

# Arrow Contraction and Expansion in Tropical Diagrams

Rostislav Matveev <sup>1,\*</sup>  and Jacobus W. Portegies <sup>2,\*</sup> 

<sup>1</sup> Max Planck Institute for Mathematics in the Sciences, 04103 Leipzig, Germany

<sup>2</sup> Department of Mathematics and Computer Science, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands

\* Correspondence: matveev@mis.mpg.de (R.M.); j.w.portegies@tue.nl (J.W.P.)

**Abstract:** Arrow contraction applied to a tropical diagram of probability spaces is a modification of the diagram, replacing one of the morphisms with an isomorphism while preserving other parts of the diagram. It is related to the rate regions introduced by Ahlswede and Körner. In a companion article, we use arrow contraction to derive information about the shape of the entropic cone. Arrow expansion is the inverse operation to the arrow contraction.

**Keywords:** tropical probability; entropic cone

## 1. Introduction

In [1], we have initiated the theory of *tropical probability spaces* for the systematic study of information optimization problems in information theory and artificial intelligence, such as those arising in robotics [2], neuroscience [3], artificial intelligence [4], variational autoencoders [5], information decomposition [6], and causal inference [7]. In [8], we applied the techniques to derive a dimension-reduction result for the entropic cone of four random variables.

Two of the main tools used for the latter are what we call *arrow contraction* and *arrow expansion*. They are formulated for tropical commutative diagrams of probability spaces. Tropical diagrams are points in the asymptotic cone of the metric space of commutative diagrams of probability spaces endowed with the asymptotic entropy distance. Arrows in diagrams of probability spaces are (equivalence classes of) measure-preserving maps.

Arrow contraction and expansion take a commutative diagram of probability spaces as input, modify it, but preserve important properties of the diagram. The precise results are formulated as Theorems 3 and 4 in the main text. Their formulation requires language, notation, and definitions that we review in Section 2.

However, to give an idea of the results in this paper, we now present two examples. For basic terminology and notations used in these examples below, the reader unfamiliar with them is referred either to Section 2 of the present article or in the introductory material in the article [9].

### 1.1. Two Examples

#### 1.1.1. Arrow Contraction and Expansion in a Two-Fan

Suppose we are given a fan  $\mathcal{Z} = (X \leftarrow Z \rightarrow Y)$ , and we would like to complete it to a diamond

$$\mathcal{Z}_\diamond = \left( \begin{array}{ccccc} & & Z & & \\ & \swarrow & & \searrow & \\ X & & & & Y \\ & \searrow & & \swarrow & \\ & & V & & \end{array} \right) \quad (1)$$

such that the entropy of  $V$ , denoted by  $[V]$ , equals the mutual information  $[X : Y]$  between  $X$  and  $Y$ , i.e., we would like to realize the mutual information between  $X$  and  $Y$  by a pair



**Citation:** Matveev, R.; Portegies, J.W. Arrow Contraction and Expansion in Tropical Diagrams. *Entropy* **2023**, *25*, 1637. <https://doi.org/10.3390/e25121637>

Academic Editor: Daniel Chicharro

Received: 6 November 2023

Revised: 4 December 2023

Accepted: 6 December 2023

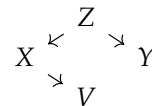
Published: 8 December 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

of reductions  $X \rightarrow V$  and  $Y \rightarrow V$ . This is not always possible, not even approximately. The Gacs-Körner Theorem [10] describes when such exact realization of mutual information is possible.

Arrow contraction instead produces another fan  $Z' = (X \leftarrow Z' \rightarrow V)$ , such that the reduction  $Z' \rightarrow X$  is an isomorphism and the relative entropy  $[X|V]$  of  $X$  given  $V$  equals  $[X|Y]$ . By collapsing this reduction, we obtain as a diagram just the reduction  $X \rightarrow V$ . If necessary, we can keep the original spaces  $Z$  and  $Y$  in the modified diagram obtaining the “broken diamond” diagram



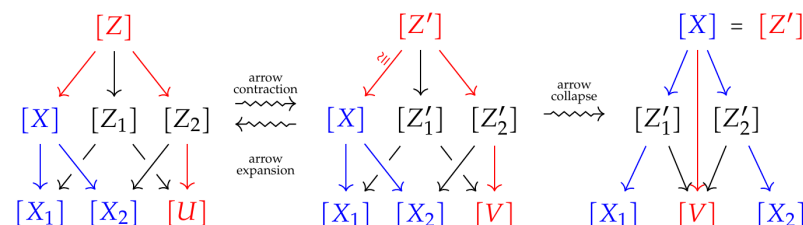
such that  $[V] = [X : Y]$ . Of course, no special technique is necessary to achieve this result since it is easy to find a reduction from a tropical space  $[X]$  to another tropical probability space with the prespecified entropy, as long as the Shannon inequalities are not violated.

However, a similar operation becomes non-trivial and in fact impossible without passing to the tropical limit, if instead of a single space  $X$ , there is a more complex sub-diagram as in the example in the next subsection.

To explain how arrow expansion works, we start with the chain of reductions  $Z \rightarrow X \rightarrow V$ . Can we extend it to a diamond, as in (1), so that  $[X : Y|V] = 0$ ? This is again not possible, in general. However, if we pass to tropical diagrams, then such an extension always exists.

### 1.1.2. One More Example of Arrow Expansion and Contraction

Consider a diagram presented in Figure 1. Such a diagram is called a  $\Lambda_3$ -diagram. We would like to find a reduction  $X \rightarrow V$  so that  $[X|U] = [X|V]$ . It is not possible to achieve this within the realm of diagrams of classical probability spaces. But once we pass to the tropical limit, the reduction  $[X] \rightarrow [V]$  can be found by contracting and then collapsing the arrow  $[Z] \rightarrow [X]$ , as shown in Figure 1.



**Figure 1.** Arrow contraction and expansion in a  $\Lambda_3$ -diagram. The fan  $([X] \leftarrow [Z] \rightarrow [U])$  (shown in red in the Figure) is admissible. Spaces  $[Z_1]$ ,  $[Z_2]$  and  $[Z]$  belong to the co-ideal  $[U]$ . After the operation the part of the diagram shown in blue in the Figure is left unmodified.

Arrow contraction is closely related to the Shannon channel coding theorem. This is perhaps most obvious from the proof. Furthermore, arrow contraction has connections with rate regions, as introduced by Ahlswede and Körner, see [11,12]. These results by Ahlswede and Körner were applied by [13], resulting in a new non-Shannon information inequality. Moreover, in [13], a new proof was given of the results; this new proof is similar to the proof of the arrow contraction result in the present paper.

The main contribution of our work lies in the fact that we prove a much stronger preservation of properties of the diagram under arrow contraction.

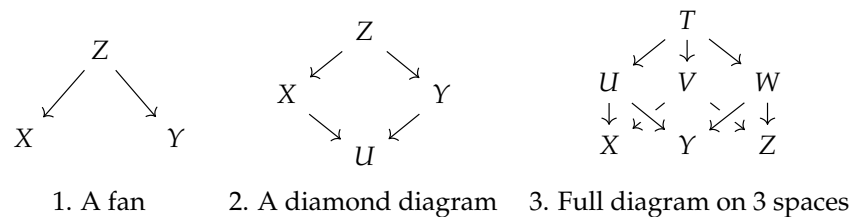
## 2. Preliminaries

### 2.1. Probability Spaces and Their Diagrams

Our main objects of study will be *commutative diagrams of probability spaces*. A *finite probability space*  $X$  is a set with a probability measure on it, supported on a finite set. We

denote by  $|X|$  the cardinality of the support of the measure. The statement  $x \in X$  means that point  $x$  is an atom with positive weight in  $X$ . For details see [1,9,14].

Examples of commutative diagrams of probability spaces are shown in Figure 2. The objects in such diagrams are finite probability spaces and morphisms are equivalence classes of measure-preserving maps. Two such maps are considered to be equivalent if they coincide on a set of full measurements. To record the combinatorial structure of a commutative diagram, i.e., the arrangement of spaces and morphisms, we use *indexing categories*, which are finite poset categories satisfying an additional property, which we describe below.



**Figure 2.** Examples of diagrams of probability spaces.

### 2.1.1. Indexing Categories

A *poset category* is a finite category such that there is at most one morphism between any two objects either way.

For a pair of objects  $k, l$  in a poset category  $\mathbf{G} = \{i; \gamma_{ij}\}$ , such that there is a morphism  $\gamma_{kl}$  in  $\mathbf{G}$ , we call  $k$  an ancestor of  $l$  and  $l$  a descendant of  $k$ . The set of all ancestors of an object  $k$  together with all the morphisms between them is itself a poset category and will be called a *co-ideal* generated by  $k$  and denoted by  $[k]$ . Co-ideals are also sometimes called *filters*. Similarly, a poset category consisting of all descendants of  $k \in \mathbf{G}$  and morphisms between them will be called an *ideal* generated by  $k$  and denoted  $[k]$ .

