

# Entropy, ultralimits and the Poisson boundary

Elad Sayag and Yehuda Shalom

**Abstract.** In this paper, we introduce for a group  $G$  the notion of ultralimit of measure class preserving actions of it, and show that its Furstenberg–Poisson boundaries can be obtained as an ultralimit of actions on itself, when equipped with appropriately chosen measures. We use this result in embarking on a systematic quantitative study of the basic question of how close to invariant one can find measures on a  $G$ -space, particularly for the action of the group on itself. As applications, we show that on amenable groups there are always “almost invariant measures” with respect to the information-theoretic Kullback–Leibler divergence (and more generally, any  $f$ -divergence), making use of the existence of measures with trivial boundary. More interestingly, for a free group  $F$  and a symmetric measure  $\lambda$  supported on its generators, one can compute explicitly the infimum over all measures  $\eta$  on  $F$  of the Furstenberg entropy  $h_\lambda(F, \eta)$ . Somewhat surprisingly, while in the case of the uniform measure on the generators the value is the same as the Furstenberg entropy of the Furstenberg–Poisson boundary of the same measure  $\lambda$ , in general it is the Furstenberg entropy of the Furstenberg–Poisson boundary of a measure on  $F$  *different* from  $\lambda$ .

*Dedicated to Slava Grigorchuk on the occasion of his 70th birthday*

## 1. Introduction

This paper has two main goals. The first is to introduce the problem of minimizing the Furstenberg entropy for group actions. We relate this problem to amenability and boundary actions. The second goal is to introduce the methods of ultralimits into ergodic theory of group actions by providing a construction of the Furstenberg–Poisson boundary via ultralimits. This construction sheds new light on known properties of the Furstenberg–Poisson boundary (see in particular the ending paragraph of the introduction for a very recent use of it in the non-commutative setting).

We use this construction and basic properties of ultralimits to establish the connection between the problem of minimizing entropy to both amenability and boundary actions.

---

*Mathematics Subject Classification 2020:* 22F10 (primary); 20F65, 28D20, 31C05, 46B08, 46M07, 60G10, 60J50, 82C41, 94A17 (secondary).

*Keywords:* Furstenberg–Poisson boundary, Furstenberg entropy, ultralimits, amenability.

### 1.1. Minimizing entropy problem

Many natural actions of discrete groups  $G \curvearrowright X$ , both in the topological and in the ergodic theoretic setting, admit no invariant probability measure. Basic examples to keep in mind are the action of an infinite discrete group on itself, the action of the free group  $F_d$  on the space  $\partial F_d$  of infinite reduced words<sup>1</sup> and the action of  $\mathrm{SL}_n(\mathbb{Z})$  on  $\mathbb{P}_{\mathbb{R}}^{n-1}$  or on  $\mathrm{SL}_n(\mathbb{R})/P(\mathbb{R})$  for any parabolic subgroup<sup>2</sup>. One can ask the following informal basic question:

*How close to invariant can measures on  $X$  get?*

The notion of Furstenberg entropy, which we now recall, enables one to formalize this question.

A standard method of measuring distance between two measures is the Kullback–Leibler (KL) divergence (see [7, Chapter 2]), which is an extremely important and well-known information-theoretic notion. Recall that given two probability measures  $\nu, m$  on  $X$ , their KL-divergence is defined by

$$D_{\mathrm{KL}}(m \parallel \nu) := \int_X -\ln\left(\frac{d\nu}{dm}\right) dm.$$

The KL-divergence is always non-negative, and it is zero exactly when  $\nu = m$ . The KL-divergence is much more sensitive than the norm distance  $\|\nu - m\|$  and is (defined to be) infinite if  $\nu, m$  are not in the same measure class; however, it can be infinite even if they are. With the help of this notion, one can quantify the lack of invariance of a given measure on  $X$  under an action of a group  $G$ . More specifically, fix a measure  $\lambda$  on  $G$  (whose support generates  $G$ ), and consider the associated Furstenberg entropy for *any* (quasi-invariant) probability measure  $\nu$  on  $X$ :

$$h_\lambda(X, \nu) = h_\lambda(\nu) := \sum_{g \in G} \lambda(g) D_{\mathrm{KL}}(g\nu \parallel \nu) = \sum_{g \in G} \lambda(g) \int_X -\ln\left(\frac{dg^{-1}\nu}{d\nu}\right) d\nu.$$

We emphasize that we consider here the Furstenberg entropy of *all* probability measures  $\nu$  and not necessarily  $\lambda$ -stationary measures (i.e.,  $\lambda * \nu = \nu$ ), as was previously done in the literature. The Furstenberg entropy vanishes exactly for invariant measures and serves as a natural quantification for non-invariance. One can also define a similar notion for the whole family of the well-known information-theoretic “distance functions” called  $f$ -divergence, which our results cover as well.

Our search for the most invariant measure(s) on  $X$  can now be formalized by defining

$$I_\lambda(X) := \inf_{\nu \in M(X)} h_\lambda(X, \nu).$$

<sup>1</sup>More generally, the action of a hyperbolic group on its Gromov boundary.

<sup>2</sup>More generally, the action of any discrete Zariski-dense subgroup of a real semi-simple Lie group with no compact factors on a flag space. For the non-existence of invariant measure, see [28] or [12].

Of course, one needs to be more precise here about the class  $M(X)$  of measures  $\nu$  we take the infimum over. Typically, it shall be a fixed measure class (consisting of all fully supported measures when  $X$  is discrete). In the case of a topological action, we consider the infimum to be over all the Borel measures on  $X$  and denote the result by  $I_\lambda^{\text{top}}(X)$ . We note that in this case, if  $X$  is compact then the infimum is achieved (see the discussion in Subsection 8.2).

A basic fact (see Corollary 3.9) is that  $I_\lambda(X) \leq I_\lambda(G)$  for any  $G$ -space  $X$ , which motivates our initial focus on the value of the latter.

These entropy-theoretic quantities give rise to many natural questions; let us state a few of them.

**Problem 1.1.** What can be said about  $I_\lambda(X)$ ? In particular:

- Find a method for calculating  $I_\lambda(X)$  or  $I_\lambda^{\text{top}}(X)$  in various cases (such as the examples at the beginning of the section).
- In the case of  $I_\lambda^{\text{top}}(X)$  what can be said about a minimizing measure? Is it unique? (Somewhat surprisingly, one of our findings is that a minimizing measure need not be  $\lambda$ -stationary.)
- What can be said about actions for which  $I_\lambda(X) = I_\lambda(G)$  or  $I_\lambda(X) = 0$ ?
- In the cases of action of lattices on flag spaces, calculate  $I_\lambda(X)$  for the canonical measure class (e.g., the Haar measure for the maximal compact subgroup) and appropriate  $\lambda$ 's. Is the infimum attained? Is it the same as  $I_\lambda^{\text{top}}(X)$ ?

This problem (especially the second one) arose naturally from an attempt to tackle the well-known problem of dependence of the Liouville property on the finite symmetric generating measure (see the discussion in Subsection 8.1).

In this paper, we are able to calculate  $I_\lambda(G)$  for amenable groups and free groups.

The main tool is the new construction of the Furstenberg–Poisson boundary (Theorem E).

For amenable groups, we use the existence of a Liouville measure on amenable groups (see [17] or [23]) to conclude.

**Theorem A** (See Theorem 6.2). *If  $G$  is amenable and  $\lambda$  has finite support, then  $I_\lambda(G) = 0$ . Namely, on any amenable group there are KL-almost invariant measures.*

It should be noted that the converse to this result is easy (and follows immediately from Pinsker's inequality [4, Theorem 4.19]). We will deduce from Theorem A an interesting fixed point property for amenable groups acting on a certain distinguished topological model of their Furstenberg–Poisson boundary (see Theorem 6.6) – as it turns out (see Corollary 2.9) every countable group admits a canonical one. This theorem was proven in [21] by entirely different method. Further study of the action on this canonical model seems interesting, in particular seeking a converse to our result.

We now formulate our results regarding  $F_d$  – the free group on  $d$  generators.

**Theorem B** (See Corollary 7.12). *Let  $\lambda$  be the symmetric uniform measure on the free generators and their inverses. Then*

$$I_\lambda(F_d) = \frac{d-1}{d} \ln(2d-1).$$

In fact, we can compute  $I_\lambda(F_d)$  for any measure  $\lambda$  which is symmetric and supported on the free generators and their inverses.

**Theorem C** (See Theorem 7.11). *Let  $\lambda$  be a symmetric measure on  $F_d$  supported on the free generators and their inverses. Then there is a symmetric measure  $\mu$ , generally different from  $\lambda$ , supported on the generators and their inverses, with*

$$I_\lambda(F_d) = h_\lambda(\partial F_d, \nu_\mu),$$

where  $\nu_\mu$  is the harmonic (hitting, stationary) measure for the  $\mu$ -random walk on the standard topological boundary of  $\partial F_d$ .

An explicit formula for the measure  $\nu_\mu$  and a complete description of the correspondence  $\lambda \mapsto \mu$  are given in Subsections 7.1 and 7.2, respectively.

More generally, one can use the well-known information-theoretic generalization  $f$ -divergence in place of the KL-divergence. Namely, we introduce

$$h_{\lambda,f}(\nu) = \sum_{g \in G} \lambda(g) D_f(g\nu \parallel \nu)$$

and a corresponding  $I_{\lambda,f}$ . Our results are formulated in this setting. For example, our generalization of Theorem B states the following.

**Theorem D** (See Corollary 7.12). *Let  $\lambda$  be the symmetric uniform measure on the free generators of  $F_d$  and their inverses. Then*

$$I_{\lambda,f}(F_d) = \frac{2d-1}{2d} f\left(\frac{1}{2d-1}\right) + \frac{1}{2d} f(2d-1).$$

Another (simple, but important) ingredient in the proof of Theorems B, C and D will be to give a general criterion<sup>3</sup> (see Proposition 3.13) for a measure  $\nu$  to satisfy  $h_{\lambda,f}(X, \nu) = I_{\lambda,f}(X)$ .

---

<sup>3</sup>It is important to note that this criterion is only about entropy in a fixed measure class. Hence, it could not be used for calculating  $I_{\lambda,f}^{\text{top}}(X)$  (see the discussion in Subsection 8.2).

## 1.2. Furstenberg–Poisson boundary as an ultralimit

The Furstenberg–Poisson boundary  $\mathcal{B}(G, \mu)$  of a group  $G$  with a probability measure  $\mu$  was first introduced by Furstenberg in the papers [10, 11] and since then it was studied extensively (cf. [2, 13, 16, 17]).

We remind here that the Furstenberg–Poisson boundary is a probability space  $(B, \nu)$  with a measurable group action  $G \curvearrowright B$  such that the measure is stationary, for example,  $\mu * \nu = \nu$ .

In this paper, we give a new construction of the Furstenberg–Poisson boundary via ultralimits. The objects of interest for us are probability spaces  $(X, \nu)$  together with a measurable  $G$ -action preserving the measure class.

Our approach toward defining ultralimits of such actions will be via ultralimits of  $C^*$ -algebras as in [15] (see Remark 4.22). More precisely, we define the ultralimit  $\mathcal{U} \lim_{\text{big}}(X_i, \nu_i)$  of “uniformly bounded-quasi-invariant”  $G$ -spaces  $(X_i, \nu_i)$  (see Definition 4.17).

Recall that given a  $G$ -space  $(X, \nu)$ , it admits a minimal measure preserving factor, called the measurable Radon–Nikodym factor (RN for short). As we observe (see Proposition 2.7), the measurable RN-factor always admits a canonical topological model (i.e., a realization on a compact metrizable space on which the group acts continuously). We denote this space by  $(X, \nu)_{\text{RN}}$  – it is the smallest model for which the Radon–Nikodym derivatives  $\frac{dg\nu}{d\nu}$  are continuous.

The RN factor of our large ultralimit space is denoted by  $\mathcal{U} \lim_{\text{RN}}(X_i, \nu_i)$ .

With this notation, we can state our new construction of the Furstenberg–Poisson boundary as an ultralimit of the action of the group on itself (taken with a sequence of “Abel” measures).

**Theorem E** (See Theorem 5.9). *Let  $G$  be a discrete countable group, and  $\mu$  a generating measure on  $G$ . Denote by  $\mu_a = (1 - a) \sum_{n=0}^{\infty} a^n \mu^{*n}$ . Fix  $\mathcal{U}$  an ultrafilter on  $(\frac{1}{2}, 1)$  converging to 1. Then  $\mathcal{U} \lim_{\text{RN}}(G, \mu_a)$  is a topological model for the Furstenberg–Poisson boundary  $\mathcal{B}(G, \mu)$ .*

From this perspective, many classical claims on the Furstenberg–Poisson boundary become clear:

- (1) The maximality of the Furstenberg–Poisson boundary among stationary actions, proved in [13]. Actually, our proof of Theorem E is by showing maximality for our construction (see Proposition 5.7).
- (2) The Kaimanovich–Vershik basic entropy formula (see [17], Corollary 5.13): 
$$h_{\mu}(\mathcal{B}(G, \mu)) = \lim_{n \rightarrow \infty} \frac{H(\mu^{*n})}{n}.$$
- (3) Weak containment of the quasi-regular- $L^2$  representation in the regular representation of the group (see [18]).

A novel aspect of Theorem E is that one can get quantitative results on the Furstenberg–Poisson boundary relating it to the action of the group on itself (see Corollary 5.11) without any assumption on the measure  $\mu$ . This is the main tool in the proof of Theorems A–D. We remark (see Remark 5.12) that while it may be possible to prove Corollary 5.11 via other more ad hoc methods, we believe that the tools of ultralimits could find use in various applications. The use of ultralimits in the sequel paper [24] allows us to handle the topological case, in particular determining  $I_{\lambda,f}^{\text{top}}(\partial F_d)$ .

Finally, we note that recently the ideas around Theorem E found another interesting application. In the paper [8] by Das and Peterson, they introduce a Poisson boundary in a non-commutative setting. They initiate entropy theory of non-commutative boundaries, and proved analogies of part of Kaimanovich–Vershik’s fundamental theorems regarding entropy [17] while the rest are left open. Generalizing our approach to the non-commutative setting, S. Zhou [27] recently completed the proof of those fundamental theorems.

## 2. Preliminaries

### 2.1. $G$ -spaces and Radon–Nikodym factor

Throughout this paper,  $G$  is a discrete countable group.

**Definition 2.1.** We introduce the following category of  $G$ -spaces:

- A *QI (quasi-invariant)  $G$ -space* is a probability space  $\mathcal{X} = (X, \nu)$  together with a measurable action of  $G$  on  $X$ , so that  $\nu$  is quasi-invariant (i.e.,  $g\nu, \nu$  are in the same measure class).
- A *factor map* between QI  $G$ -spaces  $\mathcal{X}, \mathcal{Y}$  is a measurable mapping  $T : X \rightarrow Y$  which is  $G$ -equivariant (e.g.,  $T(gx) = gT(x)$  a.e.) and with  $T_*\nu_X = \nu_Y$ . One says  $\mathcal{Y}$  is a factor of  $\mathcal{X}$ , and that  $\mathcal{X}$  is an extension of  $\mathcal{Y}$ .
- We say the  $T$  is a *measure preserving extension* if  $\frac{dg\nu_Y}{d\nu_Y} \circ T = \frac{dg\nu_X}{d\nu_X}$ .
- We say  $\mathcal{X}$  is *BQI (bounded quasi-invariant)* if  $R_\nu(g) = \frac{dg\nu}{d\nu} \in L^\infty(\mathcal{X})$  for any  $g \in G$ . In this case, a *majorant* for  $\mathcal{X}$  is a function  $M : G \rightarrow \mathbb{R}$  with  $\|\ln R_\nu(g)\|_\infty \leq M(g)$ .

**Remark 2.2.** We note that the conditions on an action are almost surely ( $G$  is countable, so one can assume the action is strict); we also do not assume our measure space is complete (although one can always replace a space by its completion).

Sometimes, when we write QI  $G$ -space, we would not like to fix the probability measure on  $X$ , only the measure class. This should not give rise to any confusion.

In this paper, we shall inherently need to work also with non-standard measure spaces, that is, ones that are not countably separated. However, we will usually “reduce” the  $\sigma$ -algebra to a countably generated one. We then would like to replace the measure space

with the new  $\sigma$ -algebra by a standard measure space. In general, this replacement will not be by an isomorphism in the sense that there are inverse factors, but it will be in the sense that there is an isomorphism of the  $L^\infty$  spaces.

In order to show that such isomorphism always comes from a factor map, we use the following well-known lemma (see [22, Theorem 2.1]).

**Lemma 2.3.** *Suppose  $\Omega$  is a measure space, and  $A \subset L^\infty(\Omega)$  is a separable sub- $C^*$ -algebra. Consider the Gelfand dual  $X$  of  $A$ . Then there is a measurable mapping  $\pi : \Omega \rightarrow X$  that induces the embedding  $A \rightarrow L^\infty(\Omega)$ . Such a mapping  $\pi$  is unique up to equality a.e.*

*If  $\Omega$  is a QI  $G$ -space and  $A \subset L^\infty(\Omega)$  is  $G$ -invariant, then  $\pi$  is a  $G$ -factor.*

**Definition 2.4.** A topological model for a QI  $G$ -space  $\mathcal{X}$  is a QI  $G$ -space  $\mathcal{Y} = (Y, \nu)$  together with a factor  $\pi : \mathcal{X} \rightarrow \mathcal{Y}$  such that:

- (1)  $Y$  is a compact metrizable (equivalently compact Hausdorff and second countable) topological space with a continuous  $G$ -action; the  $\sigma$ -algebra on  $\mathcal{Y}$  is the Borel  $\sigma$ -algebra.
- (2)  $\pi$  induces an isomorphism  $L^\infty(\mathcal{Y}) \cong L^\infty(\mathcal{X})$ .

