



Ian D. Morris · Çağrı Sert

## A converse statement to Hutchinson's theorem and a dimension gap for self-affine measures

Received June 25, 2020

**Abstract.** A well-known theorem of J. E. Hutchinson states that if an iterated function system consists of similarity transformations and satisfies the open set condition then its attractor supports a self-similar measure with Hausdorff dimension equal to the similarity dimension. In this article we prove the following result which may be regarded as a form of partial converse: if an iterated function system consists of invertible affine transformations whose linear parts do not preserve a common invariant subspace, and its attractor supports a self-affine measure with Hausdorff dimension equal to the affinity dimension, then the system necessarily consists of similarity transformations. We obtain this result by showing that the equilibrium measures of an affine iterated function system are never Bernoulli measures unless the system either is reducible or consists of similarity transformations. The proof builds on earlier results in the thermodynamic formalism of affine iterated function systems due to Bochi, Feng, Käenmäki, Shmerkin and the first named author and also relies on the work of Benoist on the spectral properties of Zariski-dense subsemigroups of reductive linear groups.

**Keywords.** Affinity dimension, iterated function system, self-affine measure, self-affine set

### 1. Introduction

An *iterated function system* is by definition a tuple  $(T_1, \dots, T_N)$  of contracting transformations of some metric space  $X$ , which in this article will be taken to be  $\mathbb{R}^d$ . To avoid trivialities it will be assumed throughout this article that  $N \geq 2$ . If  $(T_1, \dots, T_N)$  is an iterated function system acting on  $\mathbb{R}^d$  then it is well-known that there exists a unique nonempty compact subset  $Z \subset \mathbb{R}^d$  with the property  $Z = \bigcup_{i=1}^N T_i Z$ , called the *attractor* or *limit set* of the iterated function system. If additionally any probability vector  $(p_1, \dots, p_N)$  is specified then there exists a unique Borel probability measure  $m$  on  $\mathbb{R}^d$  such that  $m = \sum_{i=1}^N p_i(T_i)_*m$ . In the case where the transformations  $T_i$  are contracting

---

Ian D. Morris: School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, UK; [i.morris@qmul.ac.uk](mailto:i.morris@qmul.ac.uk)

Çağrı Sert: Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland; [cagri.sert@math.uzh.ch](mailto:cagri.sert@math.uzh.ch)

*Mathematics Subject Classification (2020):* Primary 28A80; Secondary 37D35, 37D45, 37H15

similitudes of  $\mathbb{R}^d$  we call the limit set  $Z$  a *self-similar set* and the measure  $m$  a *self-similar measure*.

For each  $x \in \mathbb{R}^d$  and  $r > 0$  let  $B_r(x)$  denote the open Euclidean ball with radius  $r$  and centre  $x$ . If  $m$  is a Borel probability measure  $m$  on  $\mathbb{R}^d$  such that the limit

$$\lim_{r \rightarrow 0} \frac{\log m(B_r(x))}{\log r}$$

exists for  $m$ -a.e.  $x$  and is constant  $m$ -a.e, we say that  $m$  is *exact-dimensional* and define the *dimension* of  $m$  to be the value of this almost-everywhere limit. We denote this dimension by  $\dim m$ . It was shown in 2009 by D.-J. Feng and H. Hu that every self-similar measure on  $\mathbb{R}^d$  is exact-dimensional [26]. We denote the Hausdorff dimension of any subset  $Z$  of  $\mathbb{R}^d$  by  $\dim_{\text{H}} Z$ .

An iterated function system is said to satisfy the *open set condition* if there exists a nonempty open set  $U$  such that  $T_i U \subseteq U$  for all  $i = 1, \dots, N$  and  $T_i U \cap T_j U = \emptyset$  whenever  $i \neq j$ , and is said to satisfy the *strong open set condition* if additionally  $U \cap Z \neq \emptyset$ . The starting point of the motivation for this article is the following landmark theorem of J. E. Hutchinson [35]:

**Theorem 1** (Hutchinson). *Let  $T_1, \dots, T_N: \mathbb{R}^d \rightarrow \mathbb{R}^d$  be contracting similitudes of the form  $T_i x := r_i O_i x + v_i$  for some  $r_i \in (0, 1)$ ,  $O_i \in O(d)$  and  $v_i \in \mathbb{R}^d$  and suppose that  $(T_1, \dots, T_N)$  satisfies the open set condition. Then the Hausdorff dimension of the attractor  $Z$  of the iterated function system  $(T_1, \dots, T_N)$  is equal to the unique real number  $s \in (0, d]$  such that  $\sum_{i=1}^N r_i^s = 1$ . Moreover, there exists a unique self-similar measure  $m$  supported on  $Z$  with dimension  $s$ .*

The extension of Theorem 1 in various directions has been an active topic of research since its original publication. One major area of research has been the problem of understanding systematically what happens when the open set condition is removed (such as in [3, 16, 31, 32, 42, 50, 53, 55]) and this line of research has focused especially on the dimensions of the resulting measures as opposed to the resulting sets. A second major direction of extension of Theorem 1 is that in which the transformations  $T_i$  are allowed to be arbitrary affine contractions instead of similitudes: this line of research dates back to the work of Bedford, McMullen and Falconer in the 1980s [8, 22, 43] and has been particularly active within the last few years (see for example [5–7, 13, 19, 25, 28, 38, 51]). It is with this second direction of extension that this article is concerned.

When  $(T_1, \dots, T_N)$  is an iterated function system consisting of affine contractions of  $\mathbb{R}^d$ , the attractor of  $(T_1, \dots, T_N)$  is referred to as a *self-affine set* and Borel probability measures satisfying  $m = \sum_{i=1}^N p_i(T_i)_* m$  are referred to as *self-affine measures*. It was shown recently by D.-J. Feng [25] that every self-affine measure is exact-dimensional; previous partial results in this direction include [4, 6, 29]. Let us now describe the most natural generalisation of Hutchinson's dimension formula  $\sum_{i=1}^N r_i^s = 1$  to the affine context. We recall that the *singular values* of a  $d \times d$  real matrix  $A$  are defined to be the square roots of the (necessarily non-negative) eigenvalues of the positive semidefinite matrix  $A^\top A$ . We denote the singular values of  $A$  by  $\sigma_1(A), \dots, \sigma_d(A)$  where it is always

understood that  $\sigma_1(A) \geq \dots \geq \sigma_d(A)$ . Following the notation of [22], given a  $d \times d$  real matrix  $A$ , for each  $s \geq 0$  we define the *singular value function*  $\varphi^s(A)$  applied to  $A$  by

$$\varphi^s(A) := \begin{cases} \sigma_1(A) \cdots \sigma_{\lfloor s \rfloor}(A) \sigma_{\lceil s \rceil}(A)^{s-\lfloor s \rfloor} & \text{if } 0 \leq s \leq d, \\ |\det A|^{s/d} & \text{if } s \geq d. \end{cases}$$

The singular value function satisfies the useful inequality  $\varphi^s(AB) \leq \varphi^s(A)\varphi^s(B)$  for all  $A, B \in \text{GL}_d(\mathbb{R})$ , as is noted in [22]. Given  $(A_1, \dots, A_N) \in \text{GL}_d(\mathbb{R})^N$  we define the *singular value pressure* of  $(A_1, \dots, A_N)$  at  $s$  to be the real number

$$P(A_1, \dots, A_N; s) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i_1, \dots, i_n=1}^N \varphi^s(A_{i_n} \cdots A_{i_1}),$$

the existence of the limit being guaranteed by subadditivity. When  $A_1, \dots, A_N \in \text{GL}_d(\mathbb{R})$  are contracting in the Euclidean norm (or indeed with respect to an arbitrary norm on  $\mathbb{R}^d$ ) it is not difficult to show that the function  $s \mapsto P(A_1, \dots, A_N; s)$  is strictly decreasing and locally Lipschitz continuous and has a unique zero in  $(0, +\infty)$  which we denote by  $\text{dim}_{\text{aff}}(A_1, \dots, A_N)$ . We observe that when every  $A_i$  has the form  $A_i = r_i O_i$  for some  $r_i \in (0, 1)$  and  $O_i \in O(d)$  as in Theorem 1, the pressure simplifies to  $P(A_1, \dots, A_N; s) = \log \sum_{i=1}^N r_i^s$  and thus in this case  $\text{dim}_{\text{aff}}(A_1, \dots, A_N)$  is simply the unique solution  $s$  to Hutchinson’s equation  $\sum_{i=1}^N r_i^s = 1$ . If  $(T_1, \dots, T_N)$  is an affine iterated function system of the form  $T_i x = A_i x + v_i$  then we will also find it useful to write  $\text{dim}_{\text{aff}}(T_1, \dots, T_N) := \text{dim}_{\text{aff}}(A_1, \dots, A_N)$ .

We note that the singular value potential and affinity dimension have a number of antecedents in the literature in the context of the dimension theory of attractors of dynamical systems: a version of the singular value potential was considered by Douady–Oesterlé [20] in the study of the Hausdorff dimensions of attractors, and in the same context the relevance of the asymptotics of singular values (in the form of Lyapunov exponents) was foreseen by Kaplan–Yorke [40] who conjectured that in generic situations the Hausdorff dimension should be related to the growth asymptotics of singular values.

An active area of research in the theory of self-affine sets is the problem of obtaining analogues of Theorem 1 for affine iterated function systems. The first general result in this direction was obtained by K. Falconer in the 1988 article [22]:

**Theorem 2** (Falconer). *Let  $A_1, \dots, A_N \in \text{GL}_d(\mathbb{R})$ . If  $\max_{1 \leq i \leq N} \|A_i\| < 1/2$  then for Lebesgue a.e.  $(v_1, \dots, v_N) \in (\mathbb{R}^d)^N$  the attractor  $Z$  of the iterated function system  $(T_1, \dots, T_N)$  defined by  $T_i x := A_i x + v_i$  satisfies*

$$\text{dim}_H Z = \min \{d, \text{dim}_{\text{aff}}(A_1, \dots, A_N)\}.$$

*If  $\max_{1 \leq i \leq N} \|A_i\| < 1$ , then for every  $(v_1, \dots, v_N) \in (\mathbb{R}^d)^N$  the attractor satisfies*

$$\text{dim}_H Z \leq \min \{d, \text{dim}_{\text{aff}}(A_1, \dots, A_N)\}.$$

Here  $\|\cdot\|$  denotes the operator norm induced by the Euclidean norm. Falconer's original argument assumed  $\max_{1 \leq i \leq N} \|A_i\| < 1/3$ , the improvement to  $1/2$  being due to Solomyak [56], who also noted that the value of  $1/2$  cannot be further improved to any  $1/2 + \varepsilon$ . We remark that the hypothesis  $\max_{1 \leq i \leq N} \|A_i\| < 1/2$  and the conclusion  $\dim_{\text{H}} Z = \min\{d, \dim_{\text{aff}}(A_1, \dots, A_N)\}$  contain a minor asymmetry: it is clear that if each  $A_i$  is replaced with  $X^{-1}A_iX$  for some fixed  $X \in \text{GL}_d(\mathbb{R})$  then the almost sure Hausdorff dimension  $\dim_{\text{H}} Z$  of the attractor does not change, but the condition  $\max_{1 \leq i \leq N} \|A_i\| < 1/2$  will in general be invalidated for certain choices of  $X$ . This asymmetry can be remedied by weakening the hypothesis to the condition  $\max_{1 \leq i \leq N} \|A_i\| < 1/2$  for the operator norm induced by *some* norm  $\|\cdot\|$  on  $\mathbb{R}^d$ , and similarly with the condition  $\max_{1 \leq i \leq N} \|A_i\| < 1$ , and under this hypothesis Falconer's proof goes through with minimal changes. Some similar remarks relating to sufficient conditions for the existence of the attractor of  $(T_1, \dots, T_N)$  were presented in [2, §6]. To avoid similar asymmetries in our results we will assume in this article that our affine iterated function systems are contracting with respect to some unspecified norm on  $\mathbb{R}^d$ .

Theorem 2 demonstrates that the affinity dimension correctly describes the Hausdorff dimension of the attractor in a large range of cases, but this result inherently does not apply to explicit, specific examples of affine iterated function systems. Since the publication of [22] an active line of research, especially in recent years, has therefore been that of extending Theorem 2 to explicit affine iterated function systems for which the vectors  $v_i$  are fixed and some version of the open set condition is satisfied (see for example [23, 34, 46]). In this direction the following powerful result was obtained recently by B. Bárány, M. Hochman and A. Rapaport [5]:

**Theorem 3** (Bárány–Hochman–Rapaport). *Let  $(T_1, \dots, T_N)$  be an affine iterated function system acting on  $\mathbb{R}^2$  and satisfying the strong open set condition, where each  $T_i$  is contracting with respect to the Euclidean norm. Let us write  $T_i x := A_i x + v_i$  for every  $i = 1, \dots, N$  and suppose that each  $A_i$  is invertible. Suppose that the linear maps  $|\det A_i|^{-1/2} A_i$  are not contained in a compact subgroup of  $\text{GL}_2(\mathbb{R})$  and do not preserve a finite union of one-dimensional subspaces of  $\mathbb{R}^2$ . Then the Hausdorff dimension of the attractor of  $(T_1, \dots, T_N)$  is equal to  $\dim_{\text{aff}}(A_1, \dots, A_N)$ .*

In dimension  $d > 2$  the problem of obtaining an analogue of Theorem 3 is substantially more challenging. At the time of writing, no explicit examples of affine iterated function systems in dimension higher than two are yet known where the Hausdorff and affinity dimensions coincide, other than those which fall within the scope of Theorem 1. On the other hand, in the broader setting of limit sets of actions of nonconformal transformations, Dufloux [21] has successfully computed the Hausdorff dimension of limit sets on the boundary  $\partial H_{\mathbb{C}}^n$  of the  $n$ -dimensional complex hyperbolic space associated to well-positioned Schottky subgroups. We also note the work of Pozzetti–Sambarino–Wienhard [47] who, under an asymptotic conformality assumption, have successfully calculated the Hausdorff dimensions of limit sets in projective spaces.

Returning to our setting of affine iterated function systems, while Theorems 2 and 3 extend the part of Theorem 1 which describes the dimension of the attractor, a feature

which has no direct parallel in Theorem 3 in particular is the question of whether or not there exists a measure supported on the attractor of the affine iterated function system  $(T_1, \dots, T_N)$  having dimension equal to the affinity dimension. While we conjecture that this should indeed be the case in the context of Theorem 3 and its presumed higher-dimensional analogues (and indeed it is known that such measures exist generically in the sense of Theorem 2 – see [36]), in this article we will focus on a narrower question: under what circumstances does an affine iterated function system  $(T_1, \dots, T_N)$  acting on  $\mathbb{R}^d$  admit a *self-affine* measure with dimension equal to the affinity dimension?

Theorem 1 indicates that this phenomenon occurs when the affine transformations are all similitudes, or more generally when they are simultaneously conjugated to similitudes by some linear transformation of  $\mathbb{R}^d$ . In this situation it was observed by P. Mattila that while the open set condition is sufficient for the existence of a self-similar measure with dimension equal to the affinity dimension, it is not necessary for it (see [54, introduction]). One may also show that self-affine measures with dimension equal to the affinity dimension can arise in certain circumstances when the linear parts of the affinities admit a common invariant subspace, or when the affinity dimension is precisely equal to  $d$ . The objective of this article is to demonstrate that these are the *only* situations in which this phenomenon occurs.

Henceforth we shall say that a subset  $A$  of  $GL_d(\mathbb{R})$  is *irreducible* if there does not exist any proper nonzero subspace of  $\mathbb{R}^d$  preserved by every  $A \in A$ , and *strongly irreducible* if a finite union of such subspaces is not preserved by every element of  $A$ . When  $A$  is not irreducible it will be called *reducible*. Clearly  $A$  is (strongly) irreducible if and only if the semigroup generated by  $A$  is. We will at times abuse notation by saying that a tuple  $(A_1, \dots, A_N)$  is (strongly) irreducible if and only if the corresponding set is. Our main result is as follows:

**Theorem 4.** *Let  $T_1, \dots, T_N$  be invertible affine transformations of  $\mathbb{R}^d$  having the form  $T_i x := A_i x + v_i$  for some  $v_1, \dots, v_N \in \mathbb{R}^d$ , where  $(A_1, \dots, A_N) \in GL_d(\mathbb{R})^N$  has the following four properties:*

- (i) *There exists a norm  $\|\cdot\|$  on  $\mathbb{R}^d$  such that  $\|A_i\| < 1$  for every  $i = 1, \dots, N$ .*
- (ii) *The affinity dimension  $\dim_{\text{aff}}(A_1, \dots, A_N)$  is strictly between 0 and  $d$ .*
- (iii) *The tuple  $(A_1, \dots, A_N)$  is irreducible.*
- (iv) *There does not exist an inner product on  $\mathbb{R}^d$  with respect to which the linear maps  $A_1, \dots, A_N$  are similitudes.*

*Then every self-affine measure  $m = \sum_{i=1}^N p_i (T_i)_* m$  satisfies  $\dim m < \dim_{\text{aff}}(A_1, \dots, A_N)$ .*

*Furthermore, this property is locally uniform in the following sense. Suppose that  $K \subset GL_d(\mathbb{R})^N$  is a compact set such that every  $(A_1, \dots, A_N) \in K$  satisfies hypotheses (i)–(iv) above. This applies in particular if  $(B_1, \dots, B_N) \in GL_d(\mathbb{R})^N$  satisfies (i)–(iv) and  $K$  is a sufficiently small compact neighbourhood of  $(B_1, \dots, B_N)$ . Then there exists  $\kappa > 0$  depending on  $K$  with the following property: if  $(A_1, \dots, A_N) \in K$ , and  $T_1, \dots, T_N: \mathbb{R}^d \rightarrow \mathbb{R}^d$  are affine transformations of the form  $T_i x = A_i x + v_i$  for*

some vectors  $v_1, \dots, v_N$ , and  $m = \sum_{i=1}^N p_i(T_i)_*m$  is a self-affine measure on  $\mathbb{R}^d$  for some probability vector  $(p_1, \dots, p_N)$ , then  $\dim m \leq \dim_{\text{aff}}(A_1, \dots, A_N) - \kappa$ .

In stating this result we have taken advantage of the fact that every self-affine measure on  $\mathbb{R}^d$  is exact-dimensional, but this result is not required in our proof. The proof of Theorem 4 in fact shows that the upper packing dimension of the measure  $m$ ,

$$\text{ess sup}_m \limsup_{r \rightarrow \infty} \frac{\log m(B_r(x))}{\log r},$$

is bounded by  $\dim_{\text{aff}}(A_1, \dots, A_N) - \kappa$ . This in turn is achieved by showing that the *Lyapunov dimension* of an appropriate measure on the coding space  $\Sigma_N := \{1, \dots, N\}^{\mathbb{N}}$  is bounded by  $\dim_{\text{aff}}(A_1, \dots, A_N) - \kappa$ . The Lyapunov dimension is relatively technical to describe and would be digressive to define in this introduction, so we defer further discussion of this point to §3 below.

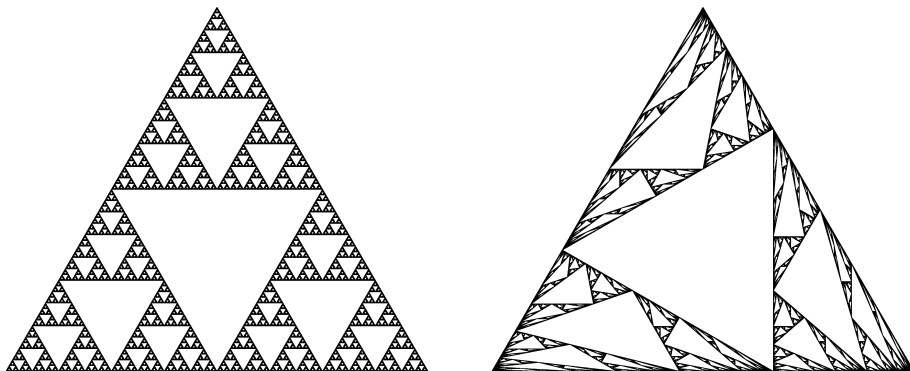
The condition that the linear maps  $A_i$  are not all similitudes with respect to some inner product on  $\mathbb{R}^d$  is equivalent to the statement that the linear maps  $|\det A_i|^{-1/d} A_i$  are not all contained in some compact subgroup of  $\text{GL}_d(\mathbb{R})$ , and we will at times prefer the latter formulation in the proofs. To see that these statements are equivalent we observe that if  $G \leq \text{GL}_d(\mathbb{R})$  is a compact group containing the linear maps  $|\det A_i|^{-1/d} A_i$ ,  $\langle \cdot, \cdot \rangle$  denotes the standard inner product on  $\mathbb{R}^d$ , and  $H$  is the normalised Haar measure on  $G$ , the formula

$$\langle u, v \rangle_G := \int_G \langle Bu, Bv \rangle dH(B)$$

may easily be verified to define an inner product on  $\mathbb{R}^d$  which is invariant under the action of elements of  $G$ . In particular, the transformations  $A_i$  are similitudes with respect to this inner product structure. The converse implication is obvious. Theorem 4 therefore admits the following corollary which motivates the title of this work:

**Corollary 1.1.** *Let  $T_1, \dots, T_N: \mathbb{R}^d \rightarrow \mathbb{R}^d$  be invertible affine transformations which are contracting with respect to some norm on  $\mathbb{R}^d$ . Write  $T_i x = A_i x + v_i$  for all  $x \in \mathbb{R}^d$  and  $i = 1, \dots, N$ , and suppose that  $\{A_1, \dots, A_N\}$  is irreducible. If there exists a self-affine measure  $m = \sum_{i=1}^N p_i(T_i)_*m$  such that  $\dim m = \dim_{\text{aff}}(A_1, \dots, A_N) \in (0, d)$ , then there exists an inner product on  $\mathbb{R}^d$  with respect to which the transformations  $T_i$  are all similitudes.*

We note that the affinity dimension of an invertible affine iterated function system is never zero and therefore the endpoint case  $\dim_{\text{aff}}(A_1, \dots, A_N) = 0$  of Theorem 4 cannot occur. In the other endpoint case  $\dim_{\text{aff}}(A_1, \dots, A_N) = d$  it is easy to construct examples in which the normalised restriction of Lebesgue measure to a convex polyhedral body in  $\mathbb{R}^d$  may be represented as a self-affine measure with respect to affine transformations which are not simultaneously conjugate to similitudes and whose linear parts do not admit an invariant proper subspace. For example, if  $U \subset \mathbb{R}^2$  is an open triangular region then up to Lebesgue measure zero it may be bisected along a line passing through one vertex and its opposite edge into the union of two smaller triangular regions  $U_1$  and  $U_2$ , each having



(a) The classical self-similar Sierpiński gasket  $X_1$ .

(b) A self-affine gasket  $X_2$  which is not self-similar.

**Fig. 1.** By Theorem 1 there exists a self-similar measure supported on the classical Sierpiński gasket  $X_1$  with dimension equal to the Hausdorff dimension of the set itself,  $\log 3/\log 2$ . This measure corresponds to that defined simply by giving measure  $1/3$  to each of the three copies of  $X_1$  with diameter half that of the original, measure  $1/9$  to each of the nine subcopies with diameter  $1/4$  that of the original, and so forth. By the combination of Theorems 3 and 4, for the self-affine gasket  $X_2$  there is a gap between the maximum possible dimension of a self-affine measure supported on  $X_2$  and the Hausdorff dimension of  $X_2$  itself.

two side lengths smaller than those of the original triangle and one side length in common with it. Taking further bisections if necessary,  $U$  may be written up to measure zero as a finite union of strictly smaller triangular regions  $V_1, \dots, V_N$  each of which is the image of  $U$  under some contracting affine transformation  $T_i$ . It is clear that if  $m$  denotes the normalised Lebesgue measure on  $U$  then it satisfies the relation  $m = \sum_{i=1}^N m(V_i)(T_i)_*m$  and hence is a self-affine measure with respect to  $(T_1, \dots, T_N)$  which has dimension 2. In general this construction may be performed in such a way as to ensure that hypotheses (i), (iii) and (iv) of Theorem 4 are satisfied; moreover the linear parts of the affinities may be taken to be *strongly* irreducible. The details of this aspect of the construction and of its generalisation to higher dimensions are left to the reader.

We remark that if in Theorem 4 instead of measures of the form  $m = \sum_{i=1}^N p_i(T_i)_*m$  we were to consider the larger category of Borel probability measures  $m$  which satisfy an equation of the form

$$m = \sum_{i_1, \dots, i_n=1}^N q_{(i_1, \dots, i_n)}(T_{i_1} \cdots T_{i_n})_*m \tag{1}$$

for some  $n \geq 1$  and some probability vector  $(q_{(1, \dots, 1)}, \dots, q_{(N, \dots, N)}) \in \mathbb{R}^{N^n}$ , then no dimension gap would occur. In two dimensions it is known that the supremum of the

Hausdorff dimensions of measures which are self-affine in the broader sense of (1) can be equal to the affinity dimension  $\text{dim}_{\text{aff}}(A_1, \dots, A_N)$  when the conditions of Theorem 4 are satisfied. Indeed, this fact played a significant role in the proof of Theorem 3 by extending the results of [46] which pertain to self-affine measures into a result concerning self-affine sets. Theorem 4 demonstrates that outside the context of similarity transformations this supremum is attained only in degenerate cases in which a common invariant subspace exists.

To conclude this introduction let us briefly outline how Theorem 4 will be proved. If  $T_1, \dots, T_N$  are contractions of  $\mathbb{R}^d$  with respect to some fixed norm then there exists a well-defined coding map  $\Pi: \{1, \dots, N\}^{\mathbb{N}} \rightarrow \mathbb{R}^d$  with the property

$$\Pi[(x_k)_{k=1}^{\infty}] = \lim_{n \rightarrow \infty} T_{x_1} \cdots T_{x_n} v$$

for all  $v \in \mathbb{R}^d$ , and whose image is precisely the attractor of  $(T_1, \dots, T_N)$ . It is a well-known result due to Hutchinson [35, §4] that a Borel probability measure  $m$  on  $\mathbb{R}^d$  satisfies  $m = \sum_{i=1}^N p_i(T_i)_* m$  if and only if it satisfies  $m = \Pi_* \mu$  where  $\mu$  is the Bernoulli measure  $(\sum_{i=1}^N p_i \delta_i)^{\mathbb{N}}$  on  $\{1, \dots, N\}^{\mathbb{N}}$ . This measure  $\mu$  is an ergodic invariant measure with respect to the shift transformation  $\sigma: \{1, \dots, N\}^{\mathbb{N}} \rightarrow \{1, \dots, N\}^{\mathbb{N}}$  defined by  $\sigma[(x_k)_{k=1}^{\infty}] := (x_{k+1})_{k=1}^{\infty}$ .

