NONLINEAR THERMODYNAMICAL FORMALISM

Jérôme Buzzi, Benoît R Kloeckner, Renaud Leplaideur

NONLINEAR THERMODYNAMICAL FORMALISM

JÉRÔME BUZZI, BENOÎT R. KLOECKNER, AND RENAUD LEPLAIDEUR

Abstract. We define a nonlinear thermodynamical formalism which translates into dynamical system theory the statistical mechanics of generalized mean-field models, extending investigation of the quadratic case by Leplaideur and Watbled.

Under suitable conditions, we prove a variational principle for the nonlinear pressure and we characterize the nonlinear equilibrium measures and relate them to specific classical equilibrium measures.

In this non-linear thermodynamical formalism, which can, e.g., model mean-field approximation of large systems, several kind of phase transitions appear, some of which cannot happen in the linear case. We use our correspondence between non-linear and linear equilibrium measures to further the understanding of phase transitions, both in previously known cases (Curie-Weiss and Potts models) and in new examples (metastable phase transition).

Finally, we apply some of the ideas introduced to the classical thermodynamical formalism, proving that freezing phase transitions can occur over *any* zero-entropy invariant compact subset of the phase space.

1. Introduction

In the 1970s, Sinai, Ruelle, Bowen, and others (see, e.g., [28, 26, 4]) developed a thermodynamical approach to dynamical systems inspired by the statistical mechanics of lattice systems. In a recent work [16], the third named author and Watbled applied this program to the Curie-Weiss mean-field theory: they introduced a new thermodynamical formalism over the full shift where the energy functional is quadratic. They obtained precise results using the specific structure of this setting.

Our goal in this paper is to understand the generality of their results. It turns out that we can define the nonlinear pressure of a measure as the sum of its entropy and its "energy", defined as *any weak-star continuous function* of the measure. We are in particular interested in the case when the energy is a smooth function of the integrals of one or several potentials, in which case we call it an *energy with potential(s)*. Assuming only that the classical thermodynamical formalism is well-behaved, we can analyze this *nonlinear* thermodynamics using suitable convex analysis.

Date: November 10, 2021.

JB was partially supported by the ISDEEC project ANR-16-CE40-0013.

RL wants to thank ERC project 692925 NUHGD for kind support for a visit to Orsay in September 2018.

We first prove a variational principle: the supremum of the nonlinear pressure of the measures is given by a combinatorial formula involving the classical separated sets for the Bowen-Dinaburg dynamical metric (Theorem A), then defining *equilibrium measure* as those measures achieving the previous supremum. It is easy to show that equilibrium measures exist and, in the expansive case, we relate them to Gibbs ensembles (Theorem B). In the case of energies with potentials we show that equilibrium measures are classical equilibrium measures for some specific linear combination of these potentials (Theorem C). When the nonlinearity is a real-anaytic function of the integral of a single potential, we obtain finiteness of the set of equilibrium measures (Theorem D). As is well-known from physics and examples including the Curie-Weiss theory, *phase transitions* can occur in this setting, e.g., there may be several equilibrium measures that may depend non-analytically on parameters giving rise to *freezing* (Theorem E and Section 5.4) or *metastable phase transitions.* (Theorem F in Section 5.2).

1.1. **Classical thermodynamical formalism.** We recall the classical definitions (see, e.g., [31]). We will sometimes call these notions *linear* to distinguish them from the ones we introduce in this paper.

Let (T, φ) be a continuous system, i.e., a continuous self-map $T : X \to X$ of a compact metric space together with a continuous function $\varphi \in C(X, \mathbb{R})$. The function φ is called the *potential*. We denote by $\mathscr{P}$ the set of Borel probability measures on X, endowed with the weak star topology, by $\mathscr{P}(T)$ the subset of T-invariant measures and by $\mathscr{P}_{\mathrm{erg}}(T)$ the subset of ergodic and invariant measures.

The weight of order n of a finite subset $C \subset M$ is:

$$w_n(C) := \sum_{x \in \mathcal{C}} \exp\left(S_n\varphi(x)\right)$$

where $S_n\varphi$ denotes a Birkhoff sum:

$$S_n\varphi(x) := \varphi(x) + \varphi(Tx) + \cdots + \varphi(T^{n-1}x).$$

Given $\varepsilon > 0$ and $n \in \mathbb{N}$, the *Bowen-Dinaburg dynamical balls* are the sets

$$B(x, \varepsilon, n) := \left\{y \in X : \forall 0 \leq k < n,\ d(T^k y, T^k x) < \varepsilon\right\}.$$

A finite set $\mathcal{C}$ is an (ε, n)-*covering* when $\bigcup_{x \in \mathcal{C}} B(x, \varepsilon, n) = X$. It is an (ε, n)-*separated* subset when for all distinct $x, x' \in \mathcal{C}$, $x' \notin B(x, \varepsilon, n)$. The *partition function* is:

$$Z(\varepsilon, n) := \sup_{\mathcal{C}} w_n(\mathcal{C})$$

where $\mathcal{C}$ ranges over the (ε, n)-separated subsets of X.

An (ε, n)-separated set $\mathcal{C}$ is said to be *adapted* when it realizes the supremum in $Z(\varepsilon, n)$, and each adapted set defines a probability measure

$$\mu_{\mathcal{C}} := \frac{1}{Z(n, \varepsilon)} \sum_{x \in \mathcal{C}} e^{S_n\phi(x)} \frac{\delta_x + \delta_{Tx} + \cdots + \delta_{T^{n-1}x}}{n}. \tag{1.1}$$

called an (ε, n)-*Gibbs ensemble.*

The (linear) topological pressure is:

$$P_{\mathrm{top}}(T,\varphi) := \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log Z(\varepsilon, n). \tag{1.2}$$

Recall that the (linear) pressure of a measure $\mu \in \mathscr{P}(T)$ with respect to the potential φ is (denoting by h the Kolmogorov-Sinai entopy):

$$P(T,\varphi,\mu) := h(T,\mu) + \int \varphi \, d\mu.$$

The variational principle states that:

$$P_{\mathrm{top}}(T,\varphi) = \sup_{\mu\in\mathscr{P}(T)} P(T,\varphi,\mu). \tag{1.3}$$

An *equilibrium measure* for (T,φ) is then an invariant probability measure μ such that $P(T,\varphi,\mu) = P_{\mathrm{top}}(T,\varphi)$, i.e., a measure that achieves the above supremum.

The (linear) pressure function is the function $\beta \mapsto P_{\mathrm{top}}(T,\beta\varphi)$ where β is a real parameter, called the *inverse of temperature*.

1.2. **Nonlinear formalism.** We propose the following generalization. It will prove convenient to write $\mu(\varphi)$ for $\int \varphi \, \mathrm{d}\mu$. We consider again a continuous map T acting on a compact metric space X.

An *energy* is a function $\mathcal{E} : \mathscr{P} \to \mathbb{R}$ which is continuous in the weak star topology; note that we will need the energy of non-invariant measures. We say that $\mathcal{E}$ is an energy *with potential* φ (a continuous function defined on X) if it can be written

$$\mathcal{E}(\mu) = F\big(\mu(\varphi)\big)$$

for some continuous function $F : I \to \mathbb{R}$ defined on an interval containing all values taken by φ. More generally, an *energy with potentials* takes the form

$$\mathcal{E}(\mu) = F\big(\mu(\varphi_1), \ldots, \mu(\varphi_d)\big) \tag{1.4}$$

where $\varphi_1, \ldots, \varphi_d$ are continuous functions defined on X and $F : U \to \mathbb{R}$ is a continuous function on some set $U \subset \mathbb{R}^d$. For $\mathcal{E}$ to be well-defined on the whole of $\mathscr{P}$, the set U must contain the convex hull of the set of values taken by $\vec{\varphi} = (\varphi_1, \ldots, \varphi_d) : X \to \mathbb{R}^d$. We add the adjective "$C^r$" ($r \geq 1$), "smooth" or "analytic" to $\mathcal{E}$ whenever the domain U of F is open and F is C^r ($r = \infty$ meaning smooth, $r = \omega$ meaning analytic) on U.

An energy is said to be *convex* when for all Borel probability measure ξ on $\mathscr{P}$ (hence, ξ is a measure of measures):

$$\mathcal{E}\Big(\int \nu \, \mathrm{d}\xi(\nu)\Big) \leq \int \mathcal{E}(\nu) \, \mathrm{d}\xi.$$

For example, if $\mathcal{E}$ is an energy with potentials, it is convex whenever F is.

Not assuming potentials, we first need to replace Birkhoff sums. Given $x \in X$ and $n \in \mathbb{N}$, we define an *empirical measure*

$$\Delta_x^n := \frac{1}{n} \sum_{i=0}^{n-1} \delta_{T^i x}.$$

Observe that for any potential φ, $\Delta_x^n(\varphi) = \frac{1}{n} S_n \varphi(x)$ is the averaged Birkhoff sum. We thus define the *nonlinear weight* of order n of a finite set $\mathcal{C}$ and the *nonlinear partition function* as

$$\omega_n(\mathcal{C}) = \sum_{x \in \mathcal{C}} e^{n\mathcal{E}(\Delta_x^n)} \qquad \zeta(\varepsilon, n) = \sup_{\mathcal{C}} \omega_n(\mathcal{C}),$$

where the supremum is taken over all (ε, n)-separated sets $\mathcal{C}$.

Again, an (ε, n)-separated set $\mathcal{C}$ is said to be *adapted* if it realizes the maximum in ζ_n and we define an *nonlinear (ε, n)-Gibbs ensemble*

$$\mu_{\mathcal{C}} := \frac{1}{\zeta(\varepsilon, n)} \sum_{x \in \mathcal{C}} e^{n\mathcal{E}(\Delta_x^n)} \Delta_x^n \quad \in \mathscr{P} \tag{1.5}$$

(note that the continuity of T ensures that the maximum in ζ is realized for all (ε, n)).

The *nonlinear topological pressure*, to be thought of as an analogue of topological entropy weighted by energy, is

$$\Pi_{\mathrm{top}}^{\mathcal{E}}(T) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \zeta(\varepsilon, n). \tag{1.6}$$

In Theorem A we will show that under suitable hypotheses, replacing the supremum limit by an infimum limit:

$$\underline{\Pi}_{\mathrm{top}}^{\mathcal{E}}(T) = \lim_{\varepsilon \to 0} \liminf_{n \to \infty} \frac{1}{n} \log \zeta(\varepsilon, n)$$

gives the same quantity as $\Pi_{\mathrm{top}}^{\mathcal{E}}(T)$. Meanwhile the *nonlinear pressure* is defined for all invariant probability measures μ by

$$\Pi^{\mathcal{E}}(T, \mu) = h(T, \mu) + \mathcal{E}(\mu).$$

1.3. **Main results.** For certain nonlinear systems $(T, \mathcal{E})$, it may happen that some measures satisfy $\Pi^{\mathcal{E}}(T, \mu) > \Pi_{\mathrm{top}}^{\mathcal{E}}(T)$; we will first give conditions excluding this.

Definition 1.7. *We will say that $(T, \mathcal{E})$ has an* abundance of ergodic measures *if for any $\mu \in \mathscr{P}(T)$ and $\varepsilon > 0$, there is an ergodic measure $\nu \in \mathscr{P}_{\mathrm{erg}}(T)$ such that $h(T, \nu) + \mathcal{E}(\nu) > h(T, \mu) + \mathcal{E}(\mu) - \varepsilon$.*

This condition is satisfied by uniformly hyperbolic diffeomorphisms that have a single basic set in their spectral decomposition as any invariant probability measure can be approximated by an ergodic one both in the weak star topology

and in entropy. It is also satisfied for arbitrary continuous systems $(T, \mathcal{E})$ with convex energy, since, in this case, for any $\mu \in \mathscr{P}(T)$,

$$\Pi^{\mathcal{E}}(T, \mu) \leq \int h(T, \mu_\xi) + \mathcal{E}(\mu_\xi)\, dP(\xi)$$

using the ergodic decomposition $\mu = \int \mu_\xi \, dP(\xi)$.

Recall that, in the invertible case, T is said to be an *expansive homeomorphism* when there exist a number $\varepsilon_0 > 0$ (called an *expansivity constant* for T) such that

$$\forall x, y \in X \quad \sup_{n \in \mathbb{Z}} d(T^n x, T^n y) \leq \varepsilon_0 \implies x = y$$

(see, e.g., [13] Definition 3.2.11; note that here we use a $\leq$ sign, making the expansivity constants possibly slightly smaller). This notion is generalized to non-necessarily invertible maps under the name of *positive expansivity* by considering only the positive orbits:

$$\forall x, y \in X \quad \sup_{n \geq 0} d(T^n x, T^n y) \leq \varepsilon_0 \implies x = y.$$

and the results we state below for expansive homeomorphisms could be extended to positively expansive map with the same proofs.

Our first result establishes a variational principle generalizing eq. (1.3) to all energies.

Theorem A (Variational principle)**.** *Let $T : X \to X$ be a continuous map of a compact space and let $\mathcal{E} : \mathscr{P} \to \mathbb{R}$ be an energy. Assume that $(T, \mathcal{E})$ has an abundance of ergodic measures,*

Then the nonlinear topological pressure satisfies:

$$\sup_{\mu \in \mathscr{P}(T)} \Pi^{\mathcal{E}}(T, \mu) = \Pi^{\mathcal{E}}_{\mathrm{top}}(T) = \underline{\Pi}^{\mathcal{E}}_{\mathrm{top}}(T) \tag{1.8}$$

If, additionally, T is an expansive homeomorphism with some constant $\varepsilon_0 > 0$, then

$$\Pi^{\mathcal{E}}_{\mathrm{top}}(T) = \lim_n \frac{1}{n} \log \zeta(\varepsilon_0, n).$$

When the conclusion $\sup_{\mathscr{P}(T)} \Pi^{\mathcal{E}}(T, \cdot) = \Pi^{\mathcal{E}}_{\mathrm{top}}(T)$ of the above theorem holds, we define a *nonlinear equilibrium measure* as any measure $m \in \mathscr{P}(T)$ realizing this supremum:

$$\Pi^{\mathcal{E}}(T, m) = \max_{\mathscr{P}(T)} \Pi^{\mathcal{E}}(T, \cdot).$$

As in the classical setting, existence of an equilibrium measure is easily obtained when entropy is upper semicontinuous, and in the expansive case equilibrium measures prescribe the asymptotic behavior of Gibbs ensembles.