From Lemma 2.3, we see that if  $L^\infty(\mathcal{X})$  is separable with respect to the weak topology (i.e., the  $\sigma$ -algebra is generated by countably many sets), then  $\mathcal{X}$  possesses a topological model. Moreover, all of them are isomorphic when considered as QI  $G$ -space factors of  $\mathcal{X}$ .

Let us recall the definition of Radon–Nikodym factor, which, using Lemma 2.3, is easily seen to agree with the definition given in the introduction.

**Definition 2.5.** Let  $\mathcal{X} = (X, \Sigma, \nu)$  be a QI  $G$ -space. Its Radon–Nikodym factor is defined to be the QI  $G$ -space which is a topological model of  $(X, \Sigma_{\text{RN}}, \nu)$  where  $\Sigma_{\text{RN}}$  is the  $\sigma$ -algebra generated by  $\left\{ \frac{dg\nu}{d\nu} \mid g \in G \right\}$ .

The next definition and proposition formalize the observation that the classical Radon–Nikodym factor has a canonical topological model. This is easily seen from the perspective of  $C^*$ -algebras.

**Definition 2.6.** An RN model of a BQI  $G$ -space  $\mathcal{X}$  is a compact metrizable space  $X$  with a Borel probability measure  $\nu$  and a continuous  $G$ -action on  $X$  such that:

- $(X, \nu)$  is isomorphic to  $\mathcal{X}$  as BQI  $G$ -spaces.
- For all  $g \in G$ , one has  $\frac{dg\nu}{d\nu} \in C(X)$ .

**Proposition 2.7.** *Let  $\mathcal{X}$  be a BQI  $G$ - $C^*$ -space and consider its (measurable) RN factor  $\mathcal{X}_{\text{RN}}$ . Then it has a topological model which is an RN model. In fact, it has a canonical such model  $X_{\text{RN}}$  which is a continuous factor of all other RN models.*

*Proof.* Consider the closure of the algebra generated by all Radon–Nikodym derivatives  $A_{\text{RN}} := C^*(R_\nu(g) = \frac{dg\nu}{d\nu} \mid g \in G) \subset L^\infty(\mathcal{X})$ . Note that  $A_{\text{RN}}$  is separable, and it is  $G$ -invariant since  $g \cdot R_\nu(h) = \frac{R_\nu(gh)}{R_\nu(g)}$ . Let  $X_{\text{RN}}$  be its Gelfand dual. The space  $X_{\text{RN}}$  is a  $\text{RN}$ -model. Namely, it is a compact metrizable space with a continuous action of  $G$  and a BQI Borel measure  $\nu$  such that  $\frac{dg\nu}{d\nu} \in C(X_{\text{RN}}) \cong A_{\text{RN}}$  for all  $g \in G$ .

By Lemma 2.3 applied to  $A_{\text{RN}} \subset L^\infty(\mathcal{X}_{\text{RN}})$ , we obtain a factor  $\pi : \mathcal{X} \rightarrow X_{\text{RN}}$  of  $G$ -spaces. Via the map  $\pi$ , the space  $X_{\text{RN}}$  is a topological model of  $\mathcal{X}_{\text{RN}}$ .

To show the universality of  $X_{\text{RN}}$ , let  $Y$  be a topological model of  $\mathcal{X}_{\text{RN}}$  which is an  $\text{RN}$ -model. Consider the inclusions  $C(Y) \subset L^\infty(Y) \cong L^\infty(\mathcal{X}_{\text{RN}}) \subset L^\infty(X)$ . As for every  $g \in G$ , we have  $\frac{dg\nu}{d\nu} \in C(Y)$ , we conclude  $A_{\text{RN}} \subset C(Y)$ . This embedding induces a continuous factor  $Y \rightarrow X_{\text{RN}}$ . ■

**Remark 2.8.** Note that  $X_{\text{RN}}$  has the property that the  $\text{RN}$ -cocycle separates the points of  $X$ . This characterizes  $X_{\text{RN}}$  among the  $\text{RN}$  models.

We finish this section with the following observation regarding the Furstenberg–Poisson boundary.

**Corollary 2.9.** *Let  $G$  be a discrete countable group. Then any BQI  $G$ -space which is equal to its (measurable)  $\text{RN}$ -factor has a topological model which is  $\text{RN}$  and minimal among such topological models.*

*In particular, this applies to the Furstenberg–Poisson boundary, hence the latter admits a “canonical” topological model  $\mathcal{B}(G, \mu)$  for a generating measure  $\mu$ .*

*Proof.* The first part is clear by Proposition 2.7. For the second, note that  $\mathcal{B}(G, \mu)$  is its own measurable Radon–Nikodym factor. Indeed, the linear span of the  $\text{RN}$ -cocycle is dense in  $L^1$ , by [3, Lemma 2.2]. (This lemma is a trivial application of Hahn–Banach and the fact that Poisson integral is an isomorphism on the Furstenberg–Poisson boundary.) ■

The interest in this model of the Furstenberg–Poisson boundary is illustrated in Theorem 6.6, as it is shown that for amenable groups, and symmetric  $\mu$ , there is always a unique fixed point there.

## 2.2. $G$ -spaces in the language of $C^*$ -algebras

**2.2.1.  $C^*$ -algebras and probability spaces.** A basic tool in our work will be commutative  $C^*$ -algebras. We review some basic notions from this theory that arise from translating the relevant concepts of measure theory to algebra.

**Notations and Recollections 2.10.** Let  $A$  be a commutative  $C^*$ -algebra with unit 1.

- $A_+$  denotes the positive elements of  $A$ .
- A positive functional/measure  $\nu$  is an element  $\nu \in A^*$  with  $\nu(A_+) \subseteq \mathbb{R}_{\geq 0}$ .



- If a measure  $\nu$  satisfies in addition that  $\nu(1) = 1$ , we say  $\nu$  is a *state/probability measure*.
- We say that a measure  $\nu$  is *faithful* if:  $\forall a \in A_+ : \nu(a) = 0 \Rightarrow a = 0$ .
- Consider the action of  $A$  on  $A^*$  by  $(a \cdot \nu)(b) = \nu(ba)$ .  
If  $a \in A_+$  and  $\nu$  is a measure, then so is  $a \cdot \nu$ .  
If  $\nu$  is a faithful measure, then we have  $a \in A, a \cdot \nu = 0 \Rightarrow a = 0$ .
- For a general measure  $\nu$ , we have  $\mathcal{N}_\nu := \{a \in A \mid \nu(a^*a) = 0\} = \{a \in A \mid a \cdot \nu = 0\}$ .  
Indeed, this follows from the Cauchy–Schwarz inequality for  $\langle a, b \rangle_\nu = \nu(b^*a)$ .
- $\mathcal{N}_\nu$  is an ideal of  $A$  and  $\nu$  induces a measure  $\bar{\nu}$  on  $\frac{A}{\mathcal{N}_\nu}$  which is faithful.  
We call  $(\frac{A}{\mathcal{N}_\nu}, \bar{\nu})$  the *reduced version* of  $(A, \nu)$  and denote it by  $(A, \nu)_{\text{red}}$ .
- We say that a measure  $m$  is *bounded absolutely continuous* with respect to a measure  $\nu$  and denote it by  $m \ll^b \nu$  iff there is  $c \in A_+$  with  $m = c \cdot \nu$ .  
In this case, we have that  $\mathcal{N}_\nu \subset \mathcal{N}_m$  and hence  $m$  defines a measure on  $\frac{A}{\mathcal{N}_\nu}$ .
- If  $\nu$  is faithful and  $m \ll^b \nu$ , then there exists a unique  $c \in A_+$  with  $m = c \cdot \nu$ . This element  $c$  is called the *Radon–Nikodym derivative*. It is denoted by  $c = \frac{dm}{d\nu}$ .  
If in addition we have  $\nu \ll^b m$ , then  $m$  is faithful and  $\frac{d\nu}{dm} = \left(\frac{dm}{d\nu}\right)^{-1}$ .

Now, we define the notion of probability space in the realm of  $C^*$ -algebras.

**Definition 2.11.** The category of  $C^*$ -probability spaces is defined as follows:

- A  $C^*$ -probability space is a pair  $(A, \nu)$  of a unital commutative  $C^*$ -algebra together with a faithful state  $\nu$ .
- A *factor map*  $T : \mathcal{B} \rightarrow \mathcal{A}$  between  $C^*$ -probability spaces  $\mathcal{B} = (B, m)$  and  $\mathcal{A} = (A, \nu)$  is a  $*$ -homomorphism  $T : B \rightarrow A$  such that  $T^*\nu = m$ . We say that  $\mathcal{B}$  is a factor of  $\mathcal{A}$  and  $\mathcal{A}$  is an extension of  $\mathcal{B}$ .

Note that in the definition of a factor,  $T$  is automatically an injection and thus an isometry.

**2.2.2.  $G$ - $C^*$ -spaces.** Let  $A$  be a  $C^*$ -algebra. By an *action* of  $G$  on  $A$ , we will mean a left action such that for any  $g \in G$ , the mapping  $a \mapsto g \cdot a = {}^g a$  is a  $C^*$ -algebra isomorphism. Such action induces an action  $G \curvearrowright A^*$  by  ${}^g \nu = g \cdot \nu = (g^{-1})^* \nu$ . The set of states is  $G$ -invariant. In addition, any finite measure  $\mu$  on  $G$  acts on  $A$  by:  $\mu * a = \sum_g \mu(g) \cdot {}^g a$ . Similarly,  $\mu$  acts on  $A^*$ .

We now define the notion of  $G$ -space in the context of  $C^*$ -algebras.

**Definition 2.12.** A *BQI  $G$ - $C^*$ -space* is a  $C^*$ -probability space  $\mathcal{A} = (A, \nu)$  together with an action of  $G$  on  $A$  such that for any  $g \in G$ , we have  ${}^g \nu \ll^b \nu$ .

We will denote  $R_{\mathcal{A}}(g) = R_\nu(g) := \frac{d{}^g \nu}{d\nu}$ , the *Radon–Nikodym cocycle*.

A *majorant* for  $\mathcal{A}$  is a function  $M : G \rightarrow \mathbb{R}$  such that  $\|\ln(R_\nu(g))\| \leq M(g)$ .

**Remark 2.13.** It is easy to see that  $R_\nu(gh) = {}^g R_\nu(h) \cdot R_\nu(g)$ , so  $R_\nu : G \rightarrow A^\times$  is indeed a cocycle.

**Remark 2.14.** Note that it is not always true that for any probability measure  $\mu$  on  $G$ , we have  $\mu * \nu \ll^b \nu$ . However, the following is true (and trivial).

Suppose  $G$  is a group that acts on a  $C^*$ -algebra  $A$  and  $\nu$  is a state on  $A$  such that  $\mathcal{N}_\nu$  is  $G$ -invariant. Then for any complex finite measure  $\mu$  on  $G$ , we have  $(\mu * \nu)(\mathcal{N}_\nu) = 0$  and thus  $\mu * \nu$  is well-defined state on  $(A, \nu)_{\text{red}}$ , that agrees with  $\mu * \bar{\nu}$ .

This observation will be used without mentioning in Subsection 4.3.

**Definition 2.15.** A factor  $p : \mathcal{B} \rightarrow \mathcal{A}$  between two BQI  $G$ - $C^*$ -spaces  $\mathcal{A}, \mathcal{B}$  is a factor of  $C^*$ -probability spaces that is  $G$ -equivariant. We say that  $\mathcal{B}$  is a factor of  $\mathcal{A}$  and  $\mathcal{A}$  is an extension of  $\mathcal{B}$ .

The factor  $p$  is called *measure preserving extension/factor* if  $p(R_{\mathcal{B}}(g)) = R_{\mathcal{A}}(g)$ .

One can define RN-factor in this setting.

**Definition 2.16.** The *Radon–Nikodym factor* of a BQI  $G$ - $C^*$ -space  $\mathcal{A}$  is the BQI  $G$ - $C^*$ -space given by  $\mathcal{A}_{\text{RN}} = (A_{\text{RN}}, \nu|_{A_{\text{RN}}})$  where  $A_{\text{RN}} = C^*(R_\nu(g) \mid g \in G) \subset A$  is the sub- $C^*$ -algebra generated by the Radon–Nikodym cocycle.

Note that  $A_{\text{RN}}$  is  $G$ -invariant as  ${}^g R_\nu(h) = \frac{R_\nu(gh)}{R_\nu(g)}$ , and also  ${}^g \nu|_{A_{\text{RN}}} \ll^b \nu|_{A_{\text{RN}}}$ , so indeed  $\mathcal{A}_{\text{RN}}$  is a BQI  $G$ - $C^*$ -space. The factor map  $p : \mathcal{A}_{\text{RN}} \rightarrow \mathcal{A}$  is a measure preserving extension, and if  $\pi : \mathcal{B} \rightarrow \mathcal{A}$  is a measure preserving extension, then  $p$  factors through  $\pi$ .

We end this section by noting the relation between our definitions for  $G$ -spaces in the algebraic and classical settings.

In one direction, let  $(X, \nu)$  be a BQI  $G$ -space, then  $(L^\infty(X), \nu)$  is a BQI  $G$ - $C^*$ -probability space.

In the opposite direction, for a BQI  $G$ - $C^*$ -probability space  $(A, \nu)$ , Gelfand's representation theorem provides a compact Hausdorff space  $X$  with  $C(X) \cong A$ . The state  $\nu$  defines a Borel measure on  $X$ . Moreover,  $X$  admits a natural continuous  $G$ -action and with those,  $(X, \nu)$  is a BQI  $G$ -space. Also, the Radon–Nikodym cocycle consists of continuous functions. Note that when  $A$  is a separable  $C^*$ -algebra,  $X$  is a RN-model.

### 3. $f$ -divergence and entropy

Throughout this section,  $f$  denotes a convex function in  $(0, \infty)$  with  $f(1) = 0$ .

We denote by  $F(x, y) := yf(\frac{x}{y})$  the corresponding homogeneous function in two variables.

It is well known that  $F : (0, \infty)^2 \rightarrow \mathbb{R}$  is convex (e.g., see Lemma 3.2).

In this section, we define the notion of entropy based on the notion of  $f$ -divergence, generalizing the Furstenberg entropy in the case of  $f(t) = t \ln(t)$ , and establish a criterion for a measure to minimize it.

### 3.1. Definition of $f$ -entropy

Recall the definition of  $f$ -divergence.

**Definition 3.1** ( $f$ -divergence). Let  $\eta, m$  be two measures on a measurable space  $X$  in the same measure class. We say that  $\eta$  has finite  $f$ -divergence with respect to  $m$  if  $f\left(\frac{d\eta}{dm}\right) \in L^1(m)$  and define

$$D_f(\eta \parallel m) := \int_X f\left(\frac{d\eta}{dm}\right) dm.$$

For a more symmetric formulation, note that we can take any measure  $\omega$  in the same measure class then the  $f$ -divergence is finite iff  $F\left(\frac{d\eta}{d\omega}, \frac{dm}{d\omega}\right) \in L^1(\omega)$ , and we have

$$D_f(\eta \parallel m) = \int_X F\left(\frac{d\eta}{d\omega}, \frac{dm}{d\omega}\right) d\omega.$$

The  $f$ -divergence does not change when we change  $f$  by a linear function vanishing at 1. Thus, by the convexity, one can always assume  $f \geq 0$ . Hence, we define  $D_f$  to be  $+\infty$  when  $f\left(\frac{d\eta}{dm}\right) \notin L^1(m)$ .

First, we have the following lemma.

**Lemma 3.2.** Let  $f : (0, \infty) \rightarrow \mathbb{R}$  be a convex function, then  $F(x, y) = yf\left(\frac{x}{y}\right)$  is convex.

Moreover, if  $f$  is strictly convex and  $x_0, x_1, y_0, y_1 \in (0, \infty)$  and  $\frac{x_0}{y_0} \neq \frac{x_1}{y_1}$ , then  $F$  restricted to the interval between  $(x_0, y_0), (x_1, y_1)$  is strictly convex.

*Proof.* Note that the convexity condition

$$F((1-t)x_0 + tx_1, (1-t)y_0 + ty_1) \leq (1-t)F(x_0, y_0) + tF(x_1, y_1)$$

is the same as  $f\left(\frac{(1-t)x_0 + tx_1}{(1-t)y_0 + ty_1}\right) \leq f\left(\frac{x_0}{y_0}\right) \cdot \frac{(1-t)y_0}{((1-t)y_0 + ty_1)} + f\left(\frac{x_1}{y_1}\right) \cdot \frac{ty_1}{((1-t)y_0 + ty_1)}$  which follows from the convexity of  $f$ . The moreover part follows too. ■

The previous lemma yields the following using (conditional) Jensen inequality.