Now, using a combination of results of A. Käenmäki [37] and T. Jordan, M. Pollicott and K. Simon [36], one may show that if an ergodic shift-invariant measure  $\mu$  on  $\{1, \dots, N\}^{\mathbb{N}}$  has the property  $\text{dim } \Pi_* \mu = \text{dim}_{\text{aff}}(T_1, \dots, T_N)$  then it necessarily maximises the quantity

$$h(\mu) + \lim_{n \rightarrow \infty} \frac{1}{n} \int \log \varphi^s(A_{x_1} \cdots A_{x_n}) d\mu[(x_k)_{k=1}^{\infty}]$$

over all shift-invariant Borel probability measures on the set  $\{1, \dots, N\}^{\mathbb{N}}$ , where  $s := \text{dim}_{\text{aff}}(T_1, \dots, T_N)$ ,  $A_i$  denotes the linear part of the affine transformation  $T_i$  and  $h(\mu)$  denotes the entropy of the measure  $\mu$  with respect to the transformation  $\sigma$ . Measures which maximise this quantity have been named *Käenmäki measures*. The critical step in proving Theorem 4 is to show that under the hypotheses of that theorem there cannot exist a Käenmäki measure which is also a Bernoulli measure. The dimension gap result then follows by relatively straightforward compactness considerations.

The proof of this statement relies on a general theorem on the structure of Käenmäki measures which was established by J. Bochi and the first named author in [13], building on the earlier works [27, 38]. Let us illustrate how this argument functions in a simple special case. Suppose that the semigroup generated by  $A_1, \dots, A_N$  is *Zariski dense* as a subgroup of  $\text{GL}_d(\mathbb{R})$ : that is, suppose that every function  $\phi: \text{GL}_d(\mathbb{R}) \rightarrow \mathbb{R}$  which corresponds to a polynomial function of the matrix entries and vanishes on the semigroup generated by  $A_1, \dots, A_N$  also vanishes identically on  $\text{GL}_d(\mathbb{R})$ . (Equivalently,  $A_1, \dots, A_N$  is not contained in any proper algebraic subgroup of  $\text{GL}_d(\mathbb{R})$ .) Then it follows by a result of A. Käenmäki and the first named author [38] that if  $\mu$  is a Käenmäki



measure for  $(T_1, \dots, T_N)$  then it satisfies

$$C^{-1} \leq \frac{\mu(\{(x_k): x_j = i_j \text{ for all } j = 1, \dots, n\})}{\varphi^s(A_{i_1} \cdots A_{i_n})} \leq C \tag{2}$$

for some constant  $C > 1$ , for all  $i_1, \dots, i_n \in \{1, \dots, N\}$  and  $n \geq 1$ . But if  $\mu$  is also a Bernoulli measure, the value of the numerator depends only on which symbols appear in the sequence  $i_1, \dots, i_n$  and not on the order in which those symbols appear. This implies that the same property must hold for  $\varphi^s(A_{i_1} \cdots A_{i_n})$  up to the introduction of a scalar multiplicative factor  $C^2$ . Using this principle one may deduce that if  $B_1$  and  $B_2$  belong to the semigroup generated by  $A_1, \dots, A_N$  then necessarily

$$C^{-3} \leq \frac{\varphi^s((B_1 B_2)^n)}{\varphi^s(B_1^n) \varphi^s(B_2^n)} \leq C^3 \tag{3}$$

for every  $n \geq 1$ . Now if  $\lambda_i(B)$  denotes the  $i$ th largest of the absolute values of the  $d$  eigenvalues of  $B \in \text{GL}_d(\mathbb{R})$ , and  $0 < s < d$ , one may show that

$$\lim_{n \rightarrow \infty} \varphi^s(B^n)^{1/n} = \lambda_1(B) \cdots \lambda_{\lfloor s \rfloor}(B) \lambda_{\lceil s \rceil}(B)^{s - \lfloor s \rfloor} =: \xi^s(B).$$

Taking the power  $1/n$  and letting  $n \rightarrow \infty$  in (3) it follows that the function  $\xi^s$  just defined satisfies  $\xi^s(B_1 B_2) = \xi^s(B_1) \xi^s(B_2)$  for all  $B_1, B_2$  in the semigroup generated by the linear maps  $A_1, \dots, A_N$ . But this turns out to be impossible for a semigroup which is Zariski dense in  $\text{GL}_d(\mathbb{R})$ , essentially by a theorem of Y. Benoist (later reproven by J.-F. Quint using a different method; see [12, Theorem 7.4 and Proposition 9.8] and additionally [11, 49]).

The extension of this argument to the more general circumstances of Theorem 4 requires us to engage with a number of complications. Similarly to the special case described above, the core of the proof operates by assuming that hypotheses (i)–(iii) of Theorem 4 hold and that a Käenmäki measure exists which is a Bernoulli measure, and proceeds to show that the linear maps  $|\det A_i|^{-1/d} A_i$  necessarily belong to a compact group, contradicting (iv). In general under the hypotheses of Theorem 4 there may be multiple inequivalent Käenmäki measures. (This remains true even under slightly stronger hypotheses: see [45].) The hypotheses imply that at least one of these measures is Bernoulli, but *a priori* other Käenmäki measures may not be. In this case the denominator of (2) will not correspond to the function  $\varphi^s(A_{i_1} \cdots A_{i_n})$  but to some more complicated function derived from the action of  $A_{i_1} \cdots A_{i_n}$  on finite unions of proper subspaces of exterior powers of  $\mathbb{R}^d$  (see [13, §2]). The more complicated structure of this function necessitates further steps in order to deduce the multiplicativity of some analogue of the function  $\xi^s$  defined above, which in general will correspond to some spectral data relating to the action of a finite-index subsemigroup of the semigroup generated by  $A_1, \dots, A_N$  on certain pairs of subspaces of exterior powers of  $\mathbb{R}^d$ . This multiplicativity will allow us to show that certain homomorphic images of a finite-index subsemigroup of the semigroup generated by  $|\det A_1|^{-1/d} A_1, \dots, |\det A_N|^{-1/d} A_N$  are contained in compact groups, and this can be applied to deduce that the elements of that finite-index

subsemigroup act as “simultaneously normal” linear maps on certain subspaces of particular exterior powers of  $\mathbb{R}^d$ ; that is, on those spaces there exists an inner product structure with respect to which the linear maps act as orthogonal direct sums of linear similitudes. An extensive additional argument is then required to show that these normal linear maps actually *are* similitudes. This additional argument makes use of the variational characterisation of Käenmäki measures to bound a weighted sum of the Lyapunov exponents of the other Käenmäki measures and so force the remaining Käenmäki measures to also be Bernoulli measures. It is then straightforward to deduce that the entire semigroup generated by  $|\det A_1|^{-1/d} A_1, \dots, |\det A_N|^{-1/d} A_N$  acts on these subspaces of exterior powers by similitudes. Still further arguments are required to deal with the possibility that these subspaces of the exterior powers may be proper. The first two parts of the argument, in which the linear maps are first shown to act normally and then shown to act by similitudes on certain subspaces of exterior powers, are dealt with in Section 5. The final part, in which the action on proper subspaces of exterior powers is related to the action on  $\mathbb{R}^d$ , forms a separate argument which is presented in Section 6.

The remainder of the article is therefore structured as follows. In the following section we review such background on the thermodynamic formalism of affine iterated function systems as is necessary to state our main technical theorem, Theorem 5, which asserts that under the hypotheses of Theorem 4 a Käenmäki measure cannot be a Bernoulli measure. In Section 3 we derive Theorem 4 from Theorem 5; this is the most technically straightforward part of the proof of Theorem 4. Section 4 then reviews key concepts from the theory of linear algebraic groups which will be used in the proof of Theorem 5. Section 5 proves a key special case of Theorem 5 in which the irreducibility of certain representations is assumed, and Section 6 applies this result to deduce the general case.

During peer review it was brought to our attention that some of the technical arguments underlying Theorem 4 may be expressed in intrinsic terms as a statement concerning potentials defined in terms of reductive linear algebraic groups. This is discussed in more detail in the appendix.

## 2. Subadditive thermodynamic formalism and the main technical theorem

Let  $\Sigma_N$  denote the set  $\{1, \dots, N\}^{\mathbb{N}}$  equipped with the infinite product topology (with respect to which it is compact and metrisable) and let  $\sigma: \Sigma_N \rightarrow \Sigma_N$  denote the shift transformation  $(x_k)_{k=1}^{\infty} \mapsto (x_{k+1})_{k=1}^{\infty}$ , which is a continuous surjection. When  $N$  is understood let  $\mathcal{M}_\sigma$  denote the set of all  $\sigma$ -invariant Borel probability measures on  $\Sigma_N$ . Via the Riesz representation theorem we identify  $\mathcal{M}_\sigma$  with a subset of  $C(\Sigma_N)^*$  equipped with the corresponding weak- $*$  topology, and in this topology it is compact and metrisable; a sequence of measures  $(\mu_n)_{n=1}^{\infty}$  in  $\mathcal{M}_\sigma$  converges to a measure  $\mu \in \mathcal{M}_\sigma$  if and only if  $\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu$  for every  $f \in C(\Sigma_N)$ .

We define  $\Sigma_N^*$  to be the set of all finite sequences  $\mathbf{i} = (i_k)_{k=1}^n \in \{1, \dots, N\}^n$ , which we refer to as *words*. If  $\mathbf{i} = (i_k)_{k=1}^n$  then we write  $|\mathbf{i}| = n$  and define this to be the *length* of the word  $\mathbf{i}$ . Given two words  $\mathbf{i} = (i_k)_{k=1}^n, \mathbf{j} = (j_k)_{k=1}^m \in \Sigma_N^*$  we define their

concatenation  $\mathbf{i}j$  to be the word of length  $|\mathbf{i}| + |j| = n + m$  with first  $n$  symbols  $i_1, \dots, i_n$  and subsequent symbols  $j_1, \dots, j_m$ . We define the concatenation of more than two words (e.g.  $\mathbf{i}jk$  where  $\mathbf{i}, j, k \in \Sigma_N^*$ ) in the obvious fashion, and if  $\mathbf{i} \in \Sigma_N^*$  and  $n \geq 1$  we let  $\mathbf{i}^n$  denote the concatenation  $\mathbf{i}\mathbf{i}\dots\mathbf{i}$  of  $n$  copies of  $\mathbf{i}$ . If  $A_1, \dots, A_N \in \text{GL}_d(\mathbb{R})$  are understood then we write  $A_{\mathbf{i}} := A_{i_1} \dots A_{i_n}$  and observe that  $A_{\mathbf{i}j} = A_{\mathbf{i}j}$  for all  $\mathbf{i}, j \in \Sigma_N^*$ . If  $x = (x_k)_{k=1}^\infty \in \Sigma_N$  then we define  $x|_n$  to be the word  $(x_k)_{k=1}^n \in \Sigma_N^*$ . If  $\mathbf{i} \in \Sigma_N^*$  then we define the *cylinder set*  $[\mathbf{i}]$  to be the set of all  $x \in \Sigma_N$  such that  $x|_n = \mathbf{i}$ . Every cylinder set is clopen and cylinder sets form a basis for the topology of  $\Sigma_N$ . The linear span of the set of all characteristic functions of cylinder sets is dense in  $C(\Sigma_N)$  and therefore a sequence  $(\mu_n)_{n=1}^\infty$  of measures in  $\mathcal{M}_\sigma$  converges to a measure  $\mu \in \mathcal{M}_\sigma$  if and only if  $\lim_{n \rightarrow \infty} \mu_n([\mathbf{i}]) = \mu([\mathbf{i}])$  for every  $\mathbf{i} \in \Sigma_N^*$ .

We will say that  $\mu \in \mathcal{M}_\sigma$  is a *Bernoulli measure* if there exists a probability vector  $(p_1, \dots, p_N)$  such that  $\mu([i_1 \dots i_n]) = p_{i_1} \dots p_{i_n}$  for all  $i_1, \dots, i_n \in \{1, \dots, N\}$  and all  $n \geq 1$ . (We permit cases in which some of the entries of the probability vector are zero.) Clearly Bernoulli measures on  $\Sigma_N$  are in one-to-one correspondence with probability vectors  $(p_1, \dots, p_N)$ . It is not difficult to see that the natural map from the  $(N - 1)$ -simplex of probability vectors to the set of corresponding Bernoulli measures on  $\Sigma_N$  is weak- $*$  continuous, and in particular the set of all Bernoulli measures on  $\Sigma_N$  is weak- $*$  compact. Every Bernoulli measure is ergodic with respect to  $\sigma$ .

Let us say that a *submultiplicative potential*, or simply a *potential*, is a function  $\Phi: \Sigma_N^* \rightarrow (0, +\infty)$  such that  $\Phi(\mathbf{i}j) \leq \Phi(\mathbf{i})\Phi(j)$  for all  $\mathbf{i}, j \in \Sigma_N^*$ . We define the *pressure* of  $\Phi$  to be the limit

$$P(\Phi) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{|\mathbf{i}|=n} \Phi(\mathbf{i})$$

and observe that this limit exists by subadditivity. If  $\Phi$  is a submultiplicative potential then we define a sequence of functions  $\Phi_n: \Sigma_N \rightarrow (0, +\infty)$  by  $\Phi_n(x) := \Phi(x|_n)$  for every  $x \in \Sigma_N$  and  $n \geq 1$ . In this case we observe that each  $\Phi_n$  is continuous (since it depends on only finitely many co-ordinates of  $x \in \Sigma_N$ ) and that the subadditivity property  $\log \Phi_{n+m}(x) \leq \log \Phi_n(\sigma^m x) + \log \Phi_m(x)$  is satisfied by the sequence of continuous functions  $\log \Phi_n: \Sigma_N \rightarrow \mathbb{R}$ . As a consequence of this property, for each ergodic  $\mu \in \mathcal{M}_\sigma$  the following limit exists and defines

$$\Lambda(\Phi, \mu) := \lim_{n \rightarrow \infty} \frac{1}{n} \int \log \Phi_n(x) d\mu(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{|\mathbf{i}|=n} \mu([\mathbf{i}]) \log \Phi(\mathbf{i}) \in [-\infty, +\infty).$$

The next result is a special case of the subadditive variational principle of Cao, Feng and Huang [17, Theorem 1.1]:

**Proposition 2.1.** *Let  $N \geq 2$  and let  $\Phi: \Sigma_N \rightarrow (0, +\infty)$  be a submultiplicative potential. Then*

$$P(\Phi) = \sup_{\mu \in \mathcal{M}_\sigma} [h(\mu) + \Lambda(\Phi, \mu)]. \tag{4}$$

When  $\mu$  attains the supremum (4) we call it an *equilibrium state* for the potential  $\Phi$ . If  $\Phi$  is a submultiplicative potential then by subadditivity

$$\Lambda(\Phi, \mu) = \inf_{n \geq 1} \frac{1}{n} \sum_{|i|=n} \mu([i]) \log \Phi(i)$$

and also

$$h(\mu) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{|i|=n} -\mu([i]) \log \mu([i]) = \inf_{n \geq 1} \frac{1}{n} \sum_{|i|=n} -\mu([i]) \log \mu([i]),$$

and since each function  $\mu \mapsto \mu([i])$  is continuous, these formulas imply that the function  $\mu \mapsto h(\mu) + \Lambda(\Phi, \mu)$  is the pointwise infimum of a family of continuous functions  $\mathcal{M}_\sigma \rightarrow \mathbb{R}$ , and hence is an upper semicontinuous function  $\mathcal{M}_\sigma \rightarrow [-\infty, +\infty)$ . In particular, it attains its maximum by the compactness of  $\mathcal{M}_\sigma$  and so at least one equilibrium state exists for any specified potential  $\Phi$ .

A submultiplicative potential  $\Phi$  will be called *quasi-multiplicative* if there exist a finite set  $F \subset \Sigma_N^*$  and a real number  $\delta > 0$  such that

$$\max_{k \in F} \Phi(ikj) \geq \delta \Phi(i)\Phi(j) \quad \text{for all } i, j \in \Sigma_N^*. \tag{5}$$

The significance of this condition is that it both guarantees the uniqueness of the equilibrium state of  $\Phi$  and provides explicit information about its structure:

**Proposition 2.2.** *Let  $\Phi: \Sigma_N^* \rightarrow \mathbb{R}$  be a submultiplicative and quasi-multiplicative potential. Then there exists a unique equilibrium state  $\mu$  for  $\Phi$ . Furthermore, there exists  $C > 0$  such that*

$$C^{-1} e^{-|i|P(\Phi)} \Phi(i) \leq \mu([i]) \leq C e^{-|i|P(\Phi)} \Phi(i) \quad \text{for all } i \in \Sigma_N^*.$$

We refer to the above inequality between  $\mu([i])$  and  $\Phi(i)$  as the *Gibbs inequality* for the potential  $\Phi$  and measure  $\mu$ . Proposition 2.2 has been proved and re-proved in various forms across a number of works: we mention for example [24, Theorem 5.5], [39, §3].

The fundamental example of a potential from the perspective of this article will be the singular value potential  $\Phi^s(i) := \varphi^s(A_i)$ , where  $A_1, \dots, A_N \in \text{GL}_d(\mathbb{R})$  are understood; this potential was investigated extensively by A. Käenmäki [37] and the properties of its equilibrium states were developed in subsequent articles [13, 27, 38]. Our argument will however require us to work with potentials which have a unique equilibrium state, and the singular value potential does not have this property unless additional constraints are imposed beyond the hypotheses of Theorem 4. In particular, although the irreducibility of  $(A_1, \dots, A_N)$  as hypothesised in Theorem 4 ensures this uniqueness for  $d = 2$ , it is not sufficient for this when  $d > 2$  and  $1 < s < d - 1$  (see for example [38, §9]). This problem cannot be alleviated by assuming strong irreducibility in place of irreducibility [45].

The core technical result of this article is the following:

**Theorem 5.** *Let  $(A_1, \dots, A_N) \in \text{GL}_d(\mathbb{R})^N$  be irreducible and define a potential  $\Phi: \Sigma_N^* \rightarrow (0, +\infty)$  by*

$$\Phi(\mathbf{i}) := \prod_{i=1}^d \sigma_i(A_{\mathbf{i}})^{\alpha_i}$$

where  $\alpha_1 \geq \dots \geq \alpha_d \geq 0$  and  $\alpha_1 > \alpha_d$ . If  $\Phi$  has an equilibrium state which is a Bernoulli measure then the linear maps  $|\det A_1|^{-1/d} A_1, \dots, |\det A_N|^{-1/d} A_N$  are all contained in a compact subgroup of  $\text{GL}_d(\mathbb{R})$ .

We observe that the submultiplicativity of the above potential  $\Phi$  follows from the inequality

$$\prod_{i=1}^k \sigma_i(AB) \leq \prod_{i=1}^k \sigma_i(A) \cdot \prod_{i=1}^k \sigma_i(B), \tag{6}$$

which is valid for all linear maps  $A, B: \mathbb{R}^d \rightarrow \mathbb{R}^d$  and all  $k = 1, \dots, d$ , since we may write

$$\prod_{i=1}^d \sigma_i(A_{\mathbf{i}})^{\alpha_i} = \prod_{k=1}^d \left( \prod_{i=1}^k \sigma_i(A_{\mathbf{i}}) \right)^{\alpha_k - \alpha_{k+1}}$$

where  $\alpha_{d+1} := 0$ . We will find it convenient to approach the inequality (6) via norms on exterior powers of  $\mathbb{R}^d$ , but an elementary proof may be found in for example [33, Theorem 3.3.4].

If  $0 < s < d$  with  $d \geq 2$  then clearly the singular value potential  $\Phi^s$  corresponds to the case  $\alpha_1 = \dots = \alpha_{\lfloor s \rfloor} = 1, \alpha_{\lceil s \rceil} = s - \lfloor s \rfloor, \alpha_{\lceil s \rceil + 1} = \dots = \alpha_d = 0$  of the above theorem. In particular, Theorem 5 implies that if  $(A_1, \dots, A_N) \in \text{GL}_d(\mathbb{R})^N$  is irreducible,  $0 < s < d$  and the singular value potential has an equilibrium state which is Bernoulli, then the linear maps  $|\det A_1|^{-1/d} A_1, \dots, |\det A_N|^{-1/d} A_N$  are all contained in a compact subgroup of  $\text{GL}_d(\mathbb{R})$ . As was indicated in the introduction, in combination with various more-or-less standard results from the literature, Theorem 5 is sufficient to prove Theorem 4. The derivation of Theorem 4 from Theorem 5 is presented in the following section, and Theorem 5 itself is proved in Sections 4 to 6.

### 3. Proof of Theorem 4 conditional on Theorem 5

We begin the process of proving Theorem 4 by collecting various results from the literature concerning the Lyapunov dimension, the affinity dimension, the natural projection from  $\Sigma_N$  to the attractor, and self-affine measures.

#### 3.1. The Lyapunov and affinity dimensions

The following result demonstrates that the affinity dimension has the properties alluded to in the introduction and introduces its counterpart for measures, the Lyapunov dimension:

**Lemma 3.1.** *Let  $A_1, \dots, A_N \in \text{GL}_d(\mathbb{R})$  with  $\max_i \|A_i\| < 1$  for some norm  $\|\cdot\|$  on  $\mathbb{R}^d$ , and for each  $s \geq 0$  define a potential  $\Phi^s$  by  $\Phi^s(\mathbf{i}) := \varphi^s(A_{\mathbf{i}})$ . Then:*

- (i) *The function  $s \mapsto P(\Phi^s) = P(A_1, \dots, A_N; s)$  is a continuous strictly decreasing function  $[0, +\infty) \rightarrow \mathbb{R}$  with a unique zero, and this zero is strictly positive.*
- (ii) *For every  $\mu \in \mathcal{M}_\sigma$  the function  $s \mapsto h(\mu) + \Lambda(\Phi^s, \mu)$  is a continuous strictly decreasing function  $[0, +\infty) \rightarrow \mathbb{R}$  with a unique zero.*

We define the affinity dimension of  $(A_1, \dots, A_N)$  to be the unique zero of the function  $s \mapsto P(\Phi^s)$ , and the Lyapunov dimension of  $\mu \in \mathcal{M}_\sigma$  relative to  $(A_1, \dots, A_N)$ , denoted  $\text{dim}_{\text{Lyap}}(\mu; A_1, \dots, A_N)$ , to be the unique zero of  $s \mapsto h(\mu) + \Lambda(\Phi^s, \mu)$ .

The proof of the above lemma is a straightforward application of the inequalities

$$\varphi^{s_1}(A_{\mathbf{i}}) \leq (C \|A_{\mathbf{i}}\|)^{s_1 - s_2} \varphi^{s_2}(A_{\mathbf{i}}) \leq C^{s_1 - s_2} \left( \max_i \|A_i\| \right)^{(s_1 - s_2)|\mathbf{i}|} \varphi^{s_2}(A_{\mathbf{i}})$$

and

$$\left( \min_i \sigma_d(A_i) \right)^{(s_1 - s_2)|\mathbf{i}|} \varphi^{s_2}(A_{\mathbf{i}}) \leq \sigma_d(A_{\mathbf{i}})^{s_1 - s_2} \varphi^{s_2}(A_{\mathbf{i}}) \leq \varphi^{s_1}(A_{\mathbf{i}}),$$

which are valid for all  $\mathbf{i} \in \Sigma_N^*$  and  $s_1 \geq s_2 \geq 0$ , where the constant  $C > 0$  depends only on  $\|\cdot\|$  and not on  $\mathbf{i}$ ,  $s_1$  or  $s_2$ . The following relationship between Lyapunov dimension and affinity dimension was observed by A. Käenmäki [37]:

**Lemma 3.2.** *Let  $A_1, \dots, A_N \in \text{GL}_d(\mathbb{R})$  with  $\max_i \|A_i\| < 1$  for some norm  $\|\cdot\|$  on  $\mathbb{R}^d$ , and let  $\mu \in \mathcal{M}_\sigma(\Sigma_N)$ . Then  $\text{dim}_{\text{Lyap}}(\mu; A_1, \dots, A_N) \leq \text{dim}_{\text{aff}}(A_1, \dots, A_N)$ , and equality holds if and only if  $\mu$  is an equilibrium state of the potential  $\Phi^s(\mathbf{i}) := \varphi^s(A_{\mathbf{i}})$  where  $s := \text{dim}_{\text{aff}}(A_1, \dots, A_N)$ .*

*Proof.* For each  $s \geq 0$  we have  $h(\mu) + \Lambda(\Phi^s, \mu) \leq P(\Phi^s)$  by the variational principle, Proposition 2.1. In particular, if  $P(\Phi^s) < 0$  for some  $s > 0$  then  $h(\mu) + \Lambda(\Phi^s, \mu) < 0$ . It follows that

$$\{s \geq 0: P(\Phi^s) < 0\} \subseteq \{s \geq 0: h(\mu) + \Lambda(\Phi^s, \mu) < 0\}$$

and since by Lemma 3.1,

$$\text{dim}_{\text{Lyap}}(\mu; A_1, \dots, A_N) = \inf \{s \geq 0: h(\mu) + \Lambda(\Phi^s, \mu) < 0\}$$

and

$$\text{dim}_{\text{aff}}(A_1, \dots, A_N) = \inf \{s \geq 0: P(\Phi^s) < 0\},$$

it follows that  $\text{dim}_{\text{Lyap}}(\mu; A_1, \dots, A_N) \leq \text{dim}_{\text{aff}}(A_1, \dots, A_N)$  as required. If these two quantities are equal to one another with common value  $s_0$ , say, then we must have

$$h(\mu) + \Lambda(\Phi^{s_0}, \mu) = 0 \quad \text{and} \quad P(\Phi^{s_0}) = 0$$

by continuity in view of Lemma 3.1, which implies that  $\mu$  is an equilibrium state for the potential  $\Phi^{s_0}$  as claimed. The converse is trivial. ■

3.2. The natural projection and the dimension of self-affine measures

If  $T_1, \dots, T_N$  are affine transformations of  $\mathbb{R}^d$  which are contractions with respect to some norm  $\|\cdot\|$  on  $\mathbb{R}^d$  then for every  $v \in \mathbb{R}^d$  and  $x = (x_k)_{k=1}^\infty \in \Sigma_N$  the limit

$$\Pi(x) := \lim_{n \rightarrow \infty} T_{x_1} \cdots T_{x_n} v$$

exists and is independent of the choice of  $v \in \mathbb{R}^d$ . Indeed, if  $\varepsilon > 0$  is chosen such that  $\|T_i u - T_i v\| \leq (1 - \varepsilon)\|u - v\|$  for all  $u, v \in \mathbb{R}^d$ , and  $v_0 \in \mathbb{R}^d$  is arbitrary, then for every  $r \geq \varepsilon^{-1} \max_i \|v_0 - T_i v_0\|$  every map  $T_i$  preserves and contracts  $\overline{B_r(v_0)}$ , the closed  $r$ -ball centred on  $v_0$  with respect to the norm  $\|\cdot\|$ . It follows easily that  $\Pi(x) = \bigcap_{n=1}^\infty T_{x_1} \cdots T_{x_n} \overline{B_r(v_0)}$ . We also deduce that the diameter of the set  $\Pi(\{i\})$  is bounded by a constant times  $(1 - \varepsilon)^{|i|}$  and so  $\Pi: \Sigma_N \rightarrow \mathbb{R}^d$  is continuous. It is not difficult to see that  $\Pi(\Sigma_N)$  is contained in the attractor of  $(T_1, \dots, T_N)$  since the initial point  $v$  may be taken to be in the attractor. It is also not difficult to see that  $\Pi(\Sigma_N)$  is precisely the attractor, although this fact will not be used. We call  $\Pi$  the natural projection associated to  $(T_1, \dots, T_N)$ .

