the dual faces $\alpha_1, \dots, \alpha_n$ of the simplex are listed in Table 1.

The covectors $\alpha_1, \dots, \alpha_{15}$ give convex coordinates in the simplex.

4.3.4. The cone $\Gamma(\mathbf{\Lambda}_4)$

The cone $\Gamma(\mathbf{\Lambda}_4)$ is squeezed between Γ_{Ab} and Γ_{sm} and the whole picture is S_4-symmetric. Thus the "unknown" part of the $\Gamma(\mathbf{\Lambda}_4)$ is the intersection $\Gamma' := \Gamma(\mathbf{\Lambda}_4) \cap \mathrm{N}$. It contains the rays spanned by vectors $a_1, \dots, a_{14}$ and therefore the whole face $\{\alpha_{15} = 0\}$. The remaining part of the boundary $\partial_+ \Gamma'$ is what we are after. From convexity of Γ' it follows that this part of the boundary is the graph of a certain function defined on the cone spanned by $a_1, \dots, a_{14}$

$$\partial_+ \Gamma' = \{\alpha_{15} = \Phi(\alpha_1, \dots, \alpha_{14})\}$$

where Φ is defined by

$$\Phi(x_1, \dots, x_{14}) := \sup \left\{ \alpha_{15}(\mathbf{x}) : \big(\alpha_1(\mathbf{x}), \dots, \alpha_{14}(\mathbf{x})\big) = (x_1, \dots, x_{14}),\ \mathbf{x} \in \Gamma(\mathbf{\Lambda}_4) \right\}.$$

Obviously, the function Φ is 1-homogeneous.

Theorem 4.1. The function Φ does not depend on the first four arguments.

After the initial submission of this article we learned that a stronger result was obtained using Matúš' convolution method in [13] and even earlier in an unpublished work by Lászlo Czirmaz.

P r o o f. For convenience, for a tropical $\mathbf{\Lambda}_4$-diagram we write

$$A_i[\mathcal{X}] := \langle \mathsf{Ent}_*[\mathcal{X}], \alpha_i \rangle$$

e. g. $A_1[\mathcal{X}] = [X_{\mathbf{1}} | X_{\mathbf{234}}]$, $A_5[\mathcal{X}] = [X_1 : X_3 | X_2]$, etc. Note that all A_i's are Lipschitz-continuous with respect to the input diagram with Lipschitz constant at most 14. In terms of functionals A_i the definition of the function Φ can be rewritten as

$$\Phi(x_1, \dots, x_{14}) := \sup \{ A_{15}[\mathcal{X}] : A_i[\mathcal{X}] = x_i \text{ for } 1 \le i \le 14;\ [\mathcal{X}] \in \mathbf{Prob}[\mathbf{\Lambda}_4]_{\mathsf{m}} \}.$$

Consider a minimal tropical $\mathbf{\Lambda}_4$-diagram $[\mathcal{X}] = \langle [X_{\mathbf{1}}], [X_{\mathbf{2}}], [X_{\mathbf{3}}], [X_{\mathbf{4}}] \rangle$. It contains an admissible sub-fan $([X_{\mathbf{234}}] \leftarrow [X_{\mathbf{1234}}] \rightarrow [X_{\mathbf{1}}])$. Applying Theorem 3.1 to $[\mathcal{X}]$ we obtain another diagram $[\mathcal{X}']$ such that

$$\begin{aligned} A_1[\mathcal{X}'] &= 0 \\ A_i[\mathcal{X}'] &= A_i[\mathcal{X}] && \text{for } i = 2, \dots, 15. \end{aligned}$$