Theorem B. *Let $T : X \to X$ be a continuous map of a compact space and let $\mathcal{E} : \mathscr{P} \to \mathbb{R}$ be an energy. Assume that $(T, \mathcal{E})$ has an abundance of ergodic measures.*

If $\mu \mapsto h(T,\mu)$ is upper semicontinuous[1]*, then there exists at least one nonlinear equilibrium measure.*

If additionally T is an expansive homeomorphism for some constant $\varepsilon_0 > 0$, then any accumulation point μ of any sequence $(\mu_{\mathcal{C}_n})_{n\in\mathbb{N}}$ of nonlinear Gibbs (ε_0, n)-ensembles belongs to the closure of the convex span of all nonlinear equilibrium measures.

The last statement means that there exists a probability measure ξ on $\mathscr{P}$ (a measure of measures), concentrated on the set $\mathscr{EM}$ of equilibrium measures, such that

$$\mu = \int_{\mathscr{EM}} \nu \,\mathrm{d}\xi(\nu)$$

(see, e.g., [24], Proposition 1.2.) The accumulation points can indeed fail to be equilibrium measures, e.g., in the Curie-Weiss model when there are two asymmetric equilibrium measures and one chooses symmetric Gibbs ensembles, see [16].

Next we study the uniqueness and nature of the nonlinear equilibrium measures in the case of an energy with potentials as in eq. (1.4). Our main point here is that we can use classical convex analysis to reduce the nonlinear thermodynamical formalism to the linear one.

More precisely, we will use the classical Legendre duality between entropy and pressure; using the vector of integral of potentials $(\mu(\varphi_1), \dots, \mu(\varphi_d))$ as intermediate coordinates, this will reduce to finite-dimensional Legendre duality. This duality holds for the class of C^r *Legendre* systems $(T, \vec{\varphi})$ (where $r \in \mathbb{N}^* \cup \{\infty, \omega\}$ and C^ω means analytic), see Definitions 4.7, 4.9. When additionally each linear combination of the (φ_i) admits a unique linear equilibrium measure, one says that $(T, \vec{\varphi})$ is C^r *Legendre with unique linear equilibrium measures.*

Let us note that classical examples fulfill these requirements: if T is a topologically transitive Anosov diffeomorphism or expanding map, and $(\varphi_1, \dots, \varphi_d)$ is a family of Hölder-continuous potentials whose linear combinations are not cohomologuous to a constant, i.e., for all $\alpha_1, \dots, \alpha_d \in \mathbb{R}^d$:

$$\Big(\exists u \in C(X,\mathbb{R}), \exists c \in \mathbb{R} \colon \sum_{i=1}^{d} \alpha_i \varphi_i = u - u \circ T + c\Big) \implies \alpha_1 = \dots = \alpha_d = 0,$$

then $(T, \vec{\varphi})$ is C^r Legendre with unique linear equilibrium measures.

Theorem C. *Assume that $(T, \vec{\varphi})$ is C^r Legendre, that $F : U \subset \mathbb{R}^d \to \mathbb{R}$ is C^r and consider the energy with potentials $\mathcal{E}(\mu) = F\big(\mu(\varphi_1), \dots, \mu(\varphi_d)\big)$. Then there is a nonempty compact subset $\mathscr{Y} \subset \mathbb{R}^d$ such that the nonlinear equilibrium measures are exactly the linear equilibrium measures with respect to each of the potentials $\sum_i y_i \varphi_i$ where $(y_1, \dots, y_d) \in \mathscr{Y}$.*

[1]This holds, e.g., if (X,T) is a subshift [31] or a C^∞ smooth map [7].

Observe that as a consequence, even though nonlinear equilibrium measures may fail to be unique, under the hypotheses of Theorem C they are ergodic as soon as linear equilibrium measures are (and more, see Corollary 1.11).

Addendum 1.9. *In the above setting, the set $\mathscr{Y}$ can be computed from the linear pressure function defined over $\mathbb{R}^d$ by $\mathsf{P}(y_1, \dots, y_d) = P_{\text{top}}(T, \sum_i y_i \varphi_i)$. More precisely $\mathscr{Y} = (\nabla \mathsf{P})^{-1}(\mathscr{V})$ where $\nabla \mathsf{P}$ is the gradient of P and*

$$\mathscr{V} = \{z \in \mathbb{R}^d \colon \mathsf{h}(z) + F(z) = \sup(\mathsf{h} + F)\}$$
$$\mathsf{h}(z) = \sup\{h(T, \mu) : \mu(\vec{\varphi}) = z\}.$$

The function h can also be computed from P, as $-\mathsf{h}$ is the Legendre dual of P.

Remarks 1.10. *Given $(T, \vec{\varphi})$ a smooth Legendre system, any compact subset of $\mathbb{R}^d$ can be realized as the set $\mathscr{Y}$ above by choosing a suitable C^∞ smooth nonlinearity F (Corollary 4.22).*

Our proof will apply to a more general notion of equilibrium measures, see eq. (4.3).

Theorem D. *If (T, φ) is a C^ω Legendre system with unique linear equilibrium measures and F is C^ω with a single potential ($d = 1$), then there are only finitely many nonlinear equilibrium measures.*

Note that we do not simply claim that $\mathscr{EM}$ is finite-dimensional, but that it is finite, even though it can contain several equilibrium measures. In fact, this failure of uniqueness can occur even for a topologically transitive subshift of finite type with a Hölder-continuous potential (see *e.g.* [16] and Section 5 below). However uniqueness holds for generic non-linearities for any $d \geq 1$ (Proposition 4.20).

The above characterization shows that for many systems with expanding or hyperbolic properties, such as mixing subshifts of finite type, the nonlinear equilibrium measures share the good ergodic properties of the classical equilibrium measures. Let us recall some of them.

Corollary 1.11 (Folklore)**.** *Let (X, T) be a mixing subshift of finite type (not reduced to a fixed point). Consider Hölder-continuous potentials $\vec{\varphi} : X \to \mathbb{R}^d$ and a C^r nonlinearity $F : U \subset \mathbb{R}^d \to \mathbb{R}$. Then, for the energy given by $\mathcal{E}(\mu) = F\big(\mu(\varphi_1), \dots, \mu(\varphi_d)\big)$, any nonlinear equilibrium measure*

- *is ergodic and mixing;*
- *has exponential decay of correlation;*
- *satisfies the almost sure invariance principle and in particular the central limit theorem.*

where the two last properties are understood to hold with respect to Hölder-continuous observables.

These results are folklore in the sense that some of them are immediate consequences of the founding results of Sinai, Ruelle, and Bowen, while others were first considered in more general settings. The following are convenient references: ergodicity, mixing, and exponential decay of correlation follow from Ruelle's Perron-Frobenius theorem (see, e.g.,[1, chapter 1]), the almost sure invariance principle, which implies many limit theorems was proved in [19] in much greater generality.

1.4. **Examples.** We will give a few examples to which the above theorems apply, mostly inspired by physics. These examples involves an additional real parameter, the inverse temperature $\beta > 0$: the energy function is then $\mathcal{E}(\mu) = \beta\mathcal{E}_1(\mu)$ where $\mathcal{E}_1$ is a reference energy and β tunes the balance between entropy and that energy, in agreement with thermodynamics.[2] This leads to the natural question of how the existence, the number, or the equilibrium measures themselves depend on this parameter β, leading to the physical notion of phase transitions.

1.4.1. *Linear case.* The classical, linear formalism is the special case where $d = 1$ and $F(z) = \beta z$ for $\beta > 0$ and taking any $\varphi \in C(X, \mathbb{R})$. The nonlinear pressure then coincides with the linear one: $h(T, \mu) + \beta \int \varphi \, \mathrm{d}\mu$, yielding a first example. Here $\mathscr{Y} = \{\beta\}$.

1.4.2. *Classical Curie-Weiss model.* Consider $X = \{-1, 1\}^{\mathbb{N}}$ of $X = \{-1, 1\}^{\mathbb{Z}}$, let T be the shift map, and set $\varphi(x_0x_1 \cdots) = x_0$ and $F(z) = \frac{\beta}{2}z^2$ for some $\beta \geq 0$; i.e., maximize $h(T, \mu) + \frac{\beta}{2}\mu(\varphi)^2$. The set $\mathscr{Y}$ can have one or two elements depending on the value of β: see [16] and Section 5.1. The notations were slightly different: $\Pi^{\mathcal{E}}_{\mathrm{top}}(T)$ here was $\mathcal{P}_2(\varphi)$ in [16], $\mathsf{h}(z)$ was $\overline{H}(z)$ and $\mathsf{h}(z) + F(z)$ was $\overline{\varphi}(z)$.

1.4.3. *Asymmetric Curie-Weiss model.* In Section 5.2 we shall give an asymmetric Curie-Weiss model, where T is again a full shift map, φ is a Bernoulli potential and $F(z) = \frac{\beta}{2}z^2$, but exhibiting a *metastable* phase transition: at each temperature there are finitely many *local* maximizers, but at some critical temperature the global maximizer jumps from one local maximizer to another.

1.4.4. *Curie-Weiss-Potts.* The consideration of several potentials is motivated by the Curie-Weiss-Potts model: $X = \{\theta_1, \dots, \theta_n\}^{\mathbb{N}}$ or $X = \{\theta_1, \dots, \theta_n\}^{\mathbb{Z}}$, T the shift map, $\varphi_i(x_0x_1 \cdots) = \mathbb{1}_{\theta_i}(x_0)$ and $F(z) = \frac{\beta}{2}\|z\|^2$ where $\|\cdot\|$ is the usual Euclidean norm, exhibiting yet another form of phase transition as β varies, see [17] and Section 5.3.

[2] In thermodynamics, the equilibrium state of a system in contact with a thermostat at inverse temperature β is such that it maximizes the entropy of the total system (combining the initial system and the thermostat), i.e., the quantity $h(T, \mu) - \beta\mathcal{E}(\mu)$, up to the addition of a constant. As is customary in dynamics, the minus sign has been included in the definition of the energy function.

1.4.5. *Wassertein distance to the maximal entropy measure.* We can go beyond the case with potentials: let us give a simple but intriguing example. Consider the map $T : x \mapsto 2x \mod 1$ on the circle, with reference energy $\mathcal{E}_1(\mu) = \mathrm{W}_p(\mu, \lambda)$ where λ denotes the Lebesgue measure, $p \in [1, +\infty)$ and W_p is the *Wasserstein distance* of exponent p.

Theorems A and B ensure that the nonlinear topological pressure is achieved by at least one invariant measure. The main question, which we leave open, is then to describe the non-empty set of equilibrium measures for $\beta\mathcal{E}_1$, in particular determine uniqueness.

For $\beta = 0$, $h(T, \mu) + \beta W_p(\mu, \lambda)$ reduces to the entropy so λ is the unique equilibrium. When $\beta \to \infty$, the set of equilibrium measures must converge to $\{\delta_0\}$, since δ_0 is the unique invariant measure maximizing $W_p(\mu, \lambda)$.

1.5. **More Phase Transitions.** A phase transition can be defined from any of a number of different phenomena that often occur simultaneously: loss of the analyticity of the pressure with respect to physical parameters, multiple equilibrium measures, or failure of the central limit theorem for example.

Sarig [27] has studied such equivalences in the setting of Markov shifts. In contrast, we see here (Section 5.1) that non-analyticity of pressure and multiplicity of equilibrium measures can occur though the central limit theorem continues to hold (Corollary 1.11). Such distinctions have been observed before in [15] and [29]. The key point of view in the definition of Legendre regular systems and the proof of Theorems C and D is to consider a certain convex set, the *entropy-potential diagram* (defined in Section 4, see figures 1, 2), which describes the pairs $(h(T, \mu); \mu(\vec{\varphi}))$ that can be achieved when μ runs over $\mathscr{P}(T)$. Phase transitions then occur when the nonlinearity "becomes more convex" than the diagram.

In Section 5.4, we shall illustrate more broadly the benefits of this diagram by considering *freezing phase transitions*, by which we mean that for all $\beta > \beta_0$ for some $\beta_0 > 0$, the set of equilibrium measures is non-empty and independent of β; its elements are called "ground states" as they must maximize the energy.

Theorem E. *Let $T : X \to X$ be a continuous dynamical system of finite, positive topological entropy, and assume that $\mu \mapsto h(T, \mu)$ is upper semi-continuous.*

- *(i) For every $\mu_0 \in \mathscr{P}_{\mathrm{erg}}(T)$ with zero entropy there exist a continuous potential $\varphi : X \to \mathbb{R}$ such that the* linear *thermodynamical formalism of (T, φ) exhibits a freezing phase transition with unique ground state μ_0.*
- *(ii) For every continuous potential $\varphi : X \to (-\infty, 0]$ such that $K = \varphi^{-1}(0)$ is T-invariant and has zero topological entropy, there exist a continuous nonlinearity $F : (-\infty, 0] \to (-\infty, 0]$ with $F(0) = 0$ such that the energy $\mathcal{E}(\mu) = F(\mu(\varphi))$ exhibits a freezing phase transition with ground states supported on K.*

The first item is not directly related to the *non-linear* thermodynamical formalism, but its analysis is a simple application of the tools developed here (more precisely, we rely on the entropy-potential diagram introduced in Section 4 which is central to our non-linear study).

