




Tropical diagrams of probability spaces

R. Matveev¹  · J. W. Portegies²

Received: 5 June 2019 / Revised: 29 January 2020 / Published online: 7 April 2020

© The Author(s) 2020

Abstract

After endowing the space of diagrams of probability spaces with an entropy distance, we study its large-scale geometry by identifying the asymptotic cone as a closed convex cone in a Banach space. We call this cone the *tropical cone*, and its elements *tropical diagrams of probability spaces*. Given that the tropical cone has a rich structure, while tropical diagrams are rather flexible objects, we expect the theory of tropical diagrams to be useful for information optimization problems in information theory and artificial intelligence. In a companion article, we give a first application to derive a statement about the entropic cone.

Keywords Tropical probability · Entropy distance · Diagrams of probability spaces · Tropical cone

1 Introduction

With [9] we started a research program aiming for a systematic approach to a class of information optimization problems in information theory and artificial intelligence. A prototypical example of such a problem, still wide open, is the characterization of the *entropic cone*: For an N -tuple of random variables, one may evaluate their entropies and the entropies of the joint variables and obtain a vector in \mathbb{R}^{2^N-1} . A vector obtained in this way is called an *entropy vector* of an N -tuple of random variables. The closure of the set of all entropy vectors of N -tuples is what we call the *entropic cone*, see also [8]. Besides the characterization of the entropic cone, other information optimization problems arise for instance in causal inference [13], artificial

✉ R. Matveev
matveev@mis.mpg.de

J. W. Portegies
j.w.portegies@tue.nl

¹ Max-Planck-Institut für Mathematik in den Naturwissenschaften, Inselstraße 22, 04103 Leipzig, Germany

² Eindhoven University of Technology, Postbus 513, 5600 MB Eindhoven, The Netherlands

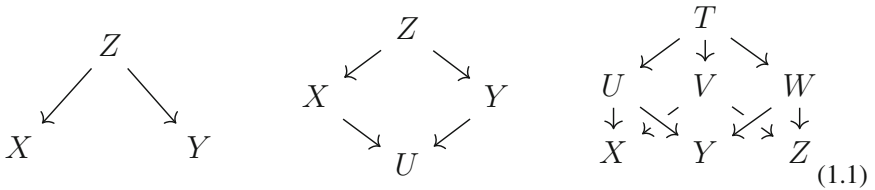
intelligence [14], information decomposition [3], robotics [1], neuroscience [5] and in variational autoencoders [7].

The global strategy of our program is roughly based on the following way of thinking. The entropic cone is clearly a very complicated object: to date, there is no explicit description of the entropic cone for four or more random variables, while it is known that it is not polyhedral [8]. Yet, perhaps, much of its complexity may be explained by it being the closure of an image under a linear map of another, simpler, higher-dimensional cone.

The purpose of this article is to construct such a higher-dimensional (infinite-dimensional, in fact) object, which we call the *tropical cone* and to derive some of its properties which are testimony to its simple structure and which help the study of information optimization problems. As an example of its use, in [11] we apply the theory to derive a statement about the entropic cone.

Before outlining the construction of the tropical cone, let us mention that for our purposes, the language of random variables proved inconvenient, which is why work with *diagrams of probability spaces* instead.

Diagrams of probability spaces are commutative diagrams in the category of probability spaces, with (equivalence classes of) measure-preserving maps as morphisms, such as



Collections of n random variables give rise to a special type of diagrams, that include, besides the target spaces of the random variables themselves, the target space of every joint variable. Such diagrams have a particular combinatorial type. The first and the last diagrams in (1.1) are examples of such special types of diagrams in case of two and three random variables respectively. The description of other diagrams, such as the diagram in the middle of (1.1), using the language of random variables is less transparent.

We will construct the tropical cone and derive its properties over several sections. In Sect. 2 we describe the construction of the asymptotic cone in the abstract setting of a metric Abelian monoid $(\Gamma, +, \mathbf{d})$. We believe that this abstract setting will make the construction more transparent and easier to follow. The results we present in that section are probably quite standard, but we find it beneficial to gather them under one roof. Such an asymptotic cone consists of *equivalence classes* of *quasi-linear sequences* in the monoid. Whereas *linear sequences* have the form $(n \cdot a)_{n \in \mathbb{N}_0}$, where a is an element of the monoid, *quasi-linear sequences* may deviate from linearity in a controlled fashion, measured by a sublinear function φ satisfying some additional conditions that we will specify later. A sequence $\gamma \in \Gamma^{\mathbb{N}_0}$ is called φ -quasi-linear if for all $m, n \in \mathbb{N}$, it satisfies

$$\mathbf{d}(\gamma(m+n), \gamma(m) + \gamma(n)) \leq \varphi(m+n)$$

and two sequences γ and γ' are equivalent if

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{d}(\gamma(n), \gamma'(n)) = 0$$

The asymptotic cone is itself again a metric Abelian monoid, but it admits additional structure. It admits a distributive action of $\mathbb{R}_{\geq 0}$ and the metric becomes homogeneous and translation invariant. As an example of this construction, A'Campo [2] constructed the real numbers as the asymptotic cone in the monoid of integers.

In Sect. 3 we show that, under certain conditions, the asymptotic cone is a complete metric space and it can be realized as a closed convex cone in a Banach space.

In Sect. 4 we apply the general construction of Sects. 2 and 3 to the monoid of diagrams of probability spaces endowed with the *intrinsic entropy distance* [6,9,15] and with the tensor product as the binary operation. We call the resulting space *tropical cone* and its elements *tropical diagrams*.¹ In Sect. 6, we give a simple characterization of the tropical cone for special types of diagrams.

For more complicated diagrams, we currently do not have an explicit description of the tropical cone, but we do show that it possesses a rich algebraic structure. In particular, one can take convex combinations of tropical diagrams. Other useful operations and constructions can be carried through for tropical diagrams, whereas they do not have an equivalent in the classical context of probability spaces, see [10]. All in all, from some perspective, tropical diagrams are easier to deal with than diagrams or probability spaces, since only rough, asymptotic relations between probability spaces are preserved under tropicalization, similar to how all complicated features of the landscape disappear when looking at the Earth from outer space.

In order to study information optimization problems, we may as well study the more malleable tropical cone. This is because the *entropic cone* is the closure of the image of the bounded linear map defined on the tropical cone that evaluates entropies of the individual spaces in a tropical diagram. More generally, we call any non-negative bounded linear functional on the tropical cone an *entropic quantity*. These include entropies of individual spaces, but also some other quantities, such as optima of some linear combinations of entropies of an extended diagram, where some extra spaces are added to the original diagram. Study of such entropic quantities is the subject of our future research.

One of the main tools in the study of entropic quantities through the tropical cone is the Asymptotic Equipartition Property for diagrams. Originally derived in [9], we cast it here into a density statement of simpler, so-called *homogeneous* tropical diagrams in the tropical cone, in terms of Theorems 5.1 and 5.2. Therefore, to prove statements

¹ The reason for the name *tropical cone* is the following. For instance in algebraic geometry, tropical varieties are, roughly speaking, divergent sequences of classical varieties, renormalized on a log scale with an increasing base. The adjective ‘tropical’ carries little semantics, but was introduced in honor of the Brazilian mathematician and computer scientist Imre Simon who worked on the subject of tropical mathematics. Analogously, we construct the asymptotic cone from certain divergent sequences with respect to the intrinsic entropy distance. As the intrinsic entropy distance is entropy-based, we achieve a similar type of renormalization as in algebraic geometry.

about entropic quantities, it suffices to study the much simpler homogeneous tropical diagrams.

2 Asymptotic cones of metric abelian monoids

In this section we define the asymptotic cone in the setting of an abstract metric Abelian monoid. In a later section, we will specify to the case of diagrams of probability spaces.

2.1 Metric and pseudo-metric spaces

A pseudo-metric space (X, \mathbf{d}) is a set X equipped with a pseudo-distance \mathbf{d} , a bivariate function satisfying all the axioms of a distance function, except that it is allowed to vanish on pairs of non-identical points. An isometry of such spaces is a distance-preserving map, such that for any point in the target space there is a point in the image at zero distance away from it. Given such a pseudo-metric space (X, \mathbf{d}) one could always construct an isometric metric space $(X/\mathbf{d}=0, \mathbf{d})$, the metric quotient, by identifying all pairs of points that are distance zero apart.

Any property formulated in terms of the pseudo-metric holds simultaneously for a pseudo-metric space and its metric quotient. It will be convenient for us to construct pseudo-metrics on spaces instead of passing to the quotient spaces.

For a pair of points $x, y \in X$ in a pseudo-metric space (X, \mathbf{d}) we will write $x \stackrel{\mathbf{d}}{=} y$ if $\mathbf{d}(x, y) = 0$. We call such a pair of points (\mathbf{d} -)metrically equivalent.

Many metric-topological notions such as (Lipschitz-)continuity, compactness, ϵ -nets, dense subsets, etc., extend to the setting of a pseudo-metric spaces and exercising certain care one may switch between a pseudo-metric space and its metric quotient replacing the $\stackrel{\mathbf{d}}{=}$ -sign with equality.

2.2 Metric abelian monoids

A monoid is a set equipped with a bivariate associative operation and a neutral element. The operation is usually called multiplication, or addition if it is commutative. We call a monoid with pseudo-distance $(\Gamma, +, \mathbf{d})$ a *metric Abelian monoid* if it satisfies:

1. For all $\gamma, \gamma' \in \Gamma$ holds

$$\gamma + \gamma' \stackrel{\mathbf{d}}{=} \gamma' + \gamma$$

2. The binary operation is 1-Lipschitz with respect to each argument: For all $\gamma, \gamma', \eta \in \Gamma$

$$\mathbf{d}(\eta + \gamma, \eta + \gamma') \leq \mathbf{d}(\gamma, \gamma')$$

In other words, the translation maps

$$T_\eta : \Gamma \rightarrow \Gamma, \quad \gamma \mapsto \eta + \gamma$$

are non-expanding for every $\eta \in \Gamma$.

The following proposition is an elementary consequence of the triangle inequality.