The following result relating Bernoulli measures to self-affine measures via the natural projection follows from a more general theorem of J. E. Hutchinson [35, §4]. Although Hutchinson’s proof assumes the probability vector  $(p_1, \dots, p_N)$  is nondegenerate, it is not difficult to check that this stipulation is unnecessary.

**Lemma 3.3.** *Let  $T_1, \dots, T_N: \mathbb{R}^d \rightarrow \mathbb{R}^d$  be affine transformations which are contractions with respect to some norm on  $\mathbb{R}^d$ , and let  $(p_1, \dots, p_N)$  be a probability vector. Then a Borel probability measure  $m$  on  $\mathbb{R}^d$  satisfies  $\sum_{i=1}^N p_i (T_i)_* m = m$  if and only if  $m = \Pi_* \mu$  where  $\mu$  is the Bernoulli measure on  $\Sigma_N$  characterised by  $\mu(\{i\}) = p_{i_1} \cdots p_{i_n}$  for all  $i = (i_k)_{k=1}^n \in \Sigma_N^*$ .*

Finally, the following result connects the Lyapunov dimension with the dimension of a measure:

**Lemma 3.4.** *Let  $T_1, \dots, T_N: \mathbb{R}^d \rightarrow \mathbb{R}^d$  be affine transformations which are contractions with respect to some norm on  $\mathbb{R}^d$  and let  $\mu \in \mathcal{M}_\sigma$ . Write  $T_i x = A_i x + v_i$  for all  $x \in \mathbb{R}^d$  and  $i = 1, \dots, N$ . Then  $\dim \Pi_* \mu \leq \dim_{\text{Lyap}}(\mu; A_1, \dots, A_N)$ .*

*Proof.* It is shown in [52, Theorem 2.2] in the more general context of a countably infinite family of transformations  $(T_i)_{i=1}^\infty$  that

$$\limsup_{r \rightarrow 0} \frac{\log \Pi_* \mu(B_r(\Pi(y)))}{\log r} \leq \dim_{\text{Lyap}}(\mu; A_1, \dots, A_N) \quad \text{for } \mu\text{-a.e. } y \in \Sigma,$$

and this obviously implies

$$\limsup_{r \rightarrow 0} \frac{\log \Pi_* \mu(B_r(x))}{\log r} \leq \dim_{\text{Lyap}}(\mu; A_1, \dots, A_N) \quad \text{for } \Pi_* \mu\text{-a.e. } x \in \mathbb{R}^d,$$

which yields the result. The result may also be derived from [36, proof of Theorem 4.3].

■

3.3. Further continuity properties of the Lyapunov and affinity dimensions

Let  $\text{Cont}(\text{GL}_d(\mathbb{R})^N)$  denote the set of all tuples  $(A_1, \dots, A_N) \in \text{GL}_d(\mathbb{R})$  with  $\max_{1 \leq i \leq N} \|A_i\| < 1$  for some norm  $\|\cdot\|$  on  $\mathbb{R}^d$  depending on  $(A_1, \dots, A_N)$ . This is clearly an open subset of  $\text{GL}_d(\mathbb{R})^N$ . The following two results will be key in proving the local uniformity of the dimension gap in Theorem 4:

**Proposition 3.5.** Define a function  $\gamma: \text{Cont}(\text{GL}_d(\mathbb{R})^N) \rightarrow \mathbb{R}$  by

$$\gamma(B_1, \dots, B_N) := \sup \{ \dim_{\text{Lyap}}(\mu; B_1, \dots, B_N) : \mu \text{ is a Bernoulli measure on } \Sigma_N \}.$$

Then  $\gamma$  is upper semicontinuous, and additionally for every tuple  $(B_1, \dots, B_N) \in \text{Cont}(\text{GL}_d(\mathbb{R})^N)$  the supremum in the definition of  $\gamma$  is attained.

*Proof.* It is sufficient to prove the following statement: given a sequence  $(A_1^{(n)}, \dots, A_N^{(n)}) \in \text{Cont}(\text{GL}_d(\mathbb{R})^N)$  which converges to a limit  $(A_1, \dots, A_N) \in \text{Cont}(\text{GL}_d(\mathbb{R})^N)$ , there exists a Bernoulli measure  $\mu$  on  $\Sigma_N$  such that

$$\dim_{\text{Lyap}}(\mu; A_1, \dots, A_N) \geq \limsup_{n \rightarrow \infty} \gamma(A_1^{(n)}, \dots, A_N^{(n)}). \tag{7}$$

Applying this result to a constant sequence of tuples  $(A_1, \dots, A_N)$  demonstrates that the supremum in the definition of  $\gamma(A_1, \dots, A_N)$  is attained; applying it to a nonconstant sequence directly implies that  $\gamma$  is upper semicontinuous.

Let us prove this claim. For each  $n \geq 1$  let  $\mu_n$  be a Bernoulli measure such that

$$\dim_{\text{Lyap}}(\mu_n; A_1^{(n)}, \dots, A_N^{(n)}) > \gamma(A_1^{(n)}, \dots, A_N^{(n)}) - 1/n.$$

By passing to a subsequence if required, we may assume that the sequences of values  $\gamma(A_1^{(n)}, \dots, A_N^{(n)})$  and  $\dim_{\text{Lyap}}(\mu_n; A_1^{(n)}, \dots, A_N^{(n)})$  are convergent in  $\mathbb{R}$ , and similarly we may assume that  $(\mu_n)$  converges to a limit  $\mu$  in the weak-\* topology. It is straightforward to verify that the set of Bernoulli measures on  $\Sigma_N$  is closed in the weak-\* topology and so the limit  $\mu$  is necessarily Bernoulli. To prove (7) it is sufficient to prove that

$$\dim_{\text{Lyap}}(\mu; A_1, \dots, A_N) \geq \lim_{n \rightarrow \infty} \dim_{\text{Lyap}}(\mu_n; A_1^{(n)}, \dots, A_N^{(n)}). \tag{8}$$

For each  $n \geq 1$  and  $s \geq 0$  define a potential  $\Phi^{s,n}: \Sigma_N^* \rightarrow (0, +\infty)$  by  $\Phi^{s,n}(\mathbf{i}) := \varphi^s(A_{\mathbf{i}}^{(n)})$ , and define also  $\Phi^s(\mathbf{i}) := \varphi^s(A_{\mathbf{i}})$  for all  $\mathbf{i} \in \Sigma_N^*$ . In the case where the limit  $\lim_{n \rightarrow \infty} \dim_{\text{Lyap}}(\mu_n; A_1^{(n)}, \dots, A_N^{(n)})$  is zero the outcome (8) holds trivially, so we assume the limit is strictly positive. In order to prove (8) it suffices to prove the following: for every positive real number  $s < \lim_{n \rightarrow \infty} \dim_{\text{Lyap}}(\mu_n; A_1^{(n)}, \dots, A_N^{(n)})$  we have  $h(\mu) + \Lambda(\Phi^s, \mu) \geq 0$ .

Let us therefore fix  $s < \lim_{n \rightarrow \infty} \dim_{\text{Lyap}}(\mu_n; A_1^{(n)}, \dots, A_N^{(n)})$ . Let  $n_0 \geq 1$  be such that  $\dim_{\text{Lyap}}(\mu_n; A_1^{(n)}, \dots, A_N^{(n)}) > s$  for all  $n \geq n_0$ . For every  $n \geq n_0$  we have  $h(\mu_n) + \Lambda(\Phi^{s,n}, \mu_n) \geq 0$  by the definition of the Lyapunov dimension. For each  $n \geq 1$  we by definition have

$$h(\mu_n) = \inf_{m \geq 1} \frac{1}{m} \sum_{|\mathbf{i}|=m} -\mu_n([\mathbf{i}]) \log \mu_n([\mathbf{i}]) = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{|\mathbf{i}|=m} -\mu_n([\mathbf{i}]) \log \mu_n([\mathbf{i}])$$



and

$$\Lambda(\Phi^{s,n}, \mu_n) = \inf_{m \geq 1} \frac{1}{m} \sum_{|\mathbf{i}|=m} \mu_n([\mathbf{i}]) \Phi^{s,n}(\mathbf{i}) = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{|\mathbf{i}|=m} \mu_n([\mathbf{i}]) \Phi^{s,n}(\mathbf{i}),$$

so for each  $n \geq n_0$  we have

$$\frac{1}{m} \sum_{|\mathbf{i}|=m} -\mu_n([\mathbf{i}]) \log \mu_n([\mathbf{i}]) + \frac{1}{m} \sum_{|\mathbf{i}|=m} \mu_n([\mathbf{i}]) \Phi^{s,n}(\mathbf{i}) \geq h(\mu_n) + \Lambda(\Phi^{s,n}, \mu_n) \geq 0$$

for every  $m \geq 1$ . We have  $\lim_{n \rightarrow \infty} \mu_n([\mathbf{i}]) = \mu([\mathbf{i}])$  for every  $\mathbf{i}$  by weak- $*$  convergence and  $\lim_{n \rightarrow \infty} \Phi^{s,n}(\mathbf{i}) = \Phi^s(\mathbf{i})$  for every  $\mathbf{i}$  by the 1-Lipschitz continuity of the singular value functions  $\sigma_k: \text{GL}_d(\mathbb{R}) \rightarrow \mathbb{R}$ . For fixed  $m \geq 1$  it is thus clear that

$$\begin{aligned} &\frac{1}{m} \sum_{|\mathbf{i}|=m} -\mu([\mathbf{i}]) \log \mu([\mathbf{i}]) + \frac{1}{m} \sum_{|\mathbf{i}|=m} \mu([\mathbf{i}]) \Phi^s(\mathbf{i}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{m} \sum_{|\mathbf{i}|=m} -\mu_n([\mathbf{i}]) \log \mu_n([\mathbf{i}]) + \frac{1}{m} \sum_{|\mathbf{i}|=m} \mu_n([\mathbf{i}]) \Phi^{s,n}(\mathbf{i}) \geq 0 \end{aligned}$$

and we deduce that

$$h(\mu) + \Lambda(\Phi^s, \mu) = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{|\mathbf{i}|=m} -\mu([\mathbf{i}]) \log \mu([\mathbf{i}]) + \frac{1}{m} \sum_{|\mathbf{i}|=m} \mu([\mathbf{i}]) \Phi^s(\mathbf{i}) \geq 0.$$

This demonstrates that  $\dim_{\text{Lyap}}(\mu; A_1, \dots, A_N) \geq s$  and the result follows. ■

We also recall the following theorem of Feng and Shmerkin, which was originally proved in [28] using thermodynamic formalism and the multiplicative ergodic theorem.<sup>1</sup> An alternative proof using linear algebra was given in [44].

**Theorem 6.** *The function  $\dim_{\text{aff}}: \text{Cont}(\text{GL}_d(\mathbb{R})^N) \rightarrow [0, +\infty)$  is continuous.*

We also require the following algebraic lemma. Although it can be deduced from the structure theory of reductive groups, we provide a brief elementary proof.

**Lemma 3.6.** *Let  $A$  be an irreducible subset of  $\text{GL}_d(\mathbb{R})$ . Suppose that for every  $A$  in the semigroup generated by  $A$ , the eigenvalues of  $A$  all have absolute value  $|\det A|^{1/d}$ . Then  $\{|\det A|^{-1/d} A : A \in A\}$  is contained in a compact subgroup of  $\text{GL}_d(\mathbb{R})$ .*

---

<sup>1</sup>The original result of Feng and Shmerkin works on the smaller space of tuples  $(A_1, \dots, A_N) \in \text{GL}_d(\mathbb{R})^N$  such that  $\max_i \|A_i\| < 1$  for the Euclidean norm on  $\mathbb{R}^d$ . If we instead assume that  $(A_1, \dots, A_N) \in \text{Cont}(\text{GL}_d(\mathbb{R})^N)$  satisfies  $\max_i \|A_i\| < 1$  for some norm  $\|\cdot\|$  on  $\mathbb{R}^d$ , then for some integer  $n \geq 1$  and all  $(B_1, \dots, B_N)$  in a small neighbourhood of  $(A_1, \dots, A_N)$ , the  $N^n$ -tuple  $(B_1^n, B_1^{n-1} B_2, \dots, B_N^{n-1} B_{N-1}, B_N^n) \in \text{GL}_d(\mathbb{R})^{N^n}$  is contracting in the Euclidean norm on  $\mathbb{R}^d$  and has affinity dimension equal to  $\dim_{\text{aff}}(B_1, \dots, B_N)$  by elementary consideration of the definition of the pressure function. In particular, Feng and Shmerkin’s result may be applied to these  $N^n$ -tuples in order to deduce the continuity of the affinity dimension with respect to  $(B_1, \dots, B_N)$  in the small neighbourhood.

*Proof.* Consider the semigroup  $\Gamma$  generated by the set  $\{|\det A|^{-1/d} A : A \in \mathbf{A}\}$ , which is clearly irreducible. We claim that  $\Gamma$  is bounded. To see this consider the closed subsemigroup  $\overline{\mathbb{R}\Gamma} := \overline{\{\beta A : A \in \Gamma \text{ and } \beta \in \mathbb{R}\}}$  of the algebra of linear endomorphisms of  $\mathbb{R}^d$ . It is clear that for every  $A \in \overline{\mathbb{R}\Gamma}$  the eigenvalues of  $A$  are also all of absolute value  $|\det A|^{1/d}$ , so in particular every element of  $\overline{\mathbb{R}\Gamma}$  is either invertible or nilpotent. It is easily seen that  $\overline{\mathbb{R}\Gamma}$  admits a nonzero nilpotent element if and only if  $\Gamma$  is unbounded, so to prove the claim we will show that the only nilpotent element of  $\overline{\mathbb{R}\Gamma}$  is zero.

For a contradiction let  $r$  be the minimal rank of a nilpotent nonzero element of  $\overline{\mathbb{R}\Gamma}$  and note that  $0 < r < d$ . Fix a nilpotent element  $B$  with rank  $r$ . Since  $\text{rank}(B^2) < \text{rank } B$  by nilpotency, we have  $\text{rank}(B^2) = 0$  by minimality of  $r$  so that  $B^2 = 0$ . The equation  $B^2 = 0$  implies that the image  $B\mathbb{R}^d$  is a subspace of the kernel of  $B$ . Since  $\Gamma$  is irreducible, the nonzero  $\Gamma$ -invariant subspace  $\text{span}\{ABv : v \in \mathbb{R}^d \text{ and } A \in \Gamma\}$  must equal  $\mathbb{R}^d$ , so in particular there exists  $A \in \Gamma$  such that  $AB\mathbb{R}^d \not\subset \ker B$ . The linear map  $AB \in \overline{\mathbb{R}\Gamma}$  has kernel equal to  $\ker B$  since  $A$  is invertible, it has rank precisely  $r$ , and it is nilpotent since every element of  $\overline{\mathbb{R}\Gamma}$  which is not invertible is nilpotent. But we have  $(AB)^2 \neq 0$  because the image of  $AB$  is not a subset of  $\ker B = \ker AB$ . This implies that  $0 < \text{rank } AB < r$ , which contradicts the minimality of  $r$ . We conclude that  $\overline{\mathbb{R}\Gamma}$  contains no nonzero nilpotents and therefore  $\Gamma$  must be bounded as claimed.

To complete the proof it is sufficient to observe that the closure  $\overline{\Gamma}$  is a group. Clearly this closure is a compact subsemigroup of  $\text{GL}_d(\mathbb{R})$ . To see that it is a group it suffices to show that every  $A \in \overline{\Gamma}$  satisfies  $A^{-1} \in \text{GL}_d(\mathbb{R})$ , which may be achieved as follows. Given  $A \in \overline{\Gamma}$  choose  $(n_k)_{k=1}^\infty$  such that  $\lim_{k \rightarrow \infty} A^{n_k}$  exists and  $n_{k+1} \geq 2 + n_k$  for all  $k \geq 1$ ; it is clear that  $\lim_{k \rightarrow \infty} A^{n_{k+1} - n_k - 1} = A^{-1} \in \overline{\Gamma}$  as required. ■

The final ingredient which we require for the proof of Theorem 4 is the following:

**Proposition 3.7.** *The set of all  $(A_1, \dots, A_N) \in \text{GL}_d(\mathbb{R})^N$  satisfying hypotheses (i)–(iv) of Theorem 4 is open.*

*Proof.* It is obvious that if  $(A_1, \dots, A_N)$  satisfies  $\max_i \|A_i\| < 1$  for some norm  $\|\cdot\|$  on  $\mathbb{R}^d$  then so does every tuple  $(B_1, \dots, B_N)$  sufficiently close to  $(A_1, \dots, A_N)$ . Similarly the set of all  $(A_1, \dots, A_N)$  satisfying (i) such that  $0 < \dim_{\text{aff}}(A_1, \dots, A_N) < d$  is open as a consequence of Theorem 6.

We claim that the set of all irreducible tuples  $(A_1, \dots, A_N) \in \text{GL}_d(\mathbb{R})^N$  is open. To see this we observe that  $(A_1, \dots, A_N)$  is *not* irreducible if and only if there exist unit vectors  $u, v \in \mathbb{R}^d$  such that  $\langle A_i u, v \rangle = 0$  for all  $i \in \Sigma_N^*$ . Indeed, if such vectors exist then  $\text{span}\{A_i u : i \in \Sigma_N^*\}$  is an invariant subspace for  $A_1, \dots, A_N$  which is clearly not the zero subspace, and is a proper subspace since it does not contain  $v$ . On the other hand, if an invariant subspace  $U$  exists for  $A_1, \dots, A_N$  then we may choose arbitrary unit vectors  $u \in U$  and  $v \in U^\perp$  and see that the preceding condition is satisfied. Now observe that if for each  $n$  the tuple  $(A_1^{(n)}, \dots, A_N^{(n)})$  and unit vectors  $u_n$  and  $v_n$  satisfy  $\langle A_i^{(n)} u_n, v_n \rangle = 0$  for all  $i \in \Sigma_N^*$ , and  $(A_1, \dots, A_N) = \lim_{n \rightarrow \infty} (A_1^{(n)}, \dots, A_N^{(n)})$ , then any accumulation point  $(u, v)$  of the sequence  $(u_n, v_n)$  satisfies  $\langle A_i u, v \rangle = 0$  for all  $i \in \Sigma_N^*$ . Thus the set of all tuples  $(A_1, \dots, A_N) \in \text{GL}_d(\mathbb{R})$  which are *not* irreducible is closed.

As was discussed immediately subsequent to the statement of Theorem 4, the tuple  $(A_1, \dots, A_N)$  satisfies (iv) if and only if the linear maps  $|\det A_i|^{-1/d} A_i$  are all contained in some compact subgroup of  $GL_d(\mathbb{R})$ . We claim that if  $(A_1, \dots, A_N) \in GL_d(\mathbb{R})$  is irreducible, then the linear maps  $|\det A_i|^{-1/d} A_i$  are all contained in a compact subgroup of  $GL_d(\mathbb{R})$  if and only if for every  $i \in \Sigma_N^*$ , every eigenvalue of  $A_i$  has absolute value equal to  $|\det A_i|^{1/d}$ . Indeed, if the first statement holds then every product  $A_i$  has the property that the sequence  $(|\det A_i|^{-n/d} A_i^n)_{n \in \mathbb{Z}}$  is bounded. Applying Gelfand’s formula as  $n \rightarrow +\infty$  it follows that  $\rho(|\det A_i|^{-1/d} A_i) = 1$  and applying Gelfand’s formula as  $n \rightarrow -\infty$  we obtain  $\rho(|\det A_i^{-1}|^{-1/d} A_i^{-1}) = 1$ . (Here and throughout this article,  $\rho(B)$  denotes the spectral radius of the linear map  $B$ .) These two identities together imply that every eigenvalue of  $|\det A_i|^{-1/d} A_i$  has modulus 1 and the second statement follows. The converse implication is given by Lemma 3.6. We conclude that for an irreducible tuple  $(A_1, \dots, A_N)$ , (iv) is equivalent to the statement that for every  $i \in \Sigma_N^*$ , every eigenvalue of  $A_i$  has absolute value equal to  $|\det A_i|^{1/d}$ .

To complete the proof of the proposition we observe that a tuple  $(A_1, \dots, A_N) \in GL_d(\mathbb{R})^N$  satisfies both (iii) and (iv) if and only if it belongs to the set of irreducible tuples (which is open) and avoids the set of tuples with the property that for every  $i \in \Sigma_N^*$ , every eigenvalue of  $A_i$  has absolute value equal to  $|\det A_i|^{1/d}$ . The latter set is obviously closed. The result follows. ■

### 3.4. Proof of Theorem 4

It is now a straightforward task to prove the main theorem. Proposition 3.7 shows that if  $(A_1, \dots, A_N)$  satisfies hypotheses (i)–(iv) of Theorem 4, and  $K$  is a sufficiently small compact neighbourhood of  $(A_1, \dots, A_N)$ , then every element of  $K$  satisfies (i)–(iv).

Fix a compact subset  $K$  of  $GL_d(\mathbb{R})^N$  such that every  $(A_1, \dots, A_N) \in K$  satisfies hypotheses (i)–(iv) of Theorem 4. By Lemma 3.2 we have  $\gamma(A_1, \dots, A_N) - \dim_{\text{aff}}(A_1, \dots, A_N) \leq 0$  for all  $(A_1, \dots, A_N) \in K$ , and by the combination of Proposition 3.5 and Theorem 6 the function  $(A_1, \dots, A_N) \mapsto \gamma(A_1, \dots, A_N) - \dim_{\text{aff}}(A_1, \dots, A_N)$  is upper semicontinuous. In particular, its supremum is attained somewhere on  $K$ , and is nonpositive.

Suppose first that this supremum is equal to some negative number  $-\kappa < 0$ . If the tuple  $(A_1, \dots, A_N)$  is in  $K$ , and  $T_1, \dots, T_N: \mathbb{R}^d \rightarrow \mathbb{R}^d$  are affine maps for which there exist  $v_1, \dots, v_N \in \mathbb{R}^d$  such that  $T_i x = A_i x + v_i$  for all  $x \in \mathbb{R}^d$ , and  $m$  is a self-affine measure with respect to  $T_1, \dots, T_N$ , then by Lemma 3.3 we have  $m = \Pi_* \mu$  for some Bernoulli measure  $\mu$  on  $\Sigma_N$ . Using Lemma 3.4 it follows that

$$\begin{aligned} \dim m &= \dim \Pi_* \mu \leq \dim_{\text{Lyp}}(\mu; A_1, \dots, A_N) \\ &\leq \gamma(A_1, \dots, A_N) \leq \dim_{\text{aff}}(A_1, \dots, A_N) - \kappa \end{aligned}$$

and we have established the conclusion of Theorem 4. To prove Theorem 4 it therefore suffices to show that the supremum

$$\sup \{ \gamma(A_1, \dots, A_N) - \dim_{\text{aff}}(A_1, \dots, A_N) : (A_1, \dots, A_N) \in K \}$$

cannot be zero. If this supremum is zero then by the upper semicontinuity of  $\gamma$ , the continuity of  $\dim_{\text{aff}}$  and the compactness of  $K$  it must be the case that  $\gamma(A_1, \dots, A_N) = \dim_{\text{aff}}(A_1, \dots, A_N)$  for some  $(A_1, \dots, A_N) \in K$ . By Proposition 3.5 we have  $\dim_{\text{Lyap}}(\mu; A_1, \dots, A_N) = \gamma(A_1, \dots, A_N) = \dim_{\text{aff}}(A_1, \dots, A_N)$  for some Bernoulli measure  $\mu$  on  $\Sigma_N$ . By Lemma 3.2 this implies that  $\mu$  is an equilibrium state of the potential  $\Phi(\mathbf{i}) := \varphi^s(A_{\mathbf{i}})$  where  $s := \dim_{\text{aff}}(A_1, \dots, A_N) \in (0, d)$ . By Theorem 5 the linear maps  $|\det A_i|^{-1/d} A_i$  are all contained in a compact subgroup of  $\text{GL}_d(\mathbb{R})$ , but as discussed subsequently to the statement of Theorem 4, this contradicts (iv). The proof of Theorem 4 is complete.

## 4. Review of linear algebraic groups

### 4.1. Reductive linear algebraic groups

Here we include a brief overview of some aspects of reductive linear algebraic groups that will be useful in the proofs of the main results. Our principal reason of interest for this class of groups is that they arise as the Zariski closures of semigroups in  $\text{GL}_d(\mathbb{R})$  that act irreducibly on  $\mathbb{R}^d$  (see below). For a more detailed exposition of the theory of reductive linear algebraic groups, we refer the reader to [12, 14, 15, 18, 41].

*4.1.1. Definition and relation to irreducible semigroups.* A linear Lie subgroup  $G$  of  $\text{GL}_d(\mathbb{R})$  is said to be reductive if it has no nontrivial normal subgroup consisting of unipotent matrices. A connected reductive linear real Lie group  $G$  is also a linear algebraic group in the sense that it is the connected component of identity  $\mathbb{G}(\mathbb{R})^o$  of the group of real points  $\mathbb{G}(\mathbb{R})$  of a (reductive) linear algebraic group  $\mathbb{G}$  defined over  $\mathbb{R}$ . The linear algebraic group  $\mathbb{G}$  admits a faithful rational representation  $\mathbb{G} \rightarrow \text{GL}_d$ . In particular, it can be seen as the set of zeros of polynomials in  $\mathbb{R}[x_{ij}, \det x^{-1}]$ , where  $x_{ij}$ 's are the entries in  $\text{Mat}(d, \mathbb{R})$ . Consequently, we can speak of the Zariski topology on  $G$ : a subset of  $G$  is said to be *Zariski closed* if it is the set of common zeros of a set of polynomial maps. This defines the Zariski topology; the notions of Zariski closure and Zariski density are defined in the obvious way. The usual Hausdorff (analytic) topology on  $G$  is finer than the Zariski topology. In the following, we shall speak of a real reductive group to mean a reductive linear real Lie group with finitely many connected components, and unless otherwise specified, topological notions refer to the analytic topology.

We will often work with semigroups in  $\text{GL}_d(\mathbb{R})$ . We recall the elementary fact that the Zariski closure of a semigroup  $\Gamma$  in  $G$  is a (Zariski closed) group, call it  $H$ . In particular, the Zariski closure of the group generated by  $\Gamma$  is also  $H$ .

Before proceeding further, let us clarify the aforementioned relationship between irreducible, or rather completely reducible, families and real reductive groups. Recall that a semigroup  $\Gamma$  in  $\text{GL}_d(\mathbb{R})$  is said to act *completely reducibly* if  $\mathbb{R}^d$  decomposes into a direct sum  $V_1 \oplus \dots \oplus V_k$  of  $\Gamma$ -invariant subspaces  $V_i$ , on which  $\Gamma$  acts irreducibly. It is equivalent to require that every  $\Gamma$ -invariant subspace has a  $\Gamma$ -invariant complement. Clearly, if  $\Gamma$  acts irreducibly on  $\mathbb{R}^d$ , then it acts completely reducibly.