1.6. **Questions.** We close this introduction with a few more open questions.

- Without assuming abundance of ergodicity, does a variational principle hold in *restriction to ergodic* measures, that is:

$$\sup_{\mu\in\mathscr{P}_{\mathrm{erg}}(T)} \Pi^F(T,\varphi,\mu) = \Pi^F_{\mathrm{top}}(T,\varphi)?$$

  (See Remark 2.3.)
- Can one find a subshift of finite type, Hölder-continuous potentials and a real-analytic nonlinearity[3] such that there exist infinitely many nonlinear equilibrium measures? What if we additionally impose the quadratic nonlinearity, i.e., $F(z) = \frac{1}{2}\|z\|^2$?
- Can one find a "natural" energy (necessarily not an energy with potentials) on some subshift of finite type such that the non-linear equilibrium measure is unique but not ergodic?
- For the doubling map and the Wasserstein energy $W_p(\cdot,\lambda)$ from Section 1.4.5, is there a finite $\beta > 0$ at which δ_0 is an equilibrium measure? Is λ still an equilibrium measure for some $\beta > 0$? What happens just after λ ceases to be an equilibrium?

2. Variational principle

In this section we prove Theorem A. We first introduce some convenient notations. We fix a compact metric space X, a map $T : X \to X$ and an energy $\mathcal{E}$. In order to be as general as possible, we do not assume T to be continuous for now, but only Borel-measurable. Note that X^n being compact, every subset is totally bounded; this ensures the finiteness of (ε, n)-separated sets even when T is not assumed to be continuous. We often omit $T, \mathcal{E}$ from the notation, i.e., $\Pi_{\mathrm{top}} = \Pi^{\mathcal{E}}_{\mathrm{top}}(T)$, $\Pi(\mu) = \Pi^{\mathcal{E}}(T,\mu)$ etc.

Recall the definitions of the empirical measures of a point $x \in X$, of the nonlinear weight of a subset $\mathcal{C} \subset X$, and of the partition function:

$$\Delta_x^n = \frac{1}{N}\sum_{k=0}^{n-1} \delta_{T^k x} \qquad \omega_n(\mathcal{C}) := \sum_{x\in\mathcal{C}} e^{n\mathcal{E}(\Delta_x^n)} \qquad \zeta(\varepsilon,n) := \sup_{\substack{\mathcal{C}\\ (\varepsilon,n)\text{-separated}}} \omega_n(\mathcal{C}).$$

We define for use in this section the following notation:

$$\Pi_{\mathrm{top}}(\varepsilon) = \limsup_{n\to\infty} \frac{1}{n}\log\zeta(\varepsilon,n) \qquad \text{so that } \Pi_{\mathrm{top}} = \lim_{\varepsilon\to 0}\Pi_{\mathrm{top}}(\varepsilon)$$

$$\underline{\Pi}_{\mathrm{top}}(\varepsilon) = \liminf_{n\to\infty} \frac{1}{n}\log\zeta(\varepsilon,n) \qquad \underline{\Pi}_{\mathrm{top}} = \lim_{\varepsilon\to 0}\underline{\Pi}_{\mathrm{top}}(\varepsilon).$$

[3]Recall that we ask that real-analytic F be defined on an *open* set containing the compact set of all possible values of $(\mu(\varphi_1),\dots,\mu(\varphi_d))$. This in particular prevents the trivial choice $F(\vec{z}) = -\sup\{h(T,\mu)\colon (\mu(\varphi_1),\dots,\mu(\varphi_d)) = \vec{z}\}$.

2.1. **Preliminaries.** We will use the Wasserstein distance on the set $\mathscr{P}$ of probability measures on X. Proofs of the statements we need can be found in many places, e.g., [30].

The distance between $\mu_1, \mu_2 \in \mathscr{P}$ can be defined as

$$\mathrm{W}(\mu_1, \mu_2) = \sup\big\{\mu_1(f) - \mu_2(f) \colon f \text{ 1-Lipschitz function } X \mapsto \mathbb{R}\big\}.$$

The "Kantorovich duality" states that this definition is equivalent to

$$\mathrm{W}(\mu_1, \mu_2) = \inf\Big\{ \int d(x,y)\, \mathrm{d}\pi(x,y) \colon \pi \in \Gamma(\mu_1, \mu_2)\Big\}$$

where d is the distance on X and $\Gamma(\mu_1, \mu_2)$ is the set of 'transport plans", i.e., Borel probability measures on $X \times X$ with marginals μ_1 and μ_2. Moreover in these definitions both the supremum and the infimum are reached; a transport plan realizing the Wasserstein distance is said to be *optimal*. The compactness of X implies that the Wassertein distance induces the weak-star topology on $\mathscr{P}$, and that Wasserstein distance can be bounded above by total variation distance:

$$\mathrm{W}(\mu_1, \mu_2) \leq \mathrm{diam}(X) \|\mu_1 - \mu_2\|_{\mathrm{TV}}$$

We will also use the following reformulation of Birkhoff's ergodic theorem.

Lemma 2.1. *Let $\mu \in \mathscr{P}(T)$ be ergodic. Then for μ-almost all $x \in X$, we have $\Delta_x^n \to \mu$ in the weak-star topology.*

Proof. Let $(f_k)_{k\in\mathbb{N}}$ be a dense sequence of the space $C(X, \mathbb{R})$ of continuous functions $X \to \mathbb{R}$, endowed with the uniform norm. There exists a set E with $\mu(E) = 1$ such that for all $x \in E$ and all $k \in \mathbb{N}$, $\Delta_x^n(f_k) \to \mu(f_k)$ as $n \to \infty$.

Let $f \in C(X, \mathbb{R})$ and $\varepsilon > 0$. There exists $k \in \mathbb{N}$ such that $\|f - f_k\|_\infty \leq \varepsilon$, and $N \in \mathbb{N}$ such that for all $n \geq N$ and all $x \in E$, $|\Delta_x^n(f_k) - \mu(f_k)| \leq \varepsilon$. We then have $|\Delta_x^n(f) - \mu(f)| \leq 3\varepsilon$. □

2.2. **First part.** Theorem A starts with the equalities:

$$\sup_{\mu\in\mathscr{P}(T)} \Pi(\mu) = \Pi_{\mathrm{top}} = \underline{\Pi}_{\mathrm{top}}. \tag{2.2}$$

We will first prove

$$\sup_{\mu\in\mathscr{R}_{\mathrm{erg}}(T)} \Pi(\mu) \underset{①}{\leq} \underline{\Pi}_{\mathrm{top}} \underset{②}{\leq} \Pi_{\mathrm{top}} \underset{③}{\leq} \sup_{\mu\in\mathscr{P}(T)} \Pi(\mu).$$

Inequality ① is proved in Proposition 2.4. Inequality ③ is proved in Proposition 2.9. Inequality ② immediatlely follows from the definitions of Π_{top} and $\underline{\Pi}_{\mathrm{top}}$.

The missing inequality

$$\sup_{\mu\in\mathscr{P}(T)} \Pi(\mu) \underset{④}{\leq} \sup_{\mu\in\mathscr{R}_{\mathrm{erg}}(T)} \Pi(\mu)$$

is proved in Proposition 2.5 assuming an abundance of ergodic measures.

Remark 2.3. *If $(T, \mathcal{E})$ is continuous but without abundance of ergodicity, the following example shows that inequality*

$$\Pi_{\mathrm{top}} < \sup_{\mu\in\mathscr{P}(T)} \Pi(\mu)$$

may hold.

Let (X,T) be the union of two distinct fixed points p,q. Let $\mathcal{E}(\mu) = F(\mu(\varphi))$ with $F(z) = -z^2$, $\varphi(p) = 1$, $\varphi(q) = -1$. Then $\Pi(\mu) = 0$ for $\mu = \frac{1}{2}(\delta_p + \delta_q)$ whereas $\Pi_{\mathrm{top}} = -1$.

2.2.1. *Bounding below the nonlinear topological pressure.* We prove Inequality ①, then Inequality ④ assuming an abundance of ergodic measures. Note that continuity of T is not needed at that stage.

Proposition 2.4 (Inequality ①)**.** *Recall that X is a compact metric space. If $T : X \to X$ is Borel-measurable, then for all* ergodic *$\mu \in \mathscr{P}(T)$, we have $\Pi(\mu) \leq \underline{\Pi}_{\mathrm{top}}$.*

Proof. Consider any $\gamma > 0$. Since $\mathcal{E}$ is continuous and $\mathscr{P}$ is compact, $\mathcal{E}$ is uniformly continuous: there exists $\delta > 0$ such that for all $\nu \in \mathscr{P}$, $\mathrm{W}(\nu,\mu) \leq 2\delta \implies \mathcal{E}(\nu) \geq \mathcal{E}(\mu) - \gamma$.

By Lemma 2.1, there is a set $A \subset X$ with $\mu(A) \geq \frac{3}{4}$ and $M_A \in \mathbb{N}$ such that for all $x \in A$ and all $n \geq M_A$ we have $\mathrm{W}(\Delta_x^n, \mu) \leq \delta$.

By the Brin-Katok entropy formula [6], there exist $B \subset X$ with $\mu(B) \geq \frac{3}{4}$ and $M_B \in \mathbb{N}$ such that for all $x \in B$ and all $n \geq M_B$ we have

$$\left|\frac{1}{n}\log \mu(B(x,2\delta,n)) + h(T,\mu)\right| \leq \gamma.$$

Consider any $n \geq \max(M_A, M_B)$ and any $0 < \varepsilon \leq \delta$. Let $\mathcal{C}$ be any (ε,n)-separated set of X that is maximal with respect to inclusion; in particular, $\mathcal{C}$ is an (ε,n)-cover, hence a (δ,n)-cover. Let $\mathcal{C}'$ be a minimal subset of $\mathcal{C}$ that is an (δ,n)-cover of $A \cap B$.

On the one hand, for all $x \in \mathcal{C}'$ by minimality $B(x,\delta,n)$ intersects B; picking any y in the intersection, we get $\mu(B(x,\delta,n)) \leq \mu(B(y,2\delta,n)) \leq e^{n(\gamma - h(T,\mu))}$. Since $\mu(A \cap B) \geq \frac{1}{2}$, it follows

$$|\mathcal{C}'| \geq \frac{1}{2} e^{n(h(T,\mu)-\gamma)}.$$

On the other hand, for all $x \in \mathcal{C}'$ by minimality $B(x,\delta,n)$ intersects A; picking any y in the intersection, we have $\mathrm{W}(\Delta_y^n,\mu) \leq \delta$ and $d(T^i x, T^i y) \leq \delta$ for all $i \in \{0,\dots,n-1\}$. By considering the transport plan $\sum_i \frac{1}{n}\delta_{T^i x} \otimes \delta_{T^i y}$, we see that $\mathrm{W}(\Delta_x^n, \Delta_y^n) \leq \delta$. The triangular inequality then ensures $\mathrm{W}(\Delta_x^n,\mu) \leq 2\delta$, and we get

$$\mathcal{E}(\Delta_x^n) \geq \mathcal{E}(\mu) - \gamma.$$

Using these two inequalities, we get

$$\omega_n(\mathcal{C}) \geq \omega_n(\mathcal{C}') \geq |\mathcal{C}'| \min_{x \in \mathcal{C}'} e^{n\mathcal{E}(\Delta_x^n)} \geq \frac{1}{2} e^{n(h(T,\mu)-\gamma)} e^{n(\mathcal{E}(\mu)-\gamma)} \geq \frac{1}{2} e^{n(\Pi(\mu)-2\gamma)}.$$

Since $\mathcal{C}$ is (ε,n)-separated, we get

$$\frac{1}{n}\log \zeta(\varepsilon,n) \geq \Pi(\mu) - 2\gamma - \frac{1}{n}\log 2.$$

Taking the infimum limit as $n \to \infty$, we obtain that for all $\gamma > 0$, there exist $\delta > 0$ such that for all $\varepsilon \in (0, \delta)$:

$$\underline{\Pi}_{\mathrm{top}}(\varepsilon) \geq \Pi(\mu) - 2\gamma,$$

and letting ε then γ go to zero ends the proof. □

Observe that we only used lower-semicontinuity for $\mathcal{E}$ here; but its upper-semicontinuity ensures it reaches its supremum, a desirable feature. This motivates the continuity requirement in the definition of an energy.

Proposition 2.5 (Inequality ④)**.** *If T is Borel-measurable and $(T, \mathcal{E})$ has an abundance of ergodic measures, then* $\sup_{\mu \in \mathscr{P}(T)} \Pi(\mu) \leq \sup_{\mu \in \mathscr{P}_{\mathrm{erg}}(T)} \Pi(\mu)$.

Proof. Let μ be any invariant probability measure. Since $(T, \mathcal{E})$ has an abundance of ergodic measures, there is a sequence of measures $\nu_n \in \mathscr{P}_{\mathrm{erg}}(T)$ such that $\lim_{n \to +\infty} h(T, \nu_n) + \mathcal{E}(\nu_n) \geq h(T, \mu) + \mathcal{E}(\mu)$; this yields that

$$h(T, \mu) + \mathcal{E}(\mu) \leq \sup_{\nu \in \mathscr{P}_{\mathrm{erg}}(T)} \Pi(\nu)$$

holds for every μ in $\mathscr{P}(T)$. □

2.2.2. *Bounding from above the nonlinear topological pressure: Inequality ③.* To end the proof of equality (1.8), it remains to construct measures almost realizing the nonlinear topological pressure. We divide the proof into several lemmas that we shall reuse in Section 3. We follow the strategy of Misiurewicz' proof of the linear variational principle, from which we extract the following result. We recall that $H_\mu(\alpha)$ stands for the entropy for the measure $\mu \in \mathscr{P}(T)$ of the partition α.