Proposition 2.1 *Let $(\Gamma, +, \mathbf{d})$ be a metric Abelian monoid. Then:*

1. *For any quadruple $\gamma_1, \gamma_2, \gamma_3, \gamma_4 \in \Gamma$ holds*

$$\mathbf{d}(\gamma_1 + \gamma_2, \gamma_3 + \gamma_4) \leq \mathbf{d}(\gamma_1, \gamma_3) + \mathbf{d}(\gamma_2, \gamma_4)$$

2. *For every $n \in \mathbb{N}$, and $\gamma_1, \gamma_2 \in \Gamma$ also holds*

$$\mathbf{d}(n \cdot \gamma_1, n \cdot \gamma_2) \leq n \cdot \mathbf{d}(\gamma_1, \gamma_2)$$

A metric Abelian monoid $(\Gamma, +, \delta)$ will be called *homogeneous* if it satisfies for all $n \in \mathbb{N}_0$

$$\delta(n \cdot \gamma_1, n \cdot \gamma_2) = n \cdot \delta(\gamma_1, \gamma_2) \tag{2.1}$$

A homogeneous metric Abelian monoid is called an $\mathbb{R}_{\geq 0}$ -*semi-module* $(\Gamma, +, \cdot, \delta)$ if in addition there is a doubly distributive $\mathbb{R}_{\geq 0}$ -action such that for any $\lambda_1, \lambda_2 \in \mathbb{R}_{\geq 0}$ and $\gamma_1, \gamma_2 \in \Gamma$ holds

$$\begin{aligned} \lambda_1 \cdot (\lambda_2 \cdot \gamma_1) &\stackrel{\delta}{=} (\lambda_1 \lambda_2) \cdot \gamma_1 \\ \lambda \cdot (\gamma_1 + \gamma_2) &\stackrel{\delta}{=} \lambda \cdot \gamma_1 + \lambda \cdot \gamma_2 \\ (\lambda + \lambda') \cdot \gamma_1 &\stackrel{\delta}{=} \lambda \cdot \gamma_1 + \lambda' \cdot \gamma_1 \\ \delta(\lambda \cdot \gamma, \lambda \cdot \gamma') &= \lambda \cdot \delta(\gamma, \gamma') \end{aligned}$$

A convex cone in a normed vector space would be a typical example of an $\mathbb{R}_{\geq 0}$ -semimodule. An intersection of a convex cone in \mathbb{R}^n with the integer lattice is an example of a monoid, that does not admit semimodule structure.

The following proposition asserts that if a metric Abelian monoid is homogeneous, then the pseudo-distance is translation invariant, and, in particular, it satisfies a cancellation property. This result was communicated to us by Tobias Fritz.

Proposition 2.2 *Let $(\Gamma, +, \delta)$ be a homogeneous metric Abelian monoid. Then the pseudo-distance function δ is translation invariant, that is it satisfies for any $\gamma_1, \gamma_2, \eta \in \Gamma$*

$$\delta(\gamma_1 + \eta, \gamma_2 + \eta) = \delta(\gamma_1, \gamma_2)$$

In particular, the following cancellation property holds in Γ

$$\text{If } \gamma_1 + \eta \stackrel{\delta}{=} \gamma_2 + \eta, \text{ then } \gamma_1 \stackrel{\delta}{=} \gamma_2.$$

The proof of this proposition is essentially the same as the proof of [9, Proposition 3.7]. Even though the latter proposition is formulated for a *specific* homogeneous metric Abelian monoid, it does not use any of its specific properties, but only defining properties of a generic homogeneous metric Abelian monoid.

2.3 Asymptotic cones (tropicalization) of monoids

In our construction points of the asymptotic cone of $(\Gamma, +, \mathbf{d})$ will be sequences of points in Γ that grow almost linearly in a certain sense described below.

2.3.1 Admissible functions

Admissible functions will be used to measure the deviation of a sequence from being linear. We call a function $\varphi : \mathbb{R}_{\geq 1} \rightarrow \mathbb{R}_{\geq 0}$ *admissible* if

1. the function φ is non-decreasing;
 2. the function $\varphi(t)/t$ is non-increasing;
 3. there exists a constant $D_\varphi \geq 0$ such that $s \cdot \int_s^\infty \frac{\varphi(t)}{t^2} dt \leq \frac{D_\varphi}{8} \cdot \varphi(s)$ for any $s \geq 1$.
- In particular the function φ is summable against dt/t^2 .

For example, the function $\varphi(t) = t^\alpha$ is admissible for any $0 \leq \alpha < 1$. Any admissible function is necessarily sub-linear, that is $\varphi(t)/t \rightarrow 0$ as $t \rightarrow \infty$. A linear combination of admissible functions with non-negative coefficients is also admissible.

Lemma 2.3 *Let φ be a positive admissible function. Then for any $\alpha \geq 0$ and $\lambda \geq 1$ there is $C > 1$ such that for any $t \geq 1$*

$$\varphi(\lambda \cdot t) + \alpha \leq C \cdot \varphi(t)$$

Proof From positivity and monotonicity of φ we have

$$\alpha = \frac{\alpha}{\varphi(1)} \cdot \varphi(1) \leq \frac{\alpha}{\varphi(1)} \cdot \varphi(t)$$

On the other hand from monotonicity of the function $\varphi(t)/t$ it follows that for any $\lambda, t \geq 1$

$$\varphi(\lambda \cdot t) \leq \lambda \cdot \varphi(t)$$

Adding the two inequalities above we obtain the conclusion of the lemma. \square

2.3.2 Quasi-linear sequences

Let $(\Gamma, +, \mathbf{d})$ be a metric Abelian monoid and φ be an admissible function. A sequence $\bar{\gamma} = \{\gamma(i)\} \in \Gamma^{\mathbb{N}_0}$ will be called quasi-linear with defect bounded by φ if for every $m, n \in \mathbb{N}$ the following bound is satisfied

$$\mathbf{d}(\gamma(m+n), \gamma(m) + \gamma(n)) \leq \varphi(m+n) \tag{2.2}$$

For technical reasons we also require $\gamma(0) = 0$. Sequences that are quasi-linear with defect bounded by $\varphi \equiv 0$ will be called *linear sequences*.

We will often need the following corollary of quasi-linearity, which follows from applying the bound (2.2) twice and using the monotonicity of φ : for all $m, n, k \in \mathbb{N}$

$$\mathbf{d}(\gamma(m+n+k), \gamma(m) + \gamma(n) + \gamma(k)) \leq 2\varphi(m+n+k) \tag{2.3}$$

For an admissible function φ we will write $QL_\varphi(\Gamma, \mathbf{d})$ for the space of all quasi-linear sequences with defect bounded by $C \cdot \varphi$ for some (depending on the sequence) constant $C \geq 0$. We will also use notation $L(\Gamma, \mathbf{d}) := QL_0(\Gamma, \mathbf{d})$ for the space of linear sequences.

2.3.3 Asymptotic distance

Given two quasi-linear sequences $\bar{\gamma}_1 \in QL_{\varphi_1}(\Gamma, \mathbf{d})$ and $\bar{\gamma}_2 \in QL_{\varphi_2}(\Gamma, \mathbf{d})$ the sequence of distances $a(n) := \mathbf{d}(\gamma_1(n), \gamma_2(n))$ is φ_3 -subadditive, where $\varphi_3 = \varphi_1 + \varphi_2$ is also admissible, i.e.

$$a(m+n) \leq a(m) + a(n) + \varphi_3(m+n)$$

for any $n, m \in \mathbb{N}$. By the generalization of Fekete’s Lemma by De Bruijn and Erdős [4, Theorem 23], it follows that the following limit exists and finite

$$\hat{\mathbf{d}}(\bar{\gamma}_1, \bar{\gamma}_2) := \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{d}(\gamma_1(n), \gamma_2(n))$$

We call the quantity $\hat{\mathbf{d}}(\bar{\gamma}_1, \bar{\gamma}_2)$ the asymptotic distance between $\bar{\gamma}_1, \bar{\gamma}_2 \in QL_\varphi(\Gamma, \mathbf{d})$. It is easy to verify that $\hat{\mathbf{d}}$ indeed satisfies all axioms of a pseudo-distance. Even if \mathbf{d} was a proper distance function, the corresponding asymptotic distance may vanish on some pairs of non-identical elements. We call two sequences $\bar{\gamma}_1 \in QL_{\varphi_1}(\Gamma, \mathbf{d})$, $\bar{\gamma}_2 \in QL_{\varphi_2}(\Gamma, \mathbf{d})$ *asymptotically equivalent* if $\hat{\mathbf{d}}(\bar{\gamma}_1, \bar{\gamma}_2) = 0$ and write

$$\bar{\gamma}_1 \stackrel{\hat{\mathbf{d}}}{=} \bar{\gamma}_2$$

2.3.4 Quasi-homogeneity

We will show that quasi-linear sequences are also quasi-homogeneous in the sense of the following lemma.

Lemma 2.4 Let $\bar{\gamma} \in \Gamma^{\mathbb{N}_0}$ be a sequence with φ -bounded defect. Then for any $m, n \in \mathbb{N}$

$$\mathbf{d}(\gamma(m \cdot n), m \cdot \gamma(n)) \leq 8 \cdot m \cdot n \cdot \int_n^{2m \cdot n} \frac{\varphi(t)}{t^2} dt$$

Proof Define the function $\psi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ related to φ as follows

$$\psi(s) := \varphi(e^s) / e^s \quad \text{or} \quad \varphi(t) =: t \cdot \psi(\ln t)$$

The conclusion of the lemma in terms of ψ then reads

$$\mathbf{d}(\gamma(m \cdot n), m \cdot \gamma(n)) \leq 8 \cdot m \cdot n \cdot \int_{\ln n}^{\ln(2 \cdot m \cdot n)} \psi(s) ds$$

and it is in that form it will be proven below.

Due to monotonicity properties of φ , the function ψ satisfies, for all $0 \leq s_0 \leq s$

$$\psi(s_0) \leq \psi(s) \cdot e^{s-s_0}$$

which integrated over s yields

$$\psi(s_0) \leq \frac{2}{\ln 2} \int_{s_0}^{s_0 + \ln 2} \psi(s) ds \leq 4 \int_{s_0}^{s_0 + \ln 2} \psi(s) ds \quad (2.4)$$

We proceed by induction with respect to m , keeping n fixed. The conclusion of the lemma is obvious for $m = 1$. For the induction step let $m = 2m' + \epsilon \geq 2$, where $m' = \lfloor m/2 \rfloor$ and $\epsilon \in \{0, 1\}$. We first use the bound in (2.3) to estimate

$$\begin{aligned} \mathbf{d}(\gamma(m \cdot n), m \cdot \gamma(n)) &= \mathbf{d}(\gamma(m' \cdot n + m' \cdot n + \epsilon \cdot n), m' \cdot \gamma(n) + m' \cdot \gamma(n) + \epsilon \cdot \gamma(n)) \\ &\leq 2\mathbf{d}(\gamma(m' \cdot n), m' \cdot \gamma(n)) + 2\varphi(m \cdot n) \end{aligned}$$

Next, we continue the estimate using bound (2.4)

$$\begin{aligned} \mathbf{d}(\gamma(m \cdot n), m \cdot \gamma(n)) &\leq 2\mathbf{d}(\gamma(m' \cdot n), m' \cdot \gamma(n)) + 2\varphi(m \cdot n) \\ &\leq 16m' \cdot n \cdot \int_{\ln n}^{\ln(2m' \cdot n)} \psi(s) ds + 2m \cdot n \cdot \psi(\ln(m \cdot n)) \\ &\leq 8m \cdot n \left(\int_{\ln n}^{\ln(2m' \cdot n)} \psi(s) ds + \int_{\ln(m \cdot n)}^{\ln(2m \cdot n)} \psi(s) ds \right) \\ &\leq 8m \cdot n \cdot \int_{\ln n}^{\ln(2m \cdot n)} \psi(s) ds \end{aligned}$$

□

Applying bound (3) in the definition of admissible functions, we obtain the following corollary.