An *indexing category*  $\mathbf{G} = \{i; \gamma_{ij}\}$  used for indexing diagrams is a poset category satisfying the following additional property: for any pair of objects  $i_1, i_2 \in \mathbf{G}$  the intersection of co-ideals is also a co-ideal generated by some object  $i_3 \in \mathbf{G}$ ,

$$[i_1] \cap [i_2] = [i_3]$$

In other words, for any pair of objects  $i_1, i_2 \in \mathbf{G}$  there exists a *least common ancestor*  $i_3$ , i.e.,  $i_3$  is an ancestor to both  $i_1$  and  $i_2$  and any other common ancestor is also an ancestor of  $i_3$ . Any indexing category is *initial*, i.e., there is a (necessarily unique) *initial* object  $\hat{i}$  in it, which is the ancestor of any other object in  $\mathbf{G}$ , in other words  $\mathbf{G} = [\hat{i}]$ .

A *fan* in a category is a pair of morphisms with the same domain. Such a diagram is also called a *span* in some literature on Category Theory. A fan  $(i \leftarrow k \rightarrow j)$  is called *minimal*, if for any other fan  $(i \leftarrow l \rightarrow j)$  included in a commutative diagram

$$\begin{array}{ccccc} & & k & & \\ & \swarrow & \downarrow & \searrow & \\ i & & l & & j \end{array}$$

the vertical morphism  $(k \rightarrow l)$  must be an isomorphism. Any indexing category also satisfies the property that, for any pair of objects in it, there exists a unique minimal fan with target objects of the given ones.

This terminology will also be applied to diagrams of probability spaces indexed by  $\mathbf{G}$ . Thus, given a space  $X$  in a  $\mathbf{G}$ -diagram, we can talk about its ancestors, descendants, co-ideal  $[X]$ , and ideal  $[X]$ . We use square brackets to denote tropical diagrams and spaces in them. For the (co-)ideals in tropical diagrams, in order to unclutter notations, we will write

$$[X] := [[X]] \quad \text{and} \quad [X] := [[X]]$$

### 2.1.2. Diagrams

For an indexing category  $\mathbf{G} = \{i; \gamma_{ij}\}$  and a category  $\mathbf{Cat}$ , a commutative  $\mathbf{G}$ -diagram  $\mathcal{X} = \{X_i; \chi_{ij}\}$  is a functor  $\mathcal{X} : \mathbf{G} \rightarrow \mathbf{Cat}$ . A diagram  $\mathcal{X}$  is called *minimal* if it maps minimal fans in  $\mathbf{G}$  to minimal fans in  $\mathbf{Cat}$ .

A *constant  $\mathbf{G}$ -diagram* denoted  $X^{\mathbf{G}}$  is a diagram where all the objects equal to  $X$ , and all morphisms are identities.

Important examples of indexing categories are a two-fan, a diamond category, a full category  $\Lambda_n$  on  $n$  spaces, chains  $\mathbf{C}_n$ . For detailed descriptions and more examples, the reader is referred to the articles cited at the beginning of this section.

### 2.2. Tropical Diagrams

#### 2.2.1. Intrinsic Entropy Distance

For a fixed indexing category  $\mathbf{G}$ , the space of commutative  $\mathbf{G}$ -diagrams will be denoted by  $\mathbf{Prob}(\mathbf{G})$ . Evaluating entropy on every space in a  $\mathbf{G}$  diagram gives a map

$$\text{Ent}_* : \mathbf{Prob}(\mathbf{G}) \rightarrow \mathbb{R}^{\mathbf{G}}$$

where the target space  $\mathbb{R}^{\mathbf{G}}$  is the space of real-valued functions on objects of  $\mathbf{G}$ . We endow this space with the  $\ell^1$ -norm. For a fan  $\mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y})$  of  $\mathbf{G}$ -diagrams we define the entropy distance between its terminal objects by

$$\text{kd}(\mathcal{F}) := \|\text{Ent}_* \mathcal{Z} - \text{Ent}_* \mathcal{X}\|_1 + \|\text{Ent}_* \mathcal{Z} - \text{Ent}_* \mathcal{Y}\|_1$$

and the intrinsic entropy distance between two arbitrary  $\mathbf{G}$ -diagrams by

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) := \inf\{\text{kd}(\mathcal{F}) : \mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y})\}$$

This intrinsic version of the entropy distance was introduced in [15,16]. The triangle inequality for  $\mathbf{k}$  and various other properties are discussed in [1].

In the same article, a useful estimate for the intrinsic entropy distance called the Slicing Lemma is also proven. The following corollary ([1], Corollary 3.10(1)) of the Slicing Lemma will be used in the next section.

**Proposition 1.** Let  $\mathbf{G}$  be an indexing category,  $\mathcal{X}, \mathcal{Y} \in \mathbf{Prob}(\mathbf{G})$  and  $U \in \mathbf{Prob}$  included in a pair of two fans

$$\mathcal{X} \leftarrow \tilde{\mathcal{X}} \rightarrow U^{\mathbf{G}} \quad U^{\mathbf{G}} \leftarrow \tilde{\mathcal{Y}} \rightarrow \mathcal{Y}$$

Then

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) \leq \int_U \mathbf{k}(\mathcal{X}|u, \mathcal{Y}|u) dp_U(u) + 2 \cdot \|\mathbf{G}\| \cdot \text{Ent}(U)$$

#### 2.2.2. Tropical Diagrams

Points in the asymptotic cone of  $(\mathbf{Prob}(\mathbf{G}), \mathbf{k})$  are called tropical  $\mathbf{G}$ -diagrams and the space of all tropical  $\mathbf{G}$ -diagrams, denoted  $\mathbf{Prob}[\mathbf{G}]$ , is endowed with the *asymptotic entropy distance*. We explain this now in more detail, and a more extensive description can be found in [14].

To describe points in  $\mathbf{Prob}[\mathbf{G}]$  we consider *quasi-linear* sequences  $\bar{\mathcal{X}} := (\mathcal{X}(n) : n \in \mathbb{N})$  of  $\mathbf{G}$ -diagrams. That is, we fix a “slowly growing” increasing function  $\varphi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$  satisfying

$$t \cdot \int_t^\infty \frac{\varphi(t)}{t^2} dt \leq D_\varphi \cdot \varphi(t)$$

for some constant  $D_\varphi > 0$  and any  $t > 1$ . We call a sequence  $\bar{\mathcal{X}} := (\mathcal{X}(n) : n \in \mathbb{N})$   $\varphi$ -quasi-linear if it satisfies the bound for all  $m, n \in \mathbb{N}$

$$\kappa(\mathcal{X}(n+m), \mathcal{X}(n) \otimes \mathcal{X}(m)) \leq C \cdot \varphi(n+m)$$

We have shown in [14] that the space  $\mathbf{Prob}[\mathbf{G}]$  does not depend on the choice of function  $\varphi$  as long as it is not zero. The space of all such sequences is endowed with the *asymptotic entropy distance* defined by

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

A tropical diagram  $[\mathcal{X}]$  is defined to be an equivalence class of such sequences, where two sequences  $\bar{\mathcal{X}}$  and  $\bar{\mathcal{Y}}$  are equivalent if  $\kappa(\bar{\mathcal{X}}, \bar{\mathcal{Y}}) = 0$ . The space  $\mathbf{Prob}[\mathbf{G}]$  carries the asymptotic entropy distance and has the structure of a  $\mathbb{R}_{\geq 0}$ -semi-module—one can take linear combinations with non-negative coefficients of tropical diagrams. The linear entropy functional  $\text{Ent}_* : \mathbf{Prob}[\mathbf{G}] \rightarrow \mathbb{R}^{\mathbf{G}}$  is defined by

$$\text{Ent}_*[\mathcal{X}] := \lim_{n \rightarrow \infty} \frac{1}{n} \text{Ent}_* \mathcal{X}(n)$$

A detailed discussion about tropical diagrams can be found in [14]. In the cited article, we show that the space  $\mathbf{Prob}[\mathbf{G}]$  is metrically complete and isometrically isomorphic to a closed convex cone in some Banach space.

For  $\mathbf{G} = \mathbf{C}_k$  a *chain category*, containing  $k$  objects  $\{1, \dots, k\}$  and unique morphism  $i \rightarrow j$  for every pair  $i \geq j$ , we have shown in [14] that the space  $\mathbf{Prob}[\mathbf{C}_k]$  is isomorphic to the following cone in  $(\mathbb{R}^k, \|\cdot\|_1)$

$$\mathbf{Prob}[\mathbf{C}_k] \cong \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} : 0 \leq x_1 \leq \dots \leq x_k \right\}$$

The isomorphism is given by the entropy function. Thus, we can identify tropical probability spaces (elements in  $\mathbf{Prob}[\mathbf{C}_1]$ ) with non-negative numbers via entropy. We will simply write  $[X]$  to mean the entropy of the space  $[X]$ . Along these lines, we also adopt the notations  $[X|Y]$ ,  $[X : Y]$  and  $[X : Y|Z]$  for the conditional entropy and mutual information for the tropical spaces included in some diagrams.