Lemma 2.6 (Misiurewicz [20])**.** *Fix $\varepsilon > 0$ and let $(\mathcal{C}_k)_{k \in \mathbb{N}}$ be a sequence of (ε, n_k)-separated sets where $n_k \to \infty$. Assume that for each k, σ_k is a probability measure concentrated on $\mathcal{C}_k$ and that*

$$\mu_k = \frac{1}{n_k} \sum_{\ell=0}^{n_k - 1} T_*^\ell \sigma_k$$

converges in the weak star topology to some measure μ_∞.

Fix any finite partition α of X into subsets of diameter less than ε and with negligible boundaries with respect to μ_∞ (such an α always exists). Then for all $m \in \mathbb{N}$,

$$H_{\mu_k}(\alpha^m) \geq \frac{m}{n_k} H_{\sigma_k}(\alpha^{n_k}) - \frac{2m^2}{n_k} \log|\alpha| \qquad \forall k \text{ such that } n_k \geq 2m$$

and $H_{\mu_k}(\alpha^m) \to H_{\mu_\infty}(\alpha^m)$ as $k \to \infty$.

The proof is not reproduced here, let us simply mention that it consists in partitioning in m different ways the integer interval $[\![0, n_k - 1]\!]$ into subintervals of length m plus a small remainder at the start and end. Note that the hypothesis that $\mathcal{C}_k$ is (ε, n_k)-separated is intended to make the computation of $H_{\sigma_k}(\alpha^{n_k})$ a formality: each element of α^{n_k} contains at most one element of $\mathcal{C}_k$.

To address the nonlinearity, we now divide the space of measures into parts where the energy is almost constant, and then split (ε, n)-separated sets according to this partition.

Lemma 2.7. *Let $\varepsilon > 0, \gamma \in (0,1)$ and $(\mathcal{C}_k)_{k\in\mathbb{N}}$ be a sequence of (ε, n_k)-separated subsets of X where $n_k \to \infty$. There exist $N = N(\gamma) \in \mathbb{N}$, real numbers $(E_i)_{1\le i\le N}$, a sequence of partitions $\mathbf{D}_k = (\mathcal{D}_{k,i})_{1\le i\le N}$ of $\mathcal{C}_k$ and $I \subset [\![1, N]\!]$ such that, up to extracting a subsequence (still denoted by $(n_k)_k$), for all k:*

(i) $\sum_{i\notin I} \omega_{n_k}(\mathcal{D}_{k,i}) \le \gamma\omega_{n_k}(\mathcal{C}_k)$,
(ii) for all $i \in I$, $\omega_{n_k}(\mathcal{D}_{k,i}) \ge \frac{\gamma}{N}\omega_{n_k}(\mathcal{C}_k)$,
(iii) for all i, for all $\mu \in \mathscr{P}$ that is a convex combination of the measures $\Delta_x^{n_k}$ where x runs over $\mathcal{D}_{k,i}$, $|\mathcal{E}(\mu) - E_i| \le \gamma$,
(iv) for all $i \in I$, $|\mathcal{D}_{k,i}| \ge \frac{\gamma}{N}\omega_{n_k}(\mathcal{C}_k)e^{-n_k(E_i+\gamma)}$.

Proof. Since $\mathcal{E}$ is continuous and $\mathscr{P}$ is compact, there exists $\delta > 0$ such that for all $\mu, \nu \in \mathscr{P}$, $\mathrm{W}(\mu,\nu) \le \delta \implies |\mathcal{E}(\mu) - \mathcal{E}(\nu)| \le \gamma$.

Let $S = \{\sigma_1, \dots, \sigma_N\}$ be a δ-covering of $(\mathscr{P}, \mathrm{W})$ and set $E_i := \mathcal{E}(\sigma_i)$. For each $\mu \in \mathscr{P}$ we can define $i(\mu) = \min\{i \mid \mathrm{W}(\mu,\sigma_i) \le \delta\}$. We then set $V_i = \{\mu \in \mathscr{P} \mid i(\mu) = i\}$; the (V_i) form a partition of $\mathscr{P}$, and for all $\mu \in V_i$ we have $|\mathcal{E}(\mu) - E_i| \le \gamma$.

For all k, i, let $\mathcal{D}_{k,i} = \{x \in \mathcal{C}_k \mid \Delta_x^{n_k} \in V_i\}$. Up to extracting a subsequence, we can assume that for each $i \in [\![1, N]\!]$, either $\omega_{n_k}(\mathcal{D}_{k,i}) \ge \frac{\gamma}{N}\omega_{n_k}(\mathcal{C}_k)$ for all k, or $\omega_{n_k}(\mathcal{D}_{k,i}) < \frac{\gamma}{N}\omega_{n_k}(\mathcal{C}_k)$ for all k. Let I be the set of indices i belonging to the first category. We have obtained the first two items; note that for any k we have $\omega_{n_k}(\mathcal{C}_k) = \sum_i \omega_{n_k}(\mathcal{D}_{k,i})$, so that I must be non empty.

Consider a probability measure $\mu = \sum_{x\in\mathcal{D}_{k,i}} a_x\Delta_x^{n_k}$; then $\mathrm{W}(\mu,\sigma_i) \le \delta$: indeed, we have for each $x \in \mathcal{D}_{k,i}$ a coupling $\pi_x \in \Gamma(\Delta_x^{n_k}, \sigma_i)$ of cost at most δ, and the cost of the coupling $\sum_x a_x\pi_x \in \Gamma(\mu,\sigma_i)$ is thus at most δ. As a consequence, $|\mathcal{E}(\mu) - E_i| \le \gamma$.

Given $i \in I$, combining both previous items yields:

$$\frac{\gamma}{N}\omega_{n_k}(\mathcal{C}_k) \le \omega_{n_k}(\mathcal{D}_{k,i}) \le |\mathcal{D}_{k,i}|e^{n_k(E_i+\gamma)}$$

so that $|\mathcal{D}_{k,i}| \ge \frac{\gamma}{N}\omega_{n_k}(\mathcal{C}_k)e^{-n_k(E_i+\gamma)}$. □

Lemma 2.8. *Using the notations of the previous lemma, fix any $i \in I$ and assume $\frac{\log(\omega_{n_k}(\mathcal{C}_k))}{n_k}$ converges as $k \to +\infty$ (this induces no loss in generality, since we already extracted subsequences and can do it once more). Define a sequence of probability measures by*

$$\tilde{\mu}_k = \frac{1}{\omega_{n_k}(\mathcal{D}_{k,i})} \sum_{x\in\mathcal{D}_{k,i}} e^{n_k\mathcal{E}(\Delta_x^{n_k})}\Delta_x^{n_k}$$

If T is continuous, then any accumulation points $\tilde{\mu}_\infty$ of this sequence is T-invariant and satisfies $\Pi(\tilde{\mu}_\infty) \ge \lim_k \frac{\log(\omega_{n_k}(\mathcal{C}_k))}{n_k} - 5\gamma$.

The sequence given by

$$\mu_k = \frac{1}{|\mathcal{D}_{k,i}|} \sum_{x \in \mathcal{D}_{k,i}} \Delta_x^{n_k}$$

could be preferred to $(\tilde{\mu}_k)_k$ for the proof of ③, and can be treated in pretty much the same way. However, we will need $(\tilde{\mu}_k)_k$ in Section 3 to describe the accumulation points of Gibbs ensembles.

Proof. Let first $\tilde{\mu}_\infty$ be an accumulation point of $(\tilde{\mu}_k)$; up to extracting a further subsequence, we assume $\tilde{\mu}_\infty = \lim_k \tilde{\mu}_k$.

To check that $\tilde{\mu}_\infty \in \mathscr{P}(T)$, first observe that $\mathrm{W}(\Delta_x^{n_k}, T_*\Delta_x^{n_k}) \leq \frac{\operatorname{diam} X}{n_k}$ by the total variation bound (i.e., using a coupling that leaves the common part $\sum_{1 \leq j < n_k} \delta_{T^j x}$ in place and moves the remaining mass $\frac{1}{n_k}$ from x to $T^{n_k}x$) and conclude using an averaged coupling as in the proof of Lemma 2.7 above that $\mathrm{W}(\tilde{\mu}_k, T_*\tilde{\mu}_k) \to 0$. Up to this point, no use was made of the continuity assumption on T. But we want to pass to the limit in the arguments of W, and the continuity ensures that $T_*\tilde{\mu}_k \to T_*\tilde{\mu}_\infty$. Then we get $\mathrm{W}(\tilde{\mu}_\infty, T_*\tilde{\mu}_\infty) = 0$, and thus $\tilde{\mu}_\infty \in \mathscr{P}(T)$. Note also that $\mathcal{E}(\tilde{\mu}_k) \leq E_i + \gamma$ for all k, so that the same holds for $\tilde{\mu}_\infty$.

Consider a partition α of X whose element have diameter at most ε and whose boundaries have zero measure with respect to $\tilde{\mu}_\infty$. Setting

$$\sigma_k = \frac{1}{\omega_{n_k}(\mathcal{D}_{k,i})} \sum_{x \in \mathcal{D}_{k,i}} e^{n_k \mathcal{E}(\Delta_x^{n_k})} \delta_x$$

we have $\tilde{\mu}_k = \frac{1}{n_k} \sum_{j=0}^{n_k - 1} T_*^j \sigma_k$ and, since $\mathcal{D}_{k,i}$ is (ε, n_k)-separated,

$$H_{\sigma_k}(\alpha^{n_k}) = \sum_{x \in \mathcal{D}_{k,i}} p_x \log \frac{1}{p_x} \qquad \text{where } p_x = \frac{e^{n_k \mathcal{E}(\Delta_x^{n_k})}}{\omega_{n_k}(\mathcal{D}_{k,i})}$$

$$\log \frac{1}{p_x} \geq \log \left(|\mathcal{D}_{k,i}| e^{n_k(E_i - \gamma)} \right) - n_k(E_i + \gamma)$$

$$H_{\sigma_k}(\alpha^{n_k}) \geq \log |\mathcal{D}_{k,i}| - 2n_k\gamma.$$

Lemma 2.6 applied to $\mathcal{D}_{k,i}$ asserts that $H_{\tilde{\mu}_k}(\alpha^m) \geq \frac{m}{n_k} \log |\mathcal{D}_{k,i}| - 2\gamma m - \frac{2m^2}{n_k} \log |\alpha|$ for all $m \in \mathbb{N}$ and all k such that $n_k \geq 2m$. It follows that for all m and all k

large enough (then taking successive limits as $k \to \infty$ and $m \to \infty$):

$$\begin{aligned}
\frac{1}{m} H_{\tilde{\mu}_k}(\alpha^m) &\geq \frac{1}{n_k} \log |\mathcal{D}_{k,i}| - 3\gamma \\
&\geq \frac{\log \omega(\mathcal{C}_k)}{n_k} - E_i - 4\gamma - \frac{\log(N/\gamma)}{n_k} \\
\frac{1}{m} H_{\tilde{\mu}_\infty}(\alpha^m) &\geq \lim_k \frac{\log \omega(\mathcal{C}_k)}{n_k} - E_i - 4\gamma \\
h(T, \tilde{\mu}_\infty) &\geq \lim_k \frac{\log \omega(\mathcal{C}_k)}{n_k} - \mathcal{E}(\mu_\infty) - 5\gamma \\
\Pi(\mu_\infty) &\geq \lim_k \frac{\log \omega(\mathcal{C}_k)}{n_k} - 5\gamma.
\end{aligned}$$

□

Proposition 2.9. *If T is continuous, then we have* $\sup_{\mu \in \mathscr{P}(T)} \Pi(\mu) \geq \Pi_{\mathrm{top}}$.

Proof. Let $\gamma > 0$, and choose $\varepsilon > 0$ small enough to ensure

$$\Pi_{\mathrm{top}}(\varepsilon) := \limsup_{n \to \infty} \frac{1}{n} \log \zeta(\varepsilon, n) \geq \Pi_{\mathrm{top}} - \gamma.$$

For each $n \in \mathbb{N}$, let $\mathcal{C}_n$ be an (ε, n)-separated subset of X realizing $\zeta(\varepsilon, n)$. Let $(n_k)_k$ be a sequence of integers such that $n_k \to \infty$ and $\frac{1}{n_k} \log \zeta(\varepsilon, n_k) \to \Pi_{\mathrm{top}}(\varepsilon)$.

We apply Lemma 2.7, fix any $i \in I$, define $\tilde{\mu}_k$ as in Lemma 2.8 and let $\tilde{\mu}_\infty$ be any of its accumulation points. We then have

$$\Pi(\tilde{\mu}_\infty) \geq \lim \frac{\log \omega_{n_k}(\mathcal{C}_k)}{n_k} - 5\gamma = \Pi_{\mathrm{top}}(\varepsilon) - 5\gamma \geq \Pi_{\mathrm{top}} - 6\gamma.$$

Letting γ go to zero ends the proof. □

Assuming T is continuous and abundance of ergodic measures, we have shown that:

$$\Pi_{\mathrm{top}} \leq \sup_{\mu \in \mathscr{P}(T)} \Pi(\mu) \leq \underline{\Pi}_{\mathrm{top}}.$$

Since, obviously, $\underline{\Pi}_{\mathrm{top}} \leq \Pi_{\mathrm{top}}$, the above inequalities must be equalities. This proves eq. (2.2) under the assumptions of Theorem A.