The action on  $\mathbb{R}^d$  of a real reductive group  $G < \mathrm{GL}_d(\mathbb{R})$  is completely reducible (see [18, Ch. 4]). Conversely, let  $\Gamma$  be a semigroup of  $\mathrm{GL}_d(\mathbb{R})$  that acts completely reducibly on  $\mathbb{R}^d$ . Let  $G$  be the Zariski closure of  $\Gamma$ . We claim that  $G$  is a real reductive group. Indeed, being algebraic,  $G$  has finitely many connected components. If it is not real reductive, then it contains a nontrivial normal subgroup  $N$  consisting of unipotent matrices. Let  $V_1$  be a  $G$ -irreducible subspace on which  $N$  acts nontrivially. By a classical result of Kolchin, the subspace  $V_0$  of fixed vectors of  $N$  in  $V_1$  is a nontrivial proper subspace of  $V_1$ . Since  $N$  is normal in  $G$ ,  $V_0$  is invariant under  $G$ , contradicting irreducibility of the  $G$ -action on  $V_1$ .

*4.1.2. Cartan space and roots.* Let  $A < G$  be a maximal connected real split torus so that it is a closed Lie subgroup of  $G$  that is isomorphic to  $(\mathbb{R}_+^*)^d$  for some  $d \in \mathbb{N}$ . The integer  $d$  is called the (real) rank of  $G$ . Let  $Z(G)$  denote the centre of  $G$ . The integer  $d_S := d - \dim Z(G)$  is called the semisimple rank of  $G$ . The Lie algebra  $\mathfrak{a}$  of  $A$  can be written as  $\mathfrak{a} = \mathfrak{a}_Z \oplus \mathfrak{a}_S$ , where  $\mathfrak{a}_Z$  is the Lie algebra of  $A \cap Z(G)$  and  $\mathfrak{a}_S$  is the Lie algebra of  $A \cap [G, G]$ . Here  $[G, G]$  denotes the closed commutator subgroup of  $G$ , which is a semisimple Lie group.

Let  $\mathfrak{g}$  be the Lie algebra of  $G$  and let  $\mathrm{Ad} : G \rightarrow \mathrm{GL}(\mathfrak{g})$  be the adjoint representation of  $G$ . A nontrivial character  $\alpha : A \rightarrow \mathbb{R}_+^*$  is said to be a root of  $G$  if it is a weight of  $A$  for the  $\mathrm{Ad}$ -representation, i.e. the subspace  $\mathfrak{g}_\alpha := \{v \in \mathfrak{g} : \mathrm{Ad}(a)v = \alpha(a)v \ \forall a \in A\}$  is nontrivial. Given a character  $\alpha$  of  $A$ , we denote by  $\bar{\alpha}$  the element of  $\mathfrak{a}^*$  satisfying  $\exp(\bar{\alpha}(x)) = \alpha(\exp(x))$  for every  $x \in \mathfrak{a}$ . The set of nonzero  $\bar{\alpha}$ 's appearing in this form from the  $\mathrm{Ad}$ -representation forms a root system that we denote by  $\Sigma$ . Let  $\{\bar{\alpha}_1, \dots, \bar{\alpha}_{d_S}\}$  be a choice of simple roots so that  $\Sigma$  splits into a disjoint union of positive roots  $\Sigma_+$  (those elements of  $\Sigma$  that can be written as a nonnegative integer linear combination of  $\bar{\alpha}_i$ 's) and negative roots  $-\Sigma_+$ .

We denote by  $\alpha^+$  the choice of a Weyl chamber in  $\mathfrak{a}$  corresponding to a choice of simple roots:  $x \in \mathfrak{a}$  belongs to  $\alpha^+$  if and only if for every  $\bar{\alpha} \in \Sigma_+$ ,  $\bar{\alpha}(x) \geq 0$ . It is a closed fundamental domain for the action of the Weyl group  $N_G(A)/Z_G(A)$ , where  $N_G(A)$  is the normaliser of  $A$  in  $G$  and  $Z_G(A)$  is the centraliser of  $A$  in  $G$ . The Weyl chamber  $\alpha^+$  is the direct sum of a salient cone  $\alpha^+ \cap \mathfrak{a}_S$  and the subspace  $\mathfrak{a}_Z$ .

An example of a real reductive group is  $G = \mathrm{GL}_d(\mathbb{R})$  itself. In this case, the maximal real split torus  $A$  can be taken to be the diagonal matrices with positive coefficients. Its Lie algebra  $\mathfrak{a}$  is the commutative Lie algebra of  $d \times d$  diagonal matrices. The rank of  $G$  is equal to  $d$ . The commutator  $[G, G]$  is  $\mathrm{SL}(d, \mathbb{R})$  so that  $\mathfrak{a}_S$  is the diagonal matrices whose coefficient sum to 0. In particular, the semisimple rank of  $G$  is  $d - 1$ . The (log) roots are the linear forms  $\bar{\alpha}_{i,j}$  with  $i \neq j \in \{1, \dots, d\}$  such that  $\alpha_{i,j}(a) = a_i/a_j$  where  $a_i$ 's are the diagonal entries of  $a$ . A base of simple roots is given by  $\bar{\alpha}_{i,i+1}$ . The corresponding choice of Weyl chamber  $\alpha^+$  is the diagonal matrices with decreasing coefficients. The Weyl group is isomorphic to the symmetric group  $S_d$  acting on  $A$  by permuting the diagonal coefficients.

4.1.3. *Cartan and Jordan projections.* Let  $G$  be a real reductive group and let  $K$  be a maximal compact subgroup of  $G$  whose Lie algebra is orthogonal to  $\mathfrak{a}$  for the Killing form. The Cartan decomposition of  $G$  says that  $G = KAK$ . Here, given an element  $g \in G$ , its factor in the Cartan decomposition corresponding to the group  $A$  is, up to the action of the Weyl group, uniquely determined. In particular, for each  $g \in G$  there exists a unique element  $a_g \in A^+ := \exp(\mathfrak{a}^+)$  such that  $g \in Ka_gK$ . Accordingly we define the Cartan projection

$$\kappa: G \rightarrow \mathfrak{a}^+$$

by setting  $\kappa(g) := a_g$ .

Every element  $g \in G$  can also be decomposed as a commuting product  $g = g_e g_h g_u$ , where  $g_e$  is an elliptic element (i.e. belonging to a compact group),  $g_u$  is a unipotent element (i.e.  $\text{Ad}(g_u)$  is a unipotent linear transformation) and  $g_h$  is a hyperbolic element (i.e. it is conjugate to an element of  $A$ ). The hyperbolic part  $g_h$  is uniquely determined and this allows us to define the *Jordan projection*

$$\lambda: G \rightarrow \mathfrak{a}^+$$

setting  $\lambda(g)$  to be the logarithm of the unique element of  $A^+$  conjugate to  $g_h$ .

When  $G = \text{GL}_d(\mathbb{R})$ , with the above choice of  $A$ , the maximal compact group  $K$  can be taken to be the orthogonal group  $O(d, \mathbb{R})$  and the Cartan decomposition is the polar decomposition: for  $g \in \text{GL}_d(\mathbb{R})$  its Cartan projection reads  $\kappa(g) = (\log \sigma_1(g), \dots, \log \sigma_d(g))$ . The factorisation  $g = g_e g_h g_u$  corresponds to Jordan block form and the Jordan projection  $\lambda(g)$  reads

$$\lambda(g) = (\log |\lambda_1(g)|, \dots, \log |\lambda_d(g)|).$$

4.1.4. *Representations and highest weights.* Let  $G$  be a connected real reductive group and let  $A < G$  and  $\Sigma$  be as above. Let  $U$  be a maximal unipotent subgroup of  $G$  normalised by  $A$  and whose Lie algebra is generated by the root spaces  $(\mathfrak{g}_\alpha)_{\alpha \in \Sigma_+}$ . Let  $V$  be a finite-dimensional real vector space and  $(\pi, V)$  an algebraic representation of  $G$ . An (algebraic) character  $\chi$  of  $A$  is said to be a *restricted weight* of  $G$  in  $(\pi, V)$  if the vector space  $V^\chi := \{v \in V: \pi(a)v = \chi(a)v \forall a \in A\}$  is nontrivial. Such a weight  $\chi$  is said to be a *parabolic weight* if it is a weight of  $A$  in the space  $V^U := \{v \in V: Uv = v\}$ . It is said to be a *dominant weight* if it belongs to the Weyl chamber  $\mathfrak{a}^+$  after the identification of  $\mathfrak{a}$  with  $\mathfrak{a}^*$  with an inner product on  $\mathfrak{a}$  extending the restriction of the Killing form on  $\mathfrak{a}_S$  and for which  $\alpha_S$  and  $\alpha_Z$  are orthogonal.

The choice of positive roots induces a partial order on the set of characters of  $A$ : we let  $\chi_1 \leq \chi_2$  if and only if  $\chi_2 - \chi_1$  is a nonnegative linear combination of positive, or equivalently simple, roots. An irreducible algebraic representation  $(\pi, V)$  of  $G$  admits a unique parabolic weight that we shall denote  $\chi_V$ . This is also the largest weight for the order induced by the choice of  $\mathfrak{a}^+$  and this dominant weight is called *the highest weight*.

We will use the following fact that serves as a bridge between the geometry of  $G$  and its representations. For its proof, see e.g. [12, Lemma 8.17].

**Lemma 4.1.** *Let  $G$  be a connected real reductive group,  $(\pi, V)$  be an irreducible linear representation of  $G$  and  $\chi$  be the highest weight. Then, for every  $g \in G$ , we have*

$$\log |\lambda_1(\pi(g))| = \bar{\chi}(\lambda(g)).$$

*4.1.5. A density result of Benoist.* In his study of asymptotic properties of linear groups and their actions on homogeneous spaces, Benoist [11] (see also [9]) introduced a notion of limit cone of a semigroup: given a semigroup  $\Gamma$  in a real reductive group  $G$ , this is the smallest closed cone in  $\mathfrak{a}^+$  containing all Jordan projections  $\lambda(\gamma)$  of elements  $\gamma \in \Gamma$ . He proved in particular that the intersection of an affine translate of this cone with  $\mathfrak{a}_S$  has nonempty interior in  $\mathfrak{a}_S$  whenever  $\Gamma$  is Zariski dense in  $G$ . The following density result of Benoist [11], later proven in a more elementary fashion by Quint [49], is a refinement of the aforementioned property of this limit cone. In the proof of our main result, it will be instrumental in deducing the compactness of the image of  $[G, G]$  under certain linear representations.

We state a version of this result that is adapted to our purposes (see [12, Proposition 9.8]):

**Theorem 7** ([11, 12, 49]). *Let  $G$  be a connected real reductive group and  $\Gamma < G$  a Zariski dense semigroup. The closed subgroup of  $\mathfrak{a}$  spanned by the elements  $\lambda(\gamma_1\gamma_2) - \lambda(\gamma_1) - \lambda(\gamma_2)$  for  $\gamma_1, \gamma_2 \in \Gamma$  is  $\mathfrak{a}_S$ .*

We remark that in the work of Quint [49], Benoist’s nonarithmeticity result was also applied in a symbolic-dynamical context.

## 5. The case of irreducible representations

### 5.1. Overview

We may now commence working in earnest on the proof of Theorem 5. We will study the potential  $\Phi(\mathfrak{i}) := \prod_{i=1}^d \sigma_i(A_i)^{\alpha_i}$  by rewriting it in the form  $\Phi(\mathfrak{i}) = \prod_{j=1}^d \|A_i^{\wedge j}\|^{\alpha_j - \alpha_{j+1}}$ , where  $\alpha_{d+1} := 0$ . Since by hypothesis the semigroup  $\Gamma := \{A_i : \mathfrak{i} \in \Sigma_N^*\}$  acts irreducibly on  $\mathbb{R}^d$ , it follows from the discussion at the beginning of §4 that the Zariski closure of  $\Gamma$  in  $\text{GL}_d(\mathbb{R})$  is a real reductive group  $G$ . We are thus in the following situation: we have a finite set of elements  $g_1, \dots, g_N$  of a real reductive group  $G$  which generate a Zariski dense subsemigroup of  $G$ , a finite collection of representations  $\pi_j$  from  $G$  to  $\text{GL}(\wedge^j \mathbb{R}^d)$ , a collection of nonnegative real numbers  $\beta_j$ , and a potential  $\Phi$  of the form  $\Phi(\mathfrak{i}) = \prod_j \|\pi_j(g_{\mathfrak{i}})\|^{\beta_j}$ , where  $g_{\mathfrak{i}} := g_{i_1} \cdots g_{i_n}$  for  $\mathfrak{i} = (i_t)_{t=1}^n$ . (Since those indices  $j$  for which  $\beta_j = 0$  have no effect on the value of  $\Phi(\mathfrak{i})$ , we discard them. The condition  $\alpha_1 > \alpha_d$  implies that at least one  $j < d$  is retained.) We wish to show that if  $\Phi$  has an equilibrium state which is a Bernoulli measure, then  $G$  must be a group of similitudes. Equivalently, we wish to show that the group  $\{|\det g|^{-1/d} g : g \in G\}$  must be compact.

In the full generality of Theorem 5 we have no reason to believe that the representations  $\pi_j$  are irreducible, which significantly complicates the argument. These rep-

representations are however completely reducible as a consequence of the reductiveness of the group  $G$ . We will therefore first prove a version of Theorem 5 in the case of irreducible representations  $\pi_j$ , and then obtain the theorem in the general case by presenting the problem as a family of sub-cases each of which corresponds to a choice of a family of irreducible subspaces, one from each exterior power. The latter task is deferred to the following section. The objective of the present section will therefore be to prove the following:

**Theorem 8.** *Let  $G$  be a real reductive group. Given a positive integer  $k$  and for each  $j = 1, \dots, k$  a real inner product space  $V_j$  of dimension  $d_j \geq 1$ , let  $\pi_j: G \rightarrow \mathrm{GL}(V_j)$  be an irreducible linear representation. Let  $g_1, \dots, g_N \in G$  and write  $g_i := g_{i_1} \cdots g_{i_n}$  for all  $i = (i_t)_{t=1}^n \in \Sigma_N^*$ . Given constants  $\beta_j > 0$ , define a potential  $\Phi: \Sigma_N^* \rightarrow (0, +\infty)$  by*

$$\Phi(\mathbf{i}) := \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})\|^{\beta_j}.$$

*Suppose that the semigroup generated by  $g_1, \dots, g_N$  is Zariski dense in  $G$ . Then the following are equivalent:*

- (i) *There exists an equilibrium state of  $\Phi$  which is a Bernoulli measure.*
- (ii) *The potential  $\Phi^{\det}: \Sigma_N^* \rightarrow (0, +\infty)$  defined by*

$$\Phi^{\det}(\mathbf{i}) := \prod_{j=1}^k |\det \pi_j(g_{\mathbf{i}})|^{\beta_j/d_j}$$

*satisfies  $P(\Phi) = P(\Phi^{\det})$ .*

- (iii) *For every  $j = 1, \dots, k$  the group*

$$\{|\det \pi_j(g)|^{-1/d_j} \pi_j(g) : g \in G\}$$

*is a compact subgroup of  $\mathrm{GL}(V_j)$ .*

The proof of the implications (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) is straightforward and almost all of the length of the proof of Theorem 8 arises from the implication (i) $\Rightarrow$ (iii). As was described briefly in §2, this proof itself consists of two somewhat separate parts, which we describe below.

*5.1.1. Comments on the proof.* The representations  $\pi_j$  are irreducible but will not in general be strongly irreducible, so in general there exists for each  $j$  a finite collection  $U_j^1, \dots, U_j^{n_j}$  of subspaces of  $V_j$  which is permuted by the action of  $G$  under the representation  $\pi_j$ . (If  $\pi_j$  is strongly irreducible then we have  $n_j = 1$  and  $U_j^1 = V_j$ .) We choose these subspaces to be of the least possible dimension and it is not difficult to deduce that they must have pairwise trivial intersection. Each  $U_j^i$  is preserved by every element of the



(Zariski) identity component<sup>2</sup>  $G^o$ , and in the first part of the proof we consider the action of  $G^o$  on each  $U_j^i$  via the restriction of  $\pi_j$  to a representation  $G^o \rightarrow \text{GL}(U_j^i)$ . By minimality of the dimension of  $U_j^i$  this action is irreducible. Using the fact that there exists a  $\Phi$ -equilibrium state which is a Bernoulli measure, a mechanism introduced in [13] for writing  $\Phi$  as the pointwise maximum of a finite collection of quasi-multiplicative potentials  $\Phi^{\mathcal{W}}$ , Proposition 2.2, and Theorem 7, we establish using the ideas outlined in the introduction that for each  $j$  and  $i$  the group  $\pi_j(G^o)|_{U_j^i}$  is a group of linear similarity transformations of  $U_j^i$  with respect to some inner product on  $U_j^i$ .

At this point we will have established that for each  $j$ , the elements of  $\pi_j(G^o)$  can be simultaneously block diagonalised (using a splitting of the form  $V_j = U_j^{i_1} \oplus \cdots \oplus U_j^{i_r}$ ) with each diagonal block equal to an orthogonal matrix times a positive real scalar. (This construction can be interpreted by saying that the elements of  $\pi_j(G^o)$  are all normal matrices with respect to some consistent inner product structure on  $V_j$ .) In order to verify that  $\pi_j(G^o)$  has the required property (iii) it remains to verify that for each fixed  $g$  these scalars are the same for every block. In this part of the proof we must use not only the existence of a potential  $\Phi^{\mathcal{W}_0}$  whose equilibrium state is a Bernoulli measure, but the fact that the pressure  $P(\Phi^{\mathcal{W}_0})$  is equal to the pressure  $P(\Phi)$  of the original potential  $\Phi$ , or equivalently, the fact that  $P(\Phi^{\mathcal{W}_0})$  is maximal among all of the pressures  $P(\Phi^{\mathcal{W}})$ . The underlying intuitive idea is that the products  $\pi_j(g_i)$  necessarily have nonseparated Lyapunov exponents with respect to the Bernoulli measure; this will be shown to imply that these products also have nonseparated Lyapunov exponents with respect to the equilibrium measures of the other potentials  $\Phi^{\mathcal{W}}$ , since if this were not the case those equilibrium states would have a larger top Lyapunov exponent than is allowed by the variational principle. In practice this argument is implemented by comparing the values of various pressure functions associated to the different potentials  $\Phi^{\mathcal{W}}$  (which are defined in terms of the growth rate of the norm of each representation and allow for separated Lyapunov exponents) and the potential  $\Phi^{\det}$ , which is defined in terms of the growth rates of determinants of representations (which does not perceive any difference between Lyapunov exponents). Once it has been shown that for each  $g \in G^o$  the scalars associated to each diagonal block in the block diagonalisation of  $\pi_j(g)$  are the same, it follows that  $\pi_j(G^o)$  is contained in a group of linear similarity transformations of  $\text{GL}(V_j)$ . The same result follows immediately for  $\pi_j(G)$  since the remaining components of  $\pi_j(G)$  form a finite collection of continuous images of  $\pi_j(G^o)$ .

The respective functions of the two parts of the proof may be illustrated by considering two opposite extreme cases of the argument as follows. If it is known *a priori* that each representation  $\pi_j$  is strongly irreducible – for example, if the group  $G$  is known to be connected – then we have  $U_j^1 = V_j$  for each  $j$  and the first part of the proof establishes directly that each  $\pi_j(G)$  is a group of linear similitudes as required. The proof

---

<sup>2</sup>For the purposes of this description of the proof it makes no difference whether  $G^o$  is taken to be the connected component with respect to the analytic topology or with respect to the Zariski topology. However, for technical reasons which will apply later, we define  $G^o$  to be the group of real points of the Zariski connected component of  $G$ .

is then complete without meaningful reference to the second part. If on the other hand it is known *a priori* that for each  $j$ , there is a basis for  $V_j$  with respect to which every  $\pi_j(g_i)$  is represented by a generalised permutation matrix (that is, a matrix with exactly one nonzero entry in each row and in each column) then the subspaces  $U_j^i$  are all one-dimensional, the action of  $G^o$  on each subspace is trivially by a similitude since no other linear transformations of a one-dimensional space exist, and the first part of the proof is entirely redundant. In this case only the second part of the proof is required.

*5.1.2. Remarks on a generalisation of Theorem 8.* Before starting the proof of Theorem 8, we lastly remark that this theorem (and Theorem 5) can be easily extended to the case of a linear Lie group  $G$  which is not necessarily reductive. Indeed, using a reductivisation argument such as [38, Proposition 6.8] one may show that the equilibrium states of an affine iterated function system are determined only by the projections of the linear parts of the affinities to a reductive Levi component (or in explicit co-ordinates, by the block diagonal parts of those linear maps when presented in block upper triangular form). This extended result does not lead to a significantly more powerful version of Theorem 4 since in general it can easily occur that the equilibrium states are determined only by a proper subset of the diagonal blocks: the existence of a Bernoulli equilibrium state in the absence of irreducibility (but in the presence of complete reducibility) consequently can be used only to deduce that *some* of the diagonal blocks of the affine transformations must consist of similitudes. Since this extended result requires few additional steps but lacks the clear interest of Theorem 4 we leave it to the reader.

*5.2. Proof of the implications (iii)⇒(ii)⇒(i)*

The implication (iii)⇒(ii) is simple: if for each  $j = 1, \dots, k$  the group

$$\{|\det \pi_j(g)|^{-1/d_j} \pi_j(g) : g \in G\}$$

is contained in a compact subset of  $GL(V_j)$ , then we may find  $K > 0$  such that

$$K^{-1}|\det \pi_j(g)|^{1/d_j} \leq \|\pi_j(g)\| \leq K|\det \pi_j(g)|^{1/d_j}$$

for all  $j = 1, \dots, k$  and all  $g \in G$ . It follows that for all  $\mathbf{i} \in \Sigma_N$  we have

$$K^{-\sum_{j=1}^k \beta_j} \Phi^{\det}(\mathbf{i}) \leq \Phi(\mathbf{i}) \leq K^{\sum_{j=1}^k \beta_j} \Phi^{\det}(\mathbf{i})$$

and we deduce that  $P(\Phi) = P(\Phi^{\det})$  by direct reference to the definition of the pressure. This proves (iii)⇒(ii).

Let us now prove (ii)⇒(i). Assuming (ii), let  $\mu$  be the Bernoulli measure on  $\Sigma_N^*$  with probability vector  $(p_1, \dots, p_N)$  given by

$$p_{i_0} := \frac{\prod_{j=1}^k |\det \pi_j(g_{i_0})|^{\beta_j/d_j}}{\sum_{i=1}^N \prod_{j=1}^k |\det \pi_j(g_i)|^{\beta_j/d_j}}$$

for every  $i_0 = 1, \dots, N$ . Since

$$\begin{aligned} P(\Phi^{\det}) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{|\mathbf{i}|=n} \Phi^{\det}(\mathbf{i}) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{|\mathbf{i}|=n} \prod_{j=1}^k |\det \pi_j(g_{i_j})|^{\beta_j/d_j} \\ &= \log \sum_{i=1}^N \prod_{j=1}^k |\det \pi_j(g_i)|^{\beta_j/d_j} \end{aligned}$$

using the multiplicativity of the determinant, we observe that

$$\mu([\mathbf{i}]) = \frac{\prod_{j=1}^k |\det \pi_j(g_{i_j})|^{\beta_j/d_j}}{(\sum_{i=1}^N \prod_{j=1}^k |\det \pi_j(g_i)|^{\beta_j/d_j})^{|\mathbf{i}|}} = \frac{\Phi^{\det}(\mathbf{i})}{e^{|\mathbf{i}|P(\Phi^{\det})}}$$

for every  $\mathbf{i} \in \Sigma_N^*$ . Now, for each  $n \geq 1$  we have

$$\begin{aligned} \sum_{|\mathbf{i}|=n} -\mu([\mathbf{i}]) \log \mu([\mathbf{i}]) + \sum_{|\mathbf{i}|=n} \mu([\mathbf{i}]) \log \Phi^{\det}(\mathbf{i}) \\ &= \sum_{|\mathbf{i}|=n} \mu([\mathbf{i}]) (nP(\Phi^{\det}) - \log \Phi^{\det}(\mathbf{i}) + \log \Phi^{\det}(\mathbf{i})) \\ &= nP(\Phi^{\det}) \sum_{|\mathbf{i}|=n} \mu([\mathbf{i}]) = nP(\Phi^{\det}) \end{aligned}$$

and since

$$h(\mu) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{|\mathbf{i}|=n} -\mu([\mathbf{i}]) \log \mu([\mathbf{i}])$$

and

$$\Lambda(\Phi^{\det}, \mu) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{|\mathbf{i}|=n} \mu([\mathbf{i}]) \log \Phi^{\det}(\mathbf{i})$$

we conclude that

$$h(\mu) + \Lambda(\Phi^{\det}, \mu) = P(\Phi^{\det}).$$

Now, clearly

$$\Phi^{\det}(\mathbf{i}) = \prod_{j=1}^k |\det \pi_j(g_{i_j})|^{\beta_j/d_j} \leq \prod_{j=1}^k \|\pi_j(g_{i_j})\|^{\beta_j} = \Phi(\mathbf{i})$$

for every  $\mathbf{i} \in \Sigma_N^*$  using the elementary bound  $|\det B| \leq \|B\|^{\dim V_j}$  valid for all  $B \in GL(V_j)$ . It follows directly that  $\Lambda(\Phi^{\det}, \mu) \leq \Lambda(\Phi, \mu)$ . We deduce that

$$P(\Phi) = P(\Phi^{\det}) = h(\mu) + \Lambda(\Phi^{\det}, \mu) \leq h(\mu) + \Lambda(\Phi, \mu) \leq P(\Phi)$$

where we have used the hypothesis (ii) and, in the final inequality, the subadditive variational principle. It follows that  $h(\mu) + \Lambda(\Phi, \mu) = P(\Phi)$  and thus the Bernoulli measure  $\mu$  is an equilibrium state for the potential  $\Phi$ . This completes the proof of (ii)  $\Rightarrow$  (i).

5.3. Proof of (i)⇒(iii)

5.3.1. *The family of subspaces with finite orbit.* For each  $j = 1, \dots, k$  let  $\ell_j \geq 1$  be the smallest possible dimension of a nonzero subspace of  $V_j$  which is invariant under  $\pi_j(g)$  for all  $g \in G^o$ , and choose  $U_j \subseteq V_j$  to be such an  $\ell_j$ -dimensional subspace. It is not difficult to see that the function  $g \mapsto \pi_j(g)U_j$  is constant on each connected component of  $G$ : if  $g_1, g_2$  belong to the same component  $G_i$  then  $g_1^{-1}g_2$  is a connected component which contains the identity, hence is  $G^o$ , hence  $g_1^{-1}g_2 \in G^o$ , so  $\pi_j(g_1^{-1}g_2)U_j = U_j$  and therefore  $\pi_j(g_1)U_j = \pi_j(g_2)U_j$ . For fixed  $j = 1, \dots, k$ , let  $U_j^1, \dots, U_j^{n_j}$  denote the complete list of subspaces of  $V_j$  having the form  $\pi_j(g)U_j$  for some  $g \in G$ .

Fix  $j \in \{1, \dots, k\}$ . We observe that  $\text{span} \bigcup_{i=1}^{n_j} U_j^i$  is a nonzero subspace of  $V_j$  which is preserved by  $\pi_j(g)$  for every  $g \in G$ , since each  $\pi_j(g)$  acts on the spaces  $U_j^i$  by permutation. By irreducibility it follows that this subspace must equal the whole of  $V_j$ .