### 2.3. Asymptotic Equipartition Property for Diagrams

#### 2.3.1. Homogeneous Diagrams

A  $\mathbf{G}$ -diagram  $\mathcal{X}$  is called *homogeneous* if the automorphism group  $\text{Aut}(\mathcal{X})$  acts transitively on every space in  $\mathcal{X}$ . Homogeneous probability spaces are uniform. For more complex indexing categories, this simple description is not sufficient.

#### 2.3.2. Tropical Homogeneous Diagrams

The subcategory of all homogeneous  $\mathbf{G}$ -diagrams will be denoted  $\mathbf{Prob}(\mathbf{G})_{\text{h}}$  and we write  $\mathbf{Prob}(\mathbf{G})_{\text{h,m}}$  for the category of minimal homogeneous  $\mathbf{G}$ -diagrams. These spaces are invariant under the tensor product. Thus, they are metric Abelian monoids.

Passing to the tropical limit, we obtain spaces of tropical (minimal) homogeneous diagrams that we denote  $\mathbf{Prob}[\mathbf{G}]_{\text{h}}$  and  $\mathbf{Prob}[\mathbf{G}]_{\text{h,m}}$ .

#### 2.3.3. Asymptotic Equipartition Property

In [1] the following theorem is proven

**Theorem 1.** Suppose  $\mathcal{X} \in \mathbf{Prob}(\mathbf{G})$  is a  $\mathbf{G}$ -diagram of probability spaces for some fixed indexing category  $\mathbf{G}$ . Then, there exists a sequence  $\bar{\mathcal{H}} = (\mathcal{H}_n)_{n=0}^{\infty}$  of homogeneous  $\mathbf{G}$ -diagrams such that

$$\frac{1}{n} \mathbf{k}(\mathcal{X}^n, \mathcal{H}_n) \leq C(|X_0|, \llbracket \mathbf{G} \rrbracket) \cdot \sqrt{\frac{\ln^3 n}{n}} \quad (2)$$

where  $C(|X_0|, \llbracket \mathbf{G} \rrbracket)$  is a constant only depending on  $|X_0|$  and  $\llbracket \mathbf{G} \rrbracket$ .

The approximating sequence of homogeneous diagrams is evidently quasi-linear with the defect bounded by the admissible function

$$\varphi(t) := 2C(|X_0|, \llbracket \mathbf{G} \rrbracket) \cdot t^{3/4} \geq 2C(|X_0|, \llbracket \mathbf{G} \rrbracket) \cdot t^{1/2} \cdot \ln^{3/2} t$$

Thus, Theorem 1 above states that  $L(\mathbf{Prob}(\mathbf{G})) \subset \mathbf{Prob}[\mathbf{G}]_h$ . On the other hand, we have shown in [14] that the space of linear sequences  $L(\mathbf{Prob}(\mathbf{G}))$  is dense in  $\mathbf{Prob}[\mathbf{G}]$ . Combining the two statements, we obtain the following theorem.

**Theorem 2.** *For any indexing category  $\mathbf{G}$ , the space  $\mathbf{Prob}[\mathbf{G}]_h$  is dense in  $\mathbf{Prob}[\mathbf{G}]$ . Similarly, the space  $\mathbf{Prob}[\mathbf{G}]_{h,m}$  is dense in  $\mathbf{Prob}[\mathbf{G}]_m$ .*

It is possible that the spaces  $\mathbf{Prob}[\mathbf{G}]_h$  and  $\mathbf{Prob}[\mathbf{G}]$  coincide. At this time, we have neither a proof nor a counterexample to this conjecture.

#### 2.4. Conditioning in Tropical Diagrams

For a tropical  $\mathbf{G}$ -diagram  $[\mathcal{X}]$  containing a space  $[U]$  we defined a conditioned diagram  $[\mathcal{X}|U]$ . It can be understood as the tropical limit of the sequence  $(\mathcal{X}(n)|u_n)$ , where  $(\mathcal{X}(n))$  is the homogeneous approximation of  $[\mathcal{X}]$ ,  $U(n)$  is the space in  $\mathcal{X}(n)$  that corresponds to  $[U]$  under combinatorial isomorphism and  $u_n$  is any atom in  $U(n)$ .

We have shown in [9] that operation of conditioning is Lipschitz-continuous with respect to the asymptotic entropy distance.

### 3. Arrow Contraction

#### 3.1. Arrow Collapse, Arrow Contraction, and Arrow Expansion

##### 3.1.1. Prime Morphisms

A morphism  $\gamma_{ij} : i \rightarrow j$  in an indexing category  $\mathbf{G} = \{i; \gamma_{ij}\}$  will be called *prime* if it cannot be factored into a composition of two non-identity morphisms in  $\mathbf{G}$ . A morphism in a  $\mathbf{G}$ -diagram indexed by a prime morphism in  $\mathbf{G}$  will also be called *prime*.

##### 3.1.2. Arrow Collapse

Suppose  $\mathcal{Z}$  is a  $\mathbf{G}$ -diagram such that for some pair  $i, j \in \mathbf{G}$ , the prime morphism  $\zeta_{ij} : Z_i \rightarrow Z_j$  is an isomorphism. *Arrow collapse* applied to  $\mathcal{Z}$  results in a new diagram  $\mathcal{Z}'$  obtained from  $\mathcal{Z}$  by identifying  $Z_i$  and  $Z_j$  via the isomorphism  $\zeta_{ij}$ . The combinatorial type of  $\mathcal{Z}'$  is different from that of  $\mathcal{Z}$ . The spaces  $Z_i$  and  $Z_j$  are replaced by a single space, and the new space will inherit all the morphisms in  $\mathcal{Z}$  with targets and domains  $Z_i$  and  $Z_j$ .

##### 3.1.3. Arrow Contraction and Expansion

*Arrow contraction and expansion* are two operations on tropical  $\mathbf{G}$ -diagrams. Roughly speaking, arrow contraction applied to a tropical  $\mathbf{G}$ -diagram  $[\mathcal{Z}]$  results in another tropical  $\mathbf{G}$ -diagram  $[\mathcal{Z}']$  such that one of the arrows becomes an isomorphism, while some parts of the diagram are not modified. Arrow expansion is an inverse operation to arrow contraction.

##### 3.1.4. Admissible and Reduced Sub-Fans

An *admissible fan* in a  $\mathbf{G}$ -diagram  $\mathcal{Z}$  is a minimal fan  $X \leftarrow Z \rightarrow U$ , such that  $Z$  is the initial space of  $\mathcal{Z}$  and any space in  $\mathcal{Z}$  belongs either to the co-ideal  $[X]$  or ideal  $[U]$ . For example, in the left-most diagram of Figure 1, the fan  $X \leftarrow Z \rightarrow U$  is admissible, while  $X_1 \leftarrow Z_1 \rightarrow U$  or  $X \leftarrow Z \rightarrow Z_2$  are not.

An admissible fan  $X \leftarrow Z \rightarrow U$  in a diagram will be called *reduced* if the morphism  $Z \rightarrow X$  is an isomorphism.

#### 3.2. The Contraction Theorem

Our aim is to prove the following theorem.

**Theorem 3.** Let  $([X] \leftarrow [Z] \rightarrow [U])$  be an admissible fan in some tropical  $\mathbf{G}$ -diagram  $[Z]$ . Then for every  $\varepsilon > 0$  there exists a  $\mathbf{G}$ -diagram  $[Z']$  containing an admissible fan  $([X'] \leftarrow [Z'] \rightarrow [U'])$ , corresponding to the original admissible fan through the combinatorial isomorphism, such that, with the notations  $\mathcal{X} = [X]$  and  $\mathcal{X}' = [X']$ , the diagram  $[Z']$  satisfies

- (i)  $\kappa([X'|U'], [X|U]) \leq \varepsilon$
- (ii)  $\kappa(\mathcal{X}', \mathcal{X}) \leq \varepsilon$
- (iii)  $[Z'|X'] \leq \varepsilon$

It is not clear that constructing diagrams  $Z'$  as in the theorem above for a sequence of values of parameter  $\varepsilon$  decreasing to 0, we can obtain a convergent sequence in  $\mathbf{Prob}[\mathbf{G}]$  with the limiting diagram satisfying conclusions of the theorem with  $\varepsilon = 0$ . If  $\mathbf{Prob}[\mathbf{G}]$  were a locally compact space, which is an open question at the moment. The convergence would be guaranteed, and then  $\varepsilon$  in the theorem above could be replaced by 0.

The proof of Theorem 3 is based on the following proposition, which will be proven in Section 5.