Vertex	Dual face	Representative	D_2-orbit
$a_1 = \begin{pmatrix} 1 \\ 1,1,1,0 \\ 1,1,1,0,0,0 \\ 1,0,0,0 \end{pmatrix}$	$\alpha_1 = [\mathbf{1}\vert\mathbf{234}]$	$l_2 \cdot \big(\mathbb{Z}_2;\ \{0\},\ \mathbb{Z}_2, \mathbb{Z}_2, \mathbb{Z}_2\big)$	a_1, a_2
$a_3 = \begin{pmatrix} 1 \\ 1,0,1,1 \\ 0,1,0,1,0,1 \\ 0,0,1,0 \end{pmatrix}$	$\alpha_3 = [\mathbf{3}\vert\mathbf{124}]$	$l_2 \cdot \big(\mathbb{Z}_2;\ \mathbb{Z}_2,\ \mathbb{Z}_2, \{0\}, \mathbb{Z}_2\big)$	a_3, a_4
$a_5 = \begin{pmatrix} 1 \\ 1,1,1,1 \\ 1,1,1,1,0,1 \\ 1,0,1,0 \end{pmatrix}$	$\alpha_5 = [\mathbf{1}:\mathbf{3}\vert\mathbf{2}]$	$l_2 \cdot \big(\mathbb{Z}_2; \{0\},\ \mathbb{Z}_2, \{0\}, \mathbb{Z}_2\big)$	$a_5, a_6,$ a_7, a_8
$a_9 = \begin{pmatrix} 1 \\ 1,1,1,1 \\ 1,1,1,1,1,1 \\ 1,1,1,0 \end{pmatrix}$	$\alpha_9 = [\mathbf{1}:\mathbf{2}\vert\mathbf{4}]$	$l_2 \cdot \big(\mathbb{Z}_2; \{0\},\ \{0\}, \{0\}, \mathbb{Z}_2\big)$	a_9, a_{10}
$a_{11} = \begin{pmatrix} 1 \\ 1,1,1,1 \\ 1,1,1,1,1,1 \\ 1,1,1,1 \end{pmatrix}$	$\alpha_{11} = [\mathbf{3}:\mathbf{4}]$	$l_2 \cdot \big(\mathbb{Z}_2; \{0\},\ \{0\}, \{0\}, \{0\}\big)$	a_{11}
$a_{12} = \begin{pmatrix} 2 \\ 2,2,2,2 \\ 1,2,2,1,1,2 \\ 1,0,1,1 \end{pmatrix}$	$\alpha_{12} = [\mathbf{3}:\mathbf{4}\vert\mathbf{1}]$	$l_2 \cdot \big((\mathbb{Z}_2)^2;\ \langle\chi_1\rangle,\ \langle\chi_1,\chi_2\rangle, \langle\chi_2\rangle,\ \langle\chi_1+\chi_2\rangle\big)$	a_{12}, a_{13}
$a_{14} = \begin{pmatrix} 3 \\ 3,3,3,3 \\ 2,2,2,2,2,2 \\ 1,1,1,1 \end{pmatrix}$	$\alpha_{14} = [\mathbf{1}:\mathbf{2}\vert\mathbf{34}]$	$l_3 \cdot \big((\mathbb{Z}_3)^3; \langle\chi_1,\chi_2\rangle,\ \langle\chi_2,\chi_3\rangle, \langle\chi_3,\chi_1\rangle,\ \langle\chi_1+\chi_2, \chi_2+\chi_3\rangle\big)$	a_{14}
$a_{15} = \begin{pmatrix} 4 \\ 4,4,4,4 \\ 3,3,3,3,3,4 \\ 2,2,2,2 \end{pmatrix}$	$\alpha_{15} = -\mathsf{ing}(\mathbf{12};\mathbf{34})$	Not representable	a_{15}

Tab. 1. The vertices and faces of the base simplex of non-Ingleton cone. The dihedral group D_2 acts on the simplex by transposing **1** and **2** and, independently, **3** and **4**, so we list only one representative in each orbit. To shorten notations we set $l_2 = (\ln 2)^{-1}$ and $l_3 := (\ln 3)^{-1}$. By $(\mathbb{Z}_n)^k$ we mean the direct product of k copies of the cyclic group of order n and $\chi_1, \dots, \chi_k$ stand for the standard generators in $(\mathbb{Z}_n)^k$.