Corollary 2.5 *Let $\tilde{\gamma}$ be a sequence with φ -bounded defect. Then for any $m, n \in \mathbb{N}$*

$$\mathbf{d}(\gamma(m \cdot n), m \cdot \gamma(n)) \leq 8 \cdot m \cdot n \cdot \int_n^\infty \frac{\varphi(t)}{t^2} dt \leq D_\varphi \cdot m \cdot \varphi(n)$$

2.3.5 The semi-module structure

The group operation $+$ on Γ induces a $\hat{\mathbf{d}}$ -continuous (in fact, 1-Lipschitz) group operation on $\text{QL}_\varphi(\Gamma, \mathbf{d})$ by adding sequences element-wise. Thus $(\text{QL}_\varphi(\Gamma, \mathbf{d}), +, \hat{\mathbf{d}})$ is also a metric Abelian monoid. In addition, if φ is positive it carries the structure of a $\mathbb{R}_{\geq 0}$ -semi-module, as explained below.

If $\varphi > 0$ is a positive admissible function, the set $\text{QL}_\varphi(\Gamma, \mathbf{d})$ admits an action of the multiplicative semigroup $(\mathbb{R}_{\geq 0}, \cdot)$ defined in the following way. Let $\lambda \in \mathbb{R}_{\geq 0}$ and $\tilde{\gamma} = \{\gamma(n)\} \in \text{QL}_\varphi(\Gamma, \mathbf{d})$. Then define the action of λ on $\tilde{\gamma}$ by

$$\lambda \cdot \tilde{\gamma} := \{\gamma(\lfloor \lambda \cdot n \rfloor)\}_{n \in \mathbb{N}_0} \tag{2.5}$$

To show that $\tilde{\gamma} := \lambda \cdot \tilde{\gamma}$ belongs to $\text{QL}_\varphi(\Gamma, \mathbf{d})$ we bound its defect as follows. Let $m, n \in \mathbb{N}_0$, and define $\epsilon := \lfloor \lambda(m+n) \rfloor - \lfloor \lambda \cdot m \rfloor - \lfloor \lambda \cdot n \rfloor \in \{0, 1\}$. In the computation below we assume that $\lambda \geq 1$. For $\lambda \in [0, 1]$ the computation is similar, but simpler. We estimate

$$\begin{aligned} &\mathbf{d}(\tilde{\gamma}(m+n), \tilde{\gamma}(m) + \tilde{\gamma}(n)) \\ &= \mathbf{d}(\gamma(\lfloor \lambda(m+n) \rfloor), \gamma(\lfloor \lambda \cdot m \rfloor) + \gamma(\lfloor \lambda \cdot n \rfloor)) \\ &= \mathbf{d}(\gamma(\lfloor \lambda \cdot m \rfloor + \lfloor \lambda \cdot n \rfloor + \epsilon), \gamma(\lfloor \lambda \cdot m \rfloor) + \gamma(\lfloor \lambda \cdot n \rfloor)) \\ &\leq \mathbf{d}(\gamma(\lfloor \lambda \cdot m \rfloor) + \gamma(\lfloor \lambda \cdot n \rfloor) + \gamma(\epsilon), \gamma(\lfloor \lambda \cdot m \rfloor) + \gamma(\lfloor \lambda \cdot n \rfloor)) \\ &\quad + 2\varphi(\lfloor \lambda \cdot m \rfloor + \lfloor \lambda \cdot n \rfloor + \epsilon) \\ &\leq \mathbf{d}(\gamma(\epsilon), 0) + 2\varphi(\lfloor \lambda(m+n) \rfloor) \\ &\leq \mathbf{d}(\gamma(1), 0) + 2\varphi(\lambda(m+n)) \\ &\leq C \cdot \varphi(m+n) \end{aligned}$$

The first inequality above is the bound (2.3) and the last inequality is obtained by applying Lemma 2.3.

The action defined above is only an action up to asymptotic equivalence. Similarly, in the constructions that follow we are tacitly assuming they are valid up to asymptotic equivalence.

The action

$$\cdot : \mathbb{R}_{\geq 0} \times \text{QL}_\varphi(\Gamma, \mathbf{d}) \rightarrow \text{QL}_\varphi(\Gamma, \mathbf{d})$$

is a homothety (dilation)

$$\hat{\mathbf{d}}(\lambda \cdot \bar{\gamma}_1, \lambda \cdot \bar{\gamma}_2) = \lambda \cdot \hat{\mathbf{d}}(\bar{\gamma}_1, \bar{\gamma}_2)$$

and therefore it is continuous with respect to \mathbf{d} .

The semigroup structure on $\text{QL}_\varphi(\Gamma, \mathbf{d})$ is distributive with respect to the $\mathbb{R}_{\geq 0}$ -action

$$\begin{aligned} \lambda \cdot (\bar{\gamma}_1 + \bar{\gamma}_2) &= \lambda \cdot \bar{\gamma}_1 + \lambda \cdot \bar{\gamma}_2 \\ (\lambda_1 + \lambda_2) \cdot \bar{\gamma} &\stackrel{\hat{\mathbf{d}}}{=} \lambda_1 \cdot \bar{\gamma} + \lambda_2 \cdot \bar{\gamma} \end{aligned}$$

In particular, for $n \in \mathbb{N}$ and $\bar{\gamma} \in \text{QL}_\varphi(\Gamma, \mathbf{d})$

$$\underbrace{\bar{\gamma} + \cdots + \bar{\gamma}}_n \stackrel{\hat{\mathbf{d}}}{=} n \cdot \bar{\gamma}$$

2.3.6 Completeness

Here, we introduce additional conditions on a metric Abelian monoid $(\Gamma, +, \mathbf{d})$, that guarantee that $(\text{QL}_\varphi(\Gamma), \hat{\mathbf{d}})$ is a complete metric space.

Suppose φ is an admissible function and $(\Gamma, +, \mathbf{d})$ is a metric Abelian monoid satisfying the following additional property: there exists a constant $C > 0$, such that for any quasi-linear sequence $\bar{\gamma} \in \text{QL}_\varphi(\Gamma, \mathbf{d})$, there exists an asymptotically equivalent quasi-linear sequence $\bar{\gamma}'$ with defect bounded by $C\varphi$. Note that, contrary to the situation in the definition of $\text{QL}_\varphi(\Gamma, \mathbf{d})$, the constant C is now not allowed to depend on the sequence. If this is the case, we say that $\text{QL}_\varphi(\Gamma, \mathbf{d})$ has the *(C-)uniformly bounded defect property*.

Proposition 2.6 *Suppose a metric Abelian monoid $(\Gamma, +, \delta)$ and an admissible function $\varphi > 0$ are such that $(\text{QL}_\varphi(\Gamma, \delta), \hat{\delta})$ has the uniformly bounded defect property and the distance function δ is homogeneous. Then the space $(\text{QL}_\varphi(\Gamma, \delta), \hat{\delta})$ is complete.*

Proof Given a Cauchy sequence $\{\bar{\gamma}_i\}$ of elements in $(\text{QL}_\varphi(\Gamma, \delta), \hat{\delta})$ we need to find a limit element $\bar{\eta} \in \text{QL}_\varphi(\Gamma, \delta)$. We will construct $\bar{\eta}$ by a diagonal argument. First we replace each element of the sequence $\{\bar{\gamma}_i\}$ by an asymptotically equivalent element with defect bounded by $C\varphi$ according to the assumption of the proposition. We will still call the new sequence $\{\bar{\gamma}_i\}$. In fact, we may without loss of generality assume that $C = 1$.

We begin by establishing a bound on the divergence of the tails of sequences $\bar{\gamma}_i$ and $\bar{\gamma}_j$. By homogeneity of δ , the triangle inequality and Corollary 2.5, it holds for any $n, k \in \mathbb{N}$ that

$$\begin{aligned} k \cdot \delta(\gamma_i(n), \gamma_j(n)) &= \delta(k \cdot \gamma_i(n), k \cdot \gamma_j(n)) \\ &\leq \delta(\gamma_i(k \cdot n), \gamma_j(k \cdot n)) + 2k \cdot D_\varphi \cdot \varphi(n) \end{aligned}$$

Dividing by k and passing to the limit $k \rightarrow \infty$, while keeping n fixed, we obtain

$$\delta(\gamma_i(n), \gamma_j(n)) \leq n \cdot \hat{\delta}(\bar{\gamma}_i, \bar{\gamma}_j) + 2D_\varphi \cdot \varphi(n)$$

Since the sequence $(\bar{\gamma}_i)_{i \in \mathbb{N}_0}$ is Cauchy, it follows that for any $n \in \mathbb{N}$ there is a number $\mathbf{i}(n) \in \mathbb{N}$ such that for any $i, j \geq \mathbf{i}(n)$ holds

$$\hat{\delta}(\bar{\gamma}_i, \bar{\gamma}_j) \leq \frac{1}{n}$$

Then for any $i, j, n \in \mathbb{N}$ with $i, j \geq \mathbf{i}(n)$ we have the following bound

$$\delta(\gamma_i(n), \gamma_j(n)) \leq 2D_\varphi \cdot \varphi(n) + 1 \tag{2.6}$$

Now we are ready to define the limiting sequence $\bar{\eta}$ by setting

$$\eta(n) := \gamma_{\mathbf{i}(n)}(n)$$

First we verify that $\bar{\eta}$ is quasi-linear. For $m, n \in \mathbb{N}$, we have

$$\begin{aligned} \delta(\eta(n+m), \eta(n) + \eta(m)) &= \delta(\gamma_{\mathbf{i}(n+m)}(n+m), \gamma_{\mathbf{i}(n)}(n) + \gamma_{\mathbf{i}(m)}(m)) \\ &\leq \delta(\gamma_{\mathbf{i}(n+m)}(n+m), \gamma_{\mathbf{i}(n+m)}(n) + \gamma_{\mathbf{i}(n+m)}(m)) \\ &\quad + \delta(\gamma_{\mathbf{i}(n+m)}(n) + \gamma_{\mathbf{i}(n+m)}(m), \gamma_{\mathbf{i}(n)}(n) + \gamma_{\mathbf{i}(m)}(m)) \\ &\leq \varphi(n+m) + 2D_\varphi \cdot \varphi(n) + 1 + 2D_\varphi \cdot \varphi(m) + 1 \\ &\leq (4D_\varphi + 1)\varphi(n+m) + 2 \leq C' \cdot \varphi(n+m) \end{aligned}$$

for some constant $C' > 0$.