**Proposition 2.** Let  $(X_0 \leftarrow Z_0 \rightarrow U)$  be an admissible fan in some homogeneous  $\mathbf{G}$ -diagram of probability spaces  $\mathcal{Z}$ . Then there exists a  $\mathbf{G}$ -diagram  $\mathcal{Z}'$  containing the admissible fan  $(X'_0 \leftarrow Z'_0 \rightarrow U')$  such that, with the notations  $\mathcal{X} := [X_0]$  and  $\mathcal{X}' := [X'_0]$ , it holds that

- (1)  $\mathcal{X}|u = \mathcal{X}'|u'$  for any  $u \in U$  and  $u' \in U'$ .
- (2)  $\kappa(\mathcal{X}, \mathcal{X}') \leq \mathbf{k}(\mathcal{X}, \mathcal{X}') \leq 20 \cdot \|\mathbf{G}\|$
- (3)  $[Z'_0|X'_0] \leq 4 \ln \ln |X_0|$

**Proof of Theorem 3.** First, we assume that  $[Z]$  is a homogeneous tropical diagram. It means that it can be represented by a quasi-linear sequence  $(\mathcal{Z}(n))_{n \in \mathbb{N}_0}$  of homogeneous diagrams, with defect of the sequence bounded by the function  $\varphi(t) := C \cdot t^{3/4}$  for some  $C \geq 0$ . This means that for any  $m, n \in \mathbb{N}$

$$\begin{aligned} \kappa(\mathcal{Z}(m) \otimes \mathcal{Z}(n), \mathcal{Z}(m+n)) &\leq \varphi(m+n) \\ \kappa(\mathcal{Z}^m(n), \mathcal{Z}(m \cdot n)) &\leq D_\varphi \cdot m \cdot \varphi(n) \end{aligned}$$

where  $D_\varphi$  is some constant depending on  $\varphi$ , see [14].

Fix a number  $n \in \mathbb{N}$  and apply Proposition 2 to the homogeneous diagram  $\mathcal{Z}(n)$ , containing the admissible fan  $X_0(n) \leftarrow Z_0(n) \rightarrow U(n)$  and sub-diagram  $\mathcal{X}(n) = [X_0(n)]$ . As a result, we obtain a diagram  $\mathcal{Z}''$  containing the fan  $X''_0 \leftarrow Z''_0 \rightarrow U''$  and the sub-diagram  $\mathcal{X}'' = [X''_0]$ , such that

$$\begin{aligned} \mathcal{X}''|u'' &= \mathcal{X}(n)|u && \text{for any } u'' \in U'' \text{ and } u \in U(n) \\ \kappa(\mathcal{X}'', \mathcal{X}(n)) &\leq 20 \|\mathbf{G}\| \\ [Z''_0|X''_0] &\leq 4 \ln \ln |X_0(n)| \end{aligned} \tag{3}$$

Recall that for a diagram  $\mathcal{A}$  of probability spaces, we denote by  $\vec{\mathcal{A}}$  the tropical diagram represented by the linear sequence  $(\mathcal{A}^k : k \in \mathbb{N}_0)$ . As an element of a closed convex cone  $\mathbf{Prob}[\mathbf{G}]$ , it can be scaled by an arbitrary non-negative real number; see, for instance, Section 2.3.5 in [14]. For example,  $\frac{1}{n} \vec{\mathcal{A}}$  is represented by the sequence  $(\mathcal{A}^{\lfloor \frac{k}{n} \rfloor} : k \in \mathbb{N}_0)$ .

Define the two tropical diagrams

$$\begin{aligned} [Z'] &:= \frac{1}{n} \vec{\mathcal{Z}''} \\ [\tilde{Z}] &:= \frac{1}{n} \vec{\mathcal{Z}(n)} \end{aligned}$$

Since  $\mathcal{X}''|u''$  does not depend on  $u''$  and  $\mathcal{X}(n)|u$  does not depend on  $u$  we have  $[\mathcal{X}'|U'] = (1/n) \cdot \overrightarrow{(\mathcal{X}''|u'')}$  and  $[\tilde{\mathcal{X}}|\tilde{U}] = (1/n) \cdot \overrightarrow{(\mathcal{X}(n)|u)}$ . From (3), we obtain

$$\begin{aligned} [\mathcal{X}'|U'] &= [\tilde{\mathcal{X}}|\tilde{U}] \\ \kappa([\mathcal{X}'], [\tilde{\mathcal{X}}]) &\leq \frac{20\|\mathbf{G}\|}{n} \\ [Z'_0|X'_0] &\leq \frac{4\ln\ln|X_0(n)|}{n} \end{aligned} \quad (4)$$

The distance between  $[\tilde{\mathcal{Z}}]$  and  $[\mathcal{Z}]$  can be bounded as follows

$$\begin{aligned} \kappa([\tilde{\mathcal{Z}}], [\mathcal{Z}]) &= \frac{1}{n} \kappa(\overrightarrow{\mathcal{Z}(n)}, n \cdot [\mathcal{Z}]) = \frac{1}{n} \lim_{m \rightarrow \infty} \frac{1}{m} \kappa(\mathcal{Z}^m(n), \mathcal{Z}(m \cdot n)) \\ &\leq \frac{1}{n} D_\varphi \cdot \varphi(n) \end{aligned} \quad (5)$$

This also implies

$$\kappa([\tilde{\mathcal{X}}], [\mathcal{X}]) \leq \frac{1}{n} D_\varphi \cdot \varphi(n) \quad (6)$$

Since conditioning is a Lipschitz-continuous operation with Lipschitz constant 2, we also have

$$\kappa([\tilde{\mathcal{X}}|\tilde{U}], [\mathcal{X}|U]) \leq \frac{2}{n} D_\varphi \cdot \varphi(n) \quad (7)$$

Combining the estimates in (4)–(7) we obtain

$$\begin{aligned} \kappa([\mathcal{X}'|U'], [\mathcal{X}|U]) &\leq 2D_\varphi \cdot \frac{\varphi(n)}{n} \\ \kappa([\mathcal{X}'], [\mathcal{X}]) &\leq \frac{20\|\mathbf{G}\|}{n} + D_\varphi \frac{\varphi(n)}{n} \\ [Z'_0|X'_0] &\leq \frac{4\ln\ln|X_0(n)|}{n} + 2D_\varphi \frac{\varphi(n)}{n} \end{aligned}$$

Please note that  $|X_0(n)|$  grows at most exponentially (it is bounded by  $e^{n([X_0]+C)}$  for some  $C$ ) and  $\varphi$  is a strictly sub-linear function. Thus, by choosing sufficiently large  $n$  depending on the given  $\varepsilon > 0$ , we obtain  $[\mathcal{Z}']$ , satisfying conclusions of the theorem for homogeneous  $[\mathcal{Z}]$ .

To prove the theorem in full generality, observe that all the quantities on the right-hand side of the inequalities are Lipschitz-continuous. Since  $\mathbf{Prob}[\mathbf{G}]_h$  is dense in  $\mathbf{Prob}[\mathbf{G}]$  the theorem extends to any  $[\mathcal{Z}]$  by first approximating it with any precision by a homogeneous configuration and applying the argument above.  $\square$

### 3.3. The Expansion Theorem

The following theorem is complementary to Theorem 3. The expansion applied to a diagram containing a reduced admissible fan produces a diagram with an admissible fan, such that the contraction of it is the original diagram. Thus, arrow expansion is a right inverse of the arrow contraction operation.

In general, contraction erases some information stored in the diagram, so there are many right inverses. We prove the theorem below by providing a simple construction of one such right inverse.

**Theorem 4.** Let  $([X] \leftarrow [Z'] \rightarrow [U'])$  be a reduced admissible fan in some tropical  $\mathbf{G}$ -diagram  $[\mathcal{Z}']$  and  $\lambda > 0$ . Let  $[\mathcal{X}] := [X]$ . Then there exists a  $\mathbf{G}$ -diagram  $[\mathcal{Z}]$  containing the copy of  $[\mathcal{X}]$ , such that the corresponding admissible fan  $([X] \leftarrow [Z] \rightarrow [U])$  has  $[Z|X] = \lambda$  and  $[\mathcal{X}|U] = [\mathcal{X}'|U']$ .

**Proof.** Let  $[W]$  be a tropical probability space with entropy equal to  $\lambda$ . For any reduction of tropical spaces  $[A] \rightarrow [B]$ , there are natural reductions

$$\begin{aligned}([A] + [W]) &\rightarrow ([B] + [W]) \\([A] + [W]) &\rightarrow [W]\end{aligned}$$

We construct the diagram  $[Z]$  by replacing every space  $[V]$  in the ideal  $[U']$  with  $[U] + [W]$ . Every morphism  $[V_1] \rightarrow [V_2]$  within  $[U']$  is replaced by

$$([V_1] + [W]) \rightarrow ([V_2] + [W])$$

And any morphism from  $[V]$  in  $[U']$  to a space  $[Y]$  in  $[X]$  is replaced by a composition

$$([V] + [W]) \rightarrow [V] \rightarrow [Y]$$

Clearly, the resulting diagram satisfies the conclusion of the theorem.  $\square$

The rest of the article is devoted to the development of the necessary tools and the proof of Proposition 2.