Repeatedly applying Theorem 3.1 to the resulting diagram after circular permutation of terminal spaces we obtain a $\mathbf{\Lambda}_4$-diagram

$$[\mathcal{X}''] = \langle [X''_{\mathbf{1}}], [X''_{\mathbf{2}}], [X''_{\mathbf{3}}], [X''_{\mathbf{4}}] \rangle$$

such that

$$\begin{aligned} A_i[\mathcal{X}''] &= 0 && \text{for } i = 1,2,3,4 \\ A_i[\mathcal{X}''] &= A_i[\mathcal{X}] && \text{for } i = 5,\dots,15. \end{aligned}$$

Therefore, for any tuple $(x_1,\dots,x_{14})$ of non-negative numbers there exists a tuple $(x''_1,\dots,x''_{14})$ such that

$$\begin{aligned} &\Phi(x_1,\dots,x_{14}) \le \Phi(x''_1,\dots,x''_{14}) \\ &x''_i = 0 \text{ for } i = 1,2,3,4 \\ &x''_i = x_i \text{ for } i = 5,\dots,14. \end{aligned}$$

In other words for any tuple $(x_1,\dots,x_{14})$ of non-negative numbers holds

$$\Phi(x_1,\dots,x_{14}) \le \Phi(0,0,0,0,x_5,\dots,x_{14}).$$

On the other hand, given a diagram $[\mathcal{Y}]$ with $A_i[\mathcal{Y}] = 0$, $i = 1,2,3,4$, and a tuple of non-negative numbers (x_1,x_2,x_3,x_4), we can expand the arrows in the four admissible fans, that we described above, to lengths (x_1,x_2,x_3,x_4). The resulting diagram $[\mathcal{Y}'']$ satisfies

$$A_i[\mathcal{Y}''] = \begin{cases} x_i & i = 1,2,3,4 \\ A_i[\mathcal{Y}] & i = 5,\dots,15. \end{cases}$$

This implies

$$\Phi(x_1,\dots,x_{14}) \ge \Phi(0,0,0,0,x_5,\dots,x_{14})$$

for any non-negative $(x_1,\dots,x_{14})$. □

Note that for $n > 4$ there are projection $\Gamma(\mathbf{\Lambda}_n) \to \Gamma(\mathbf{\Lambda}_4)$ defined by forgetting some $n-4$ random variables. These projections together with Theorem 4.1 imply similar dimension reduction results for higher-order cones $\Gamma(\mathbf{\Lambda}_n)$, $n \ge 4$.

(Received May 15, 2019)

REFERENCES

[1] R. Ahlswede and J. Körner: On common information and related characteristics of correlated information sources. Preprint, 7th Prague Conference on Information Theory, 1974.

[2] R. Ahlswede and J. Körner: On common information and related characteristics of correlated information sources. In: General Theory of Information Transfer and Combinatorics (R. Ahlswede et al., eds.), Lecture Notes in Computer Science *4123*, Springer, Berlin, Heidelberg, 2006.

[3] N. Bertschinger, J. Rauh, E. Olbrich, J. Jost, and N. Ay: Quantifying unique information. Entropy *16* (2014), 4, 2161–2183. DOI:10.3390/e16042161

[4] T. H. Chan and R. W Yeung: On a relation between information inequalities and group theory. IEEE Trans. Inform. Theory *48* (2002), 7, 1992–1995. DOI:10.1109/tit.2002.1013138

[5] R. Dougherty, Ch. Freiling, and K. Zeger: Six new non-shannon information inequalities. In: 2006 IEEE International Symposium on Information Theory, IEEE, 2006, pp. 233–236. DOI:10.1109/isit.2006.261840

[6] R. Dougherty, Ch. Freiling, and K. Zeger: Non-Shannon information inequalities in four random variables. arXiv preprint arXiv:1104.3602, 2011.