2.3. **Proof of Theorem A: the expansive case.** We assume that T is a homeomorphism admitting the expansivity constant $\varepsilon_0 > 0$. To begin with, we let $0 < \varepsilon \leq \varepsilon_0$ and show that

$$\Pi_{\mathrm{top}}(\varepsilon) = \Pi_{\mathrm{top}}(\varepsilon_0) := \limsup_{n \to \infty} \frac{1}{n} \log \zeta(\varepsilon_0, n). \tag{2.10}$$

Let us prove that $\Pi_{\mathrm{top}}(\varepsilon) \leq \Pi_{\mathrm{top}}(\varepsilon_0)$ by extracting an (ε_0, n)-separated set from an (ε, n)-separated one and comparing their weights.

We first fix $\gamma > 0$ arbitrarily small. By the uniform continuity of $\mathcal{E}$ on $\mathscr{P}$, there is $0 < \delta \leq 2\varepsilon$ such that

$$W(\mu, \nu) < \delta \implies |\mathcal{E}(\mu) - \mathcal{E}(\nu)| < \gamma. \tag{2.11}$$

We need the following version of the Theorem of uniform expansivity.

Claim 2.12. *There exists $N \geq 1$ such that for all $n \geq 2N$, for any $x \in X$,*

$$\forall N \leq k < n - N \quad \operatorname{diam}(T^k(B(x, \varepsilon_0, n))) < \delta/2 \leq \varepsilon. \tag{2.13}$$

Proof of the Claim. If this does not hold, pick for every N: $n_N \geq 2N$, $N \leq k_N \leq n_N - N$ and x_N such that

$$\operatorname{diam}(T^{k_N}(B(x_N, \varepsilon_0, n_N))) \geq \delta/2.$$

Pick N_0 and $N \geq N_0$. Note the following inclusions:

$$\begin{aligned} B(T^{k_N}(x_N), \varepsilon_0, N_0) &\supset B(T^{k_N}(x_N), \varepsilon_0, N)) \\ &\supset B(T^{k_N}(x_N), \varepsilon_0, n_N - N) \supset T^{k_N}(B(x_N, \varepsilon_0, n_N)). \end{aligned}$$

Then, consider any accumulation point y for $y_N := T^{k_N}(x_N)$. This yields

$$\forall N_0, \ \operatorname{diam}(B(y, \varepsilon_0, N_0)) \geq \delta/2.$$

This is in contraction with the fact that ε_0 is an expansivity constant. □

We now fix some finite $(\varepsilon/2, N)$-cover C_ε of X and some large enough integer $n \geq 1$ (exactly how large will be specified later on; in particular we assume equation (2.13) holds).

Given an arbitrary nonempty (ε, n)-separated subset S of X, we consider $\widehat{S}$ any (ε_0, n)-separated subset of S, maximal for inclusion.

Claim 2.14. *The following facts hold:*

(i) For every $x \in S$, $B(x, \varepsilon_0, n) \cap \widehat{S}$ is nonempty;
(ii) For every $x \in S$ and every $y \in B(x, \varepsilon_0, n)$, $|\mathcal{E}(\Delta_x^n) - \mathcal{E}(\Delta_y^n)| \leq \gamma$.
(iii) For every $y \in \widehat{S}$, $1 \leq |B(y, \varepsilon_0, n) \cap S| \leq |C_\varepsilon|^2$;

Proof of the claim. To see that (i) holds, note that, if for some $x \in S$, $B(x, \varepsilon_0, n) \cap \widehat{S} = \emptyset$, $\widehat{S} \cup \{x\}$ would still be (ε_0, n)-separated, contradicting the maximality of $\widehat{S}$.

To prove (ii), let x, y be any two points of X with $y \in B(x, \varepsilon_0, n)$. By eq. (2.13), $d(T^k x, T^k y) < \delta/2$ for all $N \leq k < n - N$, hence we get:

$$W(\Delta_x^n, \Delta_y^n) \leq \frac{1}{n} \sum_{k=0}^{n-1} d(T^k x, T^k y) \leq \frac{2N}{n} \operatorname{diam}(X) + \frac{\delta}{2} < \delta$$

for large enough n. The claim (ii) now follows from eq. (2.11).

We turn to (iii). Since $\widehat{S} \subset S$, $y \in B(y, \varepsilon_0, n) \cap S$ so this set is not empty. To prove the upper bound let $I : B(y, \varepsilon_0, n) \cap S \to C_\varepsilon \times C_\varepsilon$ satisfy $I(z) = (w, w')$ with $w \in B(z, \varepsilon/2, N)$ and $w' \in B(T^{n-N} z, \varepsilon/2, N)$. Observe that such a map exists since C_ε is a $(\varepsilon/2, N)$-cover of X and let us check that I is injective. Indeed, let $z, z' \in B(y, \varepsilon_0, n) \cap S$ with $I(z) = I(z') =: (w, w')$ and note:

- for all $0 \leq k < N$, $d(T^k z, T^k z') \leq d(T^k z, T^k w) + d(T^k w, T^k z') < \varepsilon$;
- for all $N \leq k < n - N$, $d(T^k z, T^k z') < \varepsilon$ from eq. (2.13);

- for all $n - N \leq k < n$,

$$d(T^k z, T^k z') \leq d(T^k z, T^{k-(n-N)} w') + d(T^{k-(n-N)} w', T^k z') < \varepsilon.$$

Thus $z, z' \in S$ are not (ε, n)-separated and thus must be equal, proving the injectivity of the map I, proving (iii). Claim 2.14 is established. □

We now compare the weights of S and $\widehat{S}$:

$$\begin{aligned}
\omega_n(\widehat{S}) &= \sum_{y\in\widehat{S}} e^{n\mathcal{E}(\Delta_y^n)} \geq \sum_{y\in\widehat{S}} \min_{x\in B(y,\varepsilon_0,n)\cap S} e^{n\mathcal{E}(\Delta_x^n)} && \text{since } \widehat{S}\subset S\\
&\geq \sum_{y\in\widehat{S}} \frac{e^{-\gamma n}}{|B(y,\varepsilon_0,n)\cap S|} \sum_{x\in B(y,\varepsilon_0,n)\cap S} e^{n\mathcal{E}(\Delta_x^n)} && \text{by eq. } (ii)\\
&\geq \sum_{y\in\widehat{S}} \frac{e^{-\gamma n}}{|C_\varepsilon|^2} \sum_{x\in B(y,\varepsilon_0,n)\cap S} e^{n\mathcal{E}(\Delta_x^n)} && \text{by eq. } (iii)\\
&\geq \frac{e^{-\gamma n}}{|C_\varepsilon|^2} \sum_{x\in S} |B(x,\varepsilon_0,n)\cap\widehat{S}|\, e^{n\mathcal{E}(\Delta_x^n)} && \text{exchanging the sums}\\
&\geq \frac{e^{-\gamma n}}{|C_\varepsilon|^2}\omega_n(S) && \text{by eq. } (i).
\end{aligned}$$

Therefore, $\frac{1}{n}\log\zeta(\varepsilon_0,n) \geq \frac{1}{n}\log\omega_n(\widehat{S}) \geq \frac{1}{n}\log\zeta(\varepsilon,n) - \gamma - \frac{1}{n}\log|C_\varepsilon|^2$. Hence,

$$\Pi_{\mathrm{top}}(\varepsilon) := \limsup_{n\to\infty}\frac{1}{n}\log\zeta(\varepsilon,n) \leq \limsup_{n\to\infty}\frac{1}{n}\log\zeta(\varepsilon,n) + \gamma =: \Pi_{\mathrm{top}}(\varepsilon_0) + \gamma$$

as $\gamma > 0$ was arbitrary we obtain: $\Pi_{\mathrm{top}}(\varepsilon) \leq \Pi_{\mathrm{top}}(\varepsilon_0)$ for all $0 < \varepsilon \leq \varepsilon_0$. The definitions immediately yield the inequality $\zeta(\varepsilon_0,n) \leq \zeta(\varepsilon,n)$ and therefore $\Pi_{\mathrm{top}}(\varepsilon) = \Pi_{\mathrm{top}}(\varepsilon_0)$. This proves (2.10).

The same argument applies to $\underline{\Pi}_{\mathrm{top}}(\varepsilon) := \liminf_{n\to\infty}\frac{1}{n}\log\zeta(\varepsilon,n)$ yielding: $\underline{\Pi}_{\mathrm{top}}(\varepsilon) = \underline{\Pi}_{\mathrm{top}}(\varepsilon_0)$. By eq. (2.2), $\lim_{\varepsilon\to 0}\Pi_{\mathrm{top}}(\varepsilon) = \lim_{\varepsilon\to 0}\underline{\Pi}_{\mathrm{top}}(\varepsilon)$. Thus, $\Pi_{\mathrm{top}}(\varepsilon) = \underline{\Pi}_{\mathrm{top}}(\varepsilon)$ for all $0 < \varepsilon \leq \varepsilon_0$: the upper and lower limits of $\frac{1}{n}\log\zeta(\varepsilon,n)$ as n goes to ∞ coincide. Thus, we have a true limit, independently of $\varepsilon \in (0,\varepsilon_0)$:

$$\Pi_{\mathrm{top}} = \lim_{n\to\infty}\frac{1}{n}\log\zeta(\varepsilon,n),$$

concluding the proof of Theorem A.

3. Existence of an equilibrium measure and convergence of the Gibbs ensembles

In this section we prove Theorem B. Its existence claim is a simple consequence of the variational principle we just established as Theorem A.

Lemma 3.1. *Assume that T is continuous with $\mu \mapsto h(T,\mu)$ upper semicontinuous, and that $(T,\mathcal{E})$ has an abundance of ergodic measures. Then the set $\mathscr{EM}$ of nonlinear equilibrium measures is non-empty.*

Moreover, for all $\gamma > 0$ there exists $\delta > 0$ such that invariant measures achieving $\Pi^{\mathcal{E}}_{\text{top}}(T)$ up to δ are γ-close to $\mathscr{EM}$: for all $\mu \in \mathscr{P}(T)$ such that $h(T,\mu) + \mathcal{E}(\mu) > \Pi^{\mathcal{E}}_{\text{top}}(T) - \delta$, there exist $\mu' \in \mathscr{EM}$ such that $\mathrm{W}(\mu,\mu') < \gamma$.

Proof. By assumption $\mu \mapsto h(T,\mu) + \mathcal{E}(\mu)$ is upper semi-continuous on the compact set $\mathscr{P}(T)$, it must therefore reach its maximum, which by Theorem A is $\Pi^{\mathcal{E}}_{\text{top}}(T)$. $\mathscr{EM}$ must thus be non-empty, and compact.

Given $\gamma > 0$, the upper semi-continuous function $\mu \mapsto h(T,\mu) + \mathcal{E}(\mu)$ also reaches its maximum F_γ on the compact set $\{\mu \in \mathscr{P}(T) \colon \mathrm{W}(\mu, \mathscr{EM}) \geq \gamma\}$. Since this set is disjoint from $\mathscr{EM}$, $F_\gamma < \Pi^{\mathcal{E}}_{\text{top}}(T)$. The positive number $\delta = \Pi^{\mathcal{E}}_{\text{top}}(T) - F_\gamma$ has the desired property. □

The second part of Theorem B is proven along the same lines than Proposition 2.9.

Proposition 3.2. *Assume that T is an expansive homeomorphism with $\varepsilon_0 > 0$ an expansivity constant, and that $(T,\mathcal{E})$ has an abundance of ergodic measures. Let ν be an accumulation point of (ε_0, n)-Gibbs ensembles as $n \to \infty$. Then ν can be approximated in the weak-star topology by linear combinations of nonlinear equilibrium measures.*

Proof. Note that T being an expansive homeomorphism, entropy is upper semi-continuous. By the second half of Lemma 3.1, we are reduced to approximate ν by convex combination of measures that almost achieve the nonlinear topological pressure.

By definition ν is the limit of measures of the form

$$\mu_k = \frac{1}{\zeta(\varepsilon_0, n_k)} \sum_{x \in \mathcal{C}_k} e^{n_k \mathcal{E}(\Delta_x^{n_k})} \Delta_x^{n_k}$$

where $n_k \to \infty$, $\mathcal{C}_k$ are (ε_0, n_k)-separated sets with $\omega_{n_k}(\mathcal{C}_k) = \zeta(\varepsilon_0, n_k)$. Fix some $\gamma > 0$.

Apply Lemma 2.7, providing $N \in \mathbb{N}$, partitions $\mathcal{D}_{k,1}, \ldots, \mathcal{D}_{k,N}$ of each $\mathcal{C}_k$ and $I \subset [\![1, N]\!]$ such that up to further extracting a subsequence (still denoted by $(n_k)_k$), $\sum_{i \notin I} \omega_{n_k}(\mathcal{D}_{k,i}) < \gamma \omega_{n_k}(\mathcal{C}_k)$ and for all $i \in I$, Lemma 2.8 applies.

For each i, consider

$$\tilde{\mu}_k^i = \frac{1}{\omega_{n_k}(\mathcal{D}_{k,i})} \sum_{x \in \mathcal{D}_{k,i}} e^{n_k \mathcal{E}(\Delta_x^{n_k})} \Delta_x^{n_k}$$

and assume, up to further extraction, that it converges as $k \to \infty$ to some $\tilde{\mu}^i$. Then by Lemma 2.8, whenever $i \in I$:

$$\Pi(\tilde{\mu}^i) \geq \lim \frac{\log(\omega_{n_k}(\mathcal{C}_k))}{n_k} - 5\gamma = \Pi_{\text{top}} - 5\gamma.$$

The $\tilde{\mu}^i$ with $i \in I$ are the almost equilibrium measures we are looking for.