#### 4. Local Estimate

In this section, we derive a bound, very similar to Fano's inequality, on the intrinsic entropic distance between two diagrams of probability spaces with the same underlying diagram of sets. The bound will be in terms of the total variation distance between two distributions corresponding to the diagrams of probability spaces. It will be used in the next section to prove Proposition 2.

##### 4.1. Distributions

###### 4.1.1. Distributions on Sets

For a finite set  $S$  we denote by  $\Delta S$  the collection of all probability distributions on  $S$  and by  $\|\pi_1 - \pi_2\|_1$  we denote the total variation distance between  $\pi_1, \pi_2 \in \Delta S$ .

###### 4.1.2. Distributions on Diagrams of Sets

Let **Set** denote the category of finite sets and surjective maps. For an indexing category  $\mathbf{G}$ , we denote by  $\mathbf{Set}(\mathbf{G})$  the category of  $\mathbf{G}$ -diagrams in **Set**. That is, objects in  $\mathbf{Set}(\mathbf{G})$  are commutative diagrams of sets indexed by the category  $\mathbf{G}$ , the spaces in such a diagram are finite sets, and arrows represent surjective maps, subject to commutativity relations.

For a diagram of sets  $\mathcal{S} = \{S_i; \sigma_{ij}\}$  we define the *space of distributions on the diagram  $\mathcal{S}$*  by

$$\Delta \mathcal{S} := \left\{ (\pi_i) \in \prod_i \Delta S_i : (\sigma_{ij})_* \pi_i = \pi_j \right\}$$

where  $f_* : \Delta S \rightarrow \Delta S'$  is the affine map induced by a surjective map  $f : S \rightarrow S'$ . If  $S_0$  is the initial space of  $\mathcal{S}$ , then there is an isomorphism

$$\begin{aligned}\Delta S_0 &\xrightarrow{\cong} \Delta \mathcal{S} \\ \Delta S_0 \ni \pi_0 &\mapsto \{(\sigma_{0i})_* \pi_0\} \in \Delta \mathcal{S} \\ \Delta S_0 \ni \pi_0 &\mapsto \{\pi_i\} \in \Delta\end{aligned}\tag{8}$$

Using the isomorphism (8) we define total variation distance between two distributions  $\pi, \pi' \in \Delta \mathcal{S}$  as

$$\|\pi - \pi'\|_1 := \|\pi_0 - \pi'_0\|_1$$

Given a  $\mathbf{G}$ -diagram of sets  $\mathcal{S} = \{S_i; \sigma_{ij}\}$  and an element  $\pi \in \Delta \mathcal{S}$  we can construct a  $\mathbf{G}$ -diagram of probability spaces  $(\mathcal{S}, \pi) := \{(S_i, \pi_i); \sigma_{ij}\}$ .

Below, we give the estimate of the entropy distance between two  $\mathbf{G}$ -diagrams of probability spaces  $(\mathcal{S}, \pi)$  and  $(\mathcal{S}, \pi')$  in terms of the total variation distance  $\|\pi - \pi'\|_1$  between distributions.

#### 4.2. The Estimate

The upper bound on the entropy distance, which we derive below, has two summands. One is linear in the total variation distance with the slope proportional to the log-cardinality of  $S_0$ . The second one is super-linear in the total variation distance, but it does not depend on  $\mathcal{S}$ . So, we have the following interesting observation: of course, the super-linear summand always dominates the linear one locally. However, as the cardinality of  $\mathcal{S}$  becomes large, it is the linear summand that starts playing the main role. This will be the case when we apply the bound in the next section.

For  $\alpha \in [0, 1]$  consider a binary probability space with the weight of one of the atoms equal to  $\alpha$

$$\mathbb{B}_\alpha := (\{\square, \blacksquare\}; p(\square) = 1 - \alpha, p(\blacksquare) = \alpha)$$

**Proposition 3.** For an indexing category  $\mathbf{G}$ , consider a  $\mathbf{G}$ -diagram of sets  $\mathcal{S} = \{S_i, \sigma_{ij}\} \in \mathbf{Set}(\mathbf{G})$ . Let  $\pi, \pi' \in \Delta\mathcal{S}$  be two probability distributions on  $\mathcal{S}$ . Denote  $\mathcal{X} := (\mathcal{S}, \pi)$ ,  $\mathcal{Y} := (\mathcal{S}, \pi')$  and  $\alpha := \frac{1}{2}\|\pi - \pi'\|_1$ . Then

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) \leq 2\llbracket \mathbf{G} \rrbracket (\alpha \cdot \ln |S_0| + \text{Ent}(\mathbb{B}_\alpha))$$

**Proof.** To prove the local estimate, we decompose both  $\pi$  and  $\pi'$  into a convex combination of a common part  $\hat{\pi}$  and rests  $\pi^+$  and  $\pi'^+$ . The coupling between the common parts gives no contribution to the distance and the worst possible estimate on the other parts is still enough to obtain the bound in the lemma, using Proposition 1.

Let  $S_0$  be the initial set in the diagram  $\mathcal{S}$ . We will need the following obvious rough estimate of the entropy distance that holds for any  $\pi, \pi' \in \Delta\mathcal{S}$ :

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) \leq 2\llbracket \mathbf{G} \rrbracket \cdot \ln |S_0| \quad (9)$$

It can be obtained by taking a tensor product for the coupling between  $\mathcal{X}$  and  $\mathcal{Y}$ .

Our goal now is to write  $\pi$  and  $\pi'$  as the convex combination of three other distributions  $\hat{\pi}$ ,  $\pi^+$  and  $\pi'^+$  as in

$$\begin{aligned} \pi &= (1 - \alpha) \cdot \hat{\pi} + \alpha \cdot \pi^+ \\ \pi' &= (1 - \alpha) \cdot \hat{\pi} + \alpha \cdot \pi'^+ \end{aligned}$$

with the smallest possible  $\alpha \in [0, 1]$ .

We could do it the following way. Let  $\pi_0$  and  $\pi'_0$  be the distributions on  $S_0$  that correspond to  $\pi$  and  $\pi'$  under isomorphisms (8). Let  $\alpha := \frac{1}{2}\|\pi - \pi'\|_1$ . If  $\alpha = 1$  then the proposition follows from the rough estimate (9), so from now on, we assume that  $\alpha < 1$ . Define three probability distributions  $\hat{\pi}_0$ ,  $\pi_0^+$  and  $\pi_0'^+$  on  $S_0$  by setting for every  $x \in S_0$

$$\begin{aligned} \hat{\pi}_0(x) &:= \frac{1}{1 - \alpha} \min\{\pi_0(x), \pi'_0(x)\} \\ \pi_0^+ &:= \frac{1}{\alpha} (\pi_0 - (1 - \alpha)\hat{\pi}_0) \\ \pi_0'^+ &:= \frac{1}{\alpha} (\pi'_0 - (1 - \alpha)\hat{\pi}_0) \end{aligned}$$

Denote by  $\hat{\pi}, \pi^+, \pi'^+ \in \Delta\mathcal{S}$  the distributions corresponding to  $\hat{\pi}_0, \pi_0^+, \pi_0'^+ \in \Delta S_0$  under isomorphism (8). Thus, we have

$$\begin{aligned} \pi &= (1 - \alpha) \hat{\pi} + \alpha \cdot \pi^+ \\ \pi' &= (1 - \alpha) \hat{\pi} + \alpha \cdot \pi'^+ \end{aligned}$$

Now, we construct two fans of **G**-diagrams

$$\mathcal{X} \leftarrow \tilde{\mathcal{X}} \rightarrow \mathbb{B}_\alpha \quad \mathbb{B}_\alpha \leftarrow \tilde{\mathcal{Y}} \rightarrow \mathcal{Y} \quad (10)$$

by setting

$$\begin{aligned} \tilde{X}_i &:= (S_i \times \mathbb{B}_\alpha; \tilde{\pi}_i(s, \square) = (1 - \alpha)\hat{\pi}_i(s), \tilde{\pi}_i(s, \blacksquare) = \alpha \cdot \pi_i^+(s)) \\ \tilde{Y}_i &:= (S_i \times \mathbb{B}_\alpha; \tilde{\pi}'_i(s, \square) = (1 - \alpha)\hat{\pi}_i(s), \tilde{\pi}'_i(s, \blacksquare) = \alpha \cdot \pi_i'^+(s)) \end{aligned}$$

and

$$\begin{aligned} \tilde{\mathcal{X}} &:= \{\tilde{X}_i; \sigma_{ij} \times \mathbf{id}\} \\ \tilde{\mathcal{Y}} &:= \{\tilde{Y}_i; \sigma_{ij} \times \mathbf{id}\} \end{aligned}$$

The reduction in the fans in (10) is given by coordinate projections. Note that the following isomorphisms hold

$$\begin{aligned} \mathcal{X}|_\square &\cong (\mathcal{S}, \hat{\pi}) \\ \mathcal{X}|\blacksquare &\cong (\mathcal{S}, \pi^+) \\ \mathcal{Y}|_\square &\cong (\mathcal{S}, \hat{\pi}) \cong \mathcal{X}|_\square \\ \mathcal{Y}|\blacksquare &\cong (\mathcal{S}, \pi'^+) \end{aligned}$$

Now we apply Proposition 1 along with the rough estimate in (9) to obtain the desired inequality

$$\begin{aligned} \mathbf{k}(\mathcal{X}, \mathcal{Y}) &\leq (1 - \alpha) \mathbf{k}(\mathcal{X}|_\square, \mathcal{Y}|_\square) + \alpha \cdot \mathbf{k}(\mathcal{X}|\blacksquare, \mathcal{Y}|\blacksquare) \\ &\quad + \sum_i [\text{Ent}(\mathbb{B}_\alpha|X_i) + \text{Ent}(\mathbb{B}_\alpha|Y_i)] \\ &\leq 2\llbracket \mathbf{G} \rrbracket (\alpha \cdot \ln |S_0| + \text{Ent}(\mathbb{B}_\alpha)) \end{aligned}$$

□

## 5. Proof of Proposition 2

In this section, we prove Proposition 2, which is shown below verbatim. The proof consists of the construction in Section 5.1 and estimates in Propositions 5 and 6.