**Lemma 3.3.** The  $f$ -divergence has the following properties:

- (1)  $D_f$  is a convex function in both variables and non-negative.
- (2) For a measurable mapping  $\pi : X \rightarrow Y$ , we have

$$D_f(\pi_*\eta \parallel \pi_*m) \leq D_f(\eta \parallel m).$$

- (3) In the notations of item (2), if  $D_f(\pi_*\eta \parallel \pi_*m) = D_f(\eta \parallel m) < \infty$  and  $f$  is strictly convex, then  $\frac{d\eta}{dm}$  is  $Y$ -measurable; moreover,  $\frac{d\eta}{dm} = \frac{d\pi_*\eta}{d\pi_*m} \circ \pi$ .

**Definition 3.4.** Let  $X$  be a QI  $G$ -space.

- We denote by  $M(X)$  the set of measures on  $X$  in the given measure class on it.
- Given a probability measure  $\lambda$  on  $G$ , we define the *Furstenberg  $\lambda, f$ -entropy* to be the function  $h_{\lambda,f} : M(X) \rightarrow [0, \infty]$  given by

$$h_{\lambda,f}(v) = \sum_{g \in G} \lambda(g) D_f(gv \parallel v).$$

- Define  $M_{\lambda,f}(X) = h_{\lambda,f}^{-1}[0, \infty)$ .

**Example 3.5.** For  $f(t) = t \ln(t)$ , the divergence  $D_f = D_{\text{KL}}$  is the famous KL-divergence. Notice that  $h_{\lambda,f} = h_\lambda$  is the Furstenberg entropy:

$$h_\lambda(v) = \sum_{g \in G} \lambda(g) \int_X -\ln\left(\frac{dg^{-1}v}{dv}\right) dv.$$

An immediate application of Lemma 3.3 yields the following lemma.

**Lemma 3.6.** *Let  $\lambda$  be a probability measure on  $G$ .*

- (1) *For any QI  $G$ -space  $X$ , the function  $h_{\lambda,f}$  is convex.*
- (2) *Let  $\pi : (X, v) \rightarrow (Y, m)$  be a factor of QI  $G$ -spaces. Then  $h_{\lambda,f}(X, v) \geq h_{\lambda,f}(Y, m)$ . If  $\pi$  is a measure preserving extension, we have equality.*
- (3) *Suppose  $\lambda$  is generating and  $f$  is strictly convex. If  $\pi : (X, v) \rightarrow (Y, m)$  is a factor such that  $h_{\lambda,f}(X, v) = h_{\lambda,f}(Y, m) < \infty$ , then  $\pi$  is a measure preserving extension.*

**Definition 3.7.** Let  $X$  be a QI  $G$ -space. Given a convex function  $f$  with  $f(1) = 0$  and a probability measure  $\lambda$  on  $G$ . We define the *minimal entropy number* by

$$I_{\lambda,f}(X) = \inf_{v \in M(X)} h_{\lambda,f}(X, v).$$

**Lemma 3.8.** *Let  $(X, v)$  be a QI  $G$ -space, and  $\kappa \in M(G)$ . Then  $h_{\lambda,f}(\kappa) \geq h_{\lambda,f}(\kappa * v)$ .*

*Proof.* Consider  $G \times X$  with  $G$  acting only on the first coordinate. Then  $(G \times X, \kappa \times v)$  is measure preserving extension of  $(G, \kappa)$ , and is an extension of  $(X, \kappa * v)$  via the action map. It follows from item (2) of Lemma 3.6 that

$$h_{\lambda,f}(G, \kappa) = h_{\lambda,f}(G \times X, \kappa \times v) \geq h_{\lambda,f}(X, \kappa * v). \quad \blacksquare$$

The lemma immediately implies that the action of a group on itself is largest in terms of entropy.

**Corollary 3.9.** *For any QI  $G$ -space  $X$ , we have  $I_{\lambda,f}(G) \geq I_{\lambda,f}(X)$ .*

Let us show that for ergodic measure classes, a minimizing measure for entropy is unique (if exists).

**Lemma 3.10.** *Suppose  $f$  is a strictly convex function with  $f(1) = 0$  and  $\lambda$  is a finitely supported generating probability measure on  $G$ . If  $X$  is a QI  $G$ -space which is ergodic (invariant sets are null or co-null), then  $h_{\lambda,f}$  is a strictly convex function on  $M_{\lambda,f}(X)$ .*

*In particular, there is at most one measure  $\nu \in M_{\lambda,f}(X)$  for which  $h_{\lambda,f}(X, \nu) = I_{\lambda,f}(X)$ .*

*Proof.* Fix  $\omega \in M(X)$ . Let  $\nu_0, \nu_1 \in M_{\lambda,f}(X)$  and  $\nu_t = (1-t)\nu_0 + t\nu_1$ . Note that as  $F$  is convex,

$$\begin{aligned} h_{\lambda,f}(\nu_t) &= \sum_{g \in G} \lambda(g) \int_X F\left(\frac{dg\nu_t}{d\omega}, \frac{d\nu_t}{d\omega}\right) d\omega \\ &\leq \sum_{g \in G} \lambda(g) \int_X (1-t)F\left(\frac{dg\nu_0}{d\omega}, \frac{d\nu_0}{d\omega}\right) + tF\left(\frac{dg\nu_1}{d\omega}, \frac{d\nu_1}{d\omega}\right) d\omega \\ &= (1-t)h_{\lambda,f}(\nu_0) + th_{\lambda,f}(\nu_1). \end{aligned}$$

If one has equality  $h_{\lambda,f}(\nu_t) = (1-t)h_{\lambda,f}(\nu_0) + th_{\lambda,f}(\nu_1) < \infty$ , then according to Lemma 3.2 for any  $g \in \text{Supp}(\lambda)$  we have a.e.  $\frac{dg\nu_0}{d\nu_0} = \frac{dg\nu_1}{d\nu_1}$ . However, this yields that  $\frac{d\nu_0}{d\nu_1}$  is  $G$ -invariant (since  $\lambda$  is generating) and using ergodicity we deduce  $\nu_0 = \nu_1$ . Thus,  $h_{\lambda,f}$  is strictly convex. ■

The previous lemma implies that for infinite groups, the infimum defining  $I_{\lambda,f}(G)$  is not attained.

**Corollary 3.11.** *Let  $f$  be a strictly convex function with  $f(1) = 0$  and  $\lambda$  be a generating probability measure on  $G$ . Suppose that  $\kappa$  is a probability measure on  $G$  with  $I_{\lambda,f}(G) = h_{\lambda,f}(\kappa) < \infty$ . Then  $G$  is finite and  $\kappa$  is the normalized counting measure.*

*Proof.* Since the right  $G$ -action is an automorphism of the (left)  $G$ -space  $G$ , Lemma 3.10 implies that  $\kappa$  is right  $G$ -invariant, concluding the proof. ■

### 3.2. A criterion for minimizers of entropy

Our goal is to find a criterion for a measure on a QI  $G$ -space  $X$  to have minimal entropy  $h_{\lambda,f}$ . We stress again that we consider only measures in the measure class given on  $X$ .

In this subsection, we assume in addition that  $f$  is smooth. Recall we denoted  $F(x, y) = yf\left(\frac{x}{y}\right)$ .

**Proposition 3.12.** *Let  $(X, \nu)$  be a BQI  $G$ -space. Then for any  $m \in M(X)$  and  $g \in G$ , we have*

$$D_f(gm \parallel m) \geq D_f(g\nu \parallel \nu) + \int_X \frac{\partial F}{\partial x}\left(1, \frac{dg^{-1}\nu}{d\nu}\right) + \frac{\partial F}{\partial y}\left(\frac{dg\nu}{d\nu}, 1\right) d(m - \nu).$$

*Proof.* Using the convexity of  $F$ , we get the inequality  $F(v) \geq F(u) + \nabla F(u) \cdot (v - u)$  for any  $v, u \in (0, \infty)^2$ . Fix  $\omega \in M(X)$  (e.g.,  $\omega = \nu$ ). Substituting  $v = \left(\frac{dgm}{d\omega}, \frac{dm}{d\omega}\right)$ ,  $u = \left(\frac{dgv}{d\omega}, \frac{dv}{d\omega}\right)$ , we have a.e

$$\begin{aligned} & F\left(\frac{dgm}{d\omega}, \frac{dm}{d\omega}\right) \\ & \geq F\left(\frac{dgv}{d\omega}, \frac{dv}{d\omega}\right) + \left\langle \nabla F\left(\frac{dgv}{d\omega}, \frac{dv}{d\omega}\right), \left(\frac{dgm}{d\omega}, \frac{dm}{d\omega}\right) - \left(\frac{dgv}{d\omega}, \frac{dv}{d\omega}\right) \right\rangle \\ & = F\left(\frac{dgv}{d\omega}, \frac{dv}{d\omega}\right) + \frac{\partial F}{\partial x}\left(\frac{dgv}{d\omega}, \frac{dv}{d\omega}\right) \left(\frac{d(m-v)}{d\omega}\right) \\ & \quad + \frac{\partial F}{\partial y}\left(\frac{dgv}{d\omega}, \frac{dv}{d\omega}\right) \left(\frac{d(m-v)}{d\omega}\right). \end{aligned}$$

As both partial derivatives of  $F$  are homogeneous of degree 0 and  $\nu$  is BQI, all terms in the above inequality are  $\omega$ -integrable. We conclude

$$\begin{aligned} D_f(gm \parallel m) & \geq D_f(gv \parallel v) + \int_X \frac{\partial F}{\partial x}\left(\frac{dgv}{d\omega}, \frac{dv}{d\omega}\right) dg(m-v) \\ & \quad + \int_X \frac{\partial F}{\partial y}\left(\frac{dgv}{d\omega}, \frac{dv}{d\omega}\right) d(m-v). \end{aligned}$$

Note that for  $p \in X$  we have  $\frac{dgv}{dv}(g \cdot p) = \frac{dv}{dg^{-1}v}(p)$ , and thus

$$\begin{aligned} & \int_X \frac{\partial F}{\partial x}\left(\frac{dgv}{d\omega}, \frac{dv}{d\omega}\right) dg(m-v) + \int_X \frac{\partial F}{\partial y}\left(\frac{dgv}{d\omega}, \frac{dv}{d\omega}\right) d(m-v) \\ & = \int_X \frac{\partial F}{\partial x}\left(\frac{dgv}{dv}(p), 1\right) dg(m-v)(p) + \int_X \frac{\partial F}{\partial y}\left(\frac{dgv}{dv}(p), 1\right) d(m-v)(p) \\ & = \int_X \frac{\partial F}{\partial x}\left(\frac{dgv}{dv}(g \cdot p), 1\right) d(m-v)(p) + \int_X \frac{\partial F}{\partial y}\left(\frac{dgv}{dv}, 1\right) d(m-v) \\ & = \int_X \left[ \frac{\partial F}{\partial x}\left(1, \frac{dg^{-1}v}{dv}\right) + \frac{\partial F}{\partial y}\left(\frac{dgv}{dv}, 1\right) \right] d(m-v). \end{aligned}$$

This proves the desired inequality. ■

**Proposition 3.13.** *Let  $(X, \nu)$  be a BQI  $G$ -space and  $\lambda$  is a finitely supported measure on  $G$ . Then  $h_{\lambda, f}(\nu) = I_{\lambda, f}(X)$  iff the function on  $X$*

$$\Psi_{\lambda, f}(\nu; \cdot) := \sum_{g \in G} \lambda(g) \left( \frac{\partial F}{\partial x}\left(1, \frac{dg^{-1}v}{dv}\right) + \frac{\partial F}{\partial y}\left(\frac{dgv}{dv}, 1\right) \right)$$

*is constant (a.e.).*

*Proof.* Take  $m \in M(X)$ . Using Proposition 3.12, we get

$$D_f(gm \parallel m) \geq D_f(gv \parallel v) + \int_X \frac{\partial F}{\partial x} \left( 1, \frac{dg^{-1}v}{dv} \right) + \frac{\partial F}{\partial y} \left( \frac{dgv}{dv}, 1 \right) d(m-v).$$

Averaging over  $\lambda$ , we conclude

$$h_{\lambda,f}(m) \geq h_{\lambda,f}(v) + \int_X \Psi_{\lambda,f}(v; x) d(m-v)(x).$$

If  $\Psi_{\lambda,f}(v; \cdot)$  is constant, then we get  $h_{\lambda,f}(m) \geq h_{\lambda,f}(v)$  and thus  $h_{\lambda,f}(v) = I_{\lambda,f}(X)$ .

To finish the proof, we return to the other direction of the implication in the first statement. Suppose  $v$  is BQI measure such that  $h_{\lambda,f}(v) = I_{\lambda,f}(X)$ . We will show  $\Psi_{\lambda,f}(v, \cdot)$  is constant. Consider any measure  $m$  in the bounded measure class of  $v$  and define  $\varphi(t) = h_{\lambda,f}((1-t)v + tm)$ . We claim that

$$\varphi'(0) = \int_X \Psi_{\lambda,f}(v; x) d(m-v)(x). \quad (3.1)$$

This would finish the proof. Indeed, the minimality of  $v$  implies that the above integral is non-negative. As we can take any measure  $m$  in the bounded measure class, it follows easily that  $\Psi_{\lambda,f}(v; \cdot)$  must be constant a.e.

To verify (3.1), we observe that by definition

$$\varphi(t) = \sum_g \lambda(g) \int_X F \left( (1-t) \frac{dgv}{dv} + t \frac{dgm}{dv}, (1-t) + t \frac{dm}{dv} \right) dv.$$

Since  $\lambda$  has finite support and  $m, gv$  are in the bounded measure class of  $v$ , we can differentiate under the integral sign and (3.1) follows from a direct computation (as in the proof of Proposition 3.12). ■

**Example 3.14.** For  $\lambda$  symmetric and finitely supported, we have a simpler formula for  $\Psi_{\lambda,f}$ .

Denote  $\Psi_f(z) = \frac{\partial F}{\partial x}(1, z) + \frac{\partial F}{\partial y}(z, 1) = f'(\frac{1}{z}) + f(z) - zf'(z)$ . Then we have

$$\Psi_{\lambda,f}(v; \cdot) = \sum_{g \in G} \lambda(g) \Psi_f \left( \frac{dgv}{dv} \right).$$

The next inequality will play a key role in our proof of Theorem 6.6.

**Corollary 3.15.** *Let  $X$  be a QI  $G$ -space,  $\mu$  a probability measure on  $G$  with finite support and  $v \in M(X)$  a BQI measure. Then for any  $m \in M(X)$ ,*

$$h_\mu(m) \geq \int_X \left( 1 - \frac{d\mu * v}{dv} - \sum_{g \in G} \mu(g) \ln \left( \frac{dg^{-1}v}{dv} \right) \right) dm.$$

*Proof.* We have  $F(x, y) = x \ln\left(\frac{x}{y}\right)$  so that  $\frac{\partial F}{\partial x}(1, z) = -\ln(z) + 1$ ,  $\frac{\partial F}{\partial y}(z, 1) = -z$ . By Proposition 3.12,

$$\begin{aligned} D_{\text{KL}}(gm \parallel m) &\geq D_{\text{KL}}(g\nu \parallel \nu) + \int_X \frac{\partial F}{\partial x}\left(1, \frac{dg^{-1}\nu}{d\nu}\right) + \frac{\partial F}{\partial y}\left(\frac{dg\nu}{d\nu}, 1\right) d(m - \nu) \\ &= D_{\text{KL}}(g^{-1}\nu \parallel \nu) + \int_X \left[-\ln\left(\frac{dg^{-1}\nu}{d\nu}\right) + 1 - \frac{dg\nu}{d\nu}\right] d(m - \nu) \\ &= \int_X \left[1 - \frac{dg\nu}{d\nu} - \ln\left(\frac{dg^{-1}\nu}{d\nu}\right)\right] dm. \end{aligned}$$

Averaging over  $\mu$ , we obtain the desired inequality. ■

## 4. Ultralimits

In this section, we introduce an ultralimit for  $G$ -spaces via ultralimits of  $C^*$ -algebras. We begin by recalling the basic definition of ultralimit of Banach spaces and  $C^*$ -algebras.

For this section,  $\mathcal{I}$  is a set and  $\mathcal{U}$  is an ultrafilter on  $\mathcal{I}$ . Recall that this means  $\mathcal{U} \subset 2^{\mathcal{I}}$  a collection of subsets of  $\mathcal{I}$ , satisfying:

- (1)  $I \in \mathcal{U}, \emptyset \notin \mathcal{U}$ .
- (2)  $\mathcal{U} \ni A \subset B \Rightarrow B \in \mathcal{U}$ .
- (3)  $A, B \in \mathcal{U} \Rightarrow A \cap B \in \mathcal{U}$ .
- (4) for any  $A \subset \mathcal{I}$ , either  $A \in \mathcal{U}$  or  $\mathcal{I} \setminus A \in \mathcal{U}$ .

### 4.1. Ultralimit of a sequence

We recall the usual definition of ultralimit of a sequence.

**Definition 4.1.** Let  $\mathcal{I}$  be a set and  $\mathcal{U}$  be an ultrafilter on  $\mathcal{I}$ .

- We say that a property  $P$  holds for  $\mathcal{U}$ -a.e.  $i$  iff  $\{i \in \mathcal{I} \mid P(i) \text{ is true}\} \in \mathcal{U}$ .
- Given a compact Hausdorff space  $X$  and an  $\mathcal{I}$ -sequence  $(a_i)_{i \in \mathcal{I}}$ , an element  $a \in X$  is called the  $\mathcal{U}$ -ultralimit of  $(a_i)$  iff for every open neighborhood  $W$  of  $a$ , we have  $a_i \in W$  for  $\mathcal{U}$ -a.e.  $i$ .