The convergence of $\bar{\gamma}_i$ to $\bar{\eta}$ is shown as follows. For $n, k \in \mathbb{N}$ let $q_n, r_n \in \mathbb{N}_0$ be the quotient and the remainder of the division of n by k , that is $n = q_n \cdot k + r_n$ and $0 \leq r_n < k$. Fix $k \in \mathbb{N}$ and let $i \geq \mathbf{i}(k)$, then

$$\begin{aligned} \hat{\delta}(\bar{\gamma}_i, \bar{\eta}) &= \lim_{n \rightarrow \infty} \frac{1}{n} \delta(\gamma_i(n), \eta(n)) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \delta(\gamma_i(q_n \cdot k + r_n), \gamma_{\mathbf{i}(n)}(q_n \cdot k + r_n)) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \left(q_n \cdot \delta(\gamma_i(k), \gamma_{\mathbf{i}(n)}(k)) + \delta(\gamma_i(r_n), \gamma_{\mathbf{i}(n)}(r_n)) \right. \\ &\quad \left. + 4q_n \cdot D_\varphi \cdot \varphi(k) + 2\varphi(n) \right) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \left(q_n \cdot (2D_\varphi \cdot \varphi(k) + 1) + (2D_\varphi \cdot \varphi(r_n) + 1) + \right. \\ &\quad \left. + 4q_n \cdot D_\varphi \cdot \varphi(k) + 2\varphi(n) \right) \\ &= C'' \cdot \varphi(k)/k \end{aligned}$$

Since $k \in \mathbb{N}$ is arbitrary and φ is sub-linear we have

$$\lim_{i \rightarrow \infty} \hat{\mathbf{d}}(\bar{\gamma}_i, \bar{\eta}) = 0$$

□

2.3.7 On the density of linear sequences

For a metric Abelian monoid $(\Gamma, +, \mathbf{d})$ together with an admissible function φ we say that $QL_\varphi(\Gamma, \mathbf{d})$ has the *vanishing defect property* if for every $\epsilon > 0$ and for every $\bar{\gamma} \in QL_\varphi(\Gamma, \mathbf{d})$ there exists an asymptotically equivalent quasi-linear sequence $\bar{\gamma}'$ with defect bounded by another admissible function ψ such that $\int_1^\infty \frac{\psi(t)}{t^2} dt < \epsilon$.

The proposition below gives a sufficient condition under which the linear sequences are dense in the space of quasi-linear sequences.

Proposition 2.7 *Suppose $QL_\varphi(\Gamma, +, \mathbf{d})$ has the vanishing defect property. Then $L(\Gamma, \mathbf{d})$ is dense in $QL_\varphi(\Gamma, \mathbf{d})$.*

Proof Let $\bar{\gamma} = \{\gamma(n)\}$ be a quasi-linear sequence. For any $i \in \mathbb{N}$ select a sequence $\bar{\gamma}_i$ asymptotically equivalent to $\bar{\gamma}$ with defect bounded by an admissible function φ_i such that $\int_1^\infty \frac{\varphi_i(t)}{t^2} dt < 1/i$ according to the “vanishing defect” assumption of the lemma.

Define $\bar{\eta}_i$ by

$$\eta_i(n) := n \cdot \gamma_i(1)$$

Then

$$\begin{aligned} \hat{\mathbf{d}}(\bar{\gamma}, \bar{\eta}_i) &= \hat{\mathbf{d}}(\bar{\gamma}_i, \bar{\eta}_i) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{d}(\gamma_i(n), \eta_i(n)) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{d}(\gamma_i(n), n \cdot \gamma_i(1)) \\ &\leq 8 \int_1^\infty \frac{\varphi_i(t)}{t^2} dt \leq \frac{8}{i} \end{aligned}$$

where we used Lemma 2.4 in the first inequality. Thus, any quasi-linear sequence can be approximated by linear sequences. □

2.3.8 Asymptotic distance on original monoid

Starting with an element $\gamma \in \Gamma$ one can construct a linear sequence $= \{i \cdot \gamma\}_{i \in \mathbb{N}_0}$. In view of Proposition 2.1, the map

$$\cdot : (\Gamma, \mathbf{d}) \rightarrow (L(\Gamma, \mathbf{d}), \hat{\mathbf{d}}) \tag{2.7}$$

is a contraction.

The inclusion in (2.7) induces a metric δ on Γ , satisfying for any $\gamma_1, \gamma_2 \in \Gamma$

$$\delta(\gamma_1, \gamma_2) \leq \mathbf{d}(\gamma_1, \gamma_2) \tag{2.8}$$

and the following homogeneity condition

$$\delta(n \cdot \gamma_1, n \cdot \gamma_2) = n \cdot \delta(\gamma_1, \gamma_2) \tag{2.9}$$

for all $n \in \mathbb{N}_0$.

Note that if \mathbf{d} was homogeneous to begin with, then δ coincides with \mathbf{d} on Γ .

By virtue of the bound $\delta \leq \mathbf{d}$, sequences that are quasi-linear with respect to \mathbf{d} are also quasi-linear with respect to δ . Since δ is scale-invariant, the associated asymptotic distance $\hat{\delta}$ coincides with δ on Γ . We will show (in Lemma 2.8 below) that $\hat{\delta}$ also coincides with $\hat{\mathbf{d}}$ on \mathbf{d} -quasi-linear sequences.

Let φ be an admissible function. In order to organize all these statements, and to be more precise, let us include the spaces in the following commutative diagram.

$$\begin{array}{ccc}
 & (L(\Gamma, \mathbf{d}), \hat{\mathbf{d}}) & \xrightarrow{J_1} & (QL_\varphi(\Gamma, \mathbf{d}), \hat{\mathbf{d}}) \\
 \begin{array}{c} f \\ f' \end{array} \nearrow & \downarrow \iota_1 & & \downarrow \iota_2 \\
 (\Gamma, \delta) & & & (L(\Gamma, \delta), \hat{\delta}) \xrightarrow{J_2} (QL_\varphi(\Gamma, \delta), \hat{\delta})
 \end{array} \tag{2.10}$$

The maps f, f' and ι_1 are isometries. The maps J_1 and J_2 are isometric embeddings. The next lemmas show that ι_2 is also an isometric embedding, and it has dense image.

Lemma 2.8 *Let φ be a positive, admissible function. Then, the natural inclusion*

$$\iota_2 : (QL_\varphi(\Gamma, \mathbf{d}), \hat{\mathbf{d}}) \hookrightarrow (QL_\varphi(\Gamma, \delta), \hat{\delta})$$

is an isometric embedding with the dense image.

Proof First we show that the map ι_2 is an isometric embedding. Let $\bar{\gamma}_1, \bar{\gamma}_2 \in QL_\varphi(\Gamma, \mathbf{d})$ be two φ -quasi-linear sequences with respect to the distance function \mathbf{d} . We have to show that the two numbers

$$\hat{\mathbf{d}}(\bar{\gamma}_1, \bar{\gamma}_2) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{d}(\gamma_1(n), \gamma_2(n))$$

and

$$\hat{\delta}(\bar{\gamma}_1, \bar{\gamma}_2) = \lim_{n \rightarrow \infty} \frac{1}{n} \delta(\gamma_1(n), \gamma_2(n))$$

are equal. Since shifts are non-expanding maps, we have $\delta \leq \mathbf{d}$ and it follows immediately that

$$\hat{\delta}(\bar{\gamma}_1, \bar{\gamma}_2) \leq \hat{\mathbf{d}}(\bar{\gamma}_1, \bar{\gamma}_2)$$

and we are left to show the opposite inequality. We will do it as follows. Fix $n > 0$, then

$$\begin{aligned} \hat{\mathbf{d}}(\bar{\gamma}_1, \bar{\gamma}_2) &= \lim_{k \rightarrow \infty} \frac{1}{k \cdot n} \mathbf{d}(\gamma_1(k \cdot n), \gamma_2(k \cdot n)) \\ &\leq \lim_{k \rightarrow \infty} \frac{1}{k \cdot n} \left(\mathbf{d}(k \cdot \gamma_1(n), k \cdot \gamma_2(n)) + 2k \cdot D_\varphi \cdot \varphi(n) \right) \\ &\leq \frac{1}{n} \delta(\gamma_1(n), \gamma_2(n)) + 2D_\varphi \frac{\varphi(n)}{n} \end{aligned}$$

Passing to the limit with respect to n gives the required inequality

$$\hat{\mathbf{d}}(\bar{\gamma}_1, \bar{\gamma}_2) \leq \hat{\delta}(\bar{\gamma}_1, \bar{\gamma}_2)$$

Now we will show that the image of ι_2 is dense. Given an element $\bar{\gamma} = \{\gamma(n)\}$ in $QL_\varphi(\Gamma, \hat{\mathbf{d}})$ we have to find a $\hat{\delta}$ -approximating sequence $\bar{\gamma}_i = \{\gamma_i(n)\}$ in $QL_\varphi(\Gamma, \mathbf{d})$. Define

$$\gamma_i(n) := \left\lfloor \frac{n}{i} \right\rfloor \cdot \gamma(i)$$

We have to show that each $\bar{\gamma}_i$ is \mathbf{d} -quasi-linear and that $\hat{\delta}(\bar{\gamma}_i, \bar{\gamma}) \xrightarrow{i \rightarrow \infty} 0$. These statements follow from

$$\begin{aligned} \mathbf{d}(\gamma_i(m+n), \gamma_i(m) + \gamma_i(n)) &= \mathbf{d} \left(\left\lfloor \frac{m+n}{i} \right\rfloor \cdot \gamma(i), \left\lfloor \frac{m}{i} \right\rfloor \cdot \gamma(i) + \left\lfloor \frac{n}{i} \right\rfloor \cdot \gamma(i) \right) \\ &\leq \mathbf{d}(\gamma(i), \mathbf{0}) \\ &\leq C_i \cdot \varphi(m+n) \end{aligned}$$

for some $C_i > 0$. It is worth noting that the defect of $\bar{\gamma}_i$ may not be bounded uniformly with respect to i . Finally, it holds that

$$\begin{aligned} \hat{\delta}(\bar{\gamma}_i, \bar{\gamma}) &= \lim_{n \rightarrow \infty} \frac{1}{n} \delta(\gamma_i(n), \gamma(n)) = \lim_{n \rightarrow \infty} \frac{1}{n} \delta \left(\left\lfloor \frac{n}{i} \right\rfloor \cdot \gamma(i), \gamma(n) \right) \\ &\leq \lim_{n \rightarrow \infty} \left[\frac{1}{n} \delta(\gamma(i \lfloor \frac{n}{i} \rfloor), \gamma(n)) + \frac{1}{n} \left\lfloor \frac{n}{i} \right\rfloor \cdot D_\varphi \cdot \varphi(i) \right] \\ &\leq \lim_{n \rightarrow \infty} \left[\frac{1}{n} \max_{k=0, \dots, i-1} \delta(\gamma(k), \mathbf{0}) + \frac{1}{n} \varphi(n) \right] + D_\varphi \frac{\varphi(i)}{i} = D_\varphi \frac{\varphi(i)}{i} \xrightarrow{i \rightarrow \infty} 0 \end{aligned}$$

□

The difference between the distance functions $\hat{\mathbf{d}}, \delta$ and $\hat{\delta}$ is very small: $\hat{\mathbf{d}}$ and δ are defined on the dense subset of the domain of definition of $\hat{\delta}$ and they coincide whenever are both defined. From now on we will write \mathbf{d} for the original distance function and δ for the asymptotic metric on both the monoid and its tropicalization.