### 5.1. The Construction

In this section, we fix an indexing category **G**, a minimal **G**-diagram of probability spaces  $\mathcal{Z}$  with an admissible sub-fan  $X_0 \leftarrow Z_0 \rightarrow U$ . We denote  $\mathcal{X} := [X_0]$  and by **H** we denote the combinatorial type of  $\mathcal{X} = \{X_i; \chi_{ij}\}$ .

Instead of diagram  $\mathcal{Z}$ , we consider an extended diagram, which is a two-fan of **H**-diagrams

$$\mathcal{X} \leftarrow \pi_1 \mathcal{Y} \rightarrow U^{\mathbf{H}} \quad (11)$$

where  $\mathcal{Y} = \{Y_i; \nu_{ij}\}$  consists of those spaces in  $\mathcal{Z}$ , which are initial spaces of two fans with feet in  $U$  and in some space in  $\mathcal{X}$ . That is for every  $i \in \mathbf{H}$  the space  $Y_i$  is defined to be the initial space in the minimal fan  $X_i \leftarrow Y_i \rightarrow U$  in  $\mathcal{Z}$ . It may happen that for some pair of indices  $i_1, i_2 \in \mathbf{H}$  the initial spaces of the fans with one feet  $U$  and the other  $X_{i_1}$  and  $X_{i_2}$  coincide in  $\mathcal{Z}$ . In  $\mathcal{Y}$ , however, they will be treated as separate spaces so that the

combinatorial type of  $\mathcal{Y}$  is  $\mathbf{H}$ . Starting with the diagram in (11) one can recover  $\mathcal{Z}$  by collapsing all the isomorphism arrows. The initial space of  $\mathcal{Y}$  will be denoted  $Y_0$ .

We would like to construct a new fan  $\mathcal{X}' \xrightarrow{\pi'_1} \mathcal{Y}' \rightarrow V^{\mathbf{H}}$ , such that

$$\begin{cases} \mathcal{X}|u = \mathcal{X}'|v & \text{for any } u \in U \text{ and } v \in V \\ \mathbf{k}(\mathcal{X}', \mathcal{X}) \leq 20\llbracket \mathbf{G} \rrbracket \\ [\mathcal{Y}'_0|X'_0] \leq 4 \ln \ln |X_0| \end{cases} \quad (12)$$

Once this goal is achieved, we collapse all the isomorphisms to obtain  $\mathbf{G}$ -diagram satisfying conditions in the conclusion of Proposition 2.

We start with a general description of the idea behind the construction, followed by a detailed argument. To introduce the new space  $V$  we take its points to be  $N$  atoms in  $u_1, \dots, u_N \in U$ . Ideally, we would like to choose the atoms in such a way that  $X_0|u_n$  are disjoint and cover the whole of  $X_0$ . It is not always possible to achieve this exactly. However, when  $|X_0|$  is large,  $N$  is taken slightly larger than  $e^{|X_0| \cdot |U|}$ , and  $u_1, \dots, u_N$  are chosen at random, then with high probability the spaces  $X_0|u_n$  will overlap only little and will cover most of  $X_0$ . The details of the construction follow.

We fix  $N \in \mathbb{N}$  and construct several new diagrams. For each of the new diagrams, we provide a verbal and formal description.

- The space  $U^N$ . Points in it are independent samples of length  $N$  of points in  $U$ .
- The space  $V_N = (\{1, \dots, N\}, \text{unif})$ . A point  $n \in V_N$  should be interpreted as a choice of index in a sample  $\bar{u} \in U^N$ .
- The  $\mathbf{H}$ -diagram  $\mathcal{A}$ , where

$$\begin{aligned} \mathcal{A} &= \{A_i; \alpha_{ij}\} \\ A_i &= (\{(x, n, \bar{u}) : x \in X_i|u_n\}, \text{unif}) \\ \alpha_{ij} &= (\chi_{ij}, \text{Id}, \text{Id}) \end{aligned}$$

A point  $(x, n, \bar{u})$  in  $A_i$  corresponds to the choice of a sample  $\bar{u} \in U^N$ , an independent choice of a member of the sample  $u_n$  and a point  $x \in X_i|u_n$ . Recall that the original diagram  $\mathcal{Z}$  was assumed to be homogeneous and, in particular, the distribution on  $X_i|u_n$  is uniform. Due to the assumption on homogeneity of  $\mathcal{Z}$ , the space  $X_i|u$  does not depend on  $u \in U$ . Since  $V_N$  is also equipped with the uniform distribution, it follows that the distribution on  $A_i$  will also be uniform.

- The  $\mathbf{H}$ -diagram  $\mathcal{B}$ , where

$$\begin{aligned} \mathcal{B} &= \{B_i; \beta_{ij}\} \\ B_i &= \left( \left\{ (x, \bar{u}) : x \in \bigcup_{n=1}^N X_i|u_n \right\}, p_{B_i} \right) \\ \beta_{ij} &= (\chi_{ij}, \text{Id}) \end{aligned}$$

A point  $(x, \bar{u}) \in B_i$  is the choice of a sample  $\bar{u} \in U^N$  and a point  $x$  in one of the fibers  $X_i|u_n$ ,  $n = 1, \dots, N$ . The distribution  $p_{B_i}$  on  $B_i$  is chosen so that the natural projection  $A_i \rightarrow B_i$  is the reduction of probability spaces. Given a sample  $\bar{u}$ , if the fibers  $X_i|u_n$  are not disjoint, then the distribution on  $B_i|\bar{u}$  need not to be uniform. Below, we will give an explicit description of  $p_{\mathcal{B}}$  and study the dependence of  $p_{\mathcal{B}}(\cdot|\bar{u})$  on the sample  $\bar{u} \in U^N$ .

These diagrams can be organized into a minimal diamond diagram of  $\mathbf{H}$ -diagrams, where reductions are obvious projections.

$$\begin{array}{ccccc} & & \mathcal{A} & & \\ & \swarrow & & \searrow & \\ \mathcal{B} & & & & V_N \otimes U^N \\ & \searrow & & \swarrow & \\ & & U^N & & \end{array} \quad (13)$$

To describe the probability distribution on  $\mathcal{B}$ , first we define several relevant quantities:

$$\begin{aligned} \rho &:= \frac{|X_0|u|}{|X_0|} = \mathbf{e}^{-[X_0:U]} \\ N(x, \bar{u}) &:= |\{n \in V_N : x \in X_0|u_n\}| \\ \nu(x, \bar{u}) &:= \frac{N(x, \bar{u})}{N} = p_{V_N}\{n \in V_N : x \in X_0|u_n\} \end{aligned} \quad (14)$$

Recall that the distribution  $p_{\mathcal{B}}$  is completely determined by the distribution  $p_{B_0}$  on the initial space of  $\mathcal{B}$  via isomorphism (8). From homogeneity of  $\mathcal{Z}$  it follows that distributions on both  $A_0$  and  $A|\bar{u}$  are uniform. Therefore

$$p_{B_0}(x|\bar{u}) := \frac{\nu(x, \bar{u})}{\rho \cdot |X_0|} \quad (15)$$

The desired fan  $(\mathcal{X}' \leftarrow \mathcal{Y}' \rightarrow V^H)$  mentioned in the beginning of the section is obtained from the top fan in the diagram in (13) by conditioning on  $\bar{u} \in U^N$ . We will show later that for an appropriate choice of  $N$  and for most choices of  $\bar{u}$ , the fan we obtain in this way has the required properties.