We denote  $a = \mathcal{U} \lim_{i \in \mathcal{I}} a_i$ .

We have the following well-known lemma (for a reference, see [5]).

**Lemma 4.2.** Let  $X, Y$  be compact Hausdorff spaces. Let  $(a_i)_{i \in \mathcal{I}}, (b_i)_{i \in \mathcal{I}}$  be sequences in  $X$ . Then:

- (1) An ultralimit of  $(a_i)$  exists and is unique.



- (2) If  $a_i = b_i$  for  $\mathcal{U}$ -a.e.  $i$ , then  $\mathcal{U} \lim a_i = \mathcal{U} \lim b_i$ .
- (3) If  $f : X \rightarrow Y$  is a continuous function, then  $f(\mathcal{U} \lim a_i) = \mathcal{U} \lim f(a_i)$ .

**Example 4.3.** Considering the compact Hausdorff space  $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ , we have a well-defined  $\mathcal{U}$ -ultralimit of any sequence of real numbers, and we also can define the  $\mathcal{U}$ -ultralimit of any bounded sequence of complex numbers. This ultralimit is a complex number.

By Lemma 4.2, the resulting map  $\mathcal{U} \lim_{i \in \mathcal{I}} : \ell^\infty(\mathcal{I}) \rightarrow \mathbb{C}$  is a  $C^*$ -homomorphism.

## 4.2. Ultralimits of Banach spaces and $C^*$ -algebras

We recall the basic notion from [15].

**Notation 4.4.** Let  $B_i$  be normed linear spaces (over  $\mathbb{C}$ ) for  $i \in \mathcal{I}$ , and let  $\mathcal{U}$  be an ultrafilter on  $\mathcal{I}$ .

Define the following normed space:

$$V := \ell^\infty(\mathcal{I}, \{B_i\}) := \left\{ (v_i)_{i \in \mathcal{I}} \in \prod_{i \in \mathcal{I}} B_i \mid \sup_i \|v_i\|_{B_i} < \infty \right\}, \quad \|(v_i)\|_V = \sup_i \|v_i\|_{B_i}.$$

Consider the subspace:  $\mathcal{N}_{\mathcal{U}} = \{(v_i) \in V \mid \mathcal{U} \lim \|v_i\|_{B_i} = 0\} \subset V$ .

One has the following formula for the quotient norm.

**Lemma 4.5.** *In the notations above:*

- $\mathcal{N}_{\mathcal{U}} \subset V$  is a closed subspace.
- For all  $v = (v_i) \in V$

$$\inf_{u \in \mathcal{N}_{\mathcal{U}}} \|v + u\|_V = \mathcal{U} \lim \|v_i\|_{B_i}.$$

**Definition 4.6.** Define the  $\mathcal{U}$ -ultralimit of the normed spaces  $B_i$  to be the normed space

$$B := \mathcal{U} \lim B_i := \frac{\ell^\infty(\mathcal{I}, \{B_i\})}{\mathcal{N}_{\mathcal{U}}}.$$

We will write  $\mathcal{U} \lim v_i$  for the class of  $(v_i)$  in  $B$ . By Lemma 4.5, we have  $\|\mathcal{U} \lim v_i\|_B = \mathcal{U} \lim \|v_i\|_{B_i}$ .

**Lemma 4.7.** *Suppose  $(B_i)_{i \in \mathcal{I}}$  are Banach spaces, then:*

- (1)  $\mathcal{U} \lim B_i$  is a Banach space.
- (2) Let  $a_i^{(n)} \in B_i$ , and suppose that  $M_n$  are positive numbers with  $\|a_i^{(n)}\|_{B_i} \leq M_n$  and  $\sum M_n < \infty$ . Then we have

$$\mathcal{U} \lim \sum_i a_i^{(n)} = \sum_n \mathcal{U} \lim a_i^{(n)}.$$

*Proof.* For (1), the quotient of a Banach space by a closed subspace is a Banach space.

For (2), we have equality  $(\sum_n a_i^{(n)})_i = \sum_n (a_i^{(n)})_i$  on  $V$  and the projection to  $B$  is continuous. ■

To define the ultralimit of  $C^*$  algebras, we make the following observation.

**Lemma 4.8.** *Suppose  $(B_i)_{i \in \mathcal{I}}$  are  $C^*$ -algebras. Then  $B = \mathcal{U} \lim B_i$  has a unique structure of  $C^*$ -algebra satisfying  $(\mathcal{U} \lim a_i)^* = \mathcal{U} \lim a_i^*$ ,  $(\mathcal{U} \lim a_i) \cdot (\mathcal{U} \lim b_i) = \mathcal{U} \lim a_i \cdot b_i$ .*

*If  $B_i$  are commutative, then so is  $B$ . If  $B_i$  all have unit, then  $\mathcal{U} \lim 1_{B_i}$  is a unit for  $B$ .*

*Proof.* Indeed,  $V = \ell^\infty(\mathcal{I}, \{B_i\})$  is a  $C^*$ -algebra with the operations  $(a_i)^* = (a_i^*)$ ,  $(a_i) \cdot (b_i) = (a_i \cdot b_i)$ . Clearly,  $\mathcal{N}_{\mathcal{U}}$  defined above is a  $*$ -two sided ideal. So the quotient is a  $C^*$ -algebra, the lemma follows. ■

**Definition 4.9.** Given  $C^*$ -algebras  $B_i$ , their ultralimit  $\mathcal{U} \lim B_i$  is the ultralimit as above together with the  $C^*$ -algebra structure as in Lemma 4.8.

The following basic lemma is useful.

**Lemma 4.10.** *Let  $A_i$  be  $C^*$ -algebras,  $A = \mathcal{U} \lim A_i$  be the ultralimit.*

- (1) *If  $a_i \in A_i$ ,  $\sup \|a_i\| < \infty$  are self adjoint/positive/normal, then  $a = \mathcal{U} \lim a_i$  is self adjoint/positive/normal.*
- (2) *If  $a_i \in A$  are normal and  $\|a_i\| \leq M$ , then for any continuous function  $f$  on the disc  $\overline{D}(0, M)$  we have  $f(\mathcal{U} \lim a_i) = \mathcal{U} \lim f(a_i)$ .*
- (3) *If  $a \in A$  is self adjoint/positive, then we can find  $a_i \in A_i$  self adjoint/positive with  $\|a_i\| = \|a\|$ ,  $a = \mathcal{U} \lim a_i$ .*

Finally, we would like to discuss duality: for a normed space  $B$  we denote the dual normed space by  $B^*$ , and for  $v \in B$ ,  $f \in B^*$  we denote  $\langle v, f \rangle = f(v)$ .

It is easy to see one has a canonical isometric embedding  $\Phi : \mathcal{U} \lim_i B_i^* \hookrightarrow (\mathcal{U} \lim B_i)^*$  described by the formula  $\langle \mathcal{U} \lim v_i, \Phi(\mathcal{U} \lim f_i) \rangle = \mathcal{U} \lim \langle v_i, f_i \rangle$ . Hereafter we shall omit  $\Phi$  from the notations.

### 4.3. Ultralimits of $C^*$ -probability spaces and the $G$ -equivariant case

We shall need to define the ultralimit of  $C^*$ -probability spaces (see Definition 2.11).

**Definition 4.11.** Let  $(A_i, \nu_i)_{i \in \mathcal{I}}$  be a collection of  $C^*$ -probability spaces. Define their ultralimit

$$(A, \nu) = \mathcal{U} \lim (A_i, \nu_i) := (\mathcal{U} \lim A_i, \mathcal{U} \lim \nu_i)_{\text{red}}.$$

Given  $(a_i) \in \ell^\infty(\mathcal{I}, \{A_i\})$ , we will denote by  $\mathcal{U} \lim a_i$  the element of  $A$  which is the class of  $\mathcal{U} \lim a_i \in \mathcal{U} \lim A_i$  in  $A = \frac{\mathcal{U} \lim A_i}{\mathcal{N}_{\mathcal{U} \lim \nu_i}}$ .

**Lemma 4.12.** *Let  $(A_i, \nu_i)$  be  $C^*$ -probability spaces,  $(A, \nu) = \mathcal{U} \lim(A_i, \nu_i)$ .*

- (1) *If  $a_i \in A_i$ ,  $\|a_i\| \leq M$ ,  $f \in C(\overline{D}(0, M))$ , then:*
  - $f(\mathcal{U} \lim a_i) = \mathcal{U} \lim f(a_i)$ ,
  - $\nu(\mathcal{U} \lim a_i) = \mathcal{U} \lim \nu_i(a_i)$ .
- (2) *If  $\eta_i$  are states on  $A_i$  with  $\eta_i \ll^b \nu_i$  and  $\sup \|\frac{d\eta_i}{d\nu_i}\| \leq M$ , then on  $\mathcal{U} \lim A_i$  we have  $\mathcal{U} \lim \eta_i \ll^b \mathcal{U} \lim \nu_i$ . In particular,  $\mathcal{U} \lim \eta_i$  induces a probability measure  $\eta$  on  $A$  and  $\frac{d\eta}{d\nu} = \mathcal{U} \lim \frac{d\eta_i}{d\nu_i}$ .*

*Proof.* (1) Obvious from Lemma 4.10.

(2) Obvious, since for  $a = \mathcal{U} \lim a_i \in \mathcal{U} \lim A_i$ , we have

$$\begin{aligned} (\mathcal{U} \lim \nu_i) \left( a \cdot \mathcal{U} \lim \frac{d\eta_i}{d\nu_i} \right) &= \mathcal{U} \lim \nu_i \left( a_i \cdot \frac{d\eta_i}{d\nu_i} \right) = \mathcal{U} \lim \eta_i(a_i) \\ &= (\mathcal{U} \lim \eta_i)(\mathcal{U} \lim a_i). \end{aligned}$$

■

**Lemma 4.13.** *One has the following:*

- (1) *Let  $(A_i, \nu_i)$  and  $(B_i, m_i)$  be  $C^*$ -probability spaces and  $T_i : (B_i, m_i) \rightarrow (A_i, \nu_i)$  be factors. Then there is a factor map  $T = \mathcal{U} \lim T_i : \mathcal{U} \lim(B_i, m_i) \rightarrow \mathcal{U} \lim(A_i, \nu_i)$  satisfying  $T(\mathcal{U} \lim b_i) = \mathcal{U} \lim T_i(b_i)$ .*
- (2) *For any  $C^*$ -probability space  $(A, \nu)$ , the diagonal mapping  $\Delta(a) = \mathcal{U} \lim_i a$  defines a factor  $\Delta : (A, \nu) \rightarrow \mathcal{U} \lim_i(A, \nu)$ .*

*Proof.* (1) The natural map  $T : \ell^\infty(\mathcal{I}, B_i) \rightarrow \mathcal{U} \lim(A_i, \nu_i)$  defined by  $T((b_i)) = \mathcal{U} \lim T_i(b_i)$  is a  $*$ -homomorphism. As  $T^*(\mathcal{U} \lim \nu_i)((b_i)) = (\mathcal{U} \lim \nu_i)(\mathcal{U} \lim T_i(b_i)) = p^*(\mathcal{U} \lim m_i)((b_i))$  where  $p$  is the projection  $p : \ell^\infty(\mathcal{I}, B_i) \rightarrow \mathcal{U} \lim(B_i, m_i)$ , we conclude that  $T$  gives rise to a factor map  $T : \mathcal{U} \lim(B_i, m_i) \rightarrow \mathcal{U} \lim(A_i, \nu_i)$  with the desired property.

(2)  $\Delta : A \rightarrow \frac{\mathcal{U} \lim A}{\mathcal{N}_\nu}$  is a  $*$ -homo. and  $\Delta^*(\mathcal{U} \lim_i \nu)(a) = \mathcal{U} \lim_i \nu(a) = \nu(a)$ . ■

Let  $G$  be a discrete countable group.

Consider a collection  $(\mathcal{A}_i)_{i \in \mathcal{I}}$  of BQI  $G$ - $C^*$ -spaces. We will say that  $\mathcal{A}_i$  are uniformly BQI, if there is a function  $M : G \rightarrow \mathbb{R}_+$  which is a majorant for all of them (recall the definition of a majorant from Definition 2.12).

Given such a collection, we next define its ultralimit.

**Definition 4.14.** Let  $(\mathcal{A}_i)_{i \in \mathcal{I}}$  be uniformly BQI  $G$ - $C^*$ -spaces. Consider  $\mathcal{A} = \mathcal{U} \lim \mathcal{A}_i$ . By functoriality (Lemma 4.13), we get a natural  $G$ - $C^*$ -space structure on  $\mathcal{A}$ , such that when  $a_i \in A_i$  are uniformly bounded then  ${}^g \mathcal{U} \lim a_i = \mathcal{U} \lim {}^g a_i$ . We will say that  $\mathcal{A}$  is the *ultralimit* of the uniformly BQI  $G$ - $C^*$ -spaces  $\mathcal{A}_i$ .

**Lemma 4.15.** *Let  $(\mathcal{A}_i)_{i \in \mathcal{I}}$  be uniformly  $C^*$ -BQI- $G$ -spaces, and let  $\mathcal{A} = (A, \nu) = \mathcal{U} \lim(A_i, \nu_i)$ .*

Let  $\mu$  be a probability measure on  $G$  and  $a_i \in A_i$  be uniformly bounded elements. Then:

- (1) For any  $g \in G$ , we have  $R_{\mathcal{A}}(g) = \mathcal{U} \lim R_{\mathcal{A}_i}(g)$ .
- (2)  $\mu * \mathcal{U} \lim a_i = \mathcal{U} \lim \mu * a_i$ .
- (3)  $(\mu * \nu)(\mathcal{U} \lim a_i) = \mathcal{U} \lim(\mu * \nu_i)(a_i)$ .

*Proof.* Item (1) is immediate from Lemma 4.12, whereas items (2) and (3) are immediate from Lemma 4.7 since  $\|\mu(g)^g a_i\| \leq \mu(g)(\sup_i \|a_i\|)$ ,  $\|\mu(g) \cdot^g \nu\| \leq \mu(g)$  and  $\sum_g \mu(g)$  is a convergent sum. ■

**Lemma 4.16.** *We have the following:*

- (1) Let  $(A_i, \nu_i)$  and  $(B_i, m_i)$  be uniformly BQI  $G$ - $C^*$ -spaces and  $T_i : (B_i, m_i) \rightarrow (A_i, \nu_i)$  be factors. Then there is a factor map  $T = \mathcal{U} \lim T_i : \mathcal{U} \lim(B_i, m_i) \rightarrow \mathcal{U} \lim(A_i, \nu_i)$  satisfying  $T(\mathcal{U} \lim b_i) = \mathcal{U} \lim T_i(b_i)$ .  
If each  $T_i$  is measure preserving extension, then so is  $T$ .
- (2) For any BQI  $G$ - $C^*$ -space  $(A, \nu)$ , the diagonal mapping  $\Delta(a) = \mathcal{U} \lim_i a$  defines a measure preserving extension  $\Delta : (A, \nu) \rightarrow \mathcal{U} \lim_i (A, \nu)$ .

*Proof.* Note that  $T, \Delta$  from Lemma 4.13 are  $G$ -equivariant. Also, if all  $T_i$  are measure preserving extensions, then  $T(R_{\mathcal{U} \lim m_i}(g)) = T(\mathcal{U} \lim R_{m_i}(g)) = \mathcal{U} \lim T_i(R_{m_i}(g)) = \mathcal{U} \lim R_{\nu_i}(g) = R_{\nu}(g)$ .

Since  $\Delta(R_{\nu}(g)) = \mathcal{U} \lim R_{\nu}(g) = R_{\mathcal{U} \lim \nu}(g)$ , we get that  $\Delta$  is a measure preserving extension. ■

#### 4.4. Ultralimits for $G$ -spaces via algebras

**Definition 4.17.** Let  $\mathcal{X}_i = (X_i, \nu_i)$  ( $i \in \mathcal{I}$ ) be a collection of uniformly BQI  $G$ -spaces, in the sense that there is a uniform majorant for the whole family. Consider  $\mathcal{A}_i = (L^\infty(X_i), \nu_i)$ , which are uniformly BQI  $G$ - $C^*$ -algebras, and let  $\mathcal{A} = \mathcal{U} \lim \mathcal{A}_i = (A, \nu)$ . The Gelfand space  $X$  for  $A$  is a BQI  $G$ -space that will be denoted by  $\mathcal{U} \lim_{\text{big}} \mathcal{X}_i$ . We call it a version of the ultralimit of the actions  $\mathcal{X}_i$ .

The RN factor of  $\mathcal{A}$  will be denoted by  $\mathcal{U} \lim_{\text{RN}} \mathcal{X}_i$ , we consider it is using the canonical topological RN model of Proposition 2.7, that is,  $C(\mathcal{U} \lim_{\text{RN}} \mathcal{X}_i) = \mathcal{A}_{\text{RN}}$ .

Note that if  $\varphi_i$  are uniformly bounded functions on  $X_i$ , then we have a well-defined function  $\varphi \in L^\infty(X)$  which corresponds to  $\mathcal{U} \lim \varphi_i \in A$  (for which we keep this notation).