3 Grothendieck construction

Given an Abelian monoid with the cancellation property, there is a minimal Abelian group (called the Grothendieck Group of the monoid), into which it isomorphically embeds. Similarly, an $\mathbb{R}_{\geq 0}$ -semi-module naturally embeds into a normed vector space. A nice example of this construction applied to the semi-module of convex sets in \mathbb{R}^n (with the Minkowski sum and the Hausdorff distance) can be found in [12].

Proposition 3.1 *Let $(\Gamma, +, \cdot, \delta)$ be a complete metric Abelian monoid with $\mathbb{R}_{\geq 0}$ action (an $\mathbb{R}_{\geq 0}$ -semi-module) with homogeneous pseudo-metric δ . Then there exists a Banach space $(\mathbf{B}, \|\cdot\|)$ and a distance-preserving homomorphism of monoids*

$$f : \Gamma \rightarrow \mathbf{B}$$

such that the image of f is a closed convex cone.

If δ is a proper pseudo-metric (not a metric), then the map f is not injective.

Proof By Lemma 2.2 the pseudo-metric δ is translation invariant. We can therefore apply the Grothendieck construction to define a normed vector space \mathbf{B}_0 : Define

$$\mathbf{B}_0 := \{(x, y) : x, y \in \Gamma\} / \sim$$

where $(x, y) \sim (x', y')$ if there are $z, z' \in \Gamma$, such that $x + z \stackrel{\delta}{=} x' + z'$ and $y + z \stackrel{\delta}{=} y' + z'$.

Define also addition, multiplication by a scalar and a norm on \mathbf{B}_0 by setting for all $x, y, x', y' \in \Gamma$ and $\lambda \in \mathbb{R}$

$$\begin{aligned} (x, y) + (x', y') &:= (x + x', y + y') \\ (-1) \cdot (x, y) &:= (y, x) \\ \lambda \cdot (x, y) &:= \text{sign}(\lambda) \cdot (|\lambda| \cdot x, |\lambda| \cdot y) \\ \|(x, y)\| &:= \delta(x, y) \end{aligned}$$

These operations respect the equivalence relation and turn $(\mathbf{B}_0, +, \cdot, \|\cdot\|)$ into a normed vector-space. The map f defined by

$$f : \Gamma \rightarrow \mathbf{B}_0, \quad x \mapsto (x, \mathbf{0})$$

is a well-defined distance-preserving homomorphism.

That $f(\Gamma)$ is closed immediately follows as Γ is complete and f is distance-preserving.

In general, the space \mathbf{B}_0 is not complete. We define the Banach space \mathbf{B} as the completion of the normed vector space \mathbf{B}_0 . \square

4 Tropical probability spaces and their diagrams

4.1 Diagrams of probability spaces

We will now briefly describe the construction of diagrams of probability spaces, see [9] for a more detailed discussion. By a *finite probability space* we will mean a set (not necessarily finite) with a probability measure, such that the support of the measure is finite. For such probability space X we denote by $|X|$ the cardinality of the support of probability measure and the expression $x \in X$ will mean, that x is an *atom* in X , which is a point of positive weight in the underlying set.

We will consider commutative diagrams of finite probability spaces, where arrows are equivalence classes of measure-preserving maps. Two maps are considered equivalent if they coincide on a set of full measure and such equivalence classes will be called *reductions*.

Three examples of diagrams of probability spaces are pictured in (1.1). The combinatorial structure of such a commutative diagram can be recorded by an object \mathbf{G} , which could be equivalently considered as a special type of category, a finite poset, or a directed acyclic graph (DAG) with additional properties. We will call such objects simply *indexing categories*. Below we briefly recall the definition.

An *indexing category* is a finite category such that for any pair of objects there exists at most one morphism between them in either direction, and such that it satisfies the following property. For any pair of objects i, j in an indexing category \mathbf{G} there exists a *least common ancestor*, i.e. an object k such that there are morphisms $k \rightarrow i$ and $k \rightarrow j$ in \mathbf{G} and such that for any other object l admitting morphisms $l \rightarrow i$ and $l \rightarrow j$, there is also a morphism $l \rightarrow k$.

By $[\mathbf{G}]$ we denote the number of objects in the indexing category, or equivalently the number of vertices in the DAG or the number of points in the poset \mathbf{G} . An important class of examples of indexing categories is formed by so-called *full categories* Λ_n , that correspond to the poset of non-empty subsets of a set $\{1, \dots, n\}$ ordered by coinclusion. If $n = 2$, we call the category

$$\Lambda_2 = (O_1 \leftarrow O_{\{1,2\}} \rightarrow O_2)$$

a fan. We refer to the objects O_1 and O_2 as the feet of the fan and to O_{12} as the initial object. We use the same terminology for the spaces in a diagram indexed by Λ_2 .

The space of all commutative diagrams of a fixed combinatorial type will be denoted $\mathbf{Prob}(\mathbf{G})$. A morphism between two diagrams $\mathcal{X}, \mathcal{Y} \in \mathbf{Prob}(\mathbf{G})$ is, by definition, a natural transformation between functors \mathcal{X} and \mathcal{Y} . Essentially, it is a collection of morphisms between corresponding individual spaces in \mathcal{X} and \mathcal{Y} , that commute with

morphisms within the diagrams \mathcal{X} and \mathcal{Y} . We call such morphisms *reductions of diagrams*.

The construction of forming commutative diagrams could be iterated, producing diagrams of diagrams. Especially important will be two-fans of \mathbf{G} -diagrams, the space of which will be denoted $\mathbf{Prob}(\mathbf{G})\langle\mathbf{\Lambda}_2\rangle$.

A two-fan \mathcal{X} will be called *minimal*, if for any morphism of \mathcal{X} to another two-fan \mathcal{Y} , the following holds: if the induced morphisms on the feet are isomorphisms, then the top morphism is also an isomorphism. Any \mathbf{G} -diagram will be called minimal if for any sub-diagram, which is a two-fan, it contains a minimal two-fan with the same feet.

Given an n -tuple (X_1, \dots, X_n) of finite-valued random variables, one can construct a minimal $\mathbf{\Lambda}_n$ -diagram $\mathcal{X} = \{X_I; \chi_{IJ}\}$ by setting for any $\emptyset \neq I \subset \{1, \dots, n\}$

$$X_I = \prod_{i \in I} X_i$$

where X_i is the target space of random variable X_i , and the probabilities are the induced distributions. For the diagram constructed in such a way we will write $\mathcal{X} = \langle X_1, \dots, X_n \rangle$. On the other hand, any $\mathbf{\Lambda}_n$ -diagram gives rise to the n -tuple of random variables with the domain of definition being the initial space and the targets being the spaces indexed by one-point sets.

The *constant diagram* $X^{\mathbf{G}}$ is \mathbf{G} -diagram in which all the spaces are isomorphic to a single probability space X and all the morphisms are identity maps. In particular, we denote by $\{\bullet\}^{\mathbf{G}}$ the \mathbf{G} -diagram consisting entirely of one-point spaces.

The tensor product $\mathcal{X} \otimes \mathcal{Y}$ of two \mathbf{G} -diagrams is defined by taking the tensor product of corresponding probability spaces and the Cartesian product of maps. The diagram $\{\bullet\}^{\mathbf{G}}$ is a unit with respect to the tensor product. Certain care should be exercised here, since the associativity, commutativity and unity of $\{\bullet\}^{\mathbf{G}}$ for the tensor product only hold up to isomorphism.

For a diagram $\mathcal{X} \in \mathbf{Prob}(\mathbf{G})$ one can evaluate entropies of the individual spaces. The corresponding map will be denoted

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

where the target space is the space of \mathbb{R} -valued functions on objects in \mathbf{G} and it is equipped with the ℓ^1 -norm.

For a two-fan $\mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y})$ of \mathbf{G} -diagrams define the *entropy distance*

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

We interpret $\text{kd}(\mathcal{F})$ as a measure of deviation of \mathcal{F} from being an isomorphism between the diagrams \mathcal{X} and \mathcal{Y} . Indeed, $\text{kd}(\mathcal{F}) = 0$ if and only if the two morphisms in \mathcal{F} are isomorphisms, see [9].

We define the *intrinsic entropy distance* \mathbf{k} on the space $\mathbf{Prob}(\mathbf{G})$ by

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) := \inf \{ \text{kd}(\mathcal{F}) : \mathcal{F} = (\mathcal{X} \leftarrow \mathcal{Z} \rightarrow \mathcal{Y}) \in \mathbf{Prob}(\mathbf{G})\langle\mathbf{\Lambda}_2\rangle \}$$

Note that according to the definitions used in this article any indexing category must have an initial object. In [9] such indexing categories and diagrams indexed by such categories were called *complete*. Therefore purely by a change of names, results that in [9] were said to hold for complete indexing categories, hold for the indexing categories of this article.

The tensor product is 1-Lipschitz with respect to \mathbf{k} , thus $(\mathbf{Prob}(\mathbf{G}), \otimes, \mathbf{k})$ is a metric Abelian monoid and $\text{Ent}_* : (\mathbf{Prob}(\mathbf{G}), \otimes, \mathbf{k}) \rightarrow (\mathbb{R}^{[\mathbf{G}]}, \|\cdot\|_1)$ is a 1-Lipschitz homomorphism. For proofs and more detailed discussion the reader is referred to [9].

4.2 Tropical diagrams

In this section we apply the general construction in Sects. 2 and 3 to the metric Abelian monoid $(\mathbf{Prob}(\mathbf{G}), \otimes, \cdot, \kappa)$.

We define the asymptotic distance on $\mathbf{Prob}(\mathbf{G})$ by

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

One of the main tools for the estimation of the (asymptotic) distance is the so-called Slicing Lemma, [9, Proposition 3.9]. We will only need its corollary that we formulate below in Proposition 4.1. For a diagram \mathcal{X} , a space U in it and an atom $u \in U$, we may form a conditioned diagram $\mathcal{X}|u$ by conditioning all the spaces in \mathcal{X} on u .

Proposition 4.1 *Let \mathbf{G} be an indexing category, $\mathcal{X}, \mathcal{Y} \in \mathbf{Prob}(\mathbf{G})$ and $U^{\mathbf{G}} \in \mathbf{Prob}(\mathbf{G})$.*

(1) *Let $\mathcal{X} \rightarrow U^{\mathbf{G}}$ be a reduction, then*

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) \leq \int_U \mathbf{k}(\mathcal{X}|u, \mathcal{Y}) d p_U(u) + [\mathbf{G}] \cdot \text{Ent}(U)$$

(2) *For a “co-fan” $\mathcal{X} \rightarrow U^{\mathbf{G}} \leftarrow \mathcal{Y}$ holds*

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) \leq \int_U \mathbf{k}(\mathcal{X}|u, \mathcal{Y}|u) d p_U(u)$$

The statements and the proofs of the Slicing Lemma and its consequences can be found in [9].