First, we would like to make the following observations. Fix an arbitrary  $\bar{u} \in U^N$ . Then:

- (1) The underlying set of the probability space  $B_0|\bar{u} = X_0|\bar{u}$  is  $\underline{X}_0$ .
- (2) The diagrams

$$\begin{aligned} \mathcal{Y}'_{\bar{u}} &:= \mathcal{A}|\bar{u} \\ \mathcal{X}'_{\bar{u}} &:= \mathcal{B}|\bar{u} \end{aligned}$$

are included in a two-fan of  $\mathbf{H}$ -diagrams

$$\begin{array}{ccc} & \mathcal{Y}'_{\bar{u}} & \\ \swarrow & & \searrow \\ \mathcal{X}'_{\bar{u}} & & V_N \end{array}$$

which is obtained by conditioning the top fan in the diagram in (13).

The very important observation is that diagrams  $\mathcal{X}'_{\bar{u}}|n$  and  $\mathcal{X}'|u$  are isomorphic for any choice of  $n \in V_N$  and  $u \in U$ . The isomorphism is the composition of the following sequence of isomorphisms

$$\mathcal{X}'_{\bar{u}}|n \rightarrow \mathcal{B}|(\bar{u}, n) \rightarrow \mathcal{A}|(\bar{u}, n) \rightarrow \mathcal{X}'|u_n \rightarrow \mathcal{X}'|u$$

where the first isomorphism follows from the definition of  $\mathcal{X}'_{\bar{u}}$ , the second—from minimality of the fan  $\mathcal{B} \leftarrow \mathcal{A} \rightarrow V_N$ , the third—from the definition of  $\mathcal{A}$  and the fourth—from the homogeneity of  $\mathcal{Z}$ .

## 5.2. The Estimates

We now claim and prove that one could choose a number  $N$  and  $\bar{u}$  in  $U^N$  such that

- (1)  $\mathbf{k}(\mathcal{X}'_{\bar{u}}, \mathcal{X}) \leq 20\|\mathbf{H}\|$ .
- (2)  $[Y'_{\bar{u},0}|X'_{\bar{u},0}] \leq 4 \ln \ln |X_0|$ , where  $Y'_{\bar{u},0}$  and  $X'_{\bar{u},0}$  are initial spaces in  $\mathcal{X}'_{\bar{u}}$  and  $\mathcal{Y}'_{\bar{u}}$ , respectively.

### 5.2.1. Total Variation and Entropic Distance Estimates

If we fix some  $x_0 \in X_0$ , then  $v = v(x_0, \cdot)$  is a scaled binomially distributed random variable with parameters  $N$  and  $\rho$ , which means that  $N \cdot v \sim \text{Bin}(N, \rho)$ .

First, we state the following bounds on the tails of a binomial distribution.

**Lemma 1.** *Let  $v$  be a scaled binomial random variable with parameters  $N$  and  $\rho$ , then*

(i) *for any  $t \in [0, 1]$  holds*

$$\mathbb{P}\{|v - \rho| > \rho \cdot t\} \leq 2 \cdot e^{-\frac{1}{3} \cdot N \cdot \rho \cdot t^2}$$

(ii) *for any  $t \in [0, 2]$  holds*

$$\mathbb{P}\left\{\frac{v}{\rho} \ln \frac{v}{\rho} > t\right\} \leq e^{-\frac{1}{12} \cdot N \cdot \rho \cdot t^2}$$

The proof of Lemma 1 can be found at the end of this section.

Below we use the notation  $\mathbb{P} := p_{U^N}$  for the probability distribution on  $U^N$ . For a pair of complete diagrams  $\mathcal{C}, \mathcal{C}'$  with the same underlying diagram of sets and with initial spaces  $C_0, C'_0$ , we will write  $\alpha(\mathcal{C}, \mathcal{C}')$  for the halved total variation distance between their distributions

$$\alpha(\mathcal{C}, \mathcal{C}') := \frac{1}{2} \|p_{C_0} - p_{C'_0}\|_1$$

**Proposition 4.** *In the settings above, for  $t \in [0, 1]$ , the following inequality holds*

$$\mathbb{P}\{\bar{u} \in U^N : 2\alpha(\mathcal{X}'_{\bar{u}}, \mathcal{X}) > t\} \leq 2|X_0| \cdot e^{-\frac{1}{3} \cdot N \cdot \rho \cdot t^2}$$

**Proof.** Recall that by definition  $\mathcal{X}'_{\bar{u}} = \mathcal{B}|\bar{u}$ . We use Equation (15) to expand the left-hand side of the inequality as follows

$$\begin{aligned} \mathbb{P}\{\bar{u} \in U^N : 2\alpha(\mathcal{B}|\bar{u}, \mathcal{X}) > t\} &= \mathbb{P}\left\{\bar{u} \in U^N : \sum_{x \in X_0} \left| \frac{v(x, \bar{u})}{\rho \cdot |X_0|} - \frac{1}{|X_0|} \right| > t\right\} \\ &= \mathbb{P}\left\{\bar{u} \in U^N : \sum_{x \in X_0} |v(x, \bar{u}) - \rho| > \rho \cdot |X_0| \cdot t\right\} \\ &\leq \mathbb{P}\{\bar{u} \in U^N : \text{there exists } x_0 \text{ such that } |v(x_0, \bar{u}) - \rho| > \rho \cdot t\} \\ &\leq \sum_{x \in X_0} \mathbb{P}\{\bar{u} \in U^N : |v(x, \bar{u}) - \rho| > \rho \cdot t\} \end{aligned}$$

Since by homogeneity of the original diagram, all the summands are the same, we can fix some  $x_0 \in X_0$  and estimate further:

$$\mathbb{P}\{\bar{u} \in U^N : 2\alpha(\mathcal{B}|\bar{u}, \mathcal{X}) > t\} \leq |X_0| \cdot \mathbb{P}\{\bar{u} \in U^N : |v(x_0, \bar{u}) - \rho| > \rho \cdot t\}$$

Applying Lemma 1(i), we obtain the required inequality.  $\square$

In the propositions below we assume that  $|X_0|$  is sufficiently large (larger than  $e^{20}$ ).

**Proposition 5.** *In the settings above and for any  $\frac{10}{\ln |X_0|} \leq t \leq 1$  holds:*

$$\mathbb{P}\{\bar{u} \in U^N : \mathbf{k}(\mathcal{X}'_{\bar{u}}, \mathcal{X}) > t(2 \cdot \llbracket \mathbf{G} \rrbracket \cdot \ln |X_0|)\} \leq 2|X_0| \cdot e^{-\frac{1}{3} \cdot N \cdot \rho \cdot t^2}$$

**Proof.** We will use local estimate, Proposition 3, to bound the entropy distance and then apply Proposition 4. To simplify notations, we will write simply  $\alpha$  for  $\alpha(\mathcal{X}'_{\bar{u}}, \mathcal{X}) = \alpha(\mathcal{B}|\bar{u}, \mathcal{X})$ .

$$\begin{aligned} & \mathbb{P}\{\bar{u} \in U^N : \mathbf{k}(\mathcal{B}|\bar{u}, \mathcal{X}) > (2 \cdot \llbracket \mathbf{G} \rrbracket \cdot \ln |X_0|)t\} \\ & \leq \mathbb{P}\{\bar{u} \in U^N : 2 \cdot \llbracket \mathbf{G} \rrbracket (\alpha \cdot \ln |X_0| + \text{Ent}(\Lambda_\alpha)) > (2 \cdot \llbracket \mathbf{G} \rrbracket \cdot \ln |X_0|)t\} \\ & \leq \mathbb{P}\{\bar{u} \in U^N : \alpha + \text{Ent}(\Lambda_\alpha) / \ln |X_0| > t\} \end{aligned}$$

Please note that in the chosen regime,  $t \geq 10 / \ln |X_0|$ , the first summand in the right-hand side of the inequality is larger than the second, i.e.,  $\alpha \geq \text{Ent}(\Lambda_\alpha) / \ln |X_0|$  and therefore we can write

$$\begin{aligned} & \mathbb{P}\{\bar{u} \in U^N : \mathbf{k}(\mathcal{B}|\bar{u}, \mathcal{X}) > (2 \cdot \llbracket \mathbf{G} \rrbracket \cdot \ln |X_0|)t\} \\ & \leq \mathbb{P}\{\bar{u} \in U^N : 2\alpha > t\} \\ & \leq 2|X_0| \cdot e^{-\frac{1}{3}N \cdot \rho \cdot t^2} \end{aligned}$$

□

### 5.2.2. The “Height” Estimate

Recall that for given  $N \in \mathbb{N}$  and  $\bar{u} \in U^N$  we have constructed a two-fan of  $\mathbf{H}$ -diagrams

$$\mathcal{X}'_{\bar{u}} \leftarrow \mathcal{Y}'_{\bar{u}} \rightarrow V_N^{\mathbf{H}}$$

We will now estimate the length of the arrow  $\mathcal{Y}'_{\bar{u},0} \rightarrow \mathcal{X}'_{\bar{u},0}$ .