Lemma 4.15 implies that

$$\mathcal{U} \lim \frac{dg \nu_i}{d \nu_i} = \frac{dg \nu}{d \nu}.$$

The basic result we shall need is that entropy behaves well under ultralimits.

**Lemma 4.18.** *Let  $(X_i, \nu_i)$  be uniformly BQI  $G$ -spaces and  $(X, \nu) = \mathcal{U} \lim_{\text{big}} (X_i, \nu_i)$ . If  $f \in C(0, \infty)$  is a continuous function, and  $g \in G$ , then*

$$\mathcal{U} \lim \int_{X_i} f \left( \frac{dg\nu_i}{d\nu_i} \right) d\nu_i = \int_X f \left( \frac{dg\nu}{d\nu} \right) d\nu.$$

*Proof.* Let  $M$  be a uniform majorant for  $\mathcal{X}_i$ , then  $e^{-M(g)} \leq \frac{dg\nu_i}{d\nu_i} \leq e^{M(g)}$ . From Lemma 4.12, we get

$$\begin{aligned} \mathcal{U} \lim \int_{X_i} f \left( \frac{dg\nu_i}{d\nu_i} \right) d\nu_i &= \mathcal{U} \lim \nu_i \left( f \left( \frac{dg\nu_i}{d\nu_i} \right) \right) = \nu \left( \mathcal{U} \lim f \left( \frac{dg\nu_i}{d\nu_i} \right) \right) \\ &= \nu \left( f \left( \mathcal{U} \lim \frac{dg\nu_i}{d\nu_i} \right) \right) = \nu \left( f \left( \frac{dg\nu}{d\nu} \right) \right) = \int_X f \left( \frac{dg\nu}{d\nu} \right) d\nu. \quad \blacksquare \end{aligned}$$

**Corollary 4.19.** *Let  $(X_i, \nu_i)$  be uniformly BQI  $G$ -spaces and  $(X, \nu) = \mathcal{U} \lim_{\text{big}} (X_i, \nu_i)$ . Let  $f$  be a convex function with  $f(1) = 0$ ,  $\lambda$  a finitely supported probability measure on  $G$ . Then we have*

$$\mathcal{U} \lim h_{\lambda, f}(X_i, \nu_i) = h_{\lambda, f}(X, \nu).$$

*Proof.* Follows from Lemma 4.18 by linearity. ■

**Remark 4.20.** As remarked in the introduction, in the sequel [24] the theory of ultralimit of uniformly QI  $G$ -spaces will be developed from the measure theory point of view, and this is the reason for “big” in the notation  $\mathcal{U} \lim_{\text{big}} \mathcal{X}_i$ .

**Remark 4.21.** Using the continuity of  $\mathcal{U} \lim$  for sequences, one can deduce that the equality in Corollary 4.19 holds for more general  $\lambda$ ’s by controlling the tail term. A concrete example is given in Remark 5.13.

**Remark 4.22.** In the paper [6], the authors considered a construction of ultralimit of measure preserving actions using a construction of ultralimit of probability spaces due to Loeb (for the case of finite probability spaces, see [9]). Our usage of  $C^*$ -algebras is because it was the fastest and least technical way to be able to state and prove Theorem E. However, it limits us to work with bounded functions and in particular to restrict ourselves to work with bounded-quasi-invariant actions (see Definition 2.12) and to make the assumption of uniformly bounded-quasi-invariant (see Definition 4.17). This suffices for our paper.

The sequel paper [24] elaborates on the ultralimit construction, and in particular develops a measure theoretic ultralimit for quasi-invariant actions as well, under a natural assumption of “uniformly quasi-invariance” (which is more technical). Another goal is to

show the equivalence between the construction developed here and the measure theoretic one taken there.

We believe that this method of ultralimits could be intriguing and useful in further research in the ergodic theory of group actions.

## 5. The Furstenberg–Poisson boundary as an ultralimit

In this section, we provide a novel construction of the Furstenberg–Poisson boundary  $\mathcal{B}(G, \mu)$  of a discrete countable group  $G$  together with a generating measure  $\mu$ .

**Lemma 5.1.** *Assume  $\mu$  is a generating probability on  $G$  and  $C > 0$ . Let  $\nu$  is a measure on a measurable  $G$ -space  $X$  such that  $\mu * \nu \leq C \cdot \nu$ . Then  $(X, \nu)$  is BQI with a majorant depending only on  $C, \mu$ .*

This is clear, as  $\mu(g) \frac{dgv}{d\nu} \leq \frac{d(\mu*\nu)}{d\nu} \leq C$ , hence  $\frac{dgv}{d\nu} \leq \frac{C}{\mu(g)}$ .

**Notation 5.2.** Let  $(G, \mu)$  be a measured discrete countable group, with  $\mu$  generating. For any  $0 < a < 1$ , define the Abel sum

$$\mu_a := (1 - a) \sum_{n=0}^{\infty} a^n \cdot \mu^{*n}.$$

We have the following lemma.

**Lemma 5.3.** *With the notations above:*

- $\mu_a$  is a probability measure on  $G$ .
- We have the following formula:

$$\mu * \mu_a = (1 - a) \sum_{n=0}^{\infty} a^n \cdot \mu^{*(n+1)} = \frac{\mu_a}{a} - \frac{1 - a}{a} \delta_e.$$

- For  $a > \frac{1}{2}$ , the collection of measures  $\mu_a$  is uniformly BQI.

The first two items are obvious, the last item follows from the second one and Lemma 5.1.

Now, we give the main definition of this section.

**Definition 5.4.** Let  $\mathcal{U}$  be an ultrafilter on  $(\frac{1}{2}, 1)$  which contains  $(1 - \varepsilon, 1)$  for all  $\varepsilon > 0$ . Define the  $\mathcal{U}$ -Abel  $C^*$  space of  $(G, \mu)$  to be the following ultralimit:

$$\mathcal{B}_{\mathcal{U}}^{\text{alg}}(G, \mu) = \mathcal{U} \lim_a (L^\infty(G, \mu_a), \mu_a)$$

together with the probability measure  $\nu_\mu = \mathcal{U} \lim \mu_a$  on  $\mathcal{B}_{\mathcal{U}}^{\text{alg}}(G, \mu)$ .

Denote by  $\mathcal{B}_{\mathcal{U}}(G, \mu)_{\text{RN}}$  the compact metrizable BQI  $G$ -space corresponding to  $\mathcal{B}_{\mathcal{U}}^{\text{alg}}(G, \mu)_{\text{RN}}$ , where the latter is the Radon–Nikodym factor of  $\mathcal{B}_{\mathcal{U}}^{\text{alg}}(G, \mu)$ .

Note that  $\mathcal{B}_{\mathcal{U}}(G, \mu)_{\text{RN}}$  is the canonical model of Proposition 2.7 for the RN factor of  $\mathcal{U} \lim_{\text{big}}(G, \mu_a)$ .

**Proposition 5.5.** *We have  $\mu * \nu_\mu = \nu_\mu$ . Thus,  $(\mathcal{B}_{\mathcal{U}}^{\text{alg}}(G, \mu), \nu_\mu)$ ,  $\mathcal{B}_{\mathcal{U}}(G, \mu)_{\text{RN}}$  are  $\mu$ -stationary.*

*Proof.* Using Lemma 5.3, we have the following identity in  $\mathcal{U} \lim_a (L^\infty(G)^*)$ :

$$\mu * \nu_\mu = \mathcal{U} \lim_a \mu * \mu_a = \mathcal{U} \lim_a \frac{\mu_a}{a} - \frac{1-a}{a} \delta_e = \mathcal{U} \lim_a \mu_a = \nu_\mu.$$

Thus, we get this identity when we consider  $\nu_\mu$  as a state on the underlying  $C^*$ -algebra of  $\mathcal{B}_{\mathcal{U}}^{\text{alg}}(G, \mu)$ .  $\blacksquare$

Our main goal for this section is to show that  $\mathcal{B}_{\mathcal{U}}(G, \mu)_{\text{RN}}$  is the RN model of the Furstenberg–Poisson boundary of  $(G, \mu)$  of Corollary 2.9.

**Lemma 5.6.** *The action of  $G$  on itself satisfies the following property:*

- (1) *Given a QI  $G$ -space  $(X, \nu)$  and a quasi-invariant measure  $\eta$  on  $G$ , define  $Y = G \times X$  with the  $G$ -action  $g(h, x) = (gh, x)$  and the product measure  $\eta \times \nu$ . Then  $(Y, \eta \times \nu)$  is a QI  $G$ -space. The mappings  $m(g, x) = gx$ ,  $p(g, x) = g$  define a diagram*

$$\begin{array}{ccc} (Y, \eta \times \nu) & \xrightarrow{m} & (X, \eta * \nu) \\ \downarrow p & & \\ (G, \eta) & & \end{array}$$

*in which  $m$  is a factor map and  $p$  is measure preserving extension.*

- (2) *Dually, let  $(A, \nu)$  be a BQI  $C^*$ - $G$ -space,  $\eta$  a BQI measure on  $G$ . Then we have a BQI  $C^*$ - $G$ -space  $(C, \tau)$  and a diagram*

$$\begin{array}{ccc} (A, \eta * \nu) & \xrightarrow{m} & (C, \tau) \\ & & \uparrow p \\ & & (L^\infty(G), \eta) \end{array}$$

*in which  $m$  is a factor and  $p$  is measure preserving extension.*

*Proof.* (1) It is obvious that  $p$  is  $G$ -equivariant and that  $p_*(\eta \times \nu) = \eta$ . Moreover, since the action is only on the first component, we conclude that  $\frac{dg(\eta \times \nu)}{d(\eta \times \nu)} = \frac{dg\eta}{d\eta} \circ p$  which imply that  $p$  is a measure preserving extension.

The equalities  $m(g \cdot (h, x)) = m(gh, x) = (g \cdot h) \cdot x = g \cdot (h \cdot x) = g \cdot m(h, x)$  and  $m_*(\eta \times \nu) = \eta * \nu$  implies that  $m$  is a factor.

(2) By Gelfand's representation theorem,  $A \cong C(X)$  where  $X$  is a compact Hausdorff topological space,  $G$  acts on  $X$  continuously and  $\nu$  is a regular probability measure on  $X$ .

We have that  $\frac{dgv}{dv} \in C(X)$  and in particular bounded, hence  $(X, \nu)$  is a BQI  $G$ -space. Applying the first item, take  $(C, \tau) = (L^\infty(Y), \eta \times \nu)$  and  $m$  to be the composition  $A \cong C(X) \subset L^\infty(X, \nu) \rightarrow (C, \tau)$ . The lemma follows. ■

Let us show a universality property of  $\mathcal{B}_{\mathcal{U}}^{\text{alg}}(G, \mu)$ .

**Proposition 5.7.** *Let  $(A, \nu)$  be a  $C^*$ -BQI  $G$ -space which is  $\mu$ -stationary space. Then we have  $(C, \tau)$  a  $C^*$ -BQI  $G$ -space which is  $\mu$ -stationary together with a diagram*

$$\begin{array}{ccc} (A, \nu) & \xrightarrow{m} & (C, \tau) \\ & & \uparrow p \\ \mathcal{B}_{\mathcal{U}}^{\text{alg}}(G, \mu)_{\text{RN}} & \xrightarrow{i} & \mathcal{B}_{\mathcal{U}}^{\text{alg}}(G, \mu) \end{array}$$

where  $m$  is a factor, and  $p, i$  are measure preserving extensions.

*Proof.* It is clear that  $i$  is a measure preserving extension. Consider  $\eta = \mu_a$  ( $a > \frac{1}{2}$ ), note that  $\eta * \nu = \nu$  and thus by Lemma 5.6, we have  $(C_a, \tau_a)$  with

$$\begin{array}{ccc} (A, \nu) & \xrightarrow{m_a} & (C_a, \tau_a) \\ & & \uparrow p_a \\ & & (L^\infty(G), \mu_a) \end{array}$$

Since  $\mu_a$  are uniformly BQI and  $p_a$  is measure preserving, we conclude that  $\tau_a$  are uniformly BQI. We take  $(C, \tau) := \mathcal{U} \lim (C_a, \tau_a)$ . Using functionality (Lemma 4.16), the maps  $\mathcal{U} \lim_a m_a, \mathcal{U} \lim_a p_a$  fit into the following diagram:

$$\begin{array}{ccc} (A, \nu) & \xrightarrow{\Delta} & \mathcal{U} \lim_a (A, \nu) \xrightarrow{\mathcal{U} \lim_a m_a} (C, \tau) = \mathcal{U} \lim_a (C_a, \tau_a) \\ & & \uparrow \mathcal{U} \lim_a p_a \\ & & \mathcal{B}_{\mathcal{U}}^{\text{alg}}(G, \mu) = \mathcal{U} \lim_a (L^\infty(G, \mu_a), \mu_a) \end{array}$$

such that  $m := \mathcal{U} \lim m_a \circ \Delta$  is a factor,  $p = \mathcal{U} \lim p_a$  is a measure preserving extension; in particular,  $(C, \tau)$  is  $\mu$ -stationary. ■

We shall need the following well-known result, whose proof follows by comparing the conditional measures (see [11] or [2]) to the disintegration.

**Lemma 5.8.** *Let  $\mathcal{B}(G, \mu)$  be the Furstenberg–Poisson boundary of  $(G, \mu)$ . Let  $\mathfrak{X}$  be a  $\mu$  stationary  $G$ -space that is an extension of the Furstenberg–Poisson boundary  $p : \mathfrak{X} \rightarrow \mathcal{B}(G, \mu)$ . Then  $p$  is a measure preserving extension.*

We now come to our characterization of the Furstenberg–Poisson boundary, which implies Theorem E.



**Theorem 5.9.** *Let  $G$  be a discrete countable group, and  $\mu$  a generating measure on  $G$ . Then  $\mathcal{B}_{\mathcal{U}}(G, \mu)_{\text{RN}}$  is the canonical RN model of the Furstenberg–Poisson boundary of  $(G, \mu)$ .*

**Remark 5.10.** Recall from Corollary 2.9 that the Furstenberg–Poisson boundary admits a canonical RN model.

*Proof of Theorem 5.9.* Consider  $\mathcal{B}(G, \mu)$  the canonical RN model to the Furstenberg–Poisson boundary, and consider  $(C(\mathcal{B}(G, \mu)), \nu) = (L^\infty(\mathcal{B}(G, \mu)), \nu)_{\text{RN}}$ . It is  $\mu$ -stationary, and thus by Proposition 5.7, there exists a  $\mu$ -stationary  $C^*$ - $G$ -space  $(C, \tau)$  and a diagram:

$$\begin{array}{ccc} C(\mathcal{B}(G, \mu)) & \xrightarrow{m} & (C, \tau) \\ & & \uparrow p \\ \mathcal{B}_{\mathcal{U}}^{\text{alg}}(G, \mu)_{\text{RN}} & \xrightarrow{i} & \mathcal{B}_{\mathcal{U}}^{\text{alg}}(G, \mu) \end{array}$$

Here,  $q = p \circ i$  is a measure preserving extension. Thus,

$$(C(\mathcal{B}_{\mathcal{U}}(G, \mu)_{\text{RN}}, \nu_\mu) \cong \mathcal{B}_{\mathcal{U}}^{\text{alg}}(G, \mu)_{\text{RN}} \cong (C, \tau)_{\text{RN}}.$$

Let  $\mathcal{X} = (X, \tau)$  be a Gelfand space for  $(C, \tau)$ . It is a  $\mu$ -stationary space and  $m$  gives rise to a factor map  $m : \mathcal{X} \rightarrow \mathcal{B}(G, \mu)$ . By Lemma 5.8, we conclude that  $m$  is a measure preserving extension, which implies that  $m$  induces an isomorphism  $(C, \tau)_{\text{RN}} \cong (C(\mathcal{B}(G, \mu)), \nu)$ .

In conclusion, we get  $(C(\mathcal{B}(G, \mu)), \nu) \cong (C(\mathcal{B}_{\mathcal{U}}(G, \mu)_{\text{RN}}, \nu_\mu)$ , and the proof is complete. ■

**Corollary 5.11.** *Let  $G$  be a discrete countable group,  $\mu$  a generating probability measure on  $G$ . Denote by  $(\mathcal{B}, \nu)$  the Furstenberg–Poisson boundary of  $(G, \mu)$ . Then for any  $f \in C(0, \infty)$  and  $g \in G$ , we have*

$$\lim_{a \rightarrow 1^-} \sum_{x \in G} f\left(\frac{\mu_a(g^{-1}x)}{\mu_a(x)}\right) \mu_a(x) = \int_{\mathcal{B}} f\left(\frac{dg\nu}{d\nu}(x)\right) d\nu(x).$$

*In particular, if  $f$  is convex with  $f(1) = 0$  and  $\lambda$  is a finitely supported measure on  $G$ , we have*

$$\lim_{a \rightarrow 1^-} h_{\lambda, f}(G, \mu_a) = h_{\lambda, f}(\mathcal{B}(G, \mu), \nu_\mu).$$

*Proof.* Follows immediately from Lemma 4.18 and Theorem 5.9 using the following observation:  $\lim_{a \rightarrow 1^-} R(a) = R$  iff for any ultrafilter  $\mathcal{U}$  on  $(0, 1)$  with  $\mathcal{U} \lim a = 1$  one has  $\mathcal{U} \lim R(a) = R$ . ■

**Remark 5.12.** We note that it is possible to prove the above corollary without applying our ultralimit construction, in a more direct way by considering Martin boundary (see

[25, Chapter IV section 24] and [26] and noticing that the Green function is actually  $G(g, x \mid a) = \frac{g\mu_a(x)}{1-a}$ ). It could be interesting to relate those constructions.