We have $\mu_k = \sum_{i=1}^N a_k^i \tilde{\mu}_k^i$ where $a_k^i = \frac{\omega_{n_k}(\mathcal{D}_{k,i})}{\omega_{n_k}(\mathcal{C})}$. Up to a further extraction, we can assume that for each i the sequence $(a_k^i)_k$ converges to some number

$a_i \in [0,1]$. It follows that

$$\nu = \sum_{i=0}^{N} a_i \tilde{\mu}_i = \sum_{i \in I} a_i \tilde{\mu}_i + \sum_{i \notin I} a_i \tilde{\mu}_i$$

Note that $\sum_{i \notin I} a_i \leq \gamma$, i.e., the second term above has total mass less than γ. For each $i \in I$ we set $b_i = a_i / \sum_{i \in I} a_i$, so that $\tilde{\mu} = \sum_{i \in I} b_i \tilde{\mu}_i$ is a convex combination of almost equilibrium states, and by the total variation bound $\mathrm{W}(\nu, \tilde{\mu}) = O(\gamma)$. □

Theorem B is established.

4. Convexity and nonlinear equilibrium measures

In this section, independently of Sections 2.2 and 3, we prove an extended version of Theorem C, i.e., we study the nonlinear formalism for an energy with potentials. Specifically, we consider a continuous map $T : X \to X$ with finite entropy $h_{\mathrm{top}}(T) < \infty$ together with an energy defined as

$$\mathcal{E}(\mu) = F\big(\mu(\varphi_1), \dots, \mu(\varphi_d)\big)$$

for all $\mu \in \mathscr{P}(T)$ where, for some positive integer d,

- $\varphi_1, \dots, \varphi_d : X \to \mathbb{R}$ are continuous functions called the *potentials*;
- $F : U \to \mathbb{R}$ is a smooth function called the *nonlinearity*.

Here we assume that $U \subset \mathbb{R}^d$ is an open set containing the compact *rotation set*

$$\rho(\varphi_1, \dots, \varphi_d) := \big\{ \big(\mu(\varphi_1), \dots, \mu(\varphi_d)\big) : \mu \in \mathscr{P}(T) \big\}.$$

It will sometimes be convenient to write the potentials as a single vector-valued function: $\vec{\varphi} := (\varphi_1, \dots, \varphi_d)$, $\vec{\varphi}(x) := (\varphi_1(x), \dots, \varphi_d(x))$, $\mu(\vec{\varphi}) := (\mu(\varphi_1), \dots, \mu(\varphi_d))$, etc.

We are going to study the nonlinear equilibrium measures:

$$\mathscr{EM} := \big\{ \mu \in \mathscr{P}(T) : h(\mu) + F(\mu(\vec{\varphi})) \text{ is maximal} \big\}$$

Remark 4.1. *If one would like to apply our general results (the variational principle of Theorem A and the equidistribution of Gibbs ensembles of Theorem B), then one should demand $\mathcal{E}(\mu)$ to be defined for all (non-necessarily invariant) probability measures, i.e., the open set U should contain the convex hull of $\{\vec{\varphi}(x) : x \in X\}$.*

The rest of this section is divided as follows. First, we introduce a "fully nonlinear formalism" which is the natural setting of our technique and describe the entropy-potential diagram which is a useful visualization. Second we recall the relevant background concerning Legendre duality and we set up appropriate definitions to use this duality and we provide examples of dynamical system satisfying them. Thirdly we weave all this together and apply Legendre duality in the dynamical context to reach the main goal of this section, Theorem 4.15 (which contains Theorem C). Finally we deduce some uniqueness results (Corollary 4.19, Propositions 4.20 and 4.21).

4.1. **Fully nonlinear pressure.** Our approach applies to the following more general setting:

Definition 4.2. *Given a continuous system T with potentials $\vec{\varphi}$, a* fully nonlinear pressure *is a function*

$$\mathbf{\Pi}^G(\mu, \vec{\varphi}) := G\big(h(\mu); \mu(\varphi_1), \dots, \mu(\varphi_d)\big) \tag{4.3}$$

defined for all $\mu \in \mathscr{P}(T)$ by some smooth $G : V \to \mathbb{R}$ assumed to be admissible*: it is defined on an open subset V of $\mathbb{R} \times \mathbb{R}^d$ and satisfies:*[4]

$$\partial_0 G > 0 \text{ and } V \supset \{(h(T,\mu), \mu(\varphi_1), \dots, \mu(\varphi_d)) : \mu \in \mathscr{P}(T)\}.$$

The corresponding set of fully nonlinear equilibrium measures *is then:*

$$\mathscr{EM}(T, G, \vec{\varphi}) := \{\mu \in \mathscr{P}(T) : \mathbf{\Pi}^G(\mu, \vec{\varphi}) \text{ is maximal }\}.$$

We will reduce the problem of maximizing $\mathbf{\Pi}^G$ to the classical, linear thermodynamical formalism by justifying the following claims:

(*) given $z \in \rho(\vec{\varphi})$, maximizing $\mathbf{\Pi}^G$ and maximizing the linear pressure over
$$\mathcal{M}(z) := \{\mu \in \mathscr{P}(T) : \mu(\vec{\varphi}) = z\}$$
are both equivalent to maximizing the entropy there;

(**) the values $z = \mu(\vec{\varphi})$ realized by fully nonlinear equilibrium measures μ belong to the interior of rotation set $\rho(\vec{\varphi})$;

(***) there is a diffeomorphism $\mathrm{int}(\rho(\vec{\varphi})) \to \mathbb{R}^d$, $z \mapsto y$, such that, for every $z \in \mathrm{int}(\rho(\vec{\varphi}))$, there is a linear equilibrium measure ν_y for the potential
$$y \cdot \vec{\varphi} := \sum_j y_j \varphi_j$$
with $\nu_y(\vec{\varphi}) = z$.

The first point is immediate given the assumption that $\partial_0 G > 0$. The second and third point will follow from some convex analysis; the second point more precisely follows from the assumption that the gradient of entropy diverges at the boundary in the definition of C^r Legendre systems (Definitions 4.9 and 4.7) , see the proof of Theorem 4.15.

4.2. **The entropy-potential diagram and the entropy function.** In light of the above remark (*), we will use the following geometric viewpoint. The *entropy-potential diagram*, illustrated by Figure 1, is the set

$$\mathcal{D} = \big\{(z_0; z_1, \dots, z_d) \in [0, +\infty) \times \mathbb{R}^d : \exists \mu \in \mathscr{P}(T),\ h(T,\mu) \geq z_0,\ \forall i,\ \mu(\varphi_i) = z_i\big\},$$

$\mathcal{D}$ can be seen as the hypograph of the entropy function, see the function h below.

Since the Kolmogorov-Sinai entropy is affine, $\mathcal{D}$ is a convex set, and the linear pressure associated to any linear combination $\sum_i y_i \varphi_i$ can be recovered from $\mathcal{D}$ by finding the unique[5] support hyperplane with normal vector $(1; y_1, \dots, y_d)$; this

[4]The notation $\partial_0 G$ refers to $\partial G / \partial z_0$, the derivative with respect to the first variable, corresponding to entropy since the coordinates are numbered as $(z_0, z_1, \dots, z_d)$.

[5]Since we fix the normal vector, uniqueness here does not depend on smoothness of $\mathcal{D}$; it is the contact points that may be non-unique, if strict convexity is not assumed.

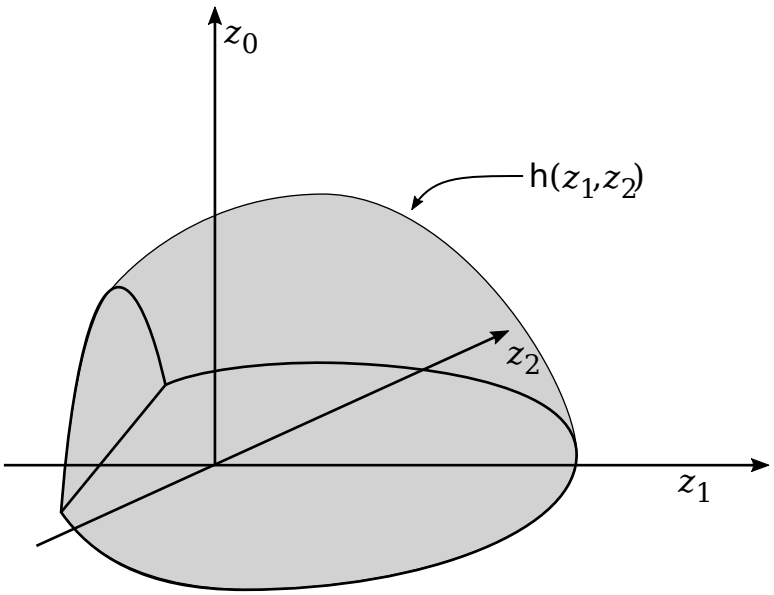

FIGURE 1. An entropy-potentials diagram in two dimensions (first coordinate represented by the vertical axis), in a case when the rotation set is not strictly convex.

has important consequences, see Proposition 4.14. Note that the convexity of $\mathcal{D}$ translates into the concavity of the following function, which is finite exactly on $\rho(\vec{\varphi})$:

Definition 4.4. *Given a continuous dynamical system T with potentials $\vec{\varphi}$, the* (finite-dimensional) entropy function $\mathsf{h} : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ *is*[6]

$$\mathsf{h}(z) := \sup_{\mu \in \mathcal{M}(z)} h(T, \mu).$$

Under our standing assumptions (X compact, $\vec{\varphi}$ continuous, and $h_{\mathrm{top}}(T) < \infty$), we have $\{z \in \mathbb{R}^d : \mathsf{h}(z) \neq -\infty\} = \rho(\vec{\varphi})$.

Remark 4.5. *To find the largest value of $\mathsf{h} + F$ is to find the largest k such that there exists $z \in \rho(\vec{\varphi})$ at which $\mathsf{h}(z) = -F(z) + k$, i.e., to find the highest vertical translate of the graph of $-F$ that touches the entropy-potential diagram. This makes apparent that the nonlinear equilibrium measures will correspond to linear equilibrium measures associated to one or several linear combinations of potentials, whose coefficients are given by the equations of the tangent hyperplanes at the touching points, see e.g., figure 3.*

4.3. **Legendre duality.** To apply the well-rounded theory of Legendre duality, let us introduce its classical assumptions, following [25].

[6]The usual convention $\sup(\varnothing) = -\infty$ is understood.

Recall that the *Legendre transform* f^* of a convex function $f : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ is the convex function:

$$f^* : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}, \quad y \mapsto \sup_{x \in \mathbb{R}^d} \langle y; x \rangle - f(x).$$

If g is a concave function, we set

$$g^\# : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}, \quad y \mapsto \sup_{x \in \mathbb{R}^d} \langle y; x \rangle + g(x),$$

i.e., $g^\# := (-g)^*$, which is convex.[7]

We will use two classical duality results from [25]. They ensure that the Legendre transform is an involution on suitable classes of semicontinuous or smooth convex functions.

Semicontinuous functions. A function is *proper* if it is finite at least at one point.

Theorem 4.6. *The Legendre transform maps bijectively the class of upper semicontinuous,*[8] *proper concave functions to the class of lower semicontinuous proper convex functions. Moreover, this restriction of the Legendre transform is an involution up to sign: for all such f, $f = -(f^\#)^*$.*

The above theorem implies that the Legendre transform is an involution over the class of lower semicontinuous proper convex functions g : $(g^*)^* = g$.

Smooth functions. We consider the smoothness classes C^r for $1 \leq r \leq \omega$, i.e., for any positive integer r as well as $r = \infty$ (infinitely differentiable) and $r = \omega$ (real-analytic). The following abuses of notation will be convenient: for $r = \infty$ or ω, C^{r-1} just means C^r; for $r = 0$, a C^r diffeomorphism is a homeomorphism.

Definition 4.7. *Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ be a function. Its* (effective) domain *is the set of points* $\mathrm{dom}(f)$ *in $\mathbb{R}^d$ where it takes a finite value:* $\mathrm{dom}(f) = f^{-1}(\mathbb{R})$. *For $1 \leq r \leq \omega$, the function f is said to be* concave of C^r Legendre type *when the following conditions are satisfied:*

- *(i) the function f is upper semicontinuous and concave;*
- *(ii) the interior* $\mathrm{int}\,\mathrm{dom}(f)$ *is not empty and, on this set, f is strictly concave and C^r smooth; when $r \geq 2$, we additionally ask that the Hessian of f is everywhere negative definite;*
- *(iii) for all sequences $(x_i)_{i \in \mathbb{N}}$ with $x_i \in \mathrm{int}(\mathrm{dom}(f))$ which converge to a boundary point of* $\mathrm{dom}(f)$*,*

$$\lim_i |\nabla f(x_i)| = +\infty.$$

We say that a function $g : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ is convex of C^r Legendre type if $-g$ is concave of C^r Legendre type.

[7] Sometimes, the Legendre transform of a concave function is defined as $-(-g)^*$ instead, so that it is again concave.

[8] In [25], lower semicontinuous convex functions are called *closed*.

Note that functions are convex of C^1 Legendre type exactly when they are convex of Legendre type in the sense of Rockafellar [25, Chap. 26]. Let us now extract the following result from the classical theory of Legendre duality.