We will show below that $(\mathbf{Prob}(\mathbf{G}), \otimes, \kappa)$ has the uniformly bounded and vanishing defect properties. For this purpose we need to develop some technical tools.

4.3 Mixtures

The input data for the mixture operation is a family of \mathbf{G} -diagrams, parameterized by a probability space. As a result one obtains another \mathbf{G} -diagram with pre-specified conditionals. One particular instance of a mixture is when one mixes two diagrams \mathcal{X} and $\{\bullet\}^{\mathbf{G}}$, the latter being a constant \mathbf{G} -diagram of one-point probability spaces. This operation will be used as a substitute for taking radicals “ $\mathcal{X}_n^{\frac{1}{n}}$ ” below.

4.3.1 Definition of mixtures

Let \mathbf{G} be an indexing category and Θ be a probability space. By $\Theta^{\mathbf{G}}$ we denote the *constant \mathbf{G} -diagram*—the diagram such that all spaces in it are Θ and all morphisms are identity morphisms. Let $\{\mathcal{X}_\theta\}_{\theta \in \Theta}$ be a family of \mathbf{G} -diagrams parameterized by Θ . The *mixture* of the family $\{\mathcal{X}_\theta\}$ is the reduction

$$\text{Mix} \{\mathcal{X}_\theta\} = (\mathcal{Y} \longrightarrow \Theta^{\mathbf{G}})$$

such that

$$\mathcal{Y}|\theta \cong \mathcal{X}_\theta \quad \text{for any } \theta \in \Theta \tag{4.1}$$

The mixture exists and is uniquely defined by property (4.1) up to an isomorphism which is identity on $\Theta^{\mathbf{G}}$.

We denote the top diagram \mathcal{Y} of the mixture by

$$\mathcal{Y} =: \bigoplus_{\theta \in \Theta} \mathcal{X}_\theta$$

and also call it the mixture of the family $\{\mathcal{X}_\theta\}$.

When

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

is a binary space we write simply

$$\mathcal{X}_{\blacksquare} \oplus_{\mathbb{B}_\alpha} \mathcal{X}_{\square}$$

for the mixture. The diagram subindexed by the \blacksquare will always be the first summand.

The entropy of the mixture can be evaluated by the following formula

$$\text{Ent}_* \left(\bigoplus_{\theta \in \Theta} \mathcal{X}_\theta \right) = \int_{\Theta} \text{Ent}_*(\mathcal{X}_\theta) \, d p(\theta) + \text{Ent}_*(\Theta^{\mathbf{G}})$$

Mixtures satisfy the distributive law with respect to the tensor product

$$\begin{aligned} \text{Mix}(\{\mathcal{X}_\theta\}_{\theta \in \Theta}) \otimes \text{Mix}(\{\mathcal{Y}_{\theta'}\}_{\theta' \in \Theta'}) &\cong \text{Mix}(\{\mathcal{X}_\theta \otimes \mathcal{Y}_{\theta'}\}_{(\theta, \theta') \in \Theta \otimes \Theta'}) \\ \left(\bigoplus_{\theta \in \Theta} \mathcal{X}_\theta \right) \otimes \left(\bigoplus_{\theta' \in \Theta'} \mathcal{Y}_{\theta'} \right) &\cong \bigoplus_{(\theta, \theta') \in \Theta \otimes \Theta'} (\mathcal{X}_\theta \otimes \mathcal{Y}_{\theta'}) \end{aligned}$$

4.3.2 The distance estimates for the mixtures

The mixture of a \mathbf{G} -diagram with the constant diagram of one-point spaces $\{\bullet\}^{\mathbf{G}}$ may serve as an substitute of taking radicals of the diagram. The following lemma provides a justification of this by some distance estimates related to mixtures and will be used below.

Lemma 4.2 *Let \mathbf{G} be a complete indexing category and $\mathcal{X}, \mathcal{Y} \in \mathbf{Prob}(\mathbf{G})$. Then*

1. $\kappa(\mathcal{X}, \mathcal{X}^n \oplus_{\mathbb{B}_{1/n}} \{\bullet\}) \leq \text{Ent}(\mathbb{B}_{1/n})$
2. $\kappa(\mathcal{X}, (\mathcal{X} \oplus_{\mathbb{B}_{1/n}} \{\bullet\})^n) \leq n \cdot \text{Ent}(\mathbb{B}_{1/n})$
3. $\kappa((\mathcal{X} \otimes \mathcal{Y}) \oplus_{\mathbb{B}_{1/n}} \{\bullet\}, (\mathcal{X} \oplus_{\mathbb{B}_{1/n}} \{\bullet\}) \otimes (\mathcal{Y} \oplus_{\mathbb{B}_{1/n}} \{\bullet\})) \leq 3\text{Ent}(\mathbb{B}_{1/n})$
4. $\kappa((\mathcal{X} \oplus_{\mathbb{B}_{1/n}} \{\bullet\}), (\mathcal{Y} \oplus_{\mathbb{B}_{1/n}} \{\bullet\})) \leq \frac{1}{n} \kappa(\mathcal{X}, \mathcal{Y})$

Note that the distance estimates in the lemma above are with respect to the asymptotic distance. This is essential, since from the perspective of the intrinsic distance mixtures are very badly behaved.

Proof For $\lambda \in \mathbb{B}_{1/n}^N$, define $\mathbf{q}(\lambda)$ to be the number of black squares in the sequence λ . It is a binomially distributed random variable with mean N/n and variance $\frac{N}{n}(1 - \frac{1}{n})$.

The first claim is then proven by the following calculation

$$\begin{aligned} &\kappa(\mathcal{X}, \mathcal{X}^n \oplus_{\mathbb{B}_{1/n}} \{\bullet\}) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{k}(\mathcal{X}^N, (\mathcal{X}^n \oplus_{\mathbb{B}_{1/n}} \{\bullet\})^N) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{k}\left(\mathcal{X}^N, \bigoplus_{\lambda \in \mathbb{B}_{1/n}^N} \mathcal{X}^{n \cdot \mathbf{q}(\lambda)}\right) \\ &\leq \text{Ent}(\mathbb{B}_{1/n}) + \lim_{N \rightarrow \infty} \frac{1}{N} \int_{\lambda \in \mathbb{B}_{1/n}^N} \mathbf{k}(\mathcal{X}^N, \mathcal{X}^{n \cdot \mathbf{q}(\lambda)}) dP(\lambda) \\ &\leq \text{Ent}(\mathbb{B}_{1/n}) + \|\text{Ent}_*(\mathcal{X})\|_1 \cdot \lim_{N \rightarrow \infty} \frac{n}{N} \cdot \int_{\lambda \in \mathbb{B}_{1/n}^N} |N/n - \mathbf{q}(\lambda)| dP(\lambda) \\ &\leq \text{Ent}(\mathbb{B}_{1/n}) + \|\text{Ent}_*(\mathcal{X})\|_1 \cdot \lim_{N \rightarrow \infty} \frac{n}{N} \cdot \sqrt{N \cdot \frac{1}{n} (1 - \frac{1}{n})} = \text{Ent}(\mathbb{B}_{1/n}) \end{aligned}$$

where we used Proposition 4.1(1) for the inequality on the third line above, and the following estimate: for any diagram \mathcal{A} and integers $0 \leq m \leq n$

$$\mathbf{k}(\mathcal{A}^n, \mathcal{A}^m) \leq \mathbf{k}(\mathcal{A}^{n-m}, 0) = \|\text{Ent}_*(\mathcal{A})\|_1 \cdot (n - m)$$

The second claim is proven similarly and the third follows from the second and the 1-Lipschitz property of the tensor product:

$$\begin{aligned} &\kappa((\mathcal{X} \otimes \mathcal{Y}) \oplus_{\mathbb{B}_{1/n}} \{\bullet\}, (\mathcal{X} \oplus_{\mathbb{B}_{1/n}} \{\bullet\}) \otimes (\mathcal{Y} \oplus_{\mathbb{B}_{1/n}} \{\bullet\})) \\ &\leq \kappa((\mathcal{X} \otimes \mathcal{Y}) \oplus_{\mathbb{B}_{1/n}} \{\bullet\}, \mathcal{X} \otimes \mathcal{Y}) + 2\text{Ent}(\mathbb{B}_{1/n}) \\ &\leq 3\text{Ent}(\mathbb{B}_{1/n}) \end{aligned}$$

Finally, the fourth follows from Proposition 4.1(2), by slicing both arguments along $\mathbb{B}_{1/n}$. □

4.4 Vanishing defect property and completeness of the tropical cone

Lemma 4.3 *For every admissible function φ , every $\bar{\mathcal{X}} \in \text{QL}_{\varphi}(\mathbf{Prob} \langle \mathbf{G} \rangle, \kappa)$ and every $k \in \mathbb{N}$, there exists an asymptotically equivalent sequence $\bar{\mathcal{Y}}$ with defect bounded by the admissible function φ_k defined by*

$$\varphi_k(s) := 3\text{Ent}(\mathbb{B}_{1/k}) + \frac{1}{k}\varphi(k \cdot s)$$

Proof Let $\bar{\mathcal{X}} = \{\mathcal{X}(i)\}$ be a quasi-linear sequence with defect bounded by φ and let $k \in \mathbb{N}$.