One can deduce from Corollary 5.11 the fundamental Theorem 3.1 in [17].

**Corollary 5.13.** *For  $\mu$  of finite entropy (entropy-  $H(\kappa) := -\sum_g \kappa(g) \ln(\kappa(g))$ ), the following holds:*

$$h_\mu(\mathfrak{B}(G, \mu), \nu_\mu) = \lim_{n \rightarrow \infty} \frac{H(\mu^{*n})}{n}.$$

*Proof.* Indeed, first note that one has  $H(\kappa * \lambda) \leq H(\kappa) + H(\lambda)$  and thus  $h = \lim_{n \rightarrow \infty} \frac{H(\mu^{*n})}{n}$  exists and  $\frac{H(\mu^{*n})}{n} \geq h$ . The desired equality follows from the following 3 steps:

*Step 1.* We have  $h \geq h_\mu(v)$  for any  $\mu$ -stationary  $(X, \nu)$ .

Let us show first  $H(\mu) \geq h_\mu(v)$ . Indeed,  $\frac{dg^{-1}v}{dv} \geq \mu(g)$  and thus  $-\ln\left(\frac{dg^{-1}v}{dv}\right) \leq -\ln(\mu(g))$  and then integrate over  $X$  and average using  $\mu$ . Since for  $\mu$ -stationary  $\nu$  one has  $h_{\mu^{*n}}(v) = n \cdot h_\mu(v)$ , we conclude Step 1.

*Step 2.* We have  $h_\mu(\mathfrak{B}(G, \mu), \nu_\mu) = \lim_{a \rightarrow 1^-} h_\mu(G, \mu_a)$ .

Corollary 5.11 yields  $h_{\delta_g}(\mu_a) \rightarrow h_{\delta_g}(\nu_\mu)$  for any  $g \in G$ . Since  $0 \leq h_{\delta_g}(\mu_a) \leq \ln\left(\frac{2}{\mu(g)}\right)$  for  $a \geq \frac{1}{2}$ , and  $\sum_g \mu(g) \ln\left(\frac{1}{\mu(g)}\right) = H(\mu) < \infty$ , we can replace sum and limit and conclude Step 2.

*Step 3.* We have  $h \leq \liminf_a h_\mu(G, \mu_a)$ .

One has the following formula  $h_\mu(G, \lambda) = -\sum_g (\mu * \lambda - \lambda)(g) \ln(\lambda(g))$  for entropy on the group.

Thus, noting that  $\mu_a(e) \geq 1 - a$  and using the concavity of  $H$ , we obtain

$$\begin{aligned} \liminf_a h_\mu(\mu_a) &= \liminf_a - \sum_g (\mu * \mu_a - \mu_a)(g) \ln(\mu_a(g)) \\ &= \liminf_a \left( \frac{1-a}{a} H(\mu_a) + \frac{1-a}{a} \ln(\mu_a(e)) \right) \\ &\geq \liminf_a (1-a)^2 \sum_n a^n H(\mu^{*n}) \\ &\geq \liminf_a (1-a)^2 \sum_n a^n \cdot n \cdot h = h. \quad \blacksquare \end{aligned}$$

The rest of the paper will consist of applications of our characterization of the Furstenberg–Poisson boundary (Theorem 5.9) and in particular its numerical consequence (Corollary 5.11). The first application will be to amenable groups (Theorem A). The second application will be to the problem of calculating entropy minimal number (see Definition 3.7) for the action of the Free group on itself (Theorems B, C and D).

## 6. Applications for amenable groups

Recall that a generating probability measure on  $G$  is called *Liouville* if the corresponding Furstenberg–Poisson boundary is a point.

**Corollary 6.1.** *If  $\mu$  is a Liouville measure on  $G$ , then for any finitely supported probability measure  $\lambda$  on  $G$  and a convex function  $f$  with  $f(1) = 0$  we have*

$$\lim_{a \rightarrow 1^-} h_{\lambda, f}(\mu_a) = 0.$$

Here,  $\mu_a = (1 - a) \sum_n a^n \mu^{*n}$ .

*Proof.* It is a special case of Corollary 5.11, as in this case,  $\mathcal{B}(G, \mu)$  is a point. ■

Using the (proven) Furstenberg conjecture (see [17]), we conclude the following characterization of amenability, which includes Theorem A.

**Theorem 6.2.** *The following are equivalent for a discrete countable group  $G$ :*

- (1)  $G$  is amenable.
- (2) For each  $S \subset G$  finite and  $\varepsilon > 0$ ,  $f \in C(0, \infty)$  with  $f(1) = 0$ , there exists a symmetric probability measure  $\eta$  on  $G$  such that

$$\forall g \in S : \left| \int_G f \left( \frac{dg\eta}{d\eta} \right) d\eta \right| < \varepsilon.$$

- (3)  $G$  has KL-almost invariant symmetric measures. That is, for any finite  $S \subset G$  and  $\varepsilon > 0$ , there is a symmetric probability measure  $\eta$  on  $G$  with

$$\forall g \in S : D_{\text{KL}}(g\eta \parallel \eta) < \varepsilon.$$

- (4) For any QI  $G$ -space  $X$ , finitely supported probability measure  $\lambda$  on  $G$  and a convex function  $f$  with  $f(1) = 0$ , one has  $I_{\lambda, f}(X) = 0$ .

Moreover, one can choose  $\eta$ 's in items (2) and (3) from a commutative subalgebra of measures.

*Proof.* (1)  $\Rightarrow$  (2): From [17, Theorem 4.4], there is a generating symmetric measure  $\mu$  such that  $(G, \mu)$  is Liouville. In particular, the Furstenberg–Poisson boundary is a point, and thus by Corollary 5.11 we conclude

$$\lim_{a \rightarrow 1^-} \int_G f \left( \frac{dg\mu_a}{d\mu_a} \right) d\mu_a = f(1) = 0$$

and thus taking  $\eta = \mu_a$  for  $a < 1$  large enough will satisfy the desired property.

For the moreover part, note that the measures  $\mu_a$  commute.

(2)  $\Rightarrow$  (4): From (2), it is clear that  $I_{\lambda, f}(G) = 0$ . From Corollary 3.9, we conclude (4).

(4)  $\Rightarrow$  (3) is clear. Also, (3) is a special case of (2) for  $f(t) = t \ln(t)$  (so the moreover part follows).

(3)  $\Rightarrow$  (1) follows from Pinsker's classical inequality  $\|m - \nu\| \leq \sqrt{2D_{\text{KL}}(m \parallel \nu)}$  (see, e.g., [4, Theorem 4.19]) and Reiter's condition for amenability. ■

**Remark 6.3.** (1) Note that in order to find  $\ell^1$ -almost invariant measures on amenable groups, one can use measures supported on finite sets (e.g., Følner sets). However, such measures cannot be the desired measures  $\eta$  given in item (3) of Theorem 6.2. Indeed, for the KL-divergence to be finite, the measure has to be supported on the whole group (if  $G$  is finitely generated).

(2) If in Theorem 6.2 one does not care about the commutativity (and symmetry) of  $\eta$ 's, it is possible to prove the result with the same machinery of ultralimits, without having to use Furstenberg's (proven) conjecture.

This will be explained in the sequel paper [24], where the framework of all amenable actions is covered.

We say that a probability measure  $\eta$  on a finitely generated group  $G$  is SAS (symmetric, adapted, smooth) if it is symmetric, generating and  $\sum \eta(g)e^{\varepsilon|g|} < \infty$  for some  $\varepsilon > 0$  and word metric  $|\cdot|$ .

Using Corollary 6.1, one sees the following property for Liouville groups.

**Corollary 6.4.** *Let  $G$  be a finitely generated group with a symmetric finitely supported Liouville measure. Then, there is a sequence  $(\eta_n)$  of SAS measures that is  $D_f$ -almost invariant, for any  $f$ -divergence. That is, for any  $g \in G$  we have  $D_f(g\eta_n \parallel \eta_n) \rightarrow 0$ .*

*Proof.* This follows from Corollary 6.1 by noting that if  $\mu$  is finitely supported and symmetric, then  $\mu_a$  are SAS for any  $a < 1$  and taking  $\eta_n = \mu_{1-1/n}$ . ■

**Remark 6.5.** Note that if a sequence of probability measures  $(\eta_n)$  on  $G$  is uniformly BQI and  $\ell^1$ -almost invariant, then it is  $D_f$ -almost invariant for any  $f$ -divergence. Indeed, from  $\ell^1$ -almost invariance, one deduces that for any non-principle ultrafilter  $\mathcal{U}$  on  $\mathbb{N}$  the ultra-limit measure  $\mathcal{U} \lim \eta_n$  is  $G$ -invariant. Using Lemma 4.18, we conclude the  $D_f$ -almost invariance.

Note that the measures built in Theorem 6.2 and Corollary 6.4 are uniformly BQI.

The next theorem concerns a fixed point result for amenable groups on RN models of stationary actions. It is a reformulation of a result from [21]. However, our proof is very different and is based on the problem of “entropy minimizing”:

**Theorem 6.6.** *Let  $G$  be an amenable group, and let  $\mu$  be a generating symmetric probability measure. Suppose  $X$  is a compact topological space, with a continuous  $G$ -action*

and that  $\nu$  is a  $\mu$ -stationary measure on  $X$  such that the Radon–Nikodym derivatives  $\frac{dg\nu}{d\nu}$  are continuous.

Then there is a continuous measure preserving factor  $\pi : X \rightarrow Y$  so that  $Y^G$ , the set of  $G$ -invariant points of  $Y$ , contains exactly one point. In particular, the action of an amenable group on the RN model of its Furstenberg–Poisson boundary admits a unique fixed point.

*Proof.* First, by changing  $\mu$  to  $\sum_{n=1}^{\infty} 2^{-n} \mu^{*n}$ , we may assume  $\mu$  is supported on all of  $G$ . By taking  $Y$  to be the RN model of the Radon–Nikodym factor of  $X$ , we may suppose that the Radon–Nikodym derivatives separate points of  $X$ , and show that  $X^G$  consists of one point.

Denote  $E = \{x \in X \mid \forall g \in G : \frac{dg\nu}{d\nu}(x) = 1\}$ . We claim that  $E = X^G$ . Indeed,  $E$  is  $G$ -invariant set, as the RN functions separate points, we see that  $E$  contains at most one point, thus  $E \subseteq X^G$ . For the other inclusion, if  $x \in X^G$ , then  $g \mapsto \frac{dg\nu}{d\nu}(x)$  is a multiplicative homomorphism; in particular,  $\frac{dg^{-1}\nu}{d\nu}(x) = \left(\frac{dg\nu}{d\nu}(x)\right)^{-1}$ , by symmetry of  $\mu$ :

$$1 = \sum_g \mu(g) \frac{dg\nu}{d\nu}(x) = \frac{1}{2} \sum_g \mu(g) \left( \frac{dg\nu}{d\nu}(x) + \left( \frac{dg\nu}{d\nu}(x) \right)^{-1} \right) \geq \sum_g \mu(g) = 1,$$

so we have equality which implies  $\frac{dg\nu}{d\nu}(x) = 1$  for any  $g \in G$ , so  $x \in E$ .

So it is enough to show  $E \neq \emptyset$ . Assume for the sake of contradiction that there is no point  $x_0$  with  $\frac{dg\nu}{d\nu}(x_0) = 1$  for all  $g$ . By compactness of  $X$ , we get there is a finite  $S_0 \subset G$  so that there is no point  $x_0$  with  $\frac{dg\nu}{d\nu}(x_0) = 1$  for all  $g \in S_0$ . We may assume  $e \in S_0$ ,  $S_0^{-1} = S_0$ . Note that in the Jensen inequality

$$-\sum_{s \in S_0} \mu(s) \ln \left( \frac{ds^{-1}\nu}{d\nu} \right) \geq -\mu(S_0) \ln \left( \sum_{s \in S_0} \frac{\mu(s)}{\mu(S_0)} \frac{ds^{-1}\nu}{d\nu} \right).$$

There is no equality, by assumption on  $S_0$ , but both sides are continuous functions, so we get some  $\varepsilon_0 > 0$  with

$$-\sum_{s \in S_0} \mu(s) \ln \left( \frac{ds^{-1}\nu}{d\nu} \right) \geq -\mu(S_0) \ln \left( \sum_{s \in S_0} \frac{\mu(s)}{\mu(S_0)} \frac{ds^{-1}\nu}{d\nu} \right) + \varepsilon_0. \quad (6.1)$$

Let  $\delta > 0$ , note that there is a symmetric probability measure  $\kappa$  on  $G$  which has finite support, agrees with  $\mu$  on  $S_0$  and  $\kappa \leq (1 + \delta) \cdot \mu$ . Indeed, take  $A \subset G \setminus S_0$  finite and symmetric with  $\mu(A) \geq \frac{\mu(G \setminus S_0)}{1 + \delta}$  and define  $\kappa = \mu|_{S_0} + \frac{1 - \mu(S_0)}{\mu(A)} \mu|_A$ .

We will show

$$I_\kappa(X) \geq \varepsilon_0 - \delta - \ln(1 + \delta). \quad (6.2)$$

Using Theorem 6.2, we see  $I_\kappa(X) = 0$ . This implies  $\varepsilon_0 \leq \delta + \ln(1 + \delta)$ , choosing an appropriate  $\delta > 0$ , we obtain a contradiction.

We now verify inequality (6.2). Using Jensen and (6.1),

$$\begin{aligned}
 & - \sum_{g \in G} \kappa(g) \ln \left( \frac{dg^{-1}v}{dv} \right) \\
 &= - \sum_{g \in S_0} \mu(g) \ln \left( \frac{dg^{-1}v}{dv} \right) - \sum_{g \in G \setminus S_0} \kappa(g) \ln \left( \frac{dg^{-1}v}{dv} \right) \\
 &\geq -\mu(S_0) \ln \left( \sum_{s \in S_0} \frac{\mu(s)}{\mu(S_0)} \frac{ds^{-1}v}{dv} \right) + \varepsilon_0 \\
 &\quad - \mu(G \setminus S_0) \ln \left( \sum_{g \in G \setminus S_0} \frac{\kappa(g)}{\kappa(G \setminus S_0)} \frac{dg^{-1}v}{dv} \right) \\
 &\geq -\ln \left( \frac{d\kappa * v}{dv} \right) + \varepsilon_0.
 \end{aligned}$$

Note that  $\kappa * v \leq (1 + \delta)\mu * v = (1 + \delta)v$  and thus  $\frac{d\kappa * v}{dv} \leq (1 + \delta)$ .

For any  $m \in M(X)$  (i.e., in the measure class of  $v$ ), we get by Corollary 3.15:

$$\begin{aligned}
 h_\kappa(m) &\geq \int_X \left( 1 - \frac{d\kappa * v}{dv} - \sum_{g \in G} \kappa(g) \ln \left( \frac{dg^{-1}v}{dv} \right) \right) dm \\
 &\geq \int_X \left( 1 - \frac{d\kappa * v}{dv} - \ln \left( \frac{d\kappa * v}{dv} \right) + \varepsilon_0 \right) dm \\
 &\geq 1 - (1 + \delta) - \ln(1 + \delta) + \varepsilon_0.
 \end{aligned}$$

Taking infimum over  $m$ , we obtain  $I_\kappa(X) \geq \varepsilon_0 - \delta - \ln(1 + \delta)$  and the proof is complete.  $\blacksquare$

**Remark 6.7.** One should note that for the lamplighter group,  $G = \mathbb{Z}/2 \wr \mathbb{Z}^d$  with  $d > 2$ , and the uniform symmetric measure  $\mu$  coming from the standard generating set  $\pm e_i, \delta_0$ , the space  $X = \prod_{\mathbb{Z}^d} \mathbb{Z}/2$  of final configuration with the measure of final configuration, is a  $\mu$ -stationary space  $(X, v)$  which is in fact the Furstenberg–Poisson boundary (see [20] for the Furstenberg–Poisson boundary of lamplighter groups). However, there are no fixed points in this action. The reason is that  $X$  is not RN model. More precisely, for  $0 \in X$ , and  $g = e_1$ , one can show that  $\frac{dg v}{dv}$  is not continuous at 0. We thank Gady Kozma for a discussion clarifying this issue.

This has the following implication: consider  $A = C^*(C(X) \cup \{\frac{dg v}{dv}\}_{g \in G}) \subset L^\infty(X)$  and let  $Y$  be the Gelfand dual. Then there is a continuous mapping  $\pi : Y \rightarrow X$  which is a measurable isomorphism. Moreover,  $Y$  is an RN-model, and thus by Theorem 6.6, we see that the  $G$ -invariant closed set  $S := \{y \mid \forall g \in G : \frac{dg v}{dv}(y) = 1\} \subset Y$  is non-empty. Since  $X$  is minimal, we conclude  $\pi(S) = X$ ; however,  $S$  has measure zero. This means that  $X$  is covered by the null-set  $S$ .