Theorem 4.8. *For each $1 \leq r \leq \omega$, the Legendre transform of any concave or convex function f of C^r Legendre type is a convex function $f^\#$ or f^* of C^r Legendre type. Moreover, the following holds for f concave:*[9]

(i) $\nabla f : \operatorname{int}(\operatorname{dom}(f)) \to \operatorname{int}(\operatorname{dom}(f^\#))$ *is a C^{r-1}-diffeomorphism;*
(ii) for all $y \in \operatorname{int}(\operatorname{dom}(f^\#))$,

$$\nabla f^\#(y) = (\nabla f)^{-1}(-y) \text{ and } f^\#(y) = z \cdot y + f(z) \text{ with } z = (\nabla f)^{-1}(-y);$$

(iii) $(f^\#)^* = -f$.

Proof. This statement follows from the results in [25, Chap. 26], except for the formula for $f^\#(y)$ in (ii). When $r = 1$, this is exactly Theorem 26.5 there applied to the convex function $g = -f$. Indeed, $f^\# = g^*$ and $\nabla f = I \circ \nabla g$ with $I(y) = -y$. In particular, $\nabla g^* = (\nabla g)^{-1}$, i.e., $\nabla f^\# = (I \circ \nabla f)^{-1} = (\nabla f)^{-1} \circ I$, proving the first formula in claim (ii).

Now, ∇f is a C^{r-1} map. From the same theorem, $\nabla f : \operatorname{dom}(f) \to \operatorname{dom}(f^\#)$ is a homeomorphism. It is a C^{r-1}-diffeomorphism, using, if $r \geq 2$, that the Hessian of f is definite. The formula for $\nabla f^\#$ ensures that this gradient is also C^{r-1}, thus $f^\#$ is C^r.

To conclude, let $y \in \operatorname{int}(\operatorname{dom}(f^\#))$. Note that $z := (\nabla f)^{-1}(-y) \in \operatorname{int}(\operatorname{dom}(f))$ satisfies $\nabla_z (y \cdot z + f(z)) = 0$. Since f is strictly concave on $\operatorname{int}(\operatorname{dom}(f))$ and concave everywhere, z must be the unique maximizer on $\operatorname{dom}(f)$, proving the second half of (ii). □

4.4. **Application to dynamical systems.** Before exploiting Legendre duality further, let us discuss how the dynamical systems on which the linear Thermodynamical formalism is well-understood fit into our framework. We start with a convenient definition.

Definition 4.9. *For $1 \leq r \leq \omega$, a continuous dynamical system with potentials $(T, \vec{\varphi})$ is C^r* Legendre *when:*

(i) the rotation set $\rho(\vec{\varphi})$ has non-empty interior in $\mathbb{R}^d$,
(ii) the topological entropy is finite: $h_{\mathrm{top}}(T) < \infty$;
(iii) the finite-dimensional entropy function $\mathsf{h} : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ is concave of C^r Legendre type.

If moreover, for every $y \in \mathbb{R}^d$, there is exactly one linear equilibrium measure ν_y for T and the potential $y \cdot \vec{\varphi}$, then we say that $(T, \vec{\varphi})$ is C^r Legendre with unique linear equilibrium measures $(\nu_y)_{y \in \mathbb{R}^d}$.

[9] For convex f, the same holds for f^* except for the minus signs: $\nabla f^*(y) = (\nabla f)^{-1}(y)$ and $f^{**} = f$.

The above classical theory of Legendre duality applied to such systems leads to the (finite-dimensional linear) *pressure function*:

$$\mathsf{P}(y) := \sup_{\mu \in \mathscr{P}(T)} h(T,\mu) + \mu(y \cdot \vec{\varphi}).$$

It is the Legendre transform of the concave finite-dimensional entropy function h:

$$\mathsf{P}(y) = \mathsf{h}^{\#}(y) := \sup_{z \in \rho(\vec{\varphi})} \mathsf{h}(z) + \langle y; z \rangle.$$

In particular, if $(T, \vec{\varphi})$ is C^r Legendre, then by applying Theorem 4.8 we obtain that the pressure is a C^r function.

In Definition 4.9, we took entropy as primary object, and then defined pressure by Legendre duality. However, it has been customary to discuss primarily the regularity of pressure – using Legendre duality, both points of view can be unified as follows.

Proposition 4.10. *If $(T, \vec{\varphi})$ is a continuous system with potentials satisfying, for some $1 \leq r \leq \omega$,*

- *the rotation set $\rho(\vec{\varphi})$ has nonempty interior in $\mathbb{R}^d$;*
- *the entropy function $h(T, \cdot)$ is upper semicontinuous and bounded over $\mathscr{P}(T)$;*
- *the finite-dimensional pressure function P is finite over $\mathbb{R}^d$, C^r smooth, strictly convex and, when $r \geq 2$, with everywhere positive definite Hessian,*

then $(T, \vec{\varphi})$ is a C^r Legendre system.

Proof. Since the Kolmogorov-Sinai entropy $h : \mathscr{P} \to \mathbb{R} \cup \{-\infty\}$ is upper semicontinuous, $\mathscr{P}$ compact, and $\vec{\varphi}$ is continuous, $\mathsf{h} : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ is upper semicontinuous. This function is also finite on its nonempty domain $\mathrm{dom}(\mathsf{h}) = \rho(\varphi)$ and concave. Therefore, by Theorem 4.6, the lower semicontinuous convex function $-\mathsf{h}$ satisfies: $-\mathsf{h} = ((-\mathsf{h})^*)^* = \mathsf{P}^*$. By assumption P is a convex C^r Legendre function. Applying now Theorem 4.8, we get that $\mathsf{P}^* = -\mathsf{h}$ is a convex C^r Legendre function, i.e., h is concave C^r Legendre. □

It is now easy to check that many classical systems satisfy the thermodynamical formalism with C^ω regularity. In many cases, the one point that needs checking is that the rotation set has non-empty interior (see Section 5.3 for an example where it does not).

Recall that a function φ is *cohomologous to a constant c* if there is a continous function u such that $\varphi = c + u - u \circ T$.

Corollary 4.11. *Let T be a mixing subshift of finite type or an Anosov diffeomorphism. Let $\vec{\varphi}$ be a finite family of Hölder-continuous potentials $\varphi_1, \ldots, \varphi_d : X \to \mathbb{R}$. Assume the following independence condition: for all $\alpha_1, \ldots, \alpha_d$ not all zero, $\sum_{i=1}^d \alpha_i \varphi_i$ is not cohomologous to a constant.*

Then $(T, \vec{\varphi})$ is a C^ω Legendre system with unique linear equilibrium measures.

Remark 4.12. *Livsič theorem applies to such systems: a function is cohomologous to a constant if and only if on each periodic orbit, the average of the function is equal to that constant. The independence condition above is therefore equivalent to the existence of $d+1$ periodic orbits with corresponding atomic measures $\mu_0, \dots, \mu_d \in \mathscr{P}(T)$ such that $\mu_0(\vec{\varphi}), \dots, \mu_d(\vec{\varphi}) \in \mathbb{R}^d$ are affinely independent.*

Proof of the corollary. Both subshifts of finite type and Anosov diffeomorphisms are Smale systems satisfying the regularity condition (SS3) in [26] in the sense of [26, 7.1, 7.11] and this will be enough for our purposes.

Since T has finite topological entropy and is expansive, the Kolmogorov-Sinai entropy function is upper semicontinuous and bounded over $\mathscr{P}(T)$.

If the rotation set, a convex set, had empty interior, it would be contained in some affine hyperplane, hence, there would be numbers $\alpha_0, \dots, \alpha_d$, not all zero, such that

$$\forall \mu \in \mathscr{P}(T) \quad \mu\left(\sum_{i=1}^{d} \alpha_i \phi_i\right) = \sum_{i=1}^{d} \alpha_i \mu(\phi_i) = \alpha_0.$$

By Livsič theorem, this implies that $\sum_{i=1}^{d} \alpha_i \phi_i$ is cohomologuous to the constant α_0, contradicting the independence assumption.

Since T is a topologically mixing Smale system, its pressure function is real-analytic [26, 7.10]. It has a semidefinite positive Hessian with kernel generated by the potentials cohomologous to constants. Hence the finite-dimensional pressure function P has definite positive Hessian in all of $\mathbb{R}^d$ under the independence assumption above. In particular, P is strictly convex.

Thus, the assumptions of Proposition 4.10 are satisfied so that $(T, \vec{\varphi})$ is a C^r Legendre system.

Finally, for each $y \in \mathbb{R}^d$, $y \cdot \vec{\varphi}$ is Hölder-continuous, hence there exists a unique linear equilibrium measure ν_y. □

The next statement follows immediately from [10, Corollary B, Theorems F & G], providing another family (intersecting the previous one) of dynamical systems to apply our framework to. We shall say that a Banach space $\mathscr{X}$ of functions $X \to \mathbb{R}$ is a *good Banach algebra of functions* when:

- $\mathscr{X}$ is stable by product and $\|fg\| \leq \|f\|\|g\|$ for all $f, g \in \mathscr{X}$,
- for every positive, bounded away from 0 function $f \in \mathscr{X}$, $\log f$ is in $\mathscr{X}$,
- the norm of $\mathscr{X}$ dominates the uniform norm (in particular the elements of $\mathscr{X}$ are bounded),
- the composition operator $f \mapsto f \circ T$ is a continuous operator on $\mathscr{X}$,
- for every equilibrium measure μ of a potential in $\mathscr{X}$ and every non-negative $f \in \mathscr{X}$, if $\int f \, \mathrm{d}\mu = 0$ then $f = 0$,
- every continuous function can be uniformly approximated by elements of $\mathscr{X}$.

(These assumptions are numerous, but many Banach spaces satisfy them, such as Hölder spaces or BV space on the interval, see [10] for some discussions of these hypotheses.) We refer to [10] for the notions of k-to-1 map, simple dominant eigenvalue, and spectral gap appearing in the following statement.

Theorem 4.13. *Assume that T is k-to-1 and $\varphi_1, \dots, \varphi_d$ belong to some good Banach algebra of functions $\mathscr{X}$ and that for all $\alpha_1, \dots \alpha_d$ not all zero, $\sum_{i=1}^d \alpha_i \varphi_i$ is not cohomologous to a constant. If for all $y \in \mathbb{R}^d$ the transfer operator defined by $\mathcal{L}f(x) = \sum_{x' \in T^{-1}(x)} e^{y \cdot \vec{\varphi}(x')} f(x')$ acts with a simple dominant eigenvalue and a spectral gap on $\mathscr{X}$, then $(T, \vec{\varphi})$ is C^ω Legendre with unique linear equilibrium measures.*

4.5. **Consequences of Legendre duality.** Now that we have seen that Theorem 4.8 applies to plenty of dynamical systems, let us note some of the consequences.

Proposition 4.14. *If $(T, \vec{\varphi})$ is a C^r Legendre system, then:*

(i) the finite-dimensional function h is continuous on the rotation set $\rho(\vec{\varphi})$,
(ii) $\nabla \mathsf{h}$ realizes a C^{r-1} diffeomorphism from the interior of $\rho(\vec{\varphi})$ onto $\mathbb{R}^d$ with inverse $y \mapsto \nabla \mathsf{P}(-y)$,
(iii) the linear pressure function P has domain $\mathbb{R}^d$ and is C^r,
(iv) for all $y \in \mathbb{R}^d$, $\nabla \mathsf{P}(y) = z_{\text{opt}}$ where z_{opt} is the unique maximizer of $\mathsf{h}(z) + \langle y; z \rangle$ over $\operatorname{int} \rho(\vec{\varphi})$.

If, additionally, $(T, \vec{\varphi})$ has unique equilibrium measures $(\nu_y)_{y \in \mathbb{R}^d}$, then

(v) for all $y \in \mathbb{R}^d$, $\nu_y(\vec{\varphi}) = \nabla \mathsf{P}(y) \in \operatorname{int}(\rho(\vec{\varphi}))$ and $h(T, \nu_y) = \mathsf{h}(\nu_y(\vec{\varphi}))$,
(vi) $\{\nu_y(\vec{\varphi}) : y \in \mathbb{R}^d\} = \operatorname{int}(\rho(\vec{\varphi}))$ and
(vii) conversely, for all $z \in \operatorname{int}(\rho(\vec{\varphi}))$, setting $y := -\nabla \mathsf{h}(z)$, $\nu_y(\vec{\varphi}) = z$ and ν_y is the unique measure of maximum entropy in $\mathcal{M}(z)$.

Proof. The function h is upper-semicontinuous, and since it is concave and finite it must be continuous on its domain, which coincides with the rotation set.

By assumption, h is a concave C^r Legendre function. Hence Theorem 4.8 ensures that the pressure $\mathsf{P} = \mathsf{h}^\#$ is C^r. Since h is upper bounded as a continuous function with a compact domain, the domain of $\mathsf{P}(y) = \sup_{z \in \rho(\vec{\varphi})} \mathsf{h}(z) + \langle y; z \rangle$ is the whole of $\mathbb{R}^d$. The same theorem tells us that $\nabla \mathsf{h}$ realizes a C^r diffeomorphism from the interior of $\rho(\vec{\varphi})$ to $\mathbb{R}^d$, the interior of the domain of P, and that, for all $y \in \operatorname{dom}(\mathsf{P})$,

$$\nabla \mathsf{P}(y) = (\nabla \mathsf{h})^{-1}(-y).$$

We further note that $\mathsf{P}(y) = \langle y \, ; z_{\text{opt}} \rangle + \mathsf{h}(z_{\text{opt}})$ with $z_{\text{opt}} := (\nabla \mathsf{h})^{-1}(-y) = \nabla P(y)$.

We now assume that $(T, \vec{\varphi})$ has unique equilibrium measures $(\nu_y)_{y \in \mathbb{R}^d}$. Let $y \in \mathbb{R}^d$.