We now make the following claim: if  $i_1, \dots, i_{t+1}$  are distinct integers in the range 1 to  $n_j$ , where  $t \geq 1$ , then  $U_j^{i_{t+1}}$  either is a subspace of the vector space  $\text{span} \bigcup_{s=1}^t U_j^{i_s}$  or has trivial intersection with it. Indeed, if neither of these statements is true then  $0 < \dim U_j^{i_{t+1}} \cap \text{span} \bigcup_{s=1}^t U_j^{i_s} < \dim U_j^{i_{t+1}} = \ell_j$ , in which case  $U_j^{i_{t+1}} \cap \text{span} \bigcup_{s=1}^t U_j^{i_s}$  is a subspace of  $V_j$  which is fixed by  $\pi_j(g)$  for all  $g \in G^o$  but has dimension strictly less than  $\ell_j$ , contradicting minimality, and we deduce the truth of the claim.

Now let  $r_j$  be the largest integer such that we can find distinct integers  $i_1, \dots, i_{r_j}$  for which the spaces  $U_j^{i_1}, \dots, U_j^{i_{r_j}}$  form a direct sum. (We observe that  $r_j$  is at least 1 and at most  $n_j$ , hence is well-defined.) If  $U_j^{i_1} \oplus \dots \oplus U_j^{i_{r_j}}$  is not equal to  $V_j$  then by the preceding observation there must be some subspace  $U_j^t$  which is not contained in it, hence has trivial intersection with it, allowing us to extend the direct sum, which is a contradiction. We therefore have  $V_j = U_j^{i_1} \oplus \dots \oplus U_j^{i_{r_j}}$  and in particular  $r_j \ell_j = d_j$ .

We now claim there exists  $C_1 > 0$  such that

$$\prod_{j=1}^k \|\pi_j(g)\|^{\beta_j} \leq C_1 \prod_{j=1}^k \max_{1 \leq i \leq n_j} \|\pi_j(g)|_{U_j^i}\|^{\beta_j} \tag{9}$$

for all  $g \in G$ . It is clearly sufficient to show that for each  $j$  there exists  $\tau_j > 0$  such that  $\max_{1 \leq i \leq n_j} \|B|_{U_j^i}\| \geq \tau_j \|B\|$  for every linear map  $B: V_j \rightarrow V_j$ , since then we may take  $C_1 := \prod_{j=1}^k \tau_j^{-\beta_j}$ . By homogeneity it is clearly sufficient to restrict to the case where  $\|B\| = 1$ . If we can show that  $\max_{1 \leq i \leq n_j} \|B|_{U_j^i}\| > 0$  for every  $B \in \text{End}(V_j)$  with  $\|B\| = 1$  then the existence of  $\tau_j$  follows by the compactness of the unit sphere of  $\text{End}(V_j)$ . But if this inequality fails for some  $B \in \text{End}(V_j)$  with  $\|B\| = 1$  then we have found a nonzero linear map from  $V_j$  to itself which is zero on every  $U_j^i$ , and this is impossible since the spaces  $U_j^i$  together span  $V_j$ . The claim is proved.

5.3.2. *Transitivity classes and the construction of quasi-multiplicative potentials.* Let  $\mathfrak{X}$  denote the set of all  $k$ -tuples  $(U_j^{i_j})_{j=1}^k$  such that  $1 \leq i_j \leq n_j$  for all  $j = 1, \dots, k$ . We observe that  $G$  acts on  $\mathfrak{X}$  by taking the pair  $(g, (U_j^{i_j})_{j=1}^k)$  to the tuple  $(\pi_j(g)U_j^{i_j})_{j=1}^k$ .

Since the value of  $(\pi_j(g)U_j^{ij})_{j=1}^k$  depends only on the connected component of  $G$  to which  $g$  belongs, the  $G$ -action on  $\mathfrak{X}$  factors through  $G^o$  and yields an action of the finite group  $G/G^o$  on  $\mathfrak{X}$ . Let us say that a *transitivity class* is a subset of  $\mathfrak{X}$  which corresponds to the orbit of a single tuple  $(U_j^{ij})_{j=1}^k$ , and denote the set of transitivity classes by  $\mathfrak{W}$ . Obviously, the number of transitivity classes is finite. For every transitivity class  $\mathcal{W} \in \mathfrak{W}$  let us define a potential  $\Phi^{\mathcal{W}}: \Sigma_N^* \rightarrow (0, +\infty)$  by

$$\Phi^{\mathcal{W}}(\mathbf{i}) := \max_{(\mathcal{W}_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})|_{\mathcal{W}_j}\|^{\beta_j}.$$

The inequality  $\Phi^{\mathcal{W}}(\mathbf{i}j) \leq \Phi^{\mathcal{W}}(\mathbf{i})\Phi^{\mathcal{W}}(j)$  follows easily from the definition. It is clear that for each  $\mathbf{i} \in \Sigma_N^*$ ,

$$\Phi(\mathbf{i}) = \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})\|^{\beta_j} \leq C_1 \prod_{j=1}^k \max_{1 \leq i \leq n_j} \|\pi_j(g_{\mathbf{i}})|_{U_j^i}\|^{\beta_j} \leq C_1 \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})\|^{\beta_j} = C_1 \Phi(\mathbf{i})$$

and also

$$\prod_{j=1}^k \max_{1 \leq i \leq n_j} \|\pi_j(g_{\mathbf{i}})|_{U_j^i}\|^{\beta_j} = \max_{(U_j^{ij})_{j=1}^k \in \mathfrak{X}} \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})|_{U_j^{ij}}\|^{\beta_j} = \max_{\mathcal{W} \in \mathfrak{W}} \Phi^{\mathcal{W}}(\mathbf{i}),$$

so that

$$C_1^{-1} \Phi(\mathbf{i}) \leq \max_{\mathcal{W} \in \mathfrak{W}} \Phi^{\mathcal{W}}(\mathbf{i}) \leq \Phi(\mathbf{i}) \tag{10}$$

for all  $\mathbf{i} \in \Sigma_N^*$ . We observe in particular that  $P(\Phi^{\mathcal{W}}) \leq P(\Phi)$  for every transitivity class  $\mathcal{W}$  by direct appeal to the definition of the pressure.

By [13, Theorem 6]<sup>3</sup> there exist  $\delta > 0$  and a finite subset  $F$  of the semigroup  $\{g_{\mathbf{i}}: \mathbf{i} \in \Sigma_N^*\}$  such that for every  $\mathbf{i}, j \in \Sigma_N^*$  we have

$$\max_{k \in F} \Phi^{\mathcal{W}}(\mathbf{i}kj) \geq \delta \Phi^{\mathcal{W}}(\mathbf{i})\Phi^{\mathcal{W}}(j).$$

By Proposition 2.2 this implies that for each transitivity class  $\mathcal{W}$  there exists a unique measure  $\nu \in \mathcal{M}_\sigma$  which is an equilibrium state for  $\Phi^{\mathcal{W}}$ , and this measure satisfies the Gibbs inequality

$$C_2^{-1} e^{-|\mathbf{i}|P(\Phi^{\mathcal{W}})} \Phi^{\mathcal{W}}(\mathbf{i}) \leq \nu([\mathbf{i}]) \leq C_2 e^{-|\mathbf{i}|P(\Phi^{\mathcal{W}})} \Phi^{\mathcal{W}}(\mathbf{i})$$

for every  $\mathbf{i} \in \Sigma_N^*$ , where  $C_2 > 0$  does not depend on  $\mathbf{i}$ . Since the number of transitivity classes is finite, we may choose  $C_2$  to be independent of the choice of  $\mathcal{W}$  also. We observe in particular that  $\nu([\mathbf{i}])$  is always nonzero.

---

<sup>3</sup>See also a predecessor of this result by Quint, based on the first property of “produit générique” in [48, Proposition I.2]; cf. Step 2 of proof of Theorem 9.

By hypothesis there exists a Bernoulli measure  $\mu \in \mathcal{M}_\sigma$  which is an equilibrium state for  $\Phi$ . Since  $\mu$  is a Bernoulli measure it is ergodic, so by the subadditive ergodic theorem we have for  $\mu$ -a.e.  $x \in \Sigma_N$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \Phi^{\mathcal{W}}(x|_n) = \Lambda(\Phi^{\mathcal{W}}, \mu)$$

for every transitivity class  $\mathcal{W}$ , and also

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \Phi(x|_n) = \Lambda(\Phi, \mu).$$

In particular, for  $\mu$ -a.e.  $x \in \Sigma_N$ ,

$$\begin{aligned} \Lambda(\Phi, \mu) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \Phi(x|_n) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \max_{\mathcal{W} \in \mathcal{W}} \Phi^{\mathcal{W}}(x|_n) \\ &= \max_{\mathcal{W} \in \mathcal{W}} \lim_{n \rightarrow \infty} \frac{1}{n} \log \Phi^{\mathcal{W}}(x|_n) = \max_{\mathcal{W} \in \mathcal{W}} \Lambda(\Phi^{\mathcal{W}}, \mu) \end{aligned} \tag{11}$$

where we have used (10) in the second equation. Choose a transitivity class  $\mathcal{W}_0$  which attains this maximum, which we fix for the remainder of the proof. We have

$$P(\Phi) = h(\mu) + \Lambda(\Phi, \mu) = h(\mu) + \Lambda(\Phi^{\mathcal{W}_0}, \mu) \leq P(\Phi^{\mathcal{W}_0}) \leq P(\Phi)$$

using the variational principle and the inequality  $P(\Phi^{\mathcal{W}}) \leq P(\Phi)$  established earlier. Since the first and last terms in this chain of inequalities are equal, the inequalities must be equations. It follows that  $\mu$  is the unique equilibrium state of the potential  $\Phi^{\mathcal{W}_0}$ .

*5.3.3. Investigation of the transitivity class  $\mathcal{W}_0$ .* We now investigate the transitivity class  $\mathcal{W}_0$  specified in the previous paragraph which attains the maximum in (11). We claim that the fact that the potential  $\Phi^{\mathcal{W}_0}$  has a Bernoulli measure as its equilibrium state implies an additional relationship between the tuples  $(W_j)_{j=1}^k$  which constitute the transitivity class  $\mathcal{W}_0$ . Specifically, we claim that there exists  $C_3 > 0$  such that for all  $\mathbf{i} \in \Sigma_N^*$  such that  $g_{\mathbf{i}} \in G^o$ ,

$$\Phi^{\mathcal{W}_0}(\mathbf{i}) \leq C_3 \min_{(W_j)_{j=1}^k \in \mathcal{W}_0} \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})|_{W_j}\|^{\beta_j}. \tag{12}$$

Before beginning the proof of the claim we make the following observation. By the Gibbs inequality established previously, there exists  $C_2 > 0$  such that for all  $\mathbf{i} \in \Sigma_N^*$ ,

$$C_2^{-1} e^{|\mathbf{i}|P(\Phi)} \mu([\mathbf{i}]) \leq \Phi^{\mathcal{W}_0}(\mathbf{i}) \leq C_2 e^{|\mathbf{i}|P(\Phi)} \mu([\mathbf{i}]).$$

If  $\mathbf{i}, \mathbf{j} \in \Sigma_N^*$  are arbitrary then we notice that  $\mu([\mathbf{i}\mathbf{j}]) = \mu([\mathbf{i}])\mu([\mathbf{j}])$  because  $\mu$  is Bernoulli, and therefore

$$\begin{aligned} \Phi^{\mathcal{W}_0}(\mathbf{i}\mathbf{j}) &\geq C_2^{-1} e^{|\mathbf{i}\mathbf{j}|P(\Phi)} \mu([\mathbf{i}\mathbf{j}]) \\ &= C_2^{-1} e^{|\mathbf{i}|P(\Phi)} \mu([\mathbf{i}]) e^{|\mathbf{j}|P(\Phi)} \mu([\mathbf{j}]) \geq C_2^{-3} \Phi^{\mathcal{W}_0}(\mathbf{i}) \Phi^{\mathcal{W}_0}(\mathbf{j}). \end{aligned} \tag{13}$$

We will use this property to prove the claim.

Let  $r$  be the number of (Zariski) connected components of  $G$ . Since the semigroup  $\{g_i : i \in \Sigma_N^*\}$  is Zariski dense in  $G$ , we may choose  $j_1, \dots, j_r \in \Sigma_N^*$  such that every connected component of  $G$  contains precisely one of the elements  $g_{j_r}, g_{j_r j_{r-1}}, \dots, g_{j_1 \dots j_r}$  and therefore the sequence  $g_{j_r} G^o, g_{j_{r-1} j_r} G^o, \dots, g_{j_1 \dots j_r} G^o$  lists the components of  $G$ . It follows that if  $(W_j)_{j=1}^k \in \mathcal{W}_0$  is arbitrary, then  $(\pi_j(g_{j_1 \dots j_r})W_j)_{j=1}^k$  lists all of the elements of  $\mathcal{W}_0$  (possibly with repetitions) as  $i$  runs through  $1, \dots, r$ .

Now let  $i \in \Sigma_N^*$  be an arbitrary word such that  $g_i \in G^o$ , and let  $(W_j)_{j=1}^k \in \mathcal{W}_0$  be such that

$$\prod_{j=1}^k \|\pi_j(g_i)|_{W_j}\|^{\beta_j} = \min_{(W_j)_{j=1}^k \in \mathcal{W}_0} \prod_{j=1}^k \|\pi_j(g_i)|_{W_j}\|^{\beta_j}.$$

Observe that by definition there exists  $(W_j)_{j=1}^k \in \mathcal{W}_0$  such that

$$\Phi^{\mathcal{W}_0}(i j_1 i j_2 i \dots j_{r-1} i j_r) = \prod_{j=1}^k \|\pi_j(g_{i j_1 i j_2 i \dots j_{r-1} i j_r})|_{W_j}\|^{\beta_j}.$$

Repeated application of (13) yields

$$\begin{aligned} \Phi^{\mathcal{W}_0}(i j_1 i j_2 i \dots j_{r-1} i j_r) &\geq C_2^{-3(2r-1)} \Phi^{\mathcal{W}_0}(i)^r \left( \prod_{t=1}^r \Phi^{\mathcal{W}_0}(j_t) \right) \\ &\geq \tau \Phi^{\mathcal{W}_0}(i)^r, \end{aligned} \tag{14}$$

say, where  $\tau > 0$  is independent of  $i$ . In the other direction we obtain

$$\begin{aligned} \Phi^{\mathcal{W}_0}(i j_1 i j_2 i \dots j_{r-1} i j_r) &= \prod_{j=1}^k \|\pi_j(g_{i j_1 i j_2 i \dots j_{r-1} i j_r})|_{W_j}\|^{\beta_j} \\ &\leq \left( \prod_{t=1}^r \prod_{j=1}^k \|\pi_j(g_{i j_t})|_{\pi_j(g_{i j_{t+1} \dots j_r})W_j}\|^{\beta_j} \right) \\ &= \left( \prod_{t=1}^r \prod_{j=1}^k \|\pi_j(g_{i j_t})|_{\pi_j(g_{j_{t+1} \dots j_r})W_j}\|^{\beta_j} \right) \end{aligned}$$

where we have used the fact that  $(\pi_j(g_i)W_j)_{j=1}^k = (W_j)_{j=1}^k$  for every  $(W_j)_{j=1}^k \in \mathcal{W}_0$  since  $g_i \in G^o$ . This is clearly bounded by

$$\left( \prod_{t=1}^r \prod_{j=1}^k \|\pi_j(g_{j_t})|_{\pi_j(g_{j_{t+1} \dots j_r})W_j}\|^{\beta_j} \right) \left( \prod_{t=1}^r \prod_{j=1}^k \|\pi_j(g_i)|_{\pi_j(g_{j_t \dots j_r})W_j}\|^{\beta_j} \right)$$

and hence by

$$K \left( \prod_{t=1}^r \prod_{j=1}^k \|\pi_j(g_i)|_{\pi_j(g_{j_t \dots j_r})W_j}\|^{\beta_j} \right)$$

where  $K := \prod_{t=1}^r \Phi^{\mathcal{W}_0}(j_t)$ , which clearly does not depend on  $\mathbf{i}$ . Thus

$$\Phi^{\mathcal{W}_0}(\mathbf{i}j_1\mathbf{i}j_2\mathbf{i}\cdots j_{r-1}\mathbf{i}j_r) \leq K \prod_{t=1}^r \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})|_{\pi_j(g_{j_t \dots j_r})W_j}\|^{\beta_j}.$$

But this in turn is clearly bounded by

$$K \left( \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})|_{W'_j}\|^{\beta_j} \right) \left( \max_{(W'_j)_{j=1}^k \in \mathcal{W}_0} \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})|_{W'_j}\|^{\beta_j} \right)^{r-1}$$

because as  $t$  ranges from 1 to  $r$ , the tuple  $(\pi_j(g_{j_t \dots j_r})W_j)_{j=1}^k$  ranges over all of the elements of  $\mathcal{W}_0$  and in particular is equal to  $(W'_j)_{j=1}^k$  for at least one value of  $t$ . Thus

$$\Phi^{\mathcal{W}_0}(\mathbf{i}j_1\mathbf{i}j_2\mathbf{i}\cdots j_{r-1}\mathbf{i}j_r) \leq K \left( \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})|_{W'_j}\|^{\beta_j} \right) \Phi^{\mathcal{W}_0}(\mathbf{i})^{r-1}. \tag{15}$$

Combining (14) and (15) yields

$$\tau \Phi^{\mathcal{W}_0}(\mathbf{i})^r \leq K \left( \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})|_{W'_j}\|^{\beta_j} \right) \Phi^{\mathcal{W}_0}(\mathbf{i})^{r-1}$$

where  $K, \tau > 0$  do not depend on  $\mathbf{i}$ , and dividing by  $\tau \Phi^{\mathcal{W}_0}(\mathbf{i})^{r-1}$  proves the claim.

5.3.4. *A multiplicativity property on a dense subsemigroup of the identity component.*

We now claim that for all  $\mathbf{i}, \mathbf{j} \in \Sigma_N^*$  such that  $g_{\mathbf{i}}, g_{\mathbf{j}} \in G^o$  and every  $(W_j)_{j=1}^k \in \mathcal{W}_0$ , we have

$$\prod_{j=1}^k \rho(\pi_j(g_{\mathbf{i}}g_{\mathbf{j}})|_{W_j})^{\beta_j} = \left( \prod_{j=1}^k \rho(\pi_j(g_{\mathbf{i}})|_{W_j})^{\beta_j} \right) \left( \prod_{j=1}^k \rho(\pi_j(g_{\mathbf{j}})|_{W_j})^{\beta_j} \right) \tag{16}$$

where  $\rho(B)$  denotes the spectral radius of the linear map  $B$ . Fix words  $\mathbf{i}$  and  $\mathbf{j}$  such that  $g_{\mathbf{i}}, g_{\mathbf{j}} \in G^o$ , and fix  $(W_j)_{j=1}^k \in \mathcal{W}_0$ . We observe that  $\pi_j(g_{\mathbf{i}})W_j = W_j$  and  $\pi_j(g_{\mathbf{j}})W_j = W_j$  for all  $j = 1, \dots, k$ . Using the fact that  $\mu$  is a Bernoulli measure we have  $\mu([(ij)^n]) = \mu([\mathbf{i}]^n)\mu([\mathbf{j}]^n) = \mu([\mathbf{i}^n])\mu([\mathbf{j}^n])$  for every  $n \geq 1$ , so by the Gibbs inequality

$$\begin{aligned} \Phi^{\mathcal{W}_0}(\mathbf{i}^n)\Phi^{\mathcal{W}_0}(\mathbf{j}^n) &\leq C_2^2 e^{n(|\mathbf{i}|+|\mathbf{j}|)P(\Phi)} \mu([\mathbf{i}^n])\mu([\mathbf{j}^n]) \\ &= C_2^2 e^{n(|\mathbf{i}|+|\mathbf{j}|)P(\Phi)} \mu([(ij)^n]) \leq C_2^3 \Phi^{\mathcal{W}_0}((ij)^n) \end{aligned}$$

and similarly

$$\Phi^{\mathcal{W}_0}((ij)^n) \leq C_2^3 \Phi^{\mathcal{W}_0}(\mathbf{i}^n)\Phi^{\mathcal{W}_0}(\mathbf{j}^n).$$

We have

$$\prod_{j=1}^k \|\pi_j(g_{\mathbf{i}\mathbf{j}}^n|_{W_j})\|^{\beta_j} \leq \Phi^{\mathcal{W}_0}((ij)^n)$$



by the definition of  $\Phi^{\mathcal{W}_0}$ , and since  $g_{i,j}^n \in G^o$  we have

$$\Phi^{\mathcal{W}_0}((i,j)^n) \leq C_3 \min_{(W'_j)_{j=1}^k \in \mathcal{W}_0} \prod_{j=1}^k \|\pi_j(g_{i,j}^n|_{W'_j})\|^{\beta_j} \leq C_3 \prod_{j=1}^k \|\pi_j(g_{i,j}^n|_{W_j})\|^{\beta_j}$$

by the previous claim. Likewise

$$\prod_{j=1}^k \|\pi_j(g_i^n|_{W_j})\|^{\beta_j} \leq \Phi^{\mathcal{W}_0}(i^n) \leq C_3 \prod_{j=1}^k \|\pi_j(g_i^n|_{W_j})\|^{\beta_j}$$

and

$$\prod_{j=1}^k \|\pi_j(g_j^n|_{W_j})\|^{\beta_j} \leq \Phi^{\mathcal{W}_0}(j^n) \leq C_3 \prod_{j=1}^k \|\pi_j(g_j^n|_{W_j})\|^{\beta_j}.$$

Thus

$$\begin{aligned} \left(\prod_{j=1}^k \|\pi_j(g_i^n|_{W_j})\|^{\beta_j}\right) \left(\prod_{j=1}^k \|\pi_j(g_j^n|_{W_j})\|^{\beta_j}\right) &\leq \Phi^{\mathcal{W}_0}(i^n) \Phi^{\mathcal{W}_0}(j^n) \leq C_2^3 \Phi^{\mathcal{W}_0}((i,j)^n) \\ &\leq C_2^3 C_3 \prod_{j=1}^k \|\pi_j(g_{i,j}^n|_{W_j})\|^{\beta_j} \end{aligned}$$

and similarly

$$\begin{aligned} \prod_{j=1}^k \|\pi_j(g_{i,j}^n|_{W_j})\|^{\beta_j} &\leq \Phi^{\mathcal{W}_0}((i,j)^n) \leq C_2^3 \Phi^{\mathcal{W}_0}(i^n) \Phi^{\mathcal{W}_0}(j^n) \\ &\leq C_2^3 C_3^2 \left(\prod_{j=1}^k \|\pi_j(g_i^n|_{W_j})\|^{\beta_j}\right) \left(\prod_{j=1}^k \|\pi_j(g_j^n|_{W_j})\|^{\beta_j}\right). \end{aligned}$$

We have obtained

$$C_2^{-3} C_3^{-1} \leq \frac{\prod_{j=1}^k \|\pi_j(g_{i,j}^n|_{W_j})\|^{\beta_j}}{(\prod_{j=1}^k \|\pi_j(g_i^n|_{W_j})\|^{\beta_j})(\prod_{j=1}^k \|\pi_j(g_j^n|_{W_j})\|^{\beta_j})} \leq C_2^3 C_3^2$$

for all  $n \geq 1$ . Taking the power  $1/n$  and letting  $n \rightarrow \infty$  we obtain by Gelfand's formula

$$\frac{\prod_{j=1}^k \rho(\pi_j(g_i g_j)|_{W_j})^{\beta_j}}{(\prod_{j=1}^k \rho(\pi_j(g_i)|_{W_j})^{\beta_j})(\prod_{j=1}^k \rho(\pi_j(g_j)|_{W_j})^{\beta_j})} = 1$$

for all  $i, j \in \Sigma_N^*$  such that  $g_i, g_j \in G^o$ , and this is precisely (16).

5.3.5. *Application of the theorem of Benoist.* We now apply the work of Benoist to show that the identity (16) severely restricts the possible structures of the groups  $\{\pi_j(g)|_{W_j} : g \in G^o\}$  for  $(W_j)_{j=1}^k \in \mathcal{W}_0$ . Fix an arbitrary tuple  $(W_j)_{j=1}^k \in \mathcal{W}_0$  and define

$$\xi(g) := \prod_{j=1}^k \rho(\pi_j(g)|_{W_j})^{\beta_j} \tag{17}$$

for all  $g \in G^\circ$ . The identity (16) asserts that  $\xi(g_i g_j) = \xi(g_i)\xi(g_j)$  for all  $i, j \in \Sigma_N^*$  such that  $g_i, g_j \in G^\circ$ .

Recall that by construction (§5.3.1), for each  $j = 1, \dots, k$ , the restriction of  $\pi_j$  to the connected reductive group  $G^\circ$  gives rise to an irreducible linear representation of  $G^\circ$  on  $W_j$ . Denote this representation by  $\hat{\pi}_j$ . We will show that for each  $j = 1, \dots, k$  the image  $\hat{\pi}_j([G^\circ, G^\circ])$  is compact. If  $[G^\circ, G^\circ]$  is itself compact then this result is trivial, so without loss of generality we assume that the semisimple group  $[G^\circ, G^\circ]$  is noncompact. For each  $j$  let  $\hat{\chi}_j$  be the highest weight of  $\hat{\pi}_j$  so that  $\bar{\chi}_j \in \alpha^*$  where  $\alpha$  is a fixed Cartan subspace in the Lie algebra  $\mathfrak{g}$  of  $G$  (see §4.1.2 and §4.1.4). By Lemma 4.1, (17) can be rewritten as

$$\log \xi(g) = \sum_{j=1}^k \beta_j \bar{\chi}_j(\lambda(g)) \tag{18}$$

where  $\lambda$  is the Jordan projection on a fixed Weyl chamber  $\alpha^+$  in  $\alpha$  (§4.1.3).

Denote by  $\Gamma$  the semigroup in  $G$  generated by  $\{g_1, \dots, g_N\}$ , and by  $\Gamma_o$  the intersection  $G^\circ \cap \Gamma$ . Since by hypothesis  $\Gamma$  is Zariski dense in  $G$ , the semigroup  $\Gamma_o$  is Zariski dense in  $G^\circ$ . Setting  $\bar{\chi} := \sum_{j=1}^k \beta_j \bar{\chi}_j$ , in view of (17) and (18), equation (16) implies that the set

$$\{\lambda(\gamma_1 \gamma_2) - \lambda(\gamma_1) - \lambda(\gamma_2) : \gamma_1, \gamma_2 \in \Gamma_o\}$$

is contained in the subspace  $\ker \bar{\chi}$ . Since the latter is closed, by Theorem 7 we deduce that the semisimple part  $\alpha_S$  of the Cartan space  $\alpha$  is contained in  $\ker \bar{\chi}$ . Furthermore, since for each  $j = 1, \dots, k$ ,  $\bar{\chi}_j$  is a dominant weight (in particular, it takes nonnegative values on the cone  $\alpha_S \cap \alpha^+$ ) and  $\beta_j > 0$ , this implies that for each  $j = 1, \dots, k$ , we have  $\alpha_S \subseteq \ker \bar{\chi}_j$ . Hence by Lemma 4.1 the spectral radius of every element of  $\hat{\pi}_j([G^\circ, G^\circ])$  is 1. The determinant of every element of  $\hat{\pi}_j([G^\circ, G^\circ])$  is also 1 as a direct consequence of the definition of  $[G^\circ, G^\circ]$  (as closure of a group generated by elements of type  $ghg^{-1}h^{-1}$ ), so every element of  $\hat{\pi}_j([G^\circ, G^\circ])$  has every eigenvalue equal to 1 in modulus. Since  $[G^\circ, G^\circ]$  is semisimple it acts completely reducibly on  $W_j$ , so by applying Lemma 3.6 to each subspace in a decomposition of  $W_j$  into invariant subspaces on which  $[G^\circ, G^\circ]$  acts irreducibly, it follows that  $\hat{\pi}_j([G^\circ, G^\circ])$  is a compact subgroup of  $GL(W_j)$  as required.