**Remark 6.8.** We believe that an existence of a fixed point on the RN-model for a Furstenberg–Poisson boundary  $\mathfrak{B}(G, \mu)$  for a symmetric generating  $\mu$  implies that  $G$  is amenable. However, we have not been able to prove this.

## 7. Free groups and entropy

Let  $F = F_d$  be a free group on  $d \geq 2$  generators, that we denote by  $a_1, \dots, a_d$ . Denote also  $a_{-i} = a_i^{-1}$ .

Our main goal in this section is, given  $f, \lambda$  where  $\lambda$  is symmetric and supported on the generators, to find  $I_{\lambda, f}(F) := \inf_{\nu \in M(F)} h_{\lambda, f}(F, \nu)$ .

### 7.1. The boundary of the free group and harmonic measures

Consider the standard topological boundary  $X = \partial F$  of  $F$  as the space of infinite length reduced words in the free generators, that is, the subset of  $\{a_{\pm 1}, \dots, a_{\pm d}\}^{\mathbb{N}}$  consisting of words  $(w_n)_{n \geq 0}$  with  $w_{n+1} \neq w_n^{-1}$ . It is clear that  $X$  is a compact metrizable space with the standard continuous  $F$  action.

For any  $\gamma \in F$ , we denote by  $X_\gamma$  the subspace of  $X$  consisting of words that begins with  $\gamma$ . The collection  $\{X_\gamma\}_{\gamma \in F}$  consists of clopen sets and gives a basis for the topology of  $X$ .

In this subsection, we fix a generating probability measure  $\mu$  on  $F$  which is supported on  $\{a_i\}_{i=\pm 1, \dots, \pm d}$  and denote  $p_i = \mu(a_i)$ .

The next lemma and most of the following discussion can be found in [19, Section 2]. We include the proof of the lemma in order to make our results (Theorem 7.11) explicit.

**Lemma 7.1.** *There exist unique  $1 > q_i > 0$  ( $i = \pm 1, \dots, \pm d$ ) such that for  $j = \pm 1, \dots, \pm d$ ,*

$$q_j = p_j + q_j \sum_{i \neq j} p_i q_{-i}.$$

*Proof.* Let  $x = 1 - \sum_i p_i q_{-i}$ . Then the equations for  $j, -j$  are equivalent to  $\frac{q_j}{p_j} = \frac{q_{-j}}{p_{-j}}$  and  $p_{-j} q_j^2 + x q_j - p_j = 0$  which yields that  $x$  determines  $q_j$  by

$$q_j = \frac{-x + \sqrt{x^2 + 4p_j p_{-j}}}{2p_{-j}}.$$

Thus, we need to show that there exist unique  $x \in (0, 1)$  with  $f(x) = 0$ , where

$$\begin{aligned} f(x) &= -x + 1 - \sum_i p_i \frac{-x + \sqrt{x^2 + 4p_{-i} p_i}}{2p_i} \\ &= (d-1)x + 1 - \sum_{j=1}^d \sqrt{x^2 + 4p_j p_{-j}}. \end{aligned}$$

Note that  $f(0) \geq 0$  (equality for symmetric measures),  $f(1) < 0$  and  $f$  is concave with  $f'(0) = d - 1 > 0$ . Thus,  $f$  has a unique zero in this interval. ■

Consider the  $\mu$ -random walk  $R_n$  on  $F$ . The random walk converges a.e. to an element of  $X$ . The hitting measure  $\nu_\mu$  is the unique  $\mu$ -stationary measure on  $X$ . The Furstenberg–Poisson boundary of  $(F, \mu)$  is isomorphic to  $(X, \nu_\mu)$  (see the Theorem in [16, Section 7.4]).

Before computing the measure  $\nu_\mu$ , we make the following observations and notation:

- The probabilities  $q_j := \mathbb{P}(\exists n : R_n = a_j)$  are the numbers of Lemma 7.1. Indeed, they are positive and  $< 1$  as the random walk is not recurrent. To show they satisfy the equation, we use the Markov property of the random walk. More precisely, we split according to the cases  $R_1 = a_i$ . The latter has probability  $p_i$ . For  $i = j$ , we get there exists such  $n$ , namely  $n = 1$ . For  $i \neq j$ , we note that in order to get to  $a_j$ , we need to get first from  $a_i$  to  $e$ , which has probability  $q_{-i}$ , and then to  $a_j$ , which has probability  $q_j$ . This shows  $q_j = p_j + q_j \sum_{i \neq j} p_i q_{-i}$ .
- Define  $v(a_i) = v_i = \nu_\mu(X_{a_i})$ . We claim that  $\frac{v_i}{1-v_{-i}} = q_i$ . Indeed, let us verify  $q_i(1 - v_{-i}) = v_i$ . For this we provide a probabilistic argument: the probability that the limit  $R_\infty$  starts with  $a_i$  is the probability that it gets to  $a_i$  for the first time at some finite  $n_0$ , and that if we consider the random walk  $R'_n = a_i^{-1} R_{n+n_0}$ , then  $R'_\infty$  does not start with  $a_{-i}$ . Now, by the Markov property, this happens in probability  $q_i(1 - v_{-i})$ , verifying the claim. From the formulas  $\frac{v_i}{1-v_{-i}} = q_i$  and  $\frac{v_{-i}}{1-v_i} = q_{-i}$ , we conclude the following direct formula for  $v_i$ :

$$v_i = \frac{q_i(1 - q_{-i})}{1 - q_{-i}q_i}.$$

Note that if  $\mu$  is symmetric, then  $p_i, q_i, v_i$  are symmetric and the above reduces to  $v_i = \frac{q_i}{1+q_i}$ .

**Lemma 7.2.** *The measure  $\nu_\mu$  is given by*

$$\nu_\mu(X_\gamma) = v(\gamma_1) \prod_{\ell=2}^m \frac{v(\gamma_\ell)}{1 - v(\gamma_{\ell-1}^{-1})} \quad \gamma = \gamma_1 \cdots \gamma_m \text{ reduced form,}$$

and

$$\frac{da_i \nu_\mu}{d\nu_\mu} = \begin{cases} q_{-i} & \text{on } X \setminus X_{a_i}, \\ \frac{1}{q_i} & \text{on } X_{a_i}. \end{cases}$$

*Proof.* Since  $\sum_i v_i = \sum_i \nu_\mu(X_{a_i}) = 1$  by the definition of  $v_i$ , the right-hand side yields a Borel probability measure  $\nu$  on  $X$ . We will show

$$\frac{da_i \nu}{d\nu} = \begin{cases} q_{-i} & \text{on } X \setminus X_{a_i}, \\ \frac{1}{q_i} & \text{on } X_{a_i}, \end{cases}$$



and conclude  $\nu$  is  $\mu$ -stationary. By the well-known uniqueness of the stationary measure, we deduce  $\nu = \nu_\mu$ .

Let  $\gamma = \gamma_1 \cdots \gamma_m$  be a reduced form. A direct computation yields

$$\begin{aligned}
 (a_i \nu)(X_\gamma) &= \nu(a_i^{-1} X_\gamma) = \begin{cases} \nu(X_{a_i^{-1} \gamma}) & \gamma_1 \neq a_i, \\ \nu(X_{\gamma_2 \cdots \gamma_m}) & \gamma_1 = a_i. \end{cases} \\
 &= \begin{cases} \nu(a_i^{-1}) \frac{\nu(\gamma_1)}{1 - \nu(a_i)} \prod_{j=2}^m \frac{\nu(\gamma_j)}{1 - \nu(\gamma_{j-1}^{-1})} & \gamma_1 \neq a_i, \\ \nu(\gamma_2) \prod_{j=3}^m \frac{\nu(\gamma_j)}{1 - \nu(\gamma_{j-1}^{-1})} & \gamma_1 = a_i. \end{cases} \\
 &= \begin{cases} \frac{\nu_{-i}}{1 - \nu_i} \cdot \nu(\gamma_1) \prod_{j=2}^m \frac{\nu(\gamma_j)}{1 - \nu(\gamma_{j-1}^{-1})} & \gamma_1 \neq a_i, \\ \frac{1 - \nu_{-i}}{\nu_i} \cdot \nu(\gamma_1) \prod_{j=2}^m \frac{\nu(\gamma_j)}{1 - \nu(\gamma_{j-1}^{-1})} & \gamma_1 = a_i. \end{cases} \\
 &= \nu(X_\gamma) \cdot \begin{cases} q_{-i} & \gamma_1 \neq a_i, \\ \frac{1}{q_i} & \gamma_1 = a_i. \end{cases}
 \end{aligned}$$

This shows the formula for  $\frac{da_i \nu}{dv}$ . Now, for  $\xi \in X$  starting with  $a_j$ , we conclude that

$$\begin{aligned}
 \frac{d\mu * \nu}{dv}(\xi) &= \sum_i p_i \cdot \frac{da_i \nu}{dv}(\xi) = \sum_{i \neq j} p_i \cdot q_{-i} + \frac{p_j}{q_j} \\
 &= \frac{1}{q_j} \left( p_j + q_j \sum_{i \neq j} p_i \cdot q_{-i} \right) = \frac{q_j}{q_j} = 1.
 \end{aligned}$$

This shows  $\mu * \nu = \nu$  which proves the lemma. ■

## 7.2. Measures on the boundary minimizing entropy in their class

We now turn to finding entropy-minimizing measures on the boundary.

**Notation 7.3.** • Let  $f$  be a strictly convex smooth function on  $(0, \infty)$  with  $f(0) = 0$ .

Denote  $\Psi_f(z) = f(z) - z f'(z) + f'(\frac{1}{z})$ .

- Denote by  $\Delta_d$  the set of generating symmetric probability measures on  $F_d$  supported on  $\{a_i\}_{i=\pm 1, \dots, \pm d}$ .

Note that  $\Psi'_f(z) = -z \cdot f''(z) - \frac{1}{z^2} f''(z) < 0$  and thus  $\Psi_f$  is a decreasing function. Given  $f$  as above, we define a mapping  $\Delta_d \xrightarrow{T} \Delta_d$  as follows.

**Definition 7.4.** Given  $\mu \in \Delta_d$ , denote  $p_i = \mu(a_i)$  and let  $q_i$  be as in Lemma 7.1. Define  $T(\mu) = \lambda \in \Delta_d$  by

$$\lambda(a_{\pm j}) = \lambda_j = \frac{c}{\Psi_f(q_j) - \Psi_f\left(\frac{1}{q_j}\right)},$$

where  $c$  is a normalization  $c = \left(2 \sum_{i=1}^d \frac{1}{\Psi_f(q_i) - \Psi_f\left(\frac{1}{q_i}\right)}\right)^{-1}$ .

Since  $\Psi_f$  is decreasing,  $T$  is well defined.

**Example 7.5.** For  $f(z) = z \ln(z)$ , we have  $\Psi(z) = -\ln(z) + 1 - z$ , and  $\Psi(q) - \Psi(q^{-1}) = 2 \ln(q^{-1}) + q^{-1} - q$ .

The next proposition gives a minimizing measure for the entropy in a specific measure classes on the boundary.

**Proposition 7.6.** Let  $f$  be as in the notation. Given  $\mu \in \Delta_d$  let  $\lambda = T(\mu)$ . Consider  $X = \partial F$  as a QI  $F$ -space with the measure class of  $\nu_\mu$ . Then

$$h_{\lambda,f}(X, \nu_\mu) = I_{\lambda,f}(X).$$

*Proof.* Note that since  $\mu$  is symmetric,  $p_i, q_i, v_i, \lambda_i$  are all symmetric.

Note that as  $\nu_\mu$  is BQI and  $\lambda$  has finite support, the conditions of Proposition 3.13 apply, and  $\nu_\mu$  is minimal for the entropy  $h_{\lambda,f}$  in its measure class iff  $\Psi_{\lambda,f}(\nu_\mu; \cdot)$  is a constant function. We now show  $\Psi_{\lambda,f}(\nu_\mu; \cdot)$  is constant. Note that as  $\lambda$  is symmetric,

$$\Psi_{\lambda,f}(\nu_\mu; \cdot) = \sum_{i=\pm 1, \dots, \pm d} \lambda(a_i) \left( \frac{\partial F}{\partial x} \left( 1, \frac{da_i \nu_\mu}{dv_\mu} \right) + \frac{\partial F}{\partial y} \left( \frac{da_i \nu_\mu}{dv_\mu}, 1 \right) \right),$$

where, as usual,  $F(x, y) := yf\left(\frac{x}{y}\right)$ . Note that

$$\frac{\partial F}{\partial x}(1, z) + \frac{\partial F}{\partial y}(z, 1) = f\left(\frac{1}{z}\right) + f(z) - zf'(z) = \Psi_f(z).$$

Now take  $\xi \in X$  that starts with  $a_j$ :

$$\begin{aligned} \Psi_{\lambda,f}(\nu_\mu; \xi) &= \sum_{i=\pm 1, \dots, \pm d} \lambda(a_i) \Psi_f \left( \frac{da_i \nu_\mu}{dv_\mu}(\xi) \right) \\ &= C + \lambda(a_j) \cdot \left( \Psi_f \left( \frac{1}{q_j} \right) - \Psi_f(q_j) \right) = C + c, \end{aligned}$$

where  $C = \sum_{i=\pm 1, \dots, \pm d} \lambda(a_i) \Psi_f(q_i)$  is independent of  $j$ , and  $c$  is the normalization in the definition of  $T$  (see Definition 7.4). This is independent of  $j$ , concluding the proof. ■

**Remark 7.7.** Note that in the infimum  $I_{\lambda,f}(X)$ , we only take measures in the measure class of  $\nu_\mu$ . In a future work, the first named author discusses the problem of minimal entropy over all measures on  $X$ , and it will turn out that the infimum remains the same, although this requires considerably more effort.

We would next like to show that we indeed cover in our analysis *all* the generating measures  $\lambda$  which are symmetric and supported on the set of free generators and their inverses.

**Lemma 7.8.** *The mapping  $T : \Delta_d \rightarrow \Delta_d$  is a bijection.*

*Proof.* We first extend  $T$  to the boundary of the simplex  $\Delta_d$ . This can be done by interpreting the boundary as elements of  $\Delta_r$ ,  $r < d$  and taking the corresponding  $T$  there. This yields a continuous map  $T : \overline{\Delta_d} \rightarrow \overline{\Delta_d}$  that sends the boundary to itself. This implies that  $T(\Delta_d) = \Delta_d \cap T(\overline{\Delta_d}) \subset \Delta_d$  is relatively closed.

We wish to show that the image is also open, which would then imply  $T$  is surjective. By invariance of domain (cf. [14, Theorem 2B.3]), it is enough to show that  $T$  is injective.

In order to show injectivity of  $T$ , we first show that  $\mu = (p_i) \mapsto (q_i)$  is injective, and then that  $(q_i) \mapsto \lambda$  is injective.

For the first mapping, from  $(q_i)$  one can reconstruct  $(p_i)$ , as the defining equation is

$$\begin{pmatrix} \frac{1}{q_1} + q_1 & 2q_2 & 2q_3 & \cdots & 2q_d \\ 2q_1 & \frac{1}{q_2} + q_2 & 2q_3 & \cdots & 2q_d \\ & & \vdots & & \\ 2q_1 & 2q_2 & 2q_3 & \cdots & \frac{1}{q_d} + q_d \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_d \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

So we will show the matrix above is non-singular. Employing row operations, this matrix has the same determinant as the matrix:

$$\begin{pmatrix} \frac{1}{q_1} + q_1 & 2q_2 & 2q_3 & \cdots & 2q_d \\ q_1 - \frac{1}{q_1} & \frac{1}{q_2} - q_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ q_1 - \frac{1}{q_1} & 0 & 0 & \cdots & \frac{1}{q_d} - q_d \end{pmatrix}.$$

We show that the determinant of the latter matrix is positive by induction. Indeed, expanding using the second row, we get determinant

$$\begin{aligned} & - \left( q_1 - \frac{1}{q_1} \right) 2q_2 \left( \frac{1}{q_3} - q_3 \right) \cdots \left( \frac{1}{q_d} - q_d \right) \\ & + \left( \frac{1}{q_2} - q_2 \right) \det \begin{pmatrix} \frac{1}{q_1} + q_1 & 2q_3 & \cdots & 2q_d \\ q_1 - \frac{1}{q_1} & \frac{1}{q_3} - q_3 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ q_1 - \frac{1}{q_1} & 0 & \cdots & \frac{1}{q_d} - q_d \end{pmatrix}. \end{aligned}$$

The second term is the same determinant but for less variables and since  $0 < q_i < 1$  the first term is  $> 0$ , by the induction assumption the second term is positive. So, we need to show only the base case, and indeed

$$\det \begin{pmatrix} \frac{1}{q_1} + q_1 & 2q_2 \\ q_1 - \frac{1}{q_1} & \frac{1}{q_2} - q_2 \end{pmatrix} = \left( \frac{1}{q_1} + q_1 \right) \left( \frac{1}{q_2} - q_2 \right) + 2q_2 \left( \frac{1}{q_1} - q_1 \right) > 0.$$

Thus  $\mu \mapsto (q_i)$  is injective. Let us show that the mapping  $(q_i) \mapsto \lambda$  is injective.