Define a new sequence $\bar{\mathcal{Y}} = \{\mathcal{Y}(i)\}$ by

$$\mathcal{Y}(i) := (\mathcal{X}(k \cdot i)) \oplus_{\mathbb{B}_{1/k}} \{\bullet\}$$

First we verify that the sequences $\bar{\mathcal{X}}$ and $\bar{\mathcal{Y}}$ are asymptotically equivalent, that is

$$\hat{\kappa}(\bar{\mathcal{X}}, \bar{\mathcal{Y}}) := \lim_{i \rightarrow \infty} \frac{1}{i} \kappa(\mathcal{X}(i), \mathcal{Y}(i)) = 0$$

We estimate the asymptotic distance between individual members of sequences $\bar{\mathcal{X}}$ and $\bar{\mathcal{Y}}$ using Lemma 4.2 and Corollary 2.5 as follows

$$\begin{aligned} \kappa(\mathcal{X}(i), \mathcal{Y}(i)) &= \kappa(\mathcal{X}(i), \mathcal{X}(k \cdot i) \oplus_{\mathbb{B}_{1/k}} \{\bullet\}) \\ &\leq \kappa(\mathcal{X}(i), \mathcal{X}(i)^k \oplus_{\mathbb{B}_{1/k}} \{\bullet\}) + \kappa(\mathcal{X}(i)^k \oplus_{\mathbb{B}_{1/k}} \{\bullet\}, \mathcal{X}(k \cdot i) \oplus_{\mathbb{B}_{1/k}} \{\bullet\}) \\ &\leq \text{Ent}(\mathbb{B}_{1/k}) + D_{\varphi} \cdot \varphi(i) \end{aligned}$$

Thus $\hat{\kappa}(\bar{\mathcal{X}}, \bar{\mathcal{Y}}) = 0$ and the two sequences are asymptotically equivalent. Next we show that the sequence $\bar{\mathcal{Y}}$ is κ -quasi-linear and evaluate its defect, also using Lemma 4.2. Let $i, j \in \mathbb{N}$, then

$$\begin{aligned} &\kappa(\mathcal{Y}(i + j), \mathcal{Y}(i) \otimes \mathcal{Y}(j)) \\ &= \kappa(\mathcal{X}(k \cdot i + k \cdot j) \oplus_{\mathbb{B}_{1/k}} \{\bullet\}, (\mathcal{X}(k \cdot i) \oplus_{\mathbb{B}_{1/k}} \{\bullet\}) \otimes (\mathcal{X}(k \cdot j) \oplus_{\mathbb{B}_{1/k}} \{\bullet\})) \\ &\leq \kappa((\mathcal{X}(k \cdot i) \otimes \mathcal{X}(k \cdot j)) \oplus_{\mathbb{B}_{1/k}} \{\bullet\}, (\mathcal{X}(k \cdot i) \oplus_{\mathbb{B}_{1/k}} \{\bullet\}) \otimes (\mathcal{X}(k \cdot j) \oplus_{\mathbb{B}_{1/k}} \{\bullet\})) \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{k} \varphi(k \cdot (i + j)) \\
 & \leq 3\text{Ent}(\mathbb{B}_{1/k}) + \frac{1}{k} \varphi(k \cdot (i + j))
 \end{aligned}$$

□

Corollary 4.4 *For any indexing category \mathbf{G} and for the admissible function φ given by $\varphi(t) = t^\alpha$, $\alpha \in [0, 1)$, $QL_\varphi(\mathbf{Prob} \langle \mathbf{G} \rangle, \kappa)$ has the uniformly bounded and vanishing defect properties.*

Proof Let $\bar{\mathcal{X}} \in QL_\varphi(\mathbf{Prob} \langle \mathbf{G} \rangle, \kappa)$. By Lemma 4.3 there exists an asymptotically equivalent sequence $\bar{\mathcal{Y}}$ with defect bounded by φ_k defined by

$$\begin{aligned}
 \varphi_k(t) & := 3\text{Ent}(\mathbb{B}_{1/k}) + \frac{1}{k} C \varphi(k \cdot t) \\
 & = 3\text{Ent}(\mathbb{B}_{1/k}) + \frac{1}{k} C (k \cdot t)^\alpha
 \end{aligned}$$

Hence there exists a sequence $c_k \rightarrow 0$ such that for all $t \geq 1$,

$$\varphi_k(t) \leq c_k t^\alpha$$

showing the uniformly bounded and vanishing defect property. □

4.5 Diagrams of tropical probability spaces

By applying the general setup in the previous section to the metric Abelian monoids $(\mathbf{Prob} \langle \mathbf{G} \rangle, \otimes, \mathbf{k})$ and $(\mathbf{Prob} \langle \mathbf{G} \rangle, \otimes, \kappa)$ and using the Corollary 4.4 we obtain the following theorem.

Theorem 4.5 *Fix an admissible function φ and consider the commutative diagram*

$$\begin{array}{ccc}
 & (L(\mathbf{Prob} \langle \mathbf{G} \rangle, \mathbf{k}), \kappa) & \xrightarrow{J_1} & (QL_\varphi(\mathbf{Prob} \langle \mathbf{G} \rangle, \mathbf{k}), \kappa) \\
 (\mathbf{Prob} \langle \mathbf{G} \rangle, \kappa) & \begin{array}{c} \xrightarrow{f} \\ \xrightarrow{f'} \end{array} & \begin{array}{c} \downarrow \iota_1 \\ \downarrow \iota_2 \end{array} & \\
 & (L(\mathbf{Prob} \langle \mathbf{G} \rangle, \kappa), \hat{\kappa}) & \xrightarrow{J_2} & (QL_\varphi(\mathbf{Prob} \langle \mathbf{G} \rangle, \kappa), \hat{\kappa})
 \end{array} \tag{4.2}$$

Then the following statements hold:

1. The maps f, f', ι_1 are isometries.
2. The maps ι_2, J_1, J_2 are isometric embeddings and each map has a dense image in the corresponding target space.
3. The space in the lower-right corner, $(QL_\varphi(\mathbf{Prob} \langle \mathbf{G} \rangle, \kappa), \hat{\kappa})$, is complete.

We would like to conjecture that all maps in the diagram above are isometries.

Since $QL_\varphi(\mathbf{Prob}\langle \mathbf{G} \rangle, \kappa)$ is complete and has $L(\mathbf{Prob}\langle \mathbf{G} \rangle, \kappa)$ as a dense subset for any $\varphi > 0$, it follows that $QL_\varphi(\mathbf{Prob}\langle \mathbf{G} \rangle, \kappa)$ does not depend (up to isometry of pseudo-metric spaces) on the choice of admissible $\varphi > 0$. From now on we will choose the particular function $\varphi(t) := t^{3/4}$. The choice will be clear when we formulate the Asymptotic Equipartition Property for diagrams. We may finally define the space of *tropical G-diagrams*, as the space in the lower-right corner of the diagram

$$\mathbf{Prob}[\mathbf{G}] := (QL_\varphi(\mathbf{Prob}\langle \mathbf{G} \rangle, \kappa), \otimes, \cdot, \hat{\kappa})$$

By Theorem 4.5 above, this space is complete.

The entropy function $\text{Ent}_* : \mathbf{Prob}\langle \mathbf{G} \rangle \rightarrow \mathbb{R}^{[\mathbf{G}]}$ extends to a linear functional

$$\text{Ent}_* : \mathbf{Prob}[\mathbf{G}] \rightarrow (\mathbb{R}^{[\mathbf{G}]}, \|\cdot\|_1)$$

of norm one, defined by

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

Applying the construction of Sect. 3 we realize $\mathbf{Prob}[\mathbf{G}]$ as a closed convex cone in some Banach space $\mathbf{Prob}[[\mathbf{G}]]$. Entropy extends to a bounded linear functional $\text{Ent}_* : \mathbf{Prob}[[\mathbf{G}]] \rightarrow (\mathbb{R}^{[\mathbf{G}]}, \|\cdot\|_1)$, whose coordinates evaluate non-negatively on the cone. At this point we would like to define an *entropic quantity* as a bounded linear functional on $\mathbf{Prob}[[\mathbf{G}]]$, which is non-negative on the cone $\mathbf{Prob}[\mathbf{G}]$. Studying such entropic quantities is the subject of our future research.

5 AEP

5.1 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. The subcategory of all homogeneous \mathbf{G} -diagrams will be denoted $\mathbf{Prob}\langle \mathbf{G} \rangle_h$. This space is invariant under the tensor product, thus it is a metric Abelian monoid.

5.1.1 Universal construction of homogeneous diagrams

Examples of homogeneous diagrams could be constructed in the following manner. Fix a finite group G and consider a \mathbf{G} -diagram $\{G_i; \alpha_{ij}\}_{i \in \mathbf{G}}$ of (not necessarily normal) subgroups of G , where morphisms α_{ij} are inclusions. The \mathbf{G} -diagram of probability spaces $\{X_i; f_{ij}\}$ is constructed by setting $X_i = (G/G_i, \text{unif})$, where G/G_i denotes the set of left cosets and unif is the uniform measure, and taking f_{ij} to be the natural projection $G/G_i \rightarrow G/G_j$, whenever $G_i \subset G_j$. The resulting diagram \mathcal{X} will be

minimal if and only if for any $i, j \in \mathbf{G}$ there is $k \in \mathbf{G}$, such that $G_k = G_i \cap G_j$. In fact, any complete homogeneous diagram arises this way, according to the following argument from [9], although the representation of homogeneous diagrams by diagrams of subgroups is highly non-unique.

Indeed, let $\mathcal{X} = \{X_i; \chi_{ij}\}$ be a homogeneous \mathbf{G} -diagram of probability spaces, such that X_0 is the initial space in \mathcal{X} . Then $\text{Aut}(\mathcal{X})$ acts transitively on every space X_i in \mathcal{X} . Let $x_0 \in X_0$ be an atom and set $x_i := \chi_{0i}x_0$. Define $G := \text{Aut}(\mathcal{X})$ to be the full automorphism group and $G_i := \text{Stab}(x_i)$ to be the stabilizer of the action of $\text{Aut}(\mathcal{X})$ on X_i at point x_i . The spaces X_i can be naturally identified with G/G_i . Note that x_i is the image of x_j under the equivariant map χ_{ji} whenever it is present in the diagram \mathcal{X} . Thus we have $G_j \subset G_i$ and a natural surjection $G/G_j \rightarrow G/G_i$, if the morphism χ_{ji} is present in the diagram \mathcal{X} . Under the identification $G/G_j \cong X_j$ the surjection $G/G_j \rightarrow G/G_i$ coincides with χ_{ji} due to the equivariance of χ_{ji} .

5.2 Asymptotic equipartition property

In [9] the following theorem is proven.

Theorem 5.1 *Suppose $\mathcal{X} \in \mathbf{Prob}(\mathbf{G})$ is a \mathbf{G} -diagram of probability spaces for some fixed complete indexing category \mathbf{G} . Then there exists a sequence $\tilde{\mathcal{H}} = (\mathcal{H}_n)_{n=0}^\infty$ of homogeneous \mathbf{G} -diagrams such that*

$$\frac{1}{n} \mathbf{k}(\mathcal{X}^{\otimes n}, \mathcal{H}_n) \leq C(|X_0|, \llbracket \mathbf{G} \rrbracket) \cdot \sqrt{\frac{\ln^3 n}{n}} \tag{5.1}$$

where $C(|X_0|, \llbracket \mathbf{G} \rrbracket)$ is a constant only depending on the cardinality $|X_0|$ of the initial space X_0 of \mathcal{X} and the number $\llbracket \mathbf{G} \rrbracket$ of objects in \mathbf{G} .

The Asymptotic Equipartition Property of Theorem 5.1 is a direct generalization of the classical Asymptotic Equipartition Property, which states that if (X_i) is a sequence of identically distributed, independent random variables, the random variables $-\frac{1}{n} \log p(X_1, \dots, X_n)$ converge as $n \rightarrow \infty$ in probability to the entropy of X_1 . Indeed, in that case the approximating sequence H_n corresponds to a sequence of uniform random variables H_n , with $\text{Ent}(H_n)/n \rightarrow \text{Ent}(X_1)$. Denote by $p(\mathbf{x}, \mathbf{h})$ the optimal coupling achieving the distance in left-hand-side of (5.1). Then

$$\begin{aligned} & \frac{1}{n} \int_{X^n} |\log p(\mathbf{x}) - \text{Ent}(H_n)| dp(\mathbf{x}) \\ &= \frac{1}{n} \int_{X^n \times H_n} \left| \log \frac{p(\mathbf{x}|\mathbf{h})}{p(\mathbf{h}|\mathbf{x})} \right| dp(\mathbf{x}, \mathbf{h}) \\ &\leq \frac{1}{n} \int_{X^n \times H_n} |\log p(\mathbf{h}|\mathbf{x})| dp(\mathbf{x}, \mathbf{h}) + \frac{1}{n} \int_{X^n \times H_n} |\log p(\mathbf{x}|\mathbf{h})| dp(\mathbf{x}, \mathbf{h}) \end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{n} \mathbf{k}(X^n, H_n) \\ &\leq C(|X_0|, \llbracket \mathbf{G} \rrbracket) \cdot \sqrt{\frac{\ln^3 n}{n}} \end{aligned}$$

which implies the classical Asymptotic Equipartition Property.