On the other hand, since  $\hat{\pi}_j$  is an irreducible representation (§5.3.1), by Schur’s lemma,  $\mathbb{R}Z(\hat{\pi}_j(G^\circ)) \leq \text{End}_{\mathbb{R}\hat{\pi}_j(G^\circ)}(W_j)$  is isomorphic to either  $\mathbb{R}$  or  $\mathbb{C}$  as a real division algebra. In the first case,  $Z(\hat{\pi}_j(G^\circ))$  is contained in the group of homotheties  $\simeq \mathbb{R}^*$  of  $W_j$  and in the latter case it is contained in a copy of  $SO(2, \mathbb{R}) \times \mathbb{R}^*$  in  $GL(W_j)$ . Finally, we recall that the connected real reductive group  $G^\circ$  is an almost direct product of its centre  $Z(G^\circ)$  and  $[G^\circ, G^\circ]$  [15, Proposition 2.2], which is to say the map  $Z(G^\circ) \times [G^\circ, G^\circ] \rightarrow G^\circ$  defined by  $(z, g) \mapsto zg$  is surjective with finite kernel. We conclude that  $\hat{\pi}_j(G^\circ)$  is contained in a compact subgroup of  $GL(W_j)$  modulo factoring out the absolute value of the determinant of each element, and therefore each of the groups  $\hat{\pi}_j(G^\circ)$  is a group of linear similarity transformations of  $W_j$  with respect to some Euclidean structure on  $W_j$ .

Now recall that, for each  $j = 1, \dots, k$ , the finite group  $G/G^\circ$  acts transitively on  $\{U_j^i : i = 1, \dots, n_j\}$ . Since for each  $j = 1, \dots, k$  we have  $W_j = U_j^i$  for some

$i \in \{1, \dots, n_j\}$ , by transitivity of  $G/G^o$ , repeating the same argument above for every  $(W_j)_{j=1}^k \in \mathcal{W}_0$ , we deduce that up to conjugation in  $\text{GL}(V_j)$ ,  $\pi_j(G^o)|_{U_j^i}$  is contained in the group of linear similarities of  $U_j^i$  for every  $i = 1, \dots, n_j$ , for every  $j = 1, \dots, k$ . In particular, passing to matrix representation by convenient choice of bases for  $U_j^{i\ell}$ 's for  $\ell = 1, \dots, r_j$  and  $j = 1, \dots, k$ ,  $\pi_j(G^o)$  is contained in the group of block diagonal matrices of the form

$$\begin{bmatrix} \gamma_1 O_1 & 0 & \dots & 0 \\ 0 & \gamma_2 O_2 & \dots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \gamma_{r_j} O_{r_j} \end{bmatrix} \tag{19}$$

where the  $\gamma_i$ 's are scalars in  $\mathbb{R}_+^*$  and  $O_i$ 's are  $\ell_j \times \ell_j$  orthogonal matrices. We have completed the first of the two parts of the proof as described in §5.1.1.

*5.3.6. The identity of the scalars.* In the second part of the proof we wish to show that for every  $g \in G^o$ , in the matrix representation (19) we have  $\gamma_1 = \dots = \gamma_{r_j}$ . Since obviously each  $\gamma_i$  is equal to  $|\det(\gamma_i O_i)|^{1/\ell_j}$ , the idea is to show that for each  $g \in G^o$  and  $j = 1, \dots, k$  the quantity  $|\det(\pi_j(g)|_{U_j^i})|^{1/\ell_j}$  is independent of  $i$ . Since  $V_j$  can be written as a direct sum of a subcollection of spaces  $U_j^{i_1}, \dots, U_j^{i_{r_j}}$ , this in turn is clearly equivalent to the identity

$$|\det(\pi_j(g)|_{U_j^i})|^{1/\ell_j} = |\det \pi_j(g)|^{1/d_j} \tag{20}$$

for every  $i = 1, \dots, n_j$  and  $j = 1, \dots, k$ , which is what shall be shown. It will then be a straightforward matter to conclude the theorem.

We therefore undertake to prove (20). To establish this equality we must use the fact that  $\Phi^{\mathcal{W}_0}$  has the greatest pressure of any  $\Phi^{\mathcal{W}}$ , which we did not previously substantially use. The key fact which we shall ultimately demonstrate is that there exists  $C > 0$  such that  $C^{-1}\Phi^{\mathcal{W}_0}(\mathbf{i}) \leq \Phi^{\mathcal{W}}(\mathbf{i}) \leq C\Phi^{\mathcal{W}_0}(\mathbf{i})$  for every  $\mathbf{i} \in \Sigma_N^*$  such that  $g_{\mathbf{i}} \in G^o$ , for every transitivity class  $\mathcal{W}$ .

*5.3.7. A first identity involving determinants.* Fix  $j \in \{1, \dots, k\}$ . If we knew that the number  $n_j$  of spaces  $U_j^i$  was equal to exactly  $d_j/\ell_j$  then we would have  $V_j = \bigoplus_{i=1}^{n_j} U_j^i$  and the identity

$$\left( \prod_{i=1}^{n_j} |\det(\pi_j(g)|_{U_j^i})|^{1/\ell_j} \right)^{1/n_j} = |\det \pi_j(g)|^{1/d_j} \tag{21}$$

would be obvious. However, in general we do not necessarily have  $n_j = d_j/\ell_j$ . Our first task will be to show that the above identity remains true even when  $n_j > d_j/\ell_j$  and the spaces  $U_j^1, \dots, U_j^{n_j}$  do not form a direct sum. The proof of this equality is conducted by exploring the combinatorial relationships between the similarity ratios  $\gamma_i(g) := |\det(\pi_j(g)|_{U_j^i})|^{1/\ell_j}$  and subspaces  $U_j^i$ . The fundamental task will be to show that the list of spaces  $U_j^1, \dots, U_j^{n_j}$  may be partitioned into equal-sized classes in such a way that

every  $g \in G^o$  has constant similarity ratio on each class, and the spans of the classes form a direct sum.

For  $i = 1, \dots, n_j$  and  $g \in G^o$ , let  $\gamma_i(g) := |\det(\pi_j(g)|_{U_j^i})|^{1/\ell_j} \in \mathbb{R}_+^*$  denote the similarity ratio of  $\pi_j(g)|_{U_j^i}$ . Define an equivalence relation  $\sim$  on  $\{1, \dots, n_j\}$  by writing  $i_1 \sim i_2$  if and only if  $\gamma_{i_1}(g) = \gamma_{i_2}(g)$  for all  $g \in G^o$ . Let  $x_1, \dots, x_p$  denote the equivalence classes under  $\sim$ . There is a natural action of  $G/G^o$  on  $\{1, \dots, n_j\}$  which takes the pair  $([g], i)$  to the unique integer  $i'$  such that  $\pi_j(g)U_j^i = U_j^{i'}$ , and this action is obviously transitive since  $G/G^o$  acts transitively on the spaces  $U_j^1, \dots, U_j^{n_j}$ . For distinct  $i_1$  and  $i_2$  and arbitrary  $g \in G^o$  and  $h \in G$  it is not difficult to see that  $\pi_j(g)$  has distinct similarity ratios on  $U_j^{i_1}$  and  $U_j^{i_2}$  if and only if  $\pi_j(hgh^{-1})$  has distinct similarity ratios on  $\pi_j(h)U_j^{i_1}$  and  $\pi_j(h)U_j^{i_2}$ , so the action on  $\{1, \dots, n_j\}$  respects the equivalence relation  $\sim$  and in particular has the effect of inducing a permutation of the equivalence classes  $x_1, \dots, x_p$ . The transitivity of the action of  $G/G^o$  on  $\{1, \dots, n_j\}$  easily implies that this action of  $G/G^o$  on the set of equivalence classes is transitive. It follows in particular that the equivalence classes must all have the same cardinality: we have  $\#x_t = n_j/p$  for every  $t = 1, \dots, p$ .

For each equivalence class  $x_t$  define  $X_t$  to be the span of the union of all the subspaces  $U_j^i$  such that  $i \in x_t$ . Arguing as in the second paragraph of §5.3.1 we note that every  $X_t$  must be equal to a direct sum  $U_j^{i_1} \oplus \dots \oplus U_j^{i_q}$  for some suitable choice of indices  $i_1, \dots, i_q \in x_t$  and for some integer  $q \geq 1$  which *a priori* might depend on  $t$ . (To see this, consider a direct sum  $U_j^{i_1} \oplus \dots \oplus U_j^{i_q} \subseteq X_t$  with  $i_1, \dots, i_q \in x_t$  which is maximal in the sense that it cannot be extended by a further direct summand  $U_j^{i_{q+1}}$  such that  $i_{q+1} \in x_t$ . If every  $U_j^i$  satisfying  $i \in x_t$  is a subspace of this direct sum then the direct sum equals  $X_t$  as required. Otherwise, there exists  $U_j^i$  satisfying  $i \in x_t$  which neither is a subspace of  $U_j^{i_1} \oplus \dots \oplus U_j^{i_q}$  nor forms a direct sum with it, in which case the intersection  $(U_j^{i_1} \oplus \dots \oplus U_j^{i_q}) \cap U_j^i$  is nonzero, has finite orbit under the action of  $\pi_j(G)$ , and has dimension smaller than  $\ell_j$ , contradicting the definition of  $\ell_j$ . We conclude that any such maximal direct sum yields a decomposition of  $X_t$  with the claimed properties.) Now, as a consequence of the result shown in §5.3.5, every  $U_j^i$  admits an inner product structure with respect to which every  $g \in G^o$  acts on  $U_j^i$  as a similarity transformation. Combined with the existence of the aforementioned direct sums this implies that for every  $t = 1, \dots, p$  there exists an inner product structure on  $X_t$  with respect to which every  $g \in G^o$  acts on  $X_t$  as a similarity transformation. For distinct  $t_1, t_2$  in the range  $1, \dots, p$ , by the definition of  $\sim$  there exists  $g \in G^o$  such that  $\pi_j(g)$  has different similarity ratios on  $X_{t_1}$  and on  $X_{t_2}$ , and this implies that necessarily  $X_{t_1} \cap X_{t_2} = \{0\}$ . We conclude that the spaces  $X_1, \dots, X_p$  form a direct sum, which is equal to the span of the spaces  $U_j^1, \dots, U_j^{n_j}$  and hence is equal to  $V_j$ . Since  $G/G^o$  transitively permutes the set of equivalence classes  $x_1, \dots, x_p$  it follows that the action  $([g], X_t) \mapsto \pi_j(g)X_t$  transitively permutes the spaces  $X_1, \dots, X_p$ . These spaces are therefore pairwise isomorphic, so  $\dim X_t$  is independent of  $t$  and therefore  $\dim X_t = d_j/p$  for every  $i = 1, \dots, p$ .

We may now prove (21). We observe that for every  $g \in G^o$  and  $t \in \{1, \dots, p\}$ ,

$$|\det(\pi_j(g)|_{X_t})|^{1/\dim X_t} = \left( \prod_{i \in x_t} |\det(\pi_j(g)|_{U_j^i})|^{1/\ell_j} \right)^{1/\#x_t}$$

because the term on the left is the similarity ratio of  $\pi_j(g)$  on  $X_t$ , which is also the similarity ratio of  $\pi_j(g)$  on  $U_j^i$  for every  $i \in x_t$ . This is to say

$$|\det(\pi_j(g)|_{X_t})|^{p/d_j} = \left( \prod_{i \in x_t} |\det(\pi_j(g)|_{U_j^i})|^{1/\ell_j} \right)^{p/n_j}$$

for every  $t = 1, \dots, p$ . Since  $V_j = \bigoplus_{t=1}^p X_t$ , we also have

$$\prod_{t=1}^p \det(\pi_j(g)|_{X_t}) = \det \pi_j(g).$$

Hence

$$\begin{aligned} |\det \pi_j(g)| &= \prod_{t=1}^p |\det(\pi_j(g)|_{X_t})| \\ &= \prod_{t=1}^p \left( \prod_{i \in x_t} |\det(\pi_j(g)|_{U_j^i})|^{1/\ell_j} \right)^{d_j/n_j} = \left( \prod_{i=1}^{n_j} |\det(\pi_j(g)|_{U_j^i})|^{1/\ell_j} \right)^{d_j/n_j} \end{aligned}$$

and this is precisely (21).

5.3.8. *A second identity involving determinants.* Here, we will apply (21) to derive a further identity: we claim that for all  $g \in G^o$  and  $\mathcal{W} \in \mathscr{W}$ ,

$$\left( \prod_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k |\det(\pi_j(g)|_{W_j})|^{\beta_j/\ell_j} \right)^{1/\#\mathcal{W}} = \prod_{j=1}^k |\det \pi_j(g)|^{\beta_j/d_j}. \tag{22}$$

To see this, fix  $g \in G^o$ , let  $\mathcal{W}$  be a transitivity class and let  $(W'_j)_{j=1}^k \in \mathcal{W}$  be arbitrary. We note that the sets

$$\{[h] \in G/G^o : (\pi_j(h)W'_j)_{j=1}^k = (W_j)_{j=1}^k\}$$

for distinct  $(W_j)_{j=1}^k \in \mathcal{W}$  form a partition of  $G/G^o$  into cosets, hence each has the same cardinality. We deduce that

$$\begin{aligned} \left( \prod_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k |\det(\pi_j(g)|_{W_j})|^{\beta_j/\ell_j} \right)^{1/\#\mathcal{W}} &= \left( \prod_{[h] \in G/G^o} \prod_{j=1}^k |\det(\pi_j(g)|_{\pi_j(h)W'_j})|^{\beta_j/\ell_j} \right)^{\frac{1}{\#G/G^o}}. \end{aligned}$$

It is therefore sufficient to show that for each  $j = 1, \dots, k$ , for every  $i_0 \in \{1, \dots, n_j\}$ ,

$$\left( \prod_{[h] \in G/G^o} |\det(\pi_j(g)|_{\pi_j(h)U_j^{i_0}})|^{1/\ell_j} \right)^{\frac{1}{\#G/G^o}} = |\det \pi_j(g)|^{1/d_j}.$$

Fix such  $j$  and  $i_0$ . As before, the sets

$$\{[h] \in G/G^o : \pi_j(h)U_j^{i_0} = U_j^i\}$$

form a partition of  $G/G^o$  into cosets and hence have equal cardinality, which implies that

$$\left( \prod_{[h] \in G/G^o} |\det(\pi_j(g)|_{\pi_j(h)U_j^{i_0}})^{1/\ell_j} \right)^{\#G/G^o} = \left( \prod_{i=1}^{n_j} |\det(\pi_j(g)|_{U_j^i})^{1/\ell_j} \right)^{1/n_j}.$$

By (21) this last expression is equal to  $|\det \pi_j(g)|^{1/d_j}$ , so combining the identities obtained so far yields (22).

5.3.9. *Two inequalities between potentials.* Let us define a new potential by

$$\Phi^{\det}(\mathbf{i}) := \prod_{j=1}^k |\det \pi_j(g_{\mathbf{i}})|^{\beta_j/d_j}$$

for all  $\mathbf{i} \in \Sigma_N^*$ . We clearly have  $\Phi^{\det}(\mathbf{i}j) = \Phi^{\det}(\mathbf{i})\Phi^{\det}(j)$  for all  $\mathbf{i}, j \in \Sigma_N^*$ . We aim to show that

$$P(\Phi) = P(\Phi^{\mathcal{W}}) = P(\Phi^{\det}) \tag{23}$$

for all transitivity classes  $\mathcal{W}$ .

In pursuit of (23) we will prove two inequalities. We first claim that there exists  $C_4 > 0$  such that for every transitivity class  $\mathcal{W}$  we have  $\Phi^{\det}(\mathbf{i}) \leq C_4 \Phi^{\mathcal{W}}(\mathbf{i})$  for all  $\mathbf{i} \in \Sigma_N^*$ . We begin by considering the case where  $\mathbf{i} \in \Sigma_N^*$  satisfies  $\mathbf{i} \in G^o$ . It follows easily from (22) that

$$\begin{aligned} \Phi^{\mathcal{W}}(\mathbf{i}) &= \max_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})|_{W_j}\|^{\beta_j} \geq \max_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k |\det(\pi_j(g_{\mathbf{i}})|_{W_j})|^{\beta_j/\ell_j} \\ &\geq \left( \prod_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k |\det(\pi_j(g_{\mathbf{i}})|_{W_j})|^{\beta_j/\ell_j} \right)^{1/\#\mathcal{W}} \\ &= |\det \pi_j(g_{\mathbf{i}})|^{\beta_j/d_j} = \Phi^{\det}(\mathbf{i}) \end{aligned} \tag{24}$$

for every transitivity class  $\mathcal{W}$  and every  $\mathbf{i} \in \Sigma_N^*$  such that  $g_{\mathbf{i}} \in G^o$ . Now observe that by the Zariski density of the semigroup  $\{g_{\mathbf{i}} : \mathbf{i} \in \Sigma_N^*\}$  in  $G$ , we may choose  $k_1, \dots, k_r$  such that every connected component of  $G$  contains one of the elements  $g_{k_t}$ . Given  $\mathbf{i} \in \Sigma_N^*$  observe that we can choose  $t_0 \in \{1, \dots, r\}$  such that  $g_{\mathbf{i}k_{t_0}} \in G^o$ . We have  $\Phi^{\mathcal{W}}(\mathbf{i}k_{t_0}) \leq \Phi^{\mathcal{W}}(\mathbf{i})\Phi^{\mathcal{W}}(k_{t_0})$  and  $\Phi^{\det}(\mathbf{i}k_{t_0}) = \Phi^{\det}(\mathbf{i})\Phi^{\det}(k_{t_0})$ , and therefore (24) yields

$$\frac{\Phi^{\det}(\mathbf{i})}{\Phi^{\mathcal{W}}(\mathbf{i})} \leq \left( \frac{\Phi^{\mathcal{W}}(k_{t_0})}{\Phi^{\mathcal{W}}(\mathbf{i}k_{t_0})} \right) \left( \frac{\Phi^{\det}(\mathbf{i}k_{t_0})}{\Phi^{\det}(k_{t_0})} \right) \leq \frac{\Phi^{\mathcal{W}}(k_{t_0})}{\Phi^{\det}(k_{t_0})} \leq C_4,$$

where

$$C_4 := \max_{\mathcal{W} \in \mathcal{W}} \max_{1 \leq t \leq r} \frac{\Phi^{\mathcal{W}}(k_t)}{\Phi^{\det}(k_t)},$$

which proves the claim.

We now establish our second inequality: we claim that there exists  $C_5 > 0$  such that  $\Phi^{\mathcal{W}_0}(\mathbf{i}) \leq C_5 \Phi^{\det}(\mathbf{i})$  for every  $\mathbf{i} \in \Sigma_N^*$ . By the inequality (12) established in §5.3.3 we have

$$\Phi^{\mathcal{W}_0}(\mathbf{i}) \leq C_3 \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})|_{W_j}\|^{\beta_j}$$

for some  $C_3 > 0$  and every  $g_{\mathbf{i}} \in G^o$  and  $(W_j)_{j=1}^k \in \mathcal{W}_0$ . It follows in particular that

$$\frac{\Phi^{\mathcal{W}_0}(\mathbf{i})}{\prod_{j=1}^k |\det(\pi_j(g_{\mathbf{i}})|_{W_j})|^{\beta_j/\ell_j}} \leq \frac{C_3 \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})|_{W_j}\|^{\beta_j}}{\prod_{j=1}^k |\det(\pi_j(g_{\mathbf{i}})|_{W_j})|^{\beta_j/\ell_j}}$$

for all  $g_{\mathbf{i}} \in G^o$  and  $(W_j)_{j=1}^k \in \mathcal{W}_0$ . Since for each  $j$ ,

$$\{|\det(\pi_j(g)|_{W_j})|^{-1/\ell_j} \pi_j(g)|_{W_j} : g \in G^o\}$$

is contained in a compact subset of  $\text{GL}(W_j)$ , it follows that there exists  $K > 0$  such that

$$\frac{\Phi^{\mathcal{W}_0}(\mathbf{i})}{\prod_{j=1}^k |\det(\pi_j(g_{\mathbf{i}})|_{W_j})|^{\beta_j/\ell_j}} \leq K$$

for all  $g_{\mathbf{i}} \in G^o$  and  $(W_j)_{j=1}^k \in \mathcal{W}_0$ . Taking the geometric mean over all  $(W_j)_{j=1}^k \in \mathcal{W}_0$  for fixed  $g_{\mathbf{i}}$  using (22) yields

$$\frac{\Phi^{\mathcal{W}_0}(\mathbf{i})}{\Phi^{\det}(\mathbf{i})} \leq K$$

for all  $\mathbf{i} \in \Sigma_N^*$  such that  $g_{\mathbf{i}} \in G^o$ . We now extend to the case of general words  $\mathbf{i}$ . Fix  $\mathbf{i} \in \Sigma_N^*$  and observe that we may choose  $t_0 \in \{1, \dots, r\}$  such that  $g_{\mathbf{i}k_{t_0}} \in G^o$ . For some  $(W_j)_{j=1}^k \in \mathcal{W}^0$  we have

$$\Phi^{\mathcal{W}_0}(\mathbf{i}) = \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})|_{W_j}\|^{\beta_j}$$

and therefore

$$\begin{aligned} \Phi^{\mathcal{W}_0}(\mathbf{i}) &= \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}}g_{k_{t_0}}g_{k_{t_0}}^{-1})|_{W_j}\|^{\beta_j} \\ &\leq \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}k_{t_0}})|_{\pi_j(g_{k_{t_0}}^{-1})W_j}\|^{\beta_j} \prod_{j=1}^k \|\pi_j(g_{k_{t_0}}^{-1})|_{W_j}\|^{\beta_j} \\ &\leq \left(\prod_{j=1}^k \|\pi_j(g_{\mathbf{i}k_{t_0}})|_{\pi_j(g_{k_{t_0}}^{-1})W_j}\|^{\beta_j}\right) \left(\max_{1 \leq t \leq r} \prod_{j=1}^k \|\pi_j(g_{k_t}^{-1})|_{W_j}\|^{\beta_j}\right) \\ &\leq C \Phi^{\mathcal{W}_0}(\mathbf{i}k_{t_0}) \leq KC \Phi^{\det}(\mathbf{i}k_{t_0}) \leq KC \left(\max_{1 \leq t \leq r} \Phi^{\det}(k_t)\right) \Phi^{\det}(\mathbf{i}) \leq C_5 \Phi^{\det}(\mathbf{i}), \end{aligned}$$

where we took  $C := \max_{1 \leq t \leq r} \prod_{j=1}^k \|\pi_j(g_{k_t}^{-1})|_{W_j}\|^{\beta_j}$  and  $C_5 := KC \max_{1 \leq t \leq r} \Phi^{\det}(k_t)$ . This proves the claim.

5.3.10. *The Gibbs property and a third inequality between potentials.* The two inequalities just proved assert that for some  $C > 0$ ,

$$C^{-1} \Phi^{\mathcal{W}_0}(\mathbf{i}) \leq \Phi^{\text{det}}(\mathbf{i}) \leq C \Phi^{\mathcal{W}}(\mathbf{i}) \tag{25}$$

for all  $\mathbf{i} \in \Sigma_N^*$  and all transitivity classes  $\mathcal{W}$ . It follows directly that

$$P(\Phi) = P(\Phi^{\mathcal{W}_0}) \leq P(\Phi^{\text{det}}) \leq P(\Phi^{\mathcal{W}}) \leq P(\Phi)$$

for all transitivity classes  $\mathcal{W}$ , and we have proved the identity (23):  $P(\Phi) = P(\Phi^{\mathcal{W}}) = P(\Phi^{\text{det}})$  for all transitivity classes  $\mathcal{W}$ .

We may now prove that  $\mu$  is the equilibrium state of  $\Phi^{\mathcal{W}}$  for every transitivity class  $\mathcal{W}$ , and is also the equilibrium state of  $\Phi^{\text{det}}$ . Indeed, for each transitivity class  $\mathcal{W}$  inequality (25) yields

$$\Lambda(\Phi^{\mathcal{W}_0}, \mu) \leq \Lambda(\Phi^{\text{det}}, \mu) \leq \Lambda(\Phi^{\mathcal{W}}, \mu)$$

and therefore

$$\begin{aligned} P(\Phi) = P(\Phi^{\text{det}}) = P(\Phi^{\mathcal{W}_0}) &= h(\mu) + \Lambda(\Phi^{\mathcal{W}_0}, \mu) \leq h(\mu) + \Lambda(\Phi^{\text{det}}, \mu) \\ &\leq h(\mu) + \Lambda(\Phi^{\mathcal{W}}, \mu) \leq P(\Phi^{\mathcal{W}}) = P(\Phi), \end{aligned}$$

so that

$$P(\Phi^{\mathcal{W}}) = h(\mu) + \Lambda(\Phi^{\mathcal{W}}, \mu) \quad \text{and} \quad P(\Phi^{\text{det}}) = h(\mu) + \Lambda(\Phi^{\text{det}}, \mu)$$

as required for  $\mu$  to be an equilibrium state of  $\Phi^{\mathcal{W}}$  and  $\Phi^{\text{det}}$  respectively.

We now make further use of the Gibbs inequality. Each  $\Phi^{\mathcal{W}}$  has a unique equilibrium state and satisfies the Gibbs inequality with respect to that equilibrium state, and the equilibrium state of each such potential is  $\mu$ . The same remarks apply to  $\mu$  and the potential  $\Phi^{\text{det}}$ . Therefore there exists  $C_6 > 0$  such that

$$C_6^{-1} \leq \frac{\Phi^{\mathcal{W}}(\mathbf{i})}{e^{-|\mathbf{i}|P(\Phi^{\mathcal{W}})}\mu([\mathbf{i}])} = \frac{\Phi^{\mathcal{W}}(\mathbf{i})}{e^{-|\mathbf{i}|P(\Phi)}\mu([\mathbf{i}])} \leq C_6$$

for all  $\mathbf{i} \in \Sigma_N^*$  and all transitivity classes  $\mathcal{W}$ , and also

$$C_6^{-1} \leq \frac{\Phi^{\text{det}}(\mathbf{i})}{e^{-|\mathbf{i}|P(\Phi^{\text{det}})}\mu([\mathbf{i}])} = \frac{\Phi^{\text{det}}(\mathbf{i})}{e^{-|\mathbf{i}|P(\Phi)}\mu([\mathbf{i}])} \leq C_6$$

for all  $\mathbf{i} \in \Sigma_N^*$ . We deduce the inequality  $\Phi^{\mathcal{W}}(\mathbf{i}) \leq C_6^2 \Phi^{\text{det}}(\mathbf{i})$  for all  $\mathbf{i} \in \Sigma_N^*$  and transitivity classes  $\mathcal{W}$ .