Consider  $\varphi(z) = \Psi_f(z) - \Psi_f(\frac{1}{z})$ ; this is a positive decreasing function (as  $\Psi_f$  is decreasing). Note that from  $\lambda$  we can reconstruct  $\frac{\varphi(q_i)}{\varphi(q_j)} = \frac{\lambda_j}{\lambda_i}$  for each  $i, j$ . Assume that  $\lambda$  corresponds to distinct  $(q_j) \neq (q'_j)$ . Without loss of generality, there is some  $j_0$  with  $q_{j_0} > q'_{j_0}$ . As  $v(z) = \frac{z}{1+z}$  is an increasing function and  $(v(q_j)), (v(q'_j))$  are probability measures (see the discussion before Lemma 7.2), we conclude that there must be  $j_1$  with  $q_{j_1} < q'_{j_1}$ . But then  $\frac{\varphi(q_{j_0})}{\varphi(q_{j_1})} < \frac{\varphi(q'_{j_0})}{\varphi(q'_{j_1})}$ , which is a contradiction. ■

**Remark 7.9.** Note that the proof of Proposition 7.6 yields that for  $\lambda \neq T(\mu)$  the measure  $\nu_\mu$  is not minimizing in its measure class for  $h_{\lambda,f}$ .

It is not difficult to see that for  $f(t) = t \ln(t)$ , the map  $T$  is not the identity (see the following example). This means that the  $\mu$ -entropy minimizer on the boundary  $X$  is, in general, *not* the  $\mu$ -stationary measure.

**Example 7.10.** Let  $d = 2$  and  $f(t) = t \ln(t)$ . Consider  $\mu(a_{\pm 1}) = \frac{1}{3}, \mu(a_{\pm 2}) = \frac{1}{6}$ . A direct computation (using computer and the proof of Lemma 7.1) shows that  $\lambda = T(\mu)$  has  $\lambda(a_{\pm 1}) \sim 0.32378$  and in particular  $\lambda \neq \mu$ . Further computation (using Lemma 7.2) yields  $h_\lambda(\nu_\mu) \sim 0.5126$ .

### 7.3. Minimizing entropy for the action of $F$ on itself

In the following theorem, we compute the minimal entropy number  $I_{\lambda,f}$  for the action of the free group  $F$  on itself. Note the next theorem includes Theorem C.

**Theorem 7.11.** *Let  $f$  be a strictly convex smooth function on  $(0, \infty)$ .*

*Let  $\lambda \in \Delta_d$  and let  $\mu = T^{-1}(\lambda) \in \Delta_d$ . Then*

$$I_{\lambda,f}(F) = h_{\lambda,f}(\partial F, \nu_\mu).$$

*Proof.* Consider  $X = \partial F$  as a QI  $G$ -space with the measure class of  $\nu_\mu$ , then by Corollary 3.9, we have  $I_{\lambda,f}(F) \geq I_{\lambda,f}(X)$ . By Proposition 7.6, we have  $I_{\lambda,f}(X) = h_{\lambda,f}(X, \nu_\mu)$ .

On the other hand, by Corollary 5.11 (the corollary of the main result),

$$h_{\lambda,f}(X, \nu_\mu) = \lim_{a \rightarrow 1^-} h_{\lambda,f}(F, \mu_a) \geq I_{\lambda,f}(F).$$

This finishes the proof. ■

As for Theorem B and the more general Theorem D, we have the following corollary.

**Corollary 7.12.** *Let  $\mu$  be the uniform symmetric measure  $\mu(a_{\pm j}) = \frac{1}{2d}$ . Then for any strictly convex and smooth  $f$ , we have*

$$I_{\mu,f}(F_d) = \frac{2d-1}{2d} f\left(\frac{1}{2d-1}\right) + \frac{1}{2d} f(2d-1).$$

In particular, taking  $f(t) = t \ln(t)$ , we conclude

$$I_{\mu, \text{KL}}(F_d) = \frac{d-1}{d} \ln(2d-1).$$

*Proof.* It is clear that  $v_i, q_i$  are independent of  $i$  and thus  $v_i = \frac{1}{2d}, \lambda = T(\mu) = \mu$ .

Hence,  $q_i = \frac{\frac{1}{2d}}{1 - \frac{1}{2d}} = \frac{1}{2d-1}$ . Thus, by Theorem 7.11,

$$\begin{aligned} I_{\mu, f}(F_d) &= h_{\lambda, f}(v_\mu) = \sum_i \frac{1}{2d} \int_{\partial F_d} f\left(\frac{da_i v_\mu}{dv_\mu}\right) dv_\mu \\ &= \frac{2d-1}{2d} f\left(\frac{1}{2d-1}\right) + \frac{1}{2d} f(2d-1). \end{aligned}$$

In particular,  $I_{\mu, \text{KL}}(F_d) = \frac{d-1}{d} \ln(2d-1)$ . ■

## 8. Final remarks

### 8.1. Possible relation with the Liouville problem

**8.1.1. Entropy and stationary measures.** One initial motivation for this work was an attempt to present stationary measures as information-theoretic minimizers. Using the approach and results established in this paper, where one approximates the entropy of all boundaries by the self action of the group, it is not difficult to see that this would have immediate implications to the fundamental open problem of (in)dependence of the Liouville property on the (symmetric, finitely supported) generating measure chosen on the group. A natural approach toward the minimization property would be to show that convolution with  $\lambda$  might reduce the  $\lambda$ -entropy. As it turns out, this is usually not true. Indeed, for the action  $F_2 \curvearrowright \partial F_2$  and  $\lambda$  a non-uniform measure on the free generators, the  $\lambda$ -harmonic measure is not in general a minimizing measure for the  $\lambda$ -entropy. Moreover, there is a minimum (in a measure class) for the entropy which is the  $\mu$ -harmonic measure for a measure  $\mu \neq \lambda$ . In fact, let us explain why in general no information-theoretic entropy can be used to detect stationary measures as minimizers.

Let  $h = h_{\kappa, f}$  (with  $\kappa$  finitely supported) be some entropy for an amenable group  $G$  with a non-Liouville measure  $\mu$  (e.g.,  $G = \mathbb{Z}/2 \wr \mathbb{Z}^3$ ). We explain why convolution by  $\mu$  cannot always decrease the entropy  $h$ .

Indeed, take  $(B, \nu)$  to be (a topological model of) the Furstenberg–Poisson boundary. We know  $0 = I_{\kappa, f}(B)$  and hence there is  $m \in M(B)$  (in the measure class) with  $h(m) < h(\nu)$ . Observe that  $\mu^{*n} * m$  converges weakly to  $\nu$ . Indeed, consider  $\Omega = G^{\mathbb{N}}$  with the  $\mu$ -random  $(X_n)$  and its probability measure  $\mathbb{P}$ . Note that for the stationary  $\sigma$ -algebra  $\mathcal{S}$  (see [17, Subsection 0.3]), one has that  $(B, \nu)$  is a topological model of  $(\Omega, \mathcal{S}, \mathbb{P})$ . Suppose

$h = \frac{dm}{dv}$ , then for any  $a \in L^\infty(B) = L^\infty(\Omega)$  we have

$$\begin{aligned} (\mu^{*n} * m)(a) &= \mathbb{E}_{\mathbb{P}}[(X_n \cdot m)(a)] \\ &= \mathbb{E}_{\mathbb{P}} \left[ \int_B a(X_n \cdot x) h(x) dv(x) \right] = \int_B \mathbb{E}_{\mathbb{P}}[a(X_n \cdot x)] h(x) dv(x). \end{aligned}$$

Note that (a.e.)  $a(X_n(\omega) \cdot x) \rightarrow a(\omega)$ . Thus,  $\mathbb{E}_{\mathbb{P}}[a(X_n \cdot x)] \rightarrow v(a)$  for a.e.  $x \in B$  which implies that  $(\mu^{*n} * m)(a) \rightarrow v(a)$  which shows weak convergence.

Since  $h$  is lower semi-continuous with respect to weak topology (follows from [1, Theorem 2.34] which shows  $f$ -divergence are such), we conclude that  $\liminf_n h(\mu^{*n} * m) \geq h(v) > h(m)$ , showing that convolution by  $\mu$  cannot always decrease entropy.

**8.1.2. Other versions of almost-invariant measures.** In view of Theorem 6.2, one can ask when there are  $\ell^\infty$ -almost-invariant measures on a group. That is, when for every  $S \subset G$  finite and  $\varepsilon > 0$ , there is a measure  $\eta$  on  $G$  such that  $\|\frac{ds\eta}{d\eta} - 1\|_\infty \leq \varepsilon$  for all  $s \in S$ . It is easy to see that for finitely generated groups, this is equivalent to having subexponential growth.

Recall that we say that a measure  $\kappa$  on a finitely generated group  $G$  is SAS if it is symmetric, generating and satisfies  $\sum \kappa(g) e^{\varepsilon|g|} < \infty$  for some  $\varepsilon > 0$  and word metric  $|\cdot|$ .

**Definition 8.1.** A finitely generated group is *tame* if there is a sequence  $(\kappa_n)$  of SAS measures that is KL-almost invariant.

Corollary 6.4 implies that groups with a symmetric finitely supported Liouville measure are tame. We find the following problem intriguing.

**Problem 8.2.** Is there a finitely generated amenable group which is not tame?

Are the lamplighters  $\mathbb{Z}/2 \wr \mathbb{Z}^d$ ,  $d \geq 3$  tame?

Our hope is that the above lamplighters are not tame. One can then make the speculation that being tame is related to the “Liouville dependence problem” concerning boundary triviality of symmetric finitely supported measures. Our proof of the fixed point Theorem 6.6 came as an attempt to employ these ideas. However, eventually it produces a fixed point without the SAS assumption.

## 8.2. Uniqueness of entropy minimizers

As explained in the introduction, for some actions  $G \curvearrowright X$ , the infimum of the entropy function is attained. The  $f$ -divergence is lower semi-continuous with respect to the weak topology<sup>4</sup>. Thus, if  $X$  is a compact  $G$ -space, the entropy function  $h_{\lambda,f} : M(X) \rightarrow [0, \infty]$

<sup>4</sup>See [1, Theorem 2.34]; there is also a proof using a variational representation of  $f$ -divergences.

is lower semi-continuous. Hence, the entropy  $h_{\lambda,f}$  attains its infimum  $I_{\lambda,f}^{\text{top}}(X)$ . It is then very natural to inquire about its uniqueness. While as we have seen (Lemma 3.10), inside a specific ergodic measure class, uniqueness holds under natural conditions on  $f$ , we have no criterion for uniqueness in the topological setting, nor can we establish any non-trivial uniqueness result.

Another problem is finding a criterion for a measure  $\nu$  to be a minimizer of entropy. Inside a specific measure class we have Proposition 7.6 yielding a complete solution; in the topological case, however, we do not have any criterion.

In the sequel paper [24], we will show that for the topological action  $F_d \curvearrowright \partial F_d$  we have under the notations of Proposition 7.6 that  $I_{\lambda,f}^{\text{top}}(\partial F_d) = h_{\lambda,f}(\partial F_d, \nu_\mu)$ .

**Funding.** The authors acknowledge the ISF support made through grant 1483/16.

## References

- [1] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*. Oxford Math. Monogr., Oxford University Press, Oxford, 2000 Zbl 0957.49001 MR 1857292
- [2] U. Bader and Y. Shalom, *Factor and normal subgroup theorems for lattices in products of groups*. *Invent. Math.* **163** (2006), no. 2, 415–454 Zbl 1085.22005 MR 2207022
- [3] M. Björklund, Five remarks about random walks on groups. 2014, arXiv:1406.0763v2
- [4] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: a nonasymptotic theory of independence*. Oxford University Press, Oxford, 2013 Zbl 1279.60005 MR 3185193
- [5] N. Bourbaki, *General topology: chapters 1–4*. Springer, Heidelberg, 2013 Zbl 0894.54001 MR 1726779
- [6] C. T. Conley, A. S. Kechris, and R. D. Tucker-Drob, *Ultraproducts of measure preserving actions and graph combinatorics*. *Ergodic Theory Dynam. Systems* **33** (2013), no. 2, 334–374 Zbl 1268.05104 MR 3035288
- [7] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, New York, 1999 Zbl 0762.94001 MR 1122806
- [8] S. Das and J. Peterson, *Poisson boundaries of  $\text{II}_1$  factors*. *Compos. Math.* **158** (2022), no. 8, 1746–1776 Zbl 1507.46041 MR 4493239
- [9] G. Elek and B. Szegedy, Limits of hypergraphs, removal and regularity lemmas. A non-standard approach. 2007, arXiv:0705.2179v1
- [10] H. Furstenberg, *A Poisson formula for semi-simple Lie groups*. *Ann. of Math. (2)* **77** (1963), no. 2, 335–386 Zbl 0192.12704 MR 0146298
- [11] H. Furstenberg, Boundary theory and stochastic processes on homogeneous spaces. In C. C. Moore (ed.), *Harmonic analysis on homogeneous spaces*, pp. 193–229, Proc. Sympos. Pure Math. 26, American Mathematical Society, Providence, RI, 1973 Zbl 0289.22011 MR 0352328
- [12] H. Furstenberg, *A note on Borel’s density theorem*. *Proc. Amer. Math. Soc.* **55** (1976), no. 1, 209–212 Zbl 0319.22010 MR 0422497

- [13] H. Furstenberg and E. Glasner, Stationary dynamical systems. In S. Kolyada, Y. Manin, M. Möller, P. Moree, and T. Ward (eds.), *Dynamical numbers: interplay between dynamical systems and number theory*, pp. 1–28, Contemp. Math. 532, American Mathematical Society, Providence, RI, 2010 Zbl [1218.22004](#) MR [2762131](#)
- [14] A. Hatcher, *Algebraic topology*. Cambridge University Press, Cambridge, 2002 Zbl [1044.55001](#) MR [1867354](#)
- [15] S. Heinrich, *Ultraproducts in Banach space theory*. *J. Reine Angew. Math.* **1980** (1980), no. 313, 72–104 Zbl [0412.46017](#) MR [0552464](#)
- [16] V. A. Kaimanovich, *The Poisson formula for groups with hyperbolic properties*. *Ann. of Math.* (2) **152** (2000), no. 3, 659–692 Zbl [0984.60088](#) MR [1815698](#)
- [17] V. A. Kaimanovich and A. M. Vershik, *Random walks on discrete groups: boundary and entropy*. *Ann. Probab.* **11** (1983), no. 3, 457–490 Zbl [0641.60009](#) MR [0704539](#)
- [18] M. G. Kuhn, *Amenable actions and weak containment of certain representations of discrete groups*. *Proc. Amer. Math. Soc.* **122** (1994), no. 3, 751–757 Zbl [0829.43003](#) MR [1209424](#)
- [19] F. Ledrappier, *Some asymptotic properties of random walks on free groups*. In J. C. Taylor (ed.), *Topics in probability and Lie groups: boundary theory*, pp. 117–152, CRM Proc. Lecture Notes 28, American Mathematical Society, Providence, RI, 2001 Zbl [0994.60073](#) MR [1832436](#)
- [20] R. Lyons and Y. Peres, *Poisson boundaries of lamplighter groups: proof of the Kaimanovich–Vershik conjecture*. *J. Eur. Math. Soc. (JEMS)* **23** (2021), no. 4, 1133–1160 Zbl [1542.20209](#) MR [4228277](#)
- [21] S. Northshield, *Amenability and superharmonic functions*. *Proc. Amer. Math. Soc.* **119** (1993), no. 2, 561–566 Zbl [0785.31006](#) MR [1164149](#)
- [22] A. Ramsay, *Virtual groups and group actions*. *Adv. Math.* **6** (1971), no. 3, 253–322 (1971) Zbl [0216.14902](#) MR [0281876](#)
- [23] J. Rosenblatt, *Ergodic and mixing random walks on locally compact groups*. *Math. Ann.* **257** (1981), no. 1, 31–42 Zbl [0451.60011](#) MR [0630645](#)
- [24] E. Sayag, *Ultralimits, amenable actions and entropy*. [v1] 2022, [v2] 2024, arXiv:[2204.11112v2](#)
- [25] W. Woess, *Random walks on infinite graphs and groups*. Cambridge Tracts in Math. 138, Cambridge University Press, Cambridge, 2000 Zbl [0951.60002](#) MR [1743100](#)
- [26] W. Woess, *Ratio limits and Martin boundary*. *Doc. Math.* **26** (2021), 1501–1528 Zbl [1482.60104](#) MR [4493568](#)
- [27] S. Zhou, *Noncommutative Poisson boundaries, ultraproducts, and entropy*. *Int. Math. Res. Not. IMRN* **2024** (2024), no. 10, 8794–8818 Zbl [07931944](#) MR [4749188](#)
- [28] R. J. Zimmer, *Ergodic theory and semisimple groups*. Monogr. Math. 81, Birkhäuser Boston, MA, 1984 Zbl [0571.58015](#) MR [0776417](#)

Received 27 December 2023.

**Elad Sayag**

School of Mathematical Sciences, Tel-Aviv University, P.O. Box 39040, 69978 Tel-Aviv, Israel;  
[eladsayag@mail.tau.ac.il](mailto:eladsayag@mail.tau.ac.il), [elad.sayag2003@gmail.com](mailto:elad.sayag2003@gmail.com)



**Yehuda Shalom**

School of Mathematical Sciences, Tel-Aviv University, P.O. Box 39040, 69978 Tel-Aviv, Israel;  
[yeshalom@tauex.tau.ac.il](mailto:yeshalom@tauex.tau.ac.il)