We define the space of *tropical homogeneous diagrams* by

$$\mathbf{Prob}[\mathbf{G}]_h := \text{QL}_\varphi(\mathbf{Prob} \langle \mathbf{G} \rangle_h, \kappa)$$

Then, the Asymptotic Equipartition Property can be reformulated as follows.

Theorem 5.2 *For any indexing category \mathbf{G} the image of the natural inclusion*

$$\mathbf{Prob}[\mathbf{G}]_h \hookrightarrow \mathbf{Prob}[\mathbf{G}]$$

is dense.

Proof By Theorem 5.1, every linear sequence can be approximated by a homogeneous sequence. It follows from the bound (5.1) that the defect of the approximating homogeneous sequence is bounded by a constant times φ , defined by $\varphi(t) = t^{3/4}$. Moreover, the linear sequences are dense by Theorem 4.5. This finishes the proof. \square

6 The tropical cone for probability spaces and chains

Although for general indexing categories \mathbf{G} the space of tropical \mathbf{G} -diagrams will typically be infinite dimensional, it has a very simple, finite-dimensional description if \mathbf{G} consists of a single object, or if it is a special type of indexing categories called a *chain*.

The chain of length k , denoted by \mathbf{C}_k , is the indexing category with k objects O_1, \dots, O_k , and a morphism from O_i to O_j whenever $i \geq j$. A \mathbf{C}_k -diagram of probability spaces is then a chain of reductions

$$X_k \rightarrow X_{k-1} \rightarrow \dots \rightarrow X_1$$

For chains we can describe the tropical cone explicitly.

Theorem 6.1 *For $k \in \mathbb{N}$, the tropical cone $\mathbf{Prob}[\mathbf{C}_k]$ is isomorphic to the following cone in $(\mathbb{R}^k, |\cdot|_1)$:*

$$\left\{ \left(\begin{array}{c} x_k \\ \vdots \\ x_1 \end{array} \right) \in \mathbb{R}^k \mid 0 \leq x_1 \leq \dots \leq x_k \right\}$$

In particular, the algebraic structure and the pseudo-distance are preserved under the isomorphism.

Recall that a homogeneous probability space is (isomorphic to) a probability space with a uniform distribution and therefore its isomorphism class is completely determined by its cardinality or entropy. A homogeneous *chain* has a very simple description as well: A chain is homogeneous if and only if the individual probability spaces are homogeneous, i.e. if and only if the individual probability spaces are (isomorphic to) probability spaces with a uniform measure. Similarly, the isomorphism class of a chain is completely determined by the cardinalities of the spaces contained in it. This allows us to construct a canonical model for any chain.

We denote by H_n the homogeneous probability space with the underlying set $\{0, \dots, n - 1\}$ with uniform measure. For $n \mid m$ the map $f_{m,n} : H_m \rightarrow H_n$ defined by

$$f_{m,n}(x) := \left\lfloor \frac{x \cdot n}{m} \right\rfloor$$

is a reduction of probability spaces. For a triple of positive integers satisfying $n \mid m \mid l$ holds

$$f_{m,n} \circ f_{l,m} = f_{l,n} \tag{6.1}$$

If $\mathcal{X} = (X_k \rightarrow \dots \rightarrow X_1)$ is a homogeneous chain, and $n_i := |X_i|$ then there is an isomorphism

$$\mathcal{X} \cong (H_{n_k} \xrightarrow{f_{n_k, n_{k-1}}} \dots \xrightarrow{f_{n_2, n_1}} H_{n_1}) \tag{6.2}$$

Let $N, m, n \in \mathbb{N}$ be such that $m \mid N$ and $n \mid N$. Consider a fan of homogeneous spaces

$$H_n \xleftarrow{f_{N,n}} H_N \xrightarrow{f_{N,m}} H_m$$

This fan is not minimal (and not homogeneous). We denote its minimization by $\mathcal{Z}_{n,m}$

$$\mathcal{Z}_{n,m} = (H_n \leftarrow Z_{n,m} \rightarrow H_m)$$

with the top space $Z_{n,m}$ in the minimization satisfying

$$|Z_{n,m}| \leq n + m$$

Also note that minimization does not depend on N , since we assumed it is a multiple of $\text{lcm}(m, n)$. Now we can estimate

$$\begin{aligned} \text{kd}(\mathcal{Z}_{n,m}) &\leq 2 \ln(n + m) - \ln(n) - \ln(m) \leq 2 \ln 2 + |\ln n - \ln m| \\ &= 2 \ln 2 + |\text{Ent}H_n - \text{Ent}H_m| \end{aligned} \tag{6.3}$$

Lemma 6.2 *Let $\mathcal{X}, \mathcal{Y} \in \mathbf{Prob} \langle \mathbf{C}_k \rangle$ be two homogeneous chains of length k . Then*

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) \leq 2k \cdot \ln 2 + \|\text{Ent}_* \mathcal{X} - \text{Ent}_* \mathcal{Y}\|_1$$

Proof Let (n_i) and (m_i) be sequences of cardinalities of spaces in \mathcal{X} and \mathcal{Y} respectively. Without loss of generality we may assume that both chains have canonical form provided by (6.2). Let $N := \text{lcm}(n_k, m_k)$. Then n_i and m_i are divisors of N for all $1 \leq i \leq k$. Consider two-fan of chains

$$\mathcal{X} \xleftarrow{l} \mathcal{H} \xrightarrow{r} \mathcal{Y}$$

where $\mathcal{H} = H_N^{\mathbf{C}_k}$ and $l_i = f_{N, n_i}, r_i = f_{N, m_i}$ for $1 \leq i \leq k$. Due to transitivity (6.1) this is indeed a two-fan of chains. Its minimization is a chain of minimal fans

$$\mathcal{Z}_i := (X_i \leftarrow Z_{n_i, m_i} \rightarrow Y_i)$$

Thus we can estimate

$$\mathbf{k}(\mathcal{X}, \mathcal{Y}) \leq \sum \text{kd}(\mathcal{Z}_i) \leq 2k \cdot \ln 2 + \|\text{Ent}_* \mathcal{X} - \text{Ent}_* \mathcal{Y}\|_1$$

□

Corollary 6.3

$$\text{Ent}_* : \mathbf{Prob}[\mathbf{C}_k]_h \rightarrow (\mathbb{R}^k, \|\cdot\|_1)$$

is an isometric embedding.

Proof Let $[\mathcal{H}_1]$ and $[\mathcal{H}_2]$ be two tropical chains of length k . Then

$$\begin{aligned} \kappa([\mathcal{H}_1], [\mathcal{H}_2]) &= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{k}(\mathcal{H}_1(n), \mathcal{H}_2(n)) \\ &\leq \lim_{n \rightarrow \infty} \frac{1}{n} (2k \ln 2 + \|\text{Ent}_* \mathcal{H}_1(n) - \text{Ent}_* \mathcal{H}_2(n)\|_1) \\ &= \|\text{Ent}_*[\mathcal{H}_1] - \text{Ent}_*[\mathcal{H}_2]\|_1 \end{aligned}$$

The opposite inequality is the 1-Lipschitz property of entropy. □

Proof of Theorem 6.1 The space $\mathbf{Prob}[\mathbf{C}_k]_h$ is dense in $\mathbf{Prob}[\mathbf{C}_k]$ by Theorem 5.2. Therefore the isomorphism in Corollary 6.3 extends to the isomorphic embedding of $\mathbf{Prob}[\mathbf{C}_k]$. To prove the surjectivity one constructs chains of probability spaces with prescribed entropies satisfying the inequalities defining the cone in the theorem. This is left to the reader. □

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Acknowledgements Open access funding provided by Projekt DEAL. The authors would like to thank the referee for the many constructive remarks that have led to a substantial improvement of the article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Ay, N., Bertschinger, N., Der, R., Güttler, F., Olbrich, E.: Predictive information and explorative behavior of autonomous robots. *Eur. Phys. J. B* **63**(3), 329–339 (2008)
2. A'Campo, N.: A natural construction for the real numbers. *Mathematics* (2003). [arXiv:math/0301015](https://arxiv.org/abs/math/0301015)
3. Bertschinger, N., Rauh, J., Olbrich, E., Jost, J., Ay, N.: Quantifying unique information. *Entropy* **16**(4), 2161–2183 (2014)
4. de Bruijn, N.G., Erdős, P.: Some linear and some quadratic recursion formulas. II. In: *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen: Series A: Mathematical Sciences*, vol. 14, pp. 152–163 (1952)
5. Friston, K.: The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* **13**(7), 293–301 (2009)
6. Kovačević, M., Stanojević, I., Šenk, V.: On the hardness of entropy minimization and related problems. In: *2012 IEEE Information Theory Workshop*, pp. 512–516. IEEE (2012)
7. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes (2013). [arXiv:1312.6114](https://arxiv.org/abs/1312.6114)
8. Matus, F.: Infinitely many information inequalities. In: *Information Theory. ISIT 2007. IEEE International Symposium on*, pp. 41–44. IEEE (2007)
9. Matveev, R., Portegies, J.W.: Asymptotic dependency structure of multiple signals. *Inf. Geom.* **1**(2), 237–285 (2018)
10. Matveev, R., Portegies, J.W.: Conditioning in tropical probability theory (2019). [arXiv:1905.05596](https://arxiv.org/abs/1905.05596)
11. Matveev, R., Portegies, J.W.: Tropical probability theory and an application to the entropic cone (2019). [arXiv:1905.05351](https://arxiv.org/abs/1905.05351)
12. Rådström, H.: An embedding theorem for spaces of convex sets. *Proc. Am. Math. Soc.* **3**(1), 165–169 (1952)
13. Steudel, B., Ay, N.: Information-theoretic inference of common ancestors. *Entropy* **17**(4), 2304–2327 (2015)
14. Van Dijk, S.G., Polani, D.: Informational constraints-driven organization in goal-directed behavior. *Adv. Complex Syst.* **16**(02n03), 1350016 (2013)
15. Vidyasagar, M.: A metric between probability distributions on finite sets of different cardinalities and applications to order reduction. *IEEE Trans. Autom. Control* **57**(10), 2464–2477 (2012)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.