[7] M. Gromov: In a search for a structure, part 1: On entropy. Preprint available at `http://www.ihes.fr/gromov`, 2012.

[8] M. Kovačević, I. Stanojević, and V. Šenk: On the hardness of entropy minimization and related problems. In: 2012 IEEE Information Theory Workshop, IEEE, 2012, pp. 512–516. DOI:10.3390/e22040407

[9] T. Leinster: Basic Category Theory, volume 143. Cambridge University Press, 2014. DOI:10.1017/cbo9781107360068.002

[10] F. Matúš: Probabilistic conditional independence structures and matroid theory: background 1. Int. J. General System *22* (1993), 2, 185–196. DOI:10.1080/03081079308935205

[11] F. Matúš: Two constructions on limits of entropy functions. IEEE Trans. Inform. Theory *53* (2006), 1, 320–330. DOI:10.1109/tit.2006.887090

[12] F. Matúš: Infinitely many information inequalities. In: IEEE International Symposium on Information Theory, ISIT 2007, IEEE, pp. 41–44. DOI:10.1109/isit.2007.4557201

[13] F. Matúš and L. Csirmaz: Entropy region and convolution. IEEE Trans. Inform. Theory *62* (2016), 11, 6007–6018. DOI:10.1109/tit.2016.2601598

[14] F. Matúš and M. Studený: Conditional independences among four random variables i. Combinat. Probab. Comput. *4* (1995), 3, 269–278. DOI:10.1017/s0963548300001644

[15] K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin: A new class of non-Shannon-type inequalities for entropies. Comm. Inform. Syst. *2* (2002), 2, 147–166. DOI:10.4310/cis.2002.v2.n2.a3

[16] R. Matveev and J. W Portegies: Asymptotic dependency structure of multiple signals. Inform. Geometry *1* (2018), 2, 237–285. DOI:10.1007/s41884-018-0013-5

[17] R. Matveev and J. W. Portegies: Arrow Contraction and Expansion in Tropical Diagrams. arXiv e-prints, page arXiv:1905.05597, 2019.

[18] R. Matveev and J. W. Portegies: Conditioning in tropical probability theory. arXiv e-prints, page arXiv:1905.05596, 2019.

[19] R. Matveev and J. W. Portegies: Tropical diagrams of probability spaces. arXiv e-prints, page arXiv:1905.04375, 2019.

[20] D. Slepian and J. Wolf: Noiseless coding of correlated information sources. IEEE Trans. Inform. Theory *19* (1973), 4, 471–480. DOI:10.1109/tit.1973.1055037

[21] M. Vidyasagar: A metric between probability distributions on finite sets of different cardinalities and applications to order reduction. IEEE Trans. Automat. Control *57* (2012), 10, 2464–2477. DOI:10.1109/tac.2012.2188423

[22] A. Wyner: The common information of two dependent random variables. IEEE Trans. Inform. Theory *21* (1975), 2, 163–179. DOI:10.1109/tit.1975.1055346

[23] R. W. Yeung: Information Theory and Network Coding. Springer Science and Business Media, 2008. DOI:10.1007/978-0-387-79234-7_1

[24] Z. Zhang and R. W. Yeung: A non-shannon-type conditional inequality of information quantities. IEEE Trans. Inform. Theory *43* (1997), 6, 1982–1986. DOI:10.1109/18.641561

[25] Z. Zhang and R. W. Yeung: On characterization of entropy function via information inequalities. IEEE Trans. Inform. Theory *44* (1998), 4, 1440–1452. DOI:10.1109/18.681320

Rostislav Matveev, Max-Planck-Institut, für Mathematik in den Naturwissenschaften, Inselstraße 22, 04103 Leipzig. Germany.
e-mail: matveev@mis.mpg.de

Jacobus W. Portegies, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven. Netherlands.
e-mail: j.w.portegies@tue.nl