5.3.11. *A final determinant identity.* Let  $\mathbf{i} \in \Sigma_N^*$  be such that  $g_{\mathbf{i}} \in G^o$ . We have

$$\begin{aligned} \max_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k |\det(\pi_j(g_{\mathbf{i}})|_{W_j})|^{\beta_j/\ell_j} &\leq \max_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k \|\pi_j(g_{\mathbf{i}})|_{W_j}\|^{\beta_j} = \Phi^{\mathcal{W}}(\mathbf{i}) \\ &\leq C_6^2 \Phi^{\text{det}}(\mathbf{i}) = C_6^2 \prod_{j=1}^k |\det \pi_j(g_{\mathbf{i}})|^{\beta_j/d_j} \end{aligned}$$



$$\begin{aligned}
 &= C_6^2 \left( \prod_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k |\det(\pi_j(g_i)|_{W_j})|^{\beta_j/\ell_j} \right)^{1/\#\mathcal{W}} \\
 &\leq C_6^2 \left( \min_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k |\det(\pi_j(g_i)|_{W_j})|^{\beta_j/\ell_j} \right)^{1/\#\mathcal{W}} \\
 &\quad \cdot \left( \max_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k |\det(\pi_j(g_i)|_{W_j})|^{\beta_j/\ell_j} \right)^{\#\mathcal{W}-1/\#\mathcal{W}}
 \end{aligned}$$

and we obtain

$$\max_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k |\det(\pi_j(g_i)|_{W_j})|^{\beta_j/\ell_j} \leq C_6^{2\#\mathcal{W}} \min_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k |\det(\pi_j(g_i)|_{W_j})|^{\beta_j/\ell_j}$$

for all transitivity classes  $\mathcal{W}$  and all  $g_i \in G^o$ . It follows that if  $(W'_j)_{j=1}^k$  is any element of any transitivity class  $\mathcal{W}$ , then for every  $g_i \in G^o$ ,

$$\begin{aligned}
 \prod_{j=1}^k |\det(\pi_j(g_i)|_{W'_j})|^{\beta_j/\ell_j} &\geq \min_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k |\det(\pi_j(g_i)|_{W_j})|^{\beta_j/\ell_j} \\
 &\geq C_6^{-2\#\mathcal{W}} \max_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k |\det(\pi_j(g_i)|_{W_j})|^{\beta_j/\ell_j} \\
 &\geq C_6^{-2\#\mathcal{W}} \left( \prod_{(W_j)_{j=1}^k \in \mathcal{W}} \left( \prod_{j=1}^k |\det(\pi_j(g_i)|_{W_j})|^{\beta_j/\ell_j} \right) \right)^{1/\#\mathcal{W}} \\
 &= C_6^{-2\#\mathcal{W}} \prod_{j=1}^k |\det \pi_j(g_i)|^{\beta_j/d_j}
 \end{aligned}$$

where we have used (22) again, and from the preceding chain of inequalities,

$$\begin{aligned}
 \prod_{j=1}^k |\det(\pi_j(g_i)|_{W'_j})|^{\beta_j/\ell_j} &\leq \max_{(W_j)_{j=1}^k \in \mathcal{W}} \prod_{j=1}^k |\det(\pi_j(g_i)|_{W_j})|^{\beta_j/\ell_j} \\
 &\leq C_6^2 \prod_{j=1}^k |\det \pi_j(g_i)|^{\beta_j/d_j}.
 \end{aligned}$$

We have found that if  $i \in \Sigma_N^*$  is such that  $g_i \in G^o$ ,  $\mathcal{W}$  is any transitivity class and  $(W_j)_{j=1}^k$  any element of  $\mathcal{W}$ , then

$$C_6^{-2(\#\mathcal{W})} \prod_{j=1}^k |\det \pi_j(g_i)|^{\beta_j/d_j} \leq \prod_{j=1}^k |\det(\pi_j(g_i)|_{W_j})|^{\beta_j/\ell_j} \leq C_6^2 \prod_{j=1}^k |\det \pi_j(g_i)|^{\beta_j/d_j}.$$

Applying this estimate to  $g_i^n = g_i^n$  in place of  $g_i$ , taking the power  $1/n$  and letting  $n \rightarrow \infty$  yields

$$\prod_{j=1}^k |\det \pi_j(g_i)|^{\beta_j/d_j} = \prod_{j=1}^k |\det(\pi_j(g_i)|_{W_j})|^{\beta_j/\ell_j} \tag{26}$$

for every  $g_i \in G^o$  and every  $(W_j)$  in any transitivity class.

*5.3.12. Conclusion of the proof.* Equation (26) suffices to get (20). Fix  $j_0 \in \{1, \dots, k\}$  and  $1 \leq i_1, i_2 \leq n_{j_0}$ . Let  $W_{j_0} := U_{j_0}^{i_1}$  and  $W'_{j_0} := U_{j_0}^{i_2}$ , and for  $j \neq j_0$ , set  $W_j := U_j^1$  and  $W'_j := U_j^1$ . Applying (26) gives

$$\frac{|\det(\pi_{j_0}(g_i)|_{U_{j_0}^{i_1}})^{\beta_{j_0}/\ell_{j_0}}}{|\det(\pi_{j_0}(g_i)|_{U_{j_0}^{i_2}})^{\beta_{j_0}/\ell_{j_0}}} = \frac{\prod_{j=1}^k |\det(\pi_j(g_i)|_{W_j})|^{\beta_j/\ell_j}}{\prod_{j=1}^k |\det(\pi_j(g_i)|_{W'_j})|^{\beta_j/\ell_j}} = \frac{\prod_{j=1}^k |\det \pi_j(g_i)|^{\beta_j/d_j}}{\prod_{j=1}^k |\det \pi_j(g_i)|^{\beta_j/d_j}} = 1$$

for every  $g_i \in G^o$ . Hence for every  $g_i \in G^o$  and every  $j \in \{1, \dots, k\}$ ,

$$|\det(\pi_j(g_i)|_{U_j^i})|^{1/\ell_j}$$

is independent of  $i \in \{1, \dots, n_j\}$  and in particular must be equal to its geometric mean with respect to  $i \in \{1, \dots, n_j\}$ , which by (21) is  $|\det \pi_j(g_i)|^{1/d_j}$ . This establishes (20), which in turn allows us to readily conclude. Indeed, together with (19), it implies that for every  $g \in G^o$  and  $j \in \{1, \dots, k\}$ ,  $\pi_j(g) = |\det \pi_j(g)|^{1/d_j} O_j(g)$  where  $O_j(g) \in O(V_j)$  for some Euclidean structure on  $V_j$  not depending on  $g$ . Therefore

$$\{|\det \pi_j(g)|^{-1/d_j} \pi_j(g) : g \in G^o\}$$

is a compact subgroup of  $GL(V_j)$  and since the index  $[G : G^o]$  is finite, the same is true of

$$\{|\det \pi_j(g)|^{-1/d_j} \pi_j(g) : g \in G\}.$$

The proof is complete.

### 6. Proof of Theorem 5

Let  $(A_1, \dots, A_N) \in GL_d(\mathbb{R})^N$  be irreducible and let  $\alpha_1 \geq \dots \geq \alpha_d \geq 0$  with  $\alpha_1 > \alpha_d$ . Let  $G \leq GL_d(\mathbb{R})$  denote the Zariski closure of the subsemigroup of  $GL_d(\mathbb{R})$  generated by  $A_1, \dots, A_N$ ; it is a real reductive group (§4.1.1). Define  $\alpha_{d+1} := 0$  and let  $k_1, \dots, k_r$  be the list of all integers  $i \in \{1, \dots, d\}$  for which the difference  $\alpha_i - \alpha_{i+1}$  is positive, where  $k_1 < \dots < k_r$ . We observe that since  $\alpha_1 > \alpha_d$  we have  $r \neq 0$  and also  $k_1 < d$ . Define  $\beta_j := \alpha_{k_j} - \alpha_{1+k_j} > 0$  for each  $j = 1, \dots, r$ , and for each  $j = 1, \dots, r$  let  $\pi_j : G \rightarrow GL(\wedge^{k_j} \mathbb{R}^d)$  denote the exterior power representation  $\pi_j(g) := g^{\wedge k_j}$ . We have

$$\prod_{j=1}^d \sigma_j(g)^{\alpha_j} = \prod_{j=1}^d \left( \prod_{i=1}^j \sigma_i(g) \right)^{\alpha_j - \alpha_{j+1}} = \prod_{j=1}^d \|g^{\wedge j}\|^{\alpha_j - \alpha_{j+1}} = \prod_{j=1}^r \|\pi_j(g)\|^{\beta_j}$$

for every  $g \in G$ , and in particular the potential  $\Phi$  defined in the statement of the theorem satisfies the description

$$\Phi(\mathbf{i}) = \prod_{j=1}^r \|\pi_j(A_{\mathbf{i}})\|^{\beta_j}.$$

Since the representations  $\pi_j: G \rightarrow \text{GL}(\bigwedge^{k_j} \mathbb{R}^d)$  are not in general irreducible, Theorem 8 is not directly applicable to the potential  $\Phi$ . We will study  $\Phi$  by writing it as the maximum of a finite collection of simpler potentials to which Theorem 8 may be applied. Since  $G$  is reductive, the rational representations  $\pi_j$  are completely reducible (§4.1.1), in other words, for each  $j = 1, \dots, r$  we may write  $\bigwedge^{k_j} \mathbb{R}^d = V_1^j \oplus \dots \oplus V_{n_j}^j$  where each  $V_i^j$  is an invariant subspace of the group  $\pi_j(G)$  on which  $\pi_j(G)$  acts irreducibly. For each  $j = 1, \dots, r$  and  $1 \leq \ell \leq n_j$  define an irreducible representation  $\pi_{j,\ell}: G \rightarrow \text{GL}(V_\ell^j)$  by  $\pi_{j,\ell}(g) := \pi_j(g)|_{V_\ell^j}$  for all  $g \in G$ . Let  $\mathcal{L}$  denote the set of all tuples of integers  $\mathbf{l} = (\ell_1, \dots, \ell_r)$  such that  $1 \leq \ell_j \leq n_j$  for each  $j = 1, \dots, r$ . For each  $\mathbf{l} = (\ell_1, \dots, \ell_r) \in \mathcal{L}$  define a potential  $\Phi_{\mathbf{l}}: \Sigma_N^* \rightarrow (0, +\infty)$  by

$$\Phi_{\mathbf{l}}(\mathbf{i}) := \prod_{j=1}^r \|\pi_j(A_{\mathbf{i}})|_{V_{\ell_j}^j}\|^{\beta_j} = \prod_{j=1}^r \|\pi_{j,\ell_j}(A_{\mathbf{i}})\|^{\beta_j}. \tag{27}$$

For each fixed  $\mathbf{l} = (\ell_1, \dots, \ell_r)$  the representations  $\pi_{j,\ell_j}$  for  $j = 1, \dots, r$  are irreducible, so each  $\Phi_{\mathbf{l}}$  satisfies the hypotheses of Theorem 8. Clearly we also have

$$\begin{aligned} \Phi(\mathbf{i}) &= \prod_{j=1}^r \|\pi_j(A_{\mathbf{i}})\|^{\beta_j} = \prod_{j=1}^r \max_{1 \leq \ell \leq n_j} \|\pi_j(A_{\mathbf{i}})|_{V_\ell^j}\|^{\beta_j} \\ &= \max_{(\ell_1, \dots, \ell_r) \in \mathcal{L}} \prod_{j=1}^r \|\pi_j(A_{\mathbf{i}})|_{V_{\ell_j}^j}\|^{\beta_j} = \max_{\mathbf{l} \in \mathcal{L}} \Phi_{\mathbf{l}}(\mathbf{i}) \end{aligned} \tag{28}$$

for every  $\mathbf{i} \in \Sigma_N^*$ . We will find it helpful to define further potentials as follows. For each  $\mathbf{l} = (\ell_1, \dots, \ell_r) \in \mathcal{L}$  define

$$\Phi_{\mathbf{l}}^{\text{det}}(\mathbf{i}) := \prod_{j=1}^r |\det \pi_{j,\ell_j}(A_{\mathbf{i}})|^{\beta_j / \dim V_{\ell_j}^j} \quad \text{for all } \mathbf{i} \in \Sigma_N^*. \tag{29}$$

Define also

$$\Phi^{\text{det}}(\mathbf{i}) = \prod_{j=1}^r |\det A_{\mathbf{i}}|^{k_j \beta_j / d} \quad \text{for all } \mathbf{i} \in \Sigma_N^*.$$

Our strategy in proving Theorem 5 will be to establish the identity

$$P(\Phi_{\mathbf{l}}) = P(\Phi_{\mathbf{l}}^{\text{det}}) \tag{30}$$

for all  $\mathbf{l} \in \mathcal{L}$ . This will permit the implication (ii)  $\Rightarrow$  (iii) of Theorem 8 to be applied, establishing that each of the groups  $\pi_{j,\ell}(G)$  is compact modulo factoring out the determinant. The compactness of each  $\pi_j(G)$  modulo factoring out the determinant will then follow

via some additional bookkeeping to ensure that for each  $j = 1, \dots, r$  the determinant which is factored out of the representation  $\pi_{j,\ell}$  is consistent across all  $\ell \in \{1, \dots, n_j\}$ , and the compactness of  $G$  modulo factoring out the determinant will follow by some simple manipulations involving singular values.

Much as in the second half of the proof of Theorem 8, before commencing the proof of (30) we must first establish an identity involving determinants. The proof of this identity is relatively long and comprises a large proportion of this section. Specifically, we make the following claim: for every  $j = 1, \dots, r$ , for all  $\ell = 1, \dots, n_j$  we have

$$|\det \pi_{j,\ell}(g)|^{1/\dim V_\ell^j} = |\det g|^{k_j/d} \tag{31}$$

for all  $g \in G$ .

To prove the claim it is sufficient to establish (31) for all  $g \in G^\circ$ , since if this has been proven then for any given  $g \in G$  we have  $g^n \in G^\circ$  for some integer  $n \geq 1$  and hence clearly

$$|\det \pi_{j,\ell}(g)|^{1/\dim V_\ell^j} = |\det \pi_{j,\ell}(g^n)|^{1/(n \cdot \dim V_\ell^j)} = |\det g^n|^{k_j/(nd)} = |\det g|^{k_j/d}$$

as required. We therefore restrict our task to proving (31) for all  $g \in G^\circ$ . To this end let us fix  $j$  and  $\ell$  and define a continuous group homomorphism  $\hat{\pi}$  from  $G^\circ$  to the multiplicative group of positive real numbers by  $\hat{\pi}(g) := |\det \pi_{j,\ell}(g)|^{1/(k_j \cdot \dim V_\ell^j)}$ . Our objective is now to show that  $\hat{\pi}(g) = |\det g|^{1/d}$  for all  $g \in G^\circ$ . The set of all  $g \in G^\circ$  satisfying this equation is obviously a group, and this set obviously includes  $[G^\circ, G^\circ]$  since by the commutativity of real multiplication we have  $\hat{\pi}(g) = 1 = |\det g|^{1/d}$  for all  $g \in [G^\circ, G^\circ]$ . Since  $G^\circ$  is equal to an almost direct product of  $Z(G^\circ)$  and  $[G^\circ, G^\circ]$ , the claim will therefore follow if we can prove that  $\hat{\pi}(z) = |\det z|^{1/d}$  for all  $z \in Z(G^\circ)$ .

We begin by analysing the action of  $Z(G^\circ)$  on  $\mathbb{R}^d$ . By Clifford’s [57, Theorem 1.7] applied to the irreducible group  $G \leq \text{GL}_d(\mathbb{R})$  and its normal subgroup  $G^\circ$ , we obtain a direct sum decomposition  $\mathbb{R}^d = X_1 \oplus \dots \oplus X_p$  consisting of the homogeneous (isotypic) components of the  $G^\circ$ -representation. By the same result of Clifford the subspaces  $X_i$  are permuted by the component group  $G/G^\circ$ . In particular, they all have the same dimension, which we denote by  $m \in \mathbb{N}$ . (Here of course we have  $m = d/p$ .) Decomposing each  $X_i$  into a sum of irreducible subspaces for the  $G^\circ$ -action and using Schur’s lemma, we deduce that there exists an inner product structure on each  $X_i$  with respect to which every  $g \in Z(G^\circ)$  acts on  $X_i$  by a similarity transformation.

Now, by [14, Proposition 8.15] there exist a maximal compact subgroup  $Z(G^\circ)_A$  and a maximal real diagonalisable subgroup  $Z(G^\circ)_D$  of  $Z(G^\circ)$  such that  $Z(G^\circ)_A \cap Z(G^\circ)_D$  is finite and  $Z(G^\circ) = Z(G^\circ)_D Z(G^\circ)_A$ , which is to say  $Z(G^\circ)$  is an almost direct product of the subgroups  $Z(G^\circ)_D$  and  $Z(G^\circ)_A$ . The group  $\hat{\pi}(Z(G^\circ)_A)$  is a compact subgroup of the positive reals and hence is equal to  $\{1\}$ , and similarly the image of  $Z(G^\circ)_A$  under the homomorphism  $z \mapsto |\det z|^{1/d}$  must also equal  $\{1\}$ , so we have  $\hat{\pi}(z) = |\det z|^{1/d}$  for all  $z \in Z(G^\circ)_A$ . Hence the claim will be proved if we can show that  $\hat{\pi}(z) = |\det z|^{1/d}$  for every  $z \in Z(G^\circ)_D$ .

Since each  $X_i$  is a sum of isomorphic irreducible representations of  $G^o$ , it follows from Schur’s lemma and (real) diagonalisability that every  $z \in Z(G^o)_D$  acts on each  $X_i$  by a scalar transformation  $v \mapsto \gamma_i(z)v$  for some nonzero real number  $\gamma_i(z)$  for  $i = 1, \dots, p$ . On the other hand, since  $V_\ell^j$  is  $Z(G^o)_D$ -invariant and  $Z(G^o)_D$  is abelian,  $V_\ell^j$  can be written as a direct sum of  $Z(G^o)_D$ -irreducible subspaces in  $\bigwedge^{k_j} \mathbb{R}^d$ . But  $Z(G^o)_D$  is also a split torus, and therefore so is its image in the exterior power representations. Hence, these  $Z(G^o)_D$ -irreducible subspaces of  $V_\ell^j$  are 1-dimensional. Each gives rise to a character of  $Z(G^o)_D$  of the form  $\gamma_1(z)^{t_1} \cdots \gamma_p(z)^{t_p}$  for some nonnegative integers  $t_1, \dots, t_p$  whose sum is equal to  $k_j$ . The quantity  $\det \pi_{j,\ell}(z) = \det z^{\wedge k_j} |_{V_\ell^j}$  is a product of precisely  $\dim V_\ell^j$  such characters, so it has the form  $\gamma_1(z)^{t'_1} \cdots \gamma_p(z)^{t'_p}$  for some nonnegative integers  $t'_1, \dots, t'_p$  such that  $\sum_{i=1}^p t'_i = k_j \cdot \dim V_\ell^j$ . Taking the absolute value and raising to the power  $1/(k_j \cdot \dim V_\ell^j)$  as in the definition of  $\hat{\pi}$ , it follows that there exist nonnegative rational numbers  $r_1, \dots, r_p$  such that  $\hat{\pi}(z) = |\gamma_1(z)|^{r_1} \cdots |\gamma_p(z)|^{r_p}$  for all  $z \in Z(G^o)_D$  and such that  $\sum_{i=1}^p r_i = 1$ . On the other hand, clearly  $\det z = \gamma_1(z)^m \cdots \gamma_p(z)^m$  for every  $z \in Z(G^o)_D$  since  $\mathbb{R}^d = \bigoplus_{i=1}^p X_i$  and  $\det(z|_{X_i}) = \gamma_i(z)^m$  for every  $i = 1, \dots, p$ , where we recall that  $m = d/p$  is the dimension of each of the spaces  $X_i$ . Hence  $|\det z|^{1/d} = |\gamma_1(z) \cdots \gamma_p(z)|^{1/p}$  for all  $z \in Z(G^o)_D$ .

Now, if  $z \in Z(G^o)_D$  and  $g \in G$  then  $gzg^{-1}$  also belongs to  $Z(G^o)$  and also acts on each  $X_i$  by a scalar transformation, which by the maximality of  $Z(G^o)_D$  as a real diagonalisable subgroup of  $Z(G^o)$  implies  $gzg^{-1} \in Z(G^o)_D$ . For every  $[g] \in G/G^o$  there exists a permutation  $\zeta$  of  $\{1, \dots, p\}$  such that  $gX_i = X_{\zeta(i)}$  for every  $i = 1, \dots, p$  and every  $g \in [g]$ , and the corresponding element  $gzg^{-1}$  of  $Z(G^o)_D$  satisfies  $\gamma_i(gzg^{-1}) = \gamma_{\zeta(i)}(z)$  for all  $i = 1, \dots, p$ . For each  $i \in \{1, \dots, p\}$  the transitivity of the action of  $G/G^o$  on  $X_1, \dots, X_p$  implies that the sets  $\{[g] \in G/G^o : gX_i = X_j\}$  for  $j = 1, \dots, p$  form a partition of  $G/G^o$  into cosets of equal cardinality  $(\#G/G^o)/p$  and therefore

$$\prod_{[g] \in G/G^o} |\gamma_i(gzg^{-1})| = \prod_{j=1}^p \left( \prod_{\substack{[g] \in G/G^o \\ gX_i = X_j}} |\gamma_j(z)| \right) = \left( \prod_{j=1}^p |\gamma_j(z)| \right)^{\frac{\#G/G^o}{p}} = |\det z|^{\frac{\#G/G^o}{d}} \tag{32}$$

for each  $i = 1, \dots, p$  and  $z \in Z(G^o)_D$ . We obviously have  $\hat{\pi}(gzg^{-1}) = \hat{\pi}(z)$  for every  $z \in Z(G^o)_D$  and  $g \in G$  by the commutativity of real multiplication. Hence for every  $z \in Z(G^o)_D$ ,

$$\begin{aligned} \hat{\pi}(z) &= \left( \prod_{[g] \in G/G^o} \hat{\pi}(gzg^{-1}) \right)^{\frac{1}{\#G/G^o}} = \left( \prod_{[g] \in G/G^o} \prod_{i=1}^p |\gamma_i(gzg^{-1})|^{r_i} \right)^{\frac{1}{\#G/G^o}} \\ &= \prod_{i=1}^p \left( \prod_{[g] \in G/G^o} |\gamma_i(gzg^{-1})| \right)^{\frac{r_i}{\#G/G^o}} = \prod_{i=1}^p |\det z|^{r_i/d} = |\det z|^{1/d} \end{aligned}$$

where we have used (32) and the equation  $r_1 + \dots + r_p = 1$ . We have obtained  $\hat{\pi}(z) = |\det z|^{1/d}$  for all  $z \in Z(G^o)_D$  and we deduce that the claimed identity (31) is valid for every  $g \in G$  as required.

We may now return to the main direction of the proof. Our first step towards the desired identity (30) is to observe that

$$P(\Phi) \geq \max_{\mathbb{I} \in \mathfrak{L}} P(\Phi_{\mathbb{I}})$$

as a direct consequence of (28) together with the definition of the pressure. Furthermore, for each  $\mathbb{I} \in \mathfrak{L}$  we have  $\Phi_{\mathbb{I}}(i) \geq \Phi_{\mathbb{I}}^{\det}(i)$  for all  $i \in \Sigma_N^*$ . This follows by comparing (27) and (29) and using the elementary inequality  $|\det B| \leq \|B\|^{\dim V}$  for all  $B \in GL(V)$ , and it entails that  $P(\Phi_{\mathbb{I}}) \geq P(\Phi_{\mathbb{I}}^{\det})$  for every  $\mathbb{I} \in \mathfrak{L}$ . We have thus far obtained

$$P(\Phi) \geq P(\Phi_{\mathbb{I}}) \geq P(\Phi_{\mathbb{I}}^{\det}) \quad \text{for every } \mathbb{I} \in \mathfrak{L}. \tag{33}$$

Using the identity (31), we immediately deduce that  $\Phi_{\mathbb{I}}^{\det}(i) = \Phi^{\det}(i)$  for all  $i \in \Sigma_N^*$  simply by applying (31) to the definition of the two potentials. Combining this observation with (33) it follows that

$$P(\Phi) \geq P(\Phi_{\mathbb{I}}) \geq P(\Phi_{\mathbb{I}}^{\det}) = P(\Phi^{\det}) \quad \text{for every } \mathbb{I} \in \mathfrak{L}. \tag{34}$$

Let us now show that  $P(\Phi) = P(\Phi^{\det})$ . By hypothesis there exists a Bernoulli measure  $\mu$  which satisfies  $h(\mu) + \Lambda(\Phi, \mu) = P(\Phi)$ . Since  $\mu$  is Bernoulli, it is ergodic, so by the subadditive ergodic theorem we have, for  $\mu$ -a.e.  $x \in \Sigma_N$ ,

$$\begin{aligned} \Lambda(\Phi, \mu) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \Phi(x|_n) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \max_{\mathbb{I} \in \mathfrak{L}} \Phi_{\mathbb{I}}(x|_n) \\ &= \max_{\mathbb{I} \in \mathfrak{L}} \lim_{n \rightarrow \infty} \frac{1}{n} \log \Phi_{\mathbb{I}}(x|_n) = \max_{\mathbb{I} \in \mathfrak{L}} \Lambda(\Phi_{\mathbb{I}}, \mu). \end{aligned}$$

Thus  $P(\Phi) = h(\mu) + \Lambda(\Phi_{\mathbb{I}_0}, \mu)$  for some particular  $\mathbb{I}_0 \in \mathfrak{L}$ , and therefore

$$P(\Phi) \geq \max_{\mathbb{I} \in \mathfrak{L}} P(\Phi_{\mathbb{I}}) \geq P(\Phi_{\mathbb{I}_0}) \geq h(\mu) + \Lambda(\Phi_{\mathbb{I}_0}, \mu) = P(\Phi)$$

where we have used the subadditive variational principle in the third inequality. We conclude that  $P(\Phi) = P(\Phi_{\mathbb{I}_0})$  and that  $\mu$  is an equilibrium state of  $\Phi_{\mathbb{I}_0}$ . By Theorem 8 applied to the potential  $\Phi_{\mathbb{I}_0}$  we have  $P(\Phi_{\mathbb{I}_0}) = P(\Phi_{\mathbb{I}_0}^{\det})$ . We have already seen that  $\Phi_{\mathbb{I}_0}^{\det}$  is identically equal to  $\Phi^{\det}$ , so

$$P(\Phi^{\det}) = P(\Phi_{\mathbb{I}_0}^{\det}) = P(\Phi_{\mathbb{I}_0}) = P(\Phi) \geq P(\Phi_{\mathbb{I}}) \geq P(\Phi_{\mathbb{I}}^{\det}) = P(\Phi^{\det})$$

for every  $\mathbb{I} \in \mathfrak{L}$ , where we have invoked (34).