**Proposition 6.** *In the settings above and for  $t \in [0, 2]$*

$$\mathbb{P}\{\bar{u} \in U^N : [Y'_{\bar{u},0} | X'_{\bar{u},0}] > \ln(N \cdot \rho) + t\} \leq |X_0| \cdot e^{-\frac{1}{12}N \cdot \rho \cdot t^2}$$

**Proof.** First, we observe that the fiber of the reduction  $\mathcal{Y}'_{\bar{u},0} \rightarrow \mathcal{X}'_{\bar{u},0}$  over a point  $x \in \mathcal{X}'_{\bar{u},0}$  is a homogeneous probability space of cardinality equal to  $N(x, \bar{u})$ , therefore its entropy is  $\ln N(x, \bar{u})$ .

$$\begin{aligned} & \mathbb{P}\{\bar{u} \in U^N : [Y'_{\bar{u},0} | X'_{\bar{u},0}] > \ln(N \cdot \rho) + t\} \\ & \mathbb{P}\left\{\bar{u} \in U^N : \int_{\mathcal{X}'_{\bar{u},0}} [Y'_{\bar{u},0} | x] dp_{\mathcal{X}'_{\bar{u},0}}(x) > \ln(N \cdot \rho) + t\right\} \\ & = \mathbb{P}\left\{\bar{u} \in U^N : \sum_{x \in X_0} \frac{\nu(x, \bar{u})}{\rho \cdot |X_0|} \ln(N \cdot \nu(x, \bar{u})) > \ln(N \cdot \rho) + t\right\} \\ & \leq \mathbb{P}\left\{\bar{u} \in U^N : \sum_{x \in X_0} \frac{\nu(x, \bar{u})}{\rho \cdot |X_0|} \ln\left(\frac{\nu(x, \bar{u})}{\rho}\right) > t\right\} \\ & \leq |X_0| \cdot \mathbb{P}\left\{\bar{u} \in U^N : \frac{\nu(x_0, \bar{u})}{\rho} \ln\left(\frac{\nu(x_0, \bar{u})}{\rho}\right) > t\right\} \\ & \leq |X_0| \cdot e^{-\frac{1}{12}N \cdot \rho \cdot t^2} \end{aligned}$$

The last inequality above follows from Lemma 1 (ii). □

### 5.3. Proof of Proposition 2

Let  $\mathcal{X}'_{\bar{u}} \leftarrow \mathcal{Y}'_{\bar{u}} \rightarrow V_N$  be the fan constructed in Section 5.1. The construction is parameterized by number  $N$  and atom  $\bar{u} \in U^N$ . Below, we will choose a particular value for  $N$  and apply estimates in Propositions 5 and 6 with particular choice of parameter  $t$  to show that there is  $\bar{u} \in U^N$ , so that the fan satisfies the conclusions of Proposition 2.

Let

$$N := \ln^3 |X_0| \cdot \rho^{-1} = \ln^3 |X_0| \cdot \mathbf{e}^{[X_0:U]}$$

$$t := \frac{10}{\ln |X_0|}$$

With these choices of  $N$  and  $t$ , Proposition 5 implies

$$\mathbb{P}\{\bar{u} \in U^N : \mathbf{k}(\mathcal{X}'_{\bar{u}}, \mathcal{X}) > 20\llbracket \mathbf{G} \rrbracket\} \leq \frac{1}{4}$$

while Proposition 6 gives

$$\mathbb{P}\{\bar{u} \in U^N : [Y'_{\bar{u},0} | X'_{\bar{u},0}] > 4 \ln \ln |X_0|\} \leq \frac{1}{4}$$

Therefore, there is a choice of  $\bar{u}$  such that the fan

$$(\mathcal{X}' \leftarrow \mathcal{Y}' \rightarrow V) := (\mathcal{X}'_{\bar{u},0} \leftarrow \mathcal{Y}'_{\bar{u},0} \rightarrow V_N)$$

satisfies conditions in (12). As we have explained at the beginning of Section 5.1, by collapsing isomorphism arrows, we obtain  $\mathbf{G}$ -diagram  $\mathcal{Z}'$  satisfying conclusions of Proposition 2.

#### 5.4. Proof of Lemma 1

The Chernoff bound for the tail of a binomially distributed random variable  $X \sim \text{Bin}(N, \rho)$  asserts that for any  $0 \leq \delta \leq 1$  holds

$$\mathbb{P}\{X < (1 - \delta)N \cdot \rho\} \leq \mathbf{e}^{-\frac{1}{2}\delta^2 N \cdot \rho}$$

$$\mathbb{P}\{X > (1 + \delta)N \cdot \rho\} \leq \mathbf{e}^{-\frac{1}{3}\delta^2 N \cdot \rho}$$

Applying the bound for the upper and lower tail for the binomially distributed random variable  $N \cdot \nu$ , we obtain the inequality in (i).

The second assertion follows from the following estimate

$$\begin{aligned} \mathbb{P}\left\{\frac{\nu}{\rho} \ln \frac{\nu}{\rho} > t\right\} &\leq \mathbb{P}\left\{\frac{\nu}{\rho} \left(\frac{\nu}{\rho} - 1\right) > t\right\} \\ &= \mathbb{P}\left\{\nu > \rho \cdot \left(\frac{\sqrt{1+4t}-1}{2} + 1\right)\right\} \end{aligned}$$

For  $0 \leq t \leq 2$  we have  $\sqrt{1+4t}-1 \geq t$ , therefore

$$\mathbb{P}\left\{\frac{\nu}{\rho} \ln \frac{\nu}{\rho} > t\right\} \leq \mathbb{P}\left\{\nu > \rho \cdot \left(\frac{t}{2} + 1\right)\right\}$$

By the Chernoff bound, we have

$$\mathbb{P}\left\{\frac{\nu}{\rho} \ln \frac{\nu}{\rho} > t\right\} \leq \mathbf{e}^{-\frac{1}{12}N \cdot \rho \cdot t^2}$$

**Author Contributions:** Investigation, R.M. and J.W.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** pen Access funding was provided by the Max Planck Society.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Matveev, R.; Portegies, J.W. Asymptotic dependency structure of multiple signals. *Inf. Geom.* **2018**, *1*, 237–285. [\[CrossRef\]](#)
2. Ay, N.; Bertschinger, N.; Der, R.; Güttler, F.; Olbrich, E. Predictive information and explorative behavior of autonomous robots. *Eur. Phys. J. B* **2008**, *63*, 329–339. [\[CrossRef\]](#)
3. Friston, K. The free-energy principle: A rough guide to the brain? *Trends Cogn. Sci.* **2009**, *13*, 293–301. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Van Dijk, S.G.; Polani, D. Informational constraints-driven organization in goal-directed behavior. *Adv. Complex Syst.* **2013**, *16*, 1350016. [\[CrossRef\]](#)
5. Kingma, D.P.; Welling, M. Auto-encoding variational Bayes. *arXiv* **2013**, arXiv:1312.6114.
6. Bertschinger, N.; Rauh, J.; Olbrich, E.; Jost, J.; Ay, N. Quantifying unique information. *Entropy* **2014**, *16*, 2161–2183. [\[CrossRef\]](#)
7. Steudel, B.; Ay, N. Information-theoretic inference of common ancestors. *Entropy* **2015**, *17*, 2304–2327. [\[CrossRef\]](#)
8. Matveev, R.; Portegies, J.W. Tropical probability theory and an application to the entropic cone. *Kybernetika* **2020**, *56*, 1133–1153. [\[CrossRef\]](#)
9. Matveev, R.; Portegies, J.W. Conditioning in tropical probability theory. *Entropy* **2023**, *25*, 1641.
10. Gács, P.; Körner, J. Common information is far less than mutual information. *Probl. Control Inf. Theory* **1973**, *2*, 149–162.
11. Ahlswede, R.; Körner, J. On the connection between the entropies of input and output distributions of discrete memoryless channels. In Proceedings of the Fifth Conference on Probability Theory, Brasov, Romania, 1–6 September 1974.
12. Ahlswede, R.; Körner, J. *Appendix: On Common Information and Related Characteristics of Correlated Information Sources*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 664–677.
13. Makarychev, K.; Makarychev, Y.; Romashchenko, A.; Vereshchagin, N. A new class of non-Shannon-type inequalities for entropies. *Commun. Inf. Syst.* **2002**, *2*, 147–166. [\[CrossRef\]](#)
14. Matveev, R.; Portegies, J.W. Tropical diagrams of probability spaces. *Inf. Geom.* **2020**, *3*, 61–88. [\[CrossRef\]](#)
15. Kovačević, M.; Stanojević, I.; Šenk, V. On the hardness of entropy minimization and related problems. In Proceedings of the 2012 IEEE Information Theory Workshop, Lausanne, Switzerland, 3–7 September 2012; pp. 512–516.
16. Vidyasagar, M. A metric between probability distributions on finite sets of different cardinalities and applications to order reduction. *IEEE Trans. Autom. Control* **2012**, *57*, 2464–2477. [\[CrossRef\]](#)

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