We have now established the desired identity

$$P(\Phi_{\mathbb{I}}) = P(\Phi_{\mathbb{I}}^{\det}) \quad \text{for every } \mathbb{I} \in \mathfrak{L}.$$

Since every  $\Phi_{\mathbb{I}}$  satisfies the hypotheses of Theorem 8 it follows from the implication (ii) $\Rightarrow$ (iii) of that theorem that for each  $\mathbb{I} = (\ell_1, \dots, \ell_r) \in \mathfrak{L}$ , for every  $j = 1, \dots, r$  the group

$$\begin{aligned} \{ |\det(\pi_{j, \ell_j}(g))|^{-1/\dim V_{\ell_j}^j} \pi_{j, \ell_j}(g) : g \in G \} &= \{ |\det g|^{-k_j/d} g^{\wedge k_j} |_{V_{\ell_j}^j} : g \in G \} \\ &= \{ (|\det g|^{-1/d} g)^{\wedge k_j} |_{V_{\ell_j}^j} : g \in G \} \end{aligned}$$

is compact, where we have again used (31). Since  $\mathcal{I}$  is arbitrary we deduce that the group

$$\{(|\det g|^{-1/d} g)^{\wedge k_j} : g \in G\}$$

is compact for every  $j = 1, \dots, r$ . In particular, it is compact for  $j = 1$ , so there exists  $K > 0$  such that for every  $g \in G$  we have  $\|(|\det g|^{-1/d} g)^{\wedge k_1}\| \leq K$ .

Let  $g \in G$  and define  $h := |\det g|^{-1/d} g$ . We observed at the beginning of the proof that  $k_1 < d$ . Since  $1 = |\det h| = \sigma_1(h) \cdots \sigma_d(h)$  we have

$$\begin{aligned} \|h\| &= \sigma_1(h) = \sigma_2(h)^{-1} \cdots \sigma_d(h)^{-1} = \sigma_1(h^{-1}) \cdots \sigma_{d-1}(h^{-1}) \\ &\leq (\sigma_1(h^{-1}) \cdots \sigma_{k_1}(h^{-1}))^{d-1/k_1} = \|(h^{-1})^{\wedge k_1}\|^{d-1/k_1} \leq K^{d-1/k_1} \end{aligned}$$

where we have used  $k_1 \leq d - 1$  in order to pass from the first line to the second. The same reasoning obviously applies to  $h^{-1}$ , and we conclude that the group

$$\{|\det g|^{-1/d} g : g \in G\} \leq \text{GL}_d(\mathbb{R})$$

is contained in the compact set

$$\{h \in \text{GL}_d(\mathbb{R}) : \max\{\|h\|, \|h^{-1}\|\} \leq K^{d-1/k_1}\}$$

and hence is compact. Since obviously that group contains all of the linear maps  $|\det A_i|^{-1/d} A_i$ , the theorem is proved.

## Appendix A. Equilibrium states of linear Cartan potentials

Our main technical results, Theorem 5 and 8, admit a counterpart that can be expressed in more intrinsic terms (with respect to the linear algebraic group given by the Zariski closure of the semigroup generated by the linear parts of the iterated function system). Beyond its relative elegance, as we shall see, this formulation will allow us to have an understanding of the structure of equilibrium states of matrix potentials from a representation-theoretic perspective. To avoid further technicalities we will restrict our considerations here to Zariski connected real reductive groups (e.g.  $\text{GL}_d(\mathbb{R})$ ).

We retain the notation used in §4 where we described some preliminary facts concerning real reductive groups. We will require some additional facts about the representation theory of those groups, further to the exposition in §4. As before we refer the reader to [12, 14, 15, 18, 41] for a more detailed exposition of this theory. For a focused account containing all the material which we will require we suggest [30, Sections 2 & 3].

Let  $G$  be a Zariski connected real reductive group. Let us say that it is *of noncompact type* if the derived group  $[G, G]$  is not compact. Let  $K$  be a maximal compact subgroup with Lie algebra  $\mathfrak{k}$  and  $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{k}^\perp$  the orthogonal decomposition of  $\mathfrak{g}$  with respect to the Killing form. Let  $\alpha < \mathfrak{k}^\perp$  be a Cartan subspace of  $\mathfrak{g}$  and  $\alpha = \alpha_S \oplus \alpha_Z$  its decomposition into the semisimple and central parts (as defined in §4). Let  $\Sigma \subset \alpha^*$  denote the nonzero (restricted) roots, and  $\Delta := \{\bar{\alpha}_1, \dots, \bar{\alpha}_{d_S}\} \subset \Sigma$  a choice of simple roots, where  $d_S \in \mathbb{N}$

is the semisimple real rank of  $G$ . We denote by  $\alpha_S^+$  the salient cone given by  $\alpha_S \cap \alpha^+$ . We fix an inner product  $(\cdot, \cdot)$  on  $\alpha$  extending the restriction of the Killing form to  $\alpha_S$  and satisfying  $\alpha_S^\perp = \alpha_Z$ . This induces an identification of  $\alpha$  with  $\alpha^*$  and we use the same notation to denote the corresponding inner product on  $\alpha^*$ . We denote by  $\mathcal{W}$  the set of restricted weights of  $G$ , which is equal to

$$\left\{ \omega \in \alpha^* : 2 \frac{(\omega, \alpha)}{(\alpha, \alpha)} \in \mathbb{Z} \ \forall \alpha \in \Sigma \right\}.$$

For a subspace  $V \leq \alpha$ , denote by  $V^0$  its annihilator subspace in  $\alpha^*$ . For each  $\alpha \in \Delta$ , we denote by  $\omega_\alpha$  or  $\omega_i$  (when  $\alpha = \alpha_i$ ) the corresponding fundamental weight, i.e. the weight in  $\mathcal{W} \cap \alpha_Z^0$  satisfying  $2 \frac{(\omega, \alpha)}{(\alpha, \alpha)} = \delta_{\alpha, \beta}$  for every  $\beta \in \Delta$ . The set  $\{\omega_i\}$  is a basis of  $\alpha_Z^0$ . By choosing linearly independent  $\omega_{d_S+1}, \dots, \omega_d \in \alpha_S^0$  (where  $d \in \mathbb{N}$  is the real rank of  $G$ ) complete  $\{\omega_i : i = 1, \dots, d_S\}$  to a basis of  $\alpha^*$ . Given  $\phi \in \alpha^*$  with  $\phi = \sum_{i=1}^d c_i \omega_i$ , write  $\phi_S = \sum_{i=1}^{d_S} c_i \omega_i$  and  $\phi_Z = \phi - \phi_S$ . Let  $\mathcal{C}$  be the cone in  $\alpha^*$  defined by

$$\mathcal{C} := \left\{ \phi \in \alpha^* : \phi_S \neq 0 \text{ and } \phi_S \text{ has only nonnegative coefficients in the basis } \{\omega_i\} \right\}.$$

Finally, as before, we denote by  $\kappa : G \rightarrow \alpha^+$  the associated Cartan projection. Theorem 5 admits the following abstract articulation:

**Theorem 9.** *Let  $G$  be a Zariski connected real reductive group of noncompact type and  $(g_1, \dots, g_N)$  be a tuple of elements such that the semigroup generated by  $\{g_1, \dots, g_N\}$  is Zariski dense in  $G$ . Let  $\Psi$  be a finite subset of  $\mathcal{C}$ . Then the potential  $\Phi^\Psi$  defined by*

$$\Phi^\Psi(\mathbf{i}) = \max_{\phi \in \Psi} e^{\phi(\kappa(g_{\mathbf{i}}))}$$

*does not have a Bernoulli equilibrium state.*

The proof of Theorem 9 follows a closely analogous path to the first half of the proof of the implication (i) $\Rightarrow$ (iii) in Theorem 8 (corresponding to §5.3.1–§5.3.5 in the main text). We therefore give the proof in outline only. The proof starts by showing that the potential  $\Phi^\Psi$  is indeed submultiplicative, justifying the use of the terminology of subadditive thermodynamic formalism in the statement above.

*Proof of Theorem 9. Step 0 (Submultiplicativity of  $\Phi^\Psi$ ).* It clearly suffices to show that given any  $\phi \in \mathcal{C}$ , the potential  $\Phi^\phi$  defined by  $\Phi^\phi(\mathbf{i}) = e^{\phi(\kappa(g_{\mathbf{i}}))}$  is submultiplicative. Since for every  $\omega \in \alpha_S^0$  and  $g, h \in G$ , we have  $\omega(\kappa(gh)) = \omega(\kappa(g)) + \omega(\kappa(h))$ ,  $\Phi^{\phi_Z}$  is a multiplicative potential. Therefore, it suffices to show that  $\Phi^{\phi_S}$  is submultiplicative. By [12, Lemma 8.15], for every  $i = 1, \dots, d_S$ , there exists a rational irreducible proximal representation  $(\rho_i, V_i)$  such that the highest weight  $\chi_i$  of  $\rho_i$  is a (positive integer) multiple of the fundamental weight  $\omega_i$ . It follows by [12, Lemma 8.17] that for each  $i = 1, \dots, d_S$ , we can choose an inner product norm  $\|\cdot\|_i$  on  $V_i$  such  $\omega_i(\kappa(g)) = \log \|\rho_i(g)\|_i$ , where we also denote by  $\|\cdot\|_i$  the associated operator norm. Since the coefficients of  $\phi_S$  in the basis  $\{\omega_i\}$  are nonnegative, it follows by submultiplicativity of the operator norms  $\|\cdot\|_i$  that the potential defined by  $\mathbf{i} \mapsto e^{\phi_S(\kappa(g_{\mathbf{i}}))}$  is submultiplicative as desired.



*Step 1* (Specialising to a maximal linear form, cf. §5.3.2). Arguing by contradiction, we suppose that the submultiplicative potential  $\Phi^\Psi$  has a Bernoulli equilibrium state  $\mu$ . It follows by the same argument as in §5.3.2 that there exists  $\phi \in \Psi$  such that  $\mu$  is an equilibrium state for the submultiplicative potential  $\Phi^\phi$  defined by  $\Phi^\phi(i) = e^{\phi(\kappa(g_i))}$ .

*Step 2* (Obtaining quasi-multiplicativity, cf. §5.3.3). Using Quint’s [48, Proposition I.2],<sup>4</sup> which is based on the aforementioned representation-theoretic ingredients [12, Lemmas 8.15 & 8.17] and the main result of Abels–Margulis–Soifer [1], we deduce that there exists a finite set  $F$  in the semigroup  $\Gamma$  generated by  $\{g_1, \dots, g_N\}$  and a constant  $K > 0$  such that for all  $g, h \in \Gamma$ , there exists  $f \in F$  satisfying

$$\|\kappa(gfh) - \kappa(g) - \kappa(h)\| \leq K.$$

It immediately follows that the potential  $\Phi^\phi$  is quasi-multiplicative in the sense of (5) — the aforementioned result of Quint is a predecessor of [13, Theorem 6] and indeed in the proof of Theorem 5, this step is analogous to where we use [13, Theorem 6] to obtain quasi-multiplicativity.

*Step 3* (Thermodynamic ingredients, cf. §5.3.3). It now follows from Proposition 2.2 that the Bernoulli measure  $\mu$  is the unique equilibrium state of the potential  $\Phi^\phi$  and satisfies the relevant Gibbs inequality.

*Step 4* (From the Gibbs inequality to the additivity of the Jordan projection, cf. §5.3.4). Denote by  $\lambda : G \rightarrow \alpha^+$  the Jordan projection (see §4.1.3). We possess the following “Gelfand formula” [12, Remark 8.7] relating the Cartan  $\kappa$  and Jordan  $\lambda$  projections: for every  $g \in G$ , we have

$$\frac{1}{n}\kappa(g^n) \rightarrow \lambda(g). \quad (35)$$

Using the same reasoning as in §5.3.4, replacing the usual Gelfand formula with (35), we find that for all  $g, h \in \Gamma$  we have

$$\phi(\lambda(gh)) = \phi(\lambda(g)) + \phi(\lambda(h)). \quad (36)$$

*Step 5* (Applying Benoist’s nonarithmeticity, cf. §5.3.5). In view of (36), the subspace  $\ker \phi$  contains the set  $\{\lambda(gh) - \lambda(g) - \lambda(h) : g, h \in \Gamma\}$ . One therefore deduces from Benoist’s Theorem 7 that  $\ker \phi$  contains  $\alpha_S$ . This contradicts the assumption that  $\phi_S \neq 0$  (or equivalently that  $\phi \notin \alpha_S^0$ ) and finishes the proof. ■

**Remark A.1** (On the set of equilibrium states of  $\Phi^\Psi$ ). 1. Since the potential  $\Phi^\Psi$  is not quasi-multiplicative in general, it is not guaranteed that it possesses a unique equilibrium state. Indeed, under the assumptions of the previous theorem it was shown in [45] that a potential of the form  $\Phi^\Psi$  can have several distinct equilibrium states. On the other hand,

---

<sup>4</sup>Namely, the first property of the “produit générique” which does not use the discreteness assumption.

the number of ergodic equilibrium states of  $\Phi^\Psi$  is always finite, as follows from the main result of [13].

2. Steps 1 and 2 of the above proof imply that the number of ergodic equilibrium states of  $\Phi^\Psi$  is bounded above by the cardinality of  $\Psi$ . In fact, these two steps can be seen as part of the proof of the main result of [13]. Indeed, the latter is proved by additionally applying a reductivisation argument (cf. [38, Proposition 6.2]) and dealing with nonconnectivity (as in §5.3.3).

3. Suppose now that  $G$  is, moreover, semisimple. For a salient cone  $C$  of non-empty interior in  $\mathfrak{a}$ , consider the partial order  $\leq_C$  on  $\mathfrak{a}^*$  defined as  $\ell \leq_C \ell'$  if and only if  $\ell(x) \leq \ell'(x)$  for every  $x \in C$ . Let  $\mathcal{BC}(\Gamma) \subseteq \mathfrak{a}^+$  be the Benoist limit cone (see [10]) of the semigroup  $\Gamma$  generated by the IFS. Then it is not hard to see that the number of ergodic equilibrium states of  $\Phi^\Psi$  is bounded above by the cardinality of the set of maximal elements in  $\Psi$  for the partial order induced by  $\mathcal{BC}(\Gamma)$ .

*Acknowledgements.* The authors are grateful to several anonymous referees for numerous helpful remarks and bibliographical suggestions, and to one referee in particular for suggesting the corollaries of this work which are described in the appendix. They also thank Emmanuel Breuillard for helpful conversations.

I. D. Morris wishes to thank Roger Tribe (who was his tutor at the University of Warwick from 1999 to 2001) for suggesting that he follow Jonathan Munn's lecture course on Lie groups in the 2000-01 academic year. This was excellent advice which he still regrets not having followed at the time.

*Funding.* The research of I. D. Morris was partially supported by the Leverhulme Trust (Research Project Grant RPG-2016-194). During the realisation of this project, Ç. Sert was supported by SNF grants 182089, 178958 and 193481.

## References

- [1] Abels, H., Margulis, G. A., Soifer, G. A.: Semigroups containing proximal linear maps. *Israel J. Math.* **91**, 1–30 (1995) Zbl [0845.22004](#) MR [1348303](#)
- [2] Atkins, R., Barnsley, M. F., Vince, A., Wilson, D. C.: A characterization of hyperbolic affine iterated function systems. *Topology Proc.* **36**, 189–211 (2010) Zbl [1196.28011](#) MR [2609349](#)
- [3] Baker, S., Fraser, J. M., Máthé, A.: Inhomogeneous self-similar sets with overlaps. *Ergodic Theory Dynam. Systems* **39**, 1–18 (2019) Zbl [1403.37024](#) MR [3881123](#)
- [4] Bárány, B.: On the Ledrappier–Young formula for self-affine measures. *Math. Proc. Cambridge Philos. Soc.* **159**, 405–432 (2015) Zbl [1371.28015](#) MR [3413884](#)
- [5] Bárány, B., Hochman, M., Rapaport, A.: Hausdorff dimension of planar self-affine sets and measures. *Invent. Math.* **216**, 601–659 (2019) Zbl [1414.28014](#) MR [3955707](#)
- [6] Bárány, B., Käenmäki, A.: Ledrappier–Young formula and exact dimensionality of self-affine measures. *Adv. Math.* **318**, 88–129 (2017) Zbl [1457.37032](#) MR [3689737](#)
- [7] Barral, J., Feng, D.-J.: Multifractal formalism for almost all self-affine measures. *Comm. Math. Phys.* **318**, 473–504 (2013) Zbl [1280.28010](#) MR [3020165](#)
- [8] Bedford, T.: Crinkly curves, Markov partitions and dimension. Ph.D. thesis, Univ. of Warwick (1984)
- [9] Benoist, Y.: Actions propres sur les espaces homogènes réductifs. *Ann. of Math. (2)* **144**, 315–347 (1996) Zbl [0868.22013](#) MR [1418901](#)

- [10] Benoist, Y.: Propriétés asymptotiques des groupes linéaires. *Geom. Funct. Anal.* **7**, 1–47 (1997) Zbl [0947.22003](#) MR [1437472](#)
- [11] Benoist, Y.: Propriétés asymptotiques des groupes linéaires. II. In: *Analysis on Homogeneous Spaces and Representation Theory of Lie Groups, Okayama–Kyoto (1997)*, *Adv. Stud. Pure Math.* **26**, Math. Soc. Japan, Tokyo, 33–48 (2000) Zbl [0960.22012](#) MR [1770716](#)
- [12] Benoist, Y., Quint, J.-F.: *Random Walks on Reductive Groups*. *Ergeb. Math. Grenzgeb.* **62**, Springer, Cham (2016) Zbl [1366.60002](#) MR [3560700](#)
- [13] Bochi, J., Morris, I. D.: Equilibrium states of generalised singular value potentials and applications to affine iterated function systems. *Geom. Funct. Anal.* **28**, 995–1028 (2018) Zbl [1420.37016](#) MR [3820437](#)
- [14] Borel, A.: *Linear Algebraic Groups*. 2nd ed., *Grad. Texts in Math.* **126**, Springer, New York (1991) Zbl [0726.20030](#) MR [1102012](#)
- [15] Borel, A., Tits, J.: Groupes réductifs. *Inst. Hautes Études Sci. Publ. Math.* **27**, 659–755 (1965) Zbl [0145.17402](#) MR [207712](#)
- [16] Broomhead, D., Montaldi, J., Sidorov, N.: Golden gaskets: variations on the Sierpiński sieve. *Nonlinearity* **17**, 1455–1480 (2004) Zbl [1054.28002](#) MR [2069714](#)
- [17] Cao, Y.-L., Feng, D.-J., Huang, W.: The thermodynamic formalism for sub-additive potentials. *Discrete Contin. Dynam. Systems* **20**, 639–657 (2008) Zbl [1140.37319](#) MR [2373208](#)
- [18] Chevalley, C.: *Théorie des groupes de Lie. Tome II. Groupes algébriques*. *Actualités Sci. Industr.* **1152**, Hermann, Paris (1951) Zbl [0054.01303](#) MR [0051242](#)
- [19] Das, T., Simmons, D.: The Hausdorff and dynamical dimensions of self-affine sponges: a dimension gap result. *Invent. Math.* **210**, 85–134 (2017) Zbl [1387.37026](#) MR [3698340](#)
- [20] Douady, A., Oesterlé, J.: Dimension de Hausdorff des attracteurs. *C. R. Acad. Sci. Paris Sér. A-B* **290**, A1135–A1138 (1980) Zbl [0443.58016](#) MR [585918](#)
- [21] Dufloux, L.: Hausdorff dimension of limit sets. *Geom. Dedicata* **191**, 1–35 (2017) Zbl [1387.37027](#) MR [3719073](#)
- [22] Falconer, K. J.: The Hausdorff dimension of self-affine fractals. *Math. Proc. Cambridge Philos. Soc.* **103**, 339–350 (1988) Zbl [0642.28005](#) MR [923687](#)
- [23] Falconer, K., Kempton, T.: Planar self-affine sets with equal Hausdorff, box and affinity dimensions. *Ergodic Theory Dynam. Systems* **38**, 1369–1388 (2018) Zbl [1388.37028](#) MR [3789169](#)
- [24] Feng, D.-J.: Equilibrium states for factor maps between subshifts. *Adv. Math.* **226**, 2470–2502 (2011) Zbl [1215.37023](#) MR [2739782](#)
- [25] Feng, D.-J.: Dimension of invariant measures for affine iterated function systems. *Duke Math. J.*, to appear; arXiv:[1901.01691](#)
- [26] Feng, D.-J., Hu, H.: Dimension theory of iterated function systems. *Comm. Pure Appl. Math.* **62**, 1435–1500 (2009) Zbl [1230.37031](#) MR [2560042](#)
- [27] Feng, D.-J., Käenmäki, A.: Equilibrium states of the pressure function for products of matrices. *Discrete Contin. Dynam. Systems* **30**, 699–708 (2011) Zbl [1219.37027](#) MR [2784616](#)
- [28] Feng, D.-J., Shmerkin, P.: Non-conformal repellers and the continuity of pressure for matrix cocycles. *Geom. Funct. Anal.* **24**, 1101–1128 (2014) Zbl [1317.37035](#) MR [3248481](#)
- [29] Fraser, J. M., Jordan, T., Jurga, N.: Dimensions of equilibrium measures on a class of planar self-affine sets. *J. Fractal Geom.* **7**, 87–111 (2020) Zbl [07196519](#) MR [4068765](#)
- [30] Guéritaud, F., Guichard, O., Kassel, F., Wienhard, A.: Anosov representations and proper actions. *Geom. Topol.* **21**, 485–584 (2017) Zbl [1373.37095](#) MR [3608719](#)
- [31] Hochman, M.: On self-similar sets with overlaps and inverse theorems for entropy. *Ann. of Math. (2)* **180**, 773–822 (2014) Zbl [1337.28015](#) MR [3224722](#)
- [32] Hochman, M.: On self-similar sets with overlaps and inverse theorems for entropy in  $\mathbb{R}^d$ . *Mem. Amer. Math. Soc.* **265**, no. 1287 (2020)

- [33] Horn, R. A., Johnson, C. R.: *Topics in Matrix Analysis*. Cambridge Univ. Press, Cambridge (1994) Zbl [0801.15001](#) MR [1288752](#)
- [34] Hueter, I., Lalley, S. P.: Falconer's formula for the Hausdorff dimension of a self-affine set in  $\mathbf{R}^2$ . *Ergodic Theory Dynam. Systems* **15**, 77–97 (1995) Zbl [0867.28006](#) MR [1314970](#)
- [35] Hutchinson, J. E.: Fractals and self-similarity. *Indiana Univ. Math. J.* **30**, 713–747 (1981) Zbl [0598.28011](#) MR [625600](#)
- [36] Jordan, T., Pollicott, M., Simon, K.: Hausdorff dimension for randomly perturbed self affine attractors. *Comm. Math. Phys.* **270**, 519–544 (2007) Zbl [1119.28004](#) MR [2276454](#)
- [37] Käenmäki, A.: On natural invariant measures on generalised iterated function systems. *Ann. Acad. Sci. Fenn. Math.* **29**, 419–458 (2004) Zbl [1078.37014](#) MR [2097242](#)
- [38] Käenmäki, A., Morris, I. D.: Structure of equilibrium states on self-affine sets and strict monotonicity of affinity dimension. *Proc. London Math. Soc. (3)* **116**, 929–956 (2018) Zbl [1391.28009](#) MR [3789836](#)
- [39] Käenmäki, A., Reeve, H. W. J.: Multifractal analysis of Birkhoff averages for typical infinitely generated self-affine sets. *J. Fractal Geom.* **1**, 83–152 (2014) Zbl [1292.28016](#) MR [3166207](#)
- [40] Kaplan, J. L., Yorke, J. A.: Chaotic behavior of multidimensional difference equations. In: *Functional Differential Equations and Approximation of Fixed Points (Bonn, 1978)*, Lecture Notes in Math. 730, Springer, Berlin, 204–227 (1979) Zbl [0448.58020](#) MR [547989](#)
- [41] Knapp, A. W.: *Lie Groups Beyond an Introduction*. Progr. Math. 140, Birkhäuser Boston, Boston, MA (1996) Zbl [1075.22501](#) MR [1399083](#)
- [42] Lindenstrauss, E., Varjú, P. P.: Random walks in the group of Euclidean isometries and self-similar measures. *Duke Math. J.* **165**, 1061–1127 (2016) Zbl [1451.60011](#) MR [3486415](#)
- [43] McMullen, C.: The Hausdorff dimension of general Sierpiński carpets. *Nagoya Math. J.* **96**, 1–9 (1984) Zbl [0539.28003](#) MR [771063](#)
- [44] Morris, I. D.: An inequality for the matrix pressure function and applications. *Adv. Math.* **302**, 280–308 (2016) Zbl [1350.15005](#) MR [3545931](#)
- [45] Morris, I. D., Sert, C.: A strongly irreducible affine iterated function system with two invariant measures of maximal dimension. *Ergodic Theory Dynam. Systems* **41**, 3417–3438 (2021) Zbl [1477.28007](#) MR [4321717](#)
- [46] Morris, I. D., Shmerkin, P.: On equality of Hausdorff and affinity dimensions, via self-affine measures on positive subsystems. *Trans. Amer. Math. Soc.* **371**, 1547–1582 (2019) Zbl [1406.28005](#) MR [3894027](#)
- [47] Pozzetti, M. B., Sambarino, A., Wienhard, A.: Conformality for a robust class of non-conformal attractors. *J. Reine Angew. Math.* **774**, 1–51 (2021) Zbl [1483.37037](#) MR [4250471](#)
- [48] Quint, J.-F.: Divergence exponentielle des sous-groupes discrets en rang supérieur. *Comment. Math. Helv.* **77**, 563–608 (2002) Zbl [1010.22018](#) MR [1933790](#)
- [49] Quint, J.-F.: Groupes de Schottky et comptage. *Ann. Inst. Fourier (Grenoble)* **55**, 373–429 (2005) Zbl [1087.22010](#) MR [2147895](#)
- [50] Rapaport, A.: A self-similar measure with dense rotations, singular projections and discrete slices. *Adv. Math.* **321**, 529–546 (2017) Zbl [1376.28009](#) MR [3715720](#)
- [51] Rapaport, A.: On self-affine measures with equal Hausdorff and Lyapunov dimensions. *Trans. Amer. Math. Soc.* **370**, 4759–4783 (2018) Zbl [1386.37021](#) MR [3812095](#)
- [52] Rossi, E.: Local dimensions of measures on infinitely generated self-affine sets. *J. Math. Anal. Appl.* **413**, 1030–1039 (2014) Zbl [1310.28008](#) MR [3159819](#)
- [53] Saglietti, S., Shmerkin, P., Solomyak, B.: Absolute continuity of non-homogeneous self-similar measures. *Adv. Math.* **335**, 60–110 (2018) Zbl [06929145](#) MR [3836658](#)
- [54] Schief, A.: Separation properties for self-similar sets. *Proc. Amer. Math. Soc.* **122**, 111–115 (1994) Zbl [0807.28005](#) MR [1191872](#)
- [55] Solomyak, B.: On the random series  $\sum \pm \lambda^n$  (an Erdős problem). *Ann. of Math. (2)* **142**, 611–625 (1995) Zbl [0837.28007](#) MR [1356783](#)

- 
- [56] Solomyak, B.: Measure and dimension for some fractal families. *Math. Proc. Cambridge Philos. Soc.* **124**, 531–546 (1998) Zbl [0927.28006](#) MR [1636589](#)
- [57] Wehrfritz, B. A. F.: *Infinite Linear Groups: An Account of the Group-Theoretic Properties of Infinite Groups of Matrices*. *Ergeb. Math. Grenzgeb.* 76, Springer, New York (1973) Zbl [0261.20038](#) MR [0335656](#)