Observe that ν_y must maximize the entropy in $\mathcal{M}(z)$ where $z = \nu_y(\vec{\varphi})$, hence $h(T, \nu_y) = \mathsf{h}(z)$. By definition the linear pressure is

$$\mathsf{P}(y) = h(T, \nu_y) + \int y \cdot \vec{\varphi} \, d\nu_y = \mathsf{h}(z) + \langle y \cdot z \rangle.$$

Therefore, in Proposition 4.14, one must have:

$$z = \nu_y(\vec{\varphi}) = \nabla \mathsf{P}(y) \text{ so } y = -\nabla \mathsf{h}(z).$$

This proves items (v) and (vii).

Note that $\{\nu_y(\vec{\varphi}) : y \in \mathbb{R}^d\} = \nabla \mathsf{P}(\mathbb{R}^d)$, which is $\operatorname{int}(\rho(\vec{\varphi}))$, proving (vi). □

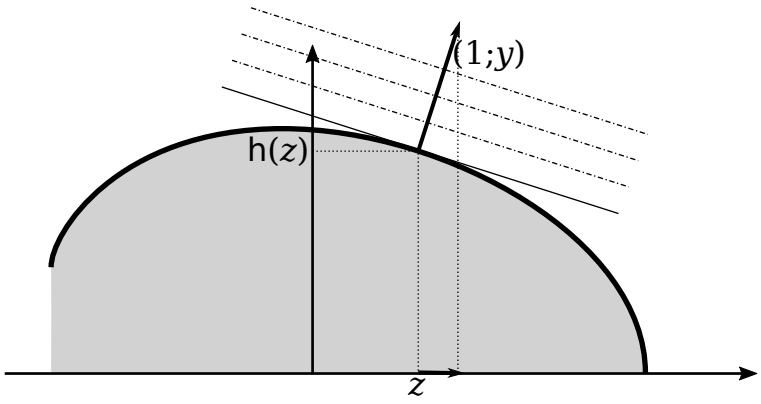

FIGURE 2. An entropy-potential diagram $\mathcal{D}$ represented in the $d = 1$ case (first coordinate z_0 represented by the vertical axis): $\mathsf{P}(y)$ is obtained by sliding a line along the normal vector $(1; y)$ until it touches the hypograph of h, which happens above some z where $\nabla \mathsf{h}(z) = -y$. At this point $\nabla \mathsf{P}(y) = z$: changing the direction y makes the touching line "roll" along the upper side of $\mathcal{D}$; this rolling combines the rotation of y and a normal translation given by scalar product with z. Changed "variation in the amount of sliding" by a hopefully clearer explanation.

4.6. **Set of nonlinear equilibrium measures.** We now identify the fully nonlinear equilibrium measures, that is, the elements of $\mathscr{EM}(T, F, \vec{\varphi})$ (or just $\mathscr{EM}$) from Definition 4.2. We define the set of $(G, \vec{\varphi})$-*equilibrium values* to be

$$\mathscr{V} := \{\mu(\vec{\varphi}) : \mu \in \mathscr{EM}\}.$$

For $z \in \rho(\vec{\varphi})$, recall the notations $\mathcal{M}(z)$ and $\mathsf{h}(z)$ from Definitions 4.4 and 4.9. We start with Theorem C, in a version generalized to fully nonlinear pressures (see Definition 4.2). We recall that G is defined on some open set $V \subset \mathbb{R} \times \mathbb{R}^d$ and in the following $\partial_i G$ stands for $\partial G/\partial z_i$, $i = 0, 1, \dots, d$.

Theorem 4.15. *Let (T, φ) be a C^r Legendre system for some $1 \le r \le \omega$ and let Π^G be a fully nonlinear pressure defined by an admissible C^r function G.*

Then the set $\mathscr{EM}$ of $(G, \vec{\varphi})$-equilibrium measures is a nonempty and compact set of linear equilibrium measures. More precisely,

(i) $\mathscr{V} = \{z \in \operatorname{int}(\rho(\vec{\varphi})) : G(\mathsf{h}(z); z)$ maximal $\}$ is a nonempty compact set on which

$$0 = \nabla G + \partial_0 G \cdot \nabla \mathsf{h} \qquad \text{where } \nabla := (\partial_1, \dots, \partial_d)\,. \tag{4.16}$$

(ii) $\mathscr{EM} = \{\nu_y \colon y \in -\nabla \mathsf{h}(\mathscr{V})\}$.

Proof. We prove assertions (i) and (ii), the rest being immediate consequences.

Let us note that a measure $\mu \in \mathscr{P}(T)$ is a fully nonlinear equilibrium measure if and only if

$$G(h(T,\mu);\mu(\vec{\varphi})) = \sup_{(z_0;z)\in\mathcal{D}} G(z_0;z) = \sup_{z\in\rho(\vec{\varphi})} g(z) \text{ where } g(z) := G(\mathsf{h}(z);z).$$

Indeed, the first equality follows from the definitions and the second one follows from the fact that $z_0 \mapsto G(z_0;z)$ is increasing for each $z \in \rho(\vec{\varphi})$. Since g is continuous on the compact set $\rho(\vec{\varphi})$, it follows that $\mathscr{V}$ is itself compact.

Claim. *Since h is concave with $|\nabla \mathsf{h}| \to \infty$ at the boundary of $\rho(\vec{\varphi})$, we must have $\mathscr{V} \subset \mathrm{int}(\rho(\vec{\varphi}))$.*

Proof of the claim. Consider a point z_0 on the boundary of $\rho(\vec{\varphi})$, and let us prove that it cannot maximize g. Let $\vec{u}$ be any vector such that $z_0 + \vec{u} \in \mathrm{int}(\rho(\vec{\varphi}))$ and consider the function defined on $[0,1]$ by $f(t) = \mathsf{h}(z_0 + t\vec{u})$. By concavity its derivative has a limit, finite or infinite, as $t \to 0$. For all small enough $t > 0$, we have $f'(t) = \langle \nabla \mathsf{h}(z_0 + t\vec{u}), \vec{u}\rangle$. We know that $|\nabla \mathsf{h}| \to \infty$ at the boundary, but it could *a priori* be that $\nabla \mathsf{h}$ becomes orthogonal to $\vec{u}$ as $t \to 0$; we now prove that this cannot be the case.

At each small enough $t > 0$, the tangent space H_t to the upper boundary of $\mathcal{D}$ has $(1, -\nabla \mathsf{h})$ as normal vector. As $t \to 0$, $|\nabla \mathsf{h}| \to \infty$ so that any accumulation point H_0 of H_t is vertical, of the form $\mathbb{R} \times L$ where L is a hyperplane of $\mathbb{R}^d$ (normal to an accumulation point of the direction of $\nabla \mathsf{h}(z_0 + t\vec{u})$). Since $\mathcal{D}$ is contained in a half-space delimited by H_0, L must be a supporting hyperplane of $\rho(\vec{\varphi})$ at z_0. Since $\vec{u}$ has been chosen pointing to the interior of $\rho(\vec{\varphi})$, the angle between $\vec{u}$ and L is bounded away from 0. It follows that for some constant $c > 0$ and all $t > 0$, $\langle \nabla \mathsf{h}(z_0 + t\vec{u}), \vec{u}\rangle \geq c|\nabla \mathsf{h}(z_0 + t\vec{u})||\vec{u}| \to \infty$.

We deduce that $f'(t) \to +\infty$ as $t \to 0$. Since $\partial_0 G > 0$, it is bounded away from 0 on the segment of endpoints z_0 and $z_0 + \vec{u}$ and it follows that $g(z_0 + t\vec{u}) - g(z_0) \gg t$ as $t \to 0$. In particular there exists $t > 0$ such that $g(z_0 + t\vec{u}) > g(z_0)$. □

It follows that $\nabla g = 0$ on $\mathscr{V}$. Now,

$$\nabla g = \nabla G + \frac{\partial G}{\partial z_0} \nabla \mathsf{h}$$

and eq. (4.16) follows and assertion (i) is established.

Let $\nu \in \mathscr{EM}$. The above remarks show that ν maximizes the entropy in $\mathcal{M}(z)$ where $z := \nu(\vec{\varphi})$. By Proposition 4.14, this implies that $\nu = \nu_y$ where $y := -\nabla \mathsf{h}(z)$, yielding the inclusion

$$\mathscr{EM} \subset \{\nu_y : y \in -\nabla \mathsf{h}(\mathscr{V})\}.$$

To check the converse inclusion, let $z \in \mathscr{V}$ and apply Proposition 4.14. Setting $y := -\nabla \mathsf{h}(z)$ so $z := \nabla \mathsf{P}(y)$, we get $G(h(T,\nu_y);\nu_y(\vec{\varphi})) = g(z)$ which is maximum since $z \in \mathscr{V}$. Hence $\nu_y \in \mathscr{EM}$. Assertion (ii) is established. □

Remark 4.17. *The value $\max_{\mathscr{P}(T)} \mathbf{\Pi}$ is a generalization of our previous definition of nonlinear pressure. Of course, one could decide to study the variational principle for full general G without any restriction. Nevertheless we point out that:*

(i) *Assumption* $\inf \partial_0 G > 0$ *is crucial: a change of sign would modify the nature of the problem,*
(ii) *the case* $G(z_0; z) = z_0 + F(z)$ *is of particular interest: in the classical variational principle, the term* $h(T, \mu)$ *comes from the summation over* (ε, n)*-covers in the Gibbs measures (see Formula* (1.6)*), and there is at the moment no candidate to replace this summation and define a* topological *pressure in the case of a general* G*.*

To state our next result, we recall that a *subvariety* of an open set $W \subset \mathbb{R}^d$ is a subset defined by finitely many functions $h_1, \dots, h_k \in C^r(W)$ as $\{x \in W : h_1(x) = \cdots = h_k(x)\}$. If $r = \omega$, it is easy to see that any nontrivial subvariety has zero Lebesgue measure (see, e.g., [21] for a simple proof).

The previous theorem implies the following, which in particular contains Theorem D.

Corollary 4.18. *Let* $(T, \vec{\varphi})$ *be a* C^ω *Legendre system and* G *be a* C^ω *admissible function defined on an open set* $V \subset \mathbb{R}^{1+d}$*. Then the set* $\mathscr{V}$ *of* $(G, \vec{\varphi})$*-equilibrium values is a compact subset of an analytic sub-variety of* $\mathbb{R}^d$*.*

In particular, it is a closed set with empty interior which is Lebesgue negligible.

Since a proper analytic sub-variety of a compact line segment is finite:

Corollary 4.19. *Let* (T, φ) *be a* C^ω *Legendre system and* G *be a* C^ω *admissible with* $d = 1$*, then the set* $\mathscr{EM}$ *of equilibrium measures is finite.*

In full generality, we have a *generic* uniqueness:

Proposition 4.20. *Let* $(T, \vec{\varphi})$ *be a Legendre* C^r *regular system for some* $2 \leq r \leq \omega$*. There is a unique nonlinear equilibrium measure in both of the following settings:*

(i) *For* G *in some open and dense subset of* $\{G \in C^r(V) : \partial_0 G > 0\}$ *where* V *is a given admissible open subset of* $\mathbb{R} \times \mathbb{R}^d$*;*
(ii) *For* $G(z_0; z) = z_0 + F(z)$ *with* F *in some open and dense subset of* $C^r(U)$ *where* U *is a given open neighborhood of* $\rho(\vec{\varphi})$ *in* $\mathbb{R}^d$*.*

Claim (*ii*) above means that, for a generic nonlinearity F, there is a unique nonlinear equilibrium measure. It is not implied by the fully nonlinear case (*i*) since the corresponding set of Gs has empty interior. It would be interesting to determine conditions on a fixed non-linearity F or G under which a generic $\vec{\varphi}$ leads to a unique equilibrium measure

In higher dimension $d \geq 2$, we do not know any example with C^ω regularity where finiteness does not hold. Beyond the real analytic case, even finiteness fails to hold for arbitrary nonlinearity:

Proposition 4.21. *Let* $(T, \vec{\varphi})$ *be a* C^r *Legendre system for some* $2 \leq r \leq \infty$*. For all compact* $E \subset \operatorname{int} \rho(\vec{\varphi})$*, there exists a* C^r *nonlinearity* F *such that the set of equilibrium values* $\mathscr{V}$ *equals* E*. In particular the set of equilibrium measures can be infinite, even uncountable.*

Before proving these two propositions, we recall some well-known facts about Morse functions. Given any open subset $U \subset \mathbb{R}^d$, a function $g \in C^r(U)$ with $2 \leq r \leq \omega$ is Morse on $K \subset U$ if no critical point in K is degenerate and it is nonresonant if it takes distinct values at each of its critical points in K [22, Def. 1.1.7 and 1.2.11]. In particular, it has at most one maximizer on K. Finally, the set of nonresonant Morse C^r functions on a compact set is open and dense (see the proofs in [22, Sect. 1.2]). This is to be understood with respect to the classical uniform topologies on $C^r(U)$ with finite r, or the limit topology for $C^\infty(U)$, or the more complicated standard topology of $C^\omega(U)$ (see, e.g., [14, p. 53]).

Proof of Proposition 4.20. We prove Claim (i). The proof of Claim (ii) is entirely similar. Note that it is enough to prove the claim under the auxiliary assumptions $\partial_0 G > 1/C$ and $|\nabla G| < C$ for $C > 0$ arbitrary.

First, note that $0 = \nabla g$ implies that $|\nabla h| \leq C^2$. Hence, it is enough to ensure that G is nonresonant Morse on the compact subset:

$$K := \{z \in V : |\nabla \mathsf{h}| \leq C^2\}.$$

Second, observe that $G \mapsto g$ is continuous from $C^r(V) \to C^r(\operatorname{int} \rho(\vec{\varphi}))$.Therefore the set $\mathcal{G}$ of $G \in C^r(V)$ such that g is nonresonant and Morse on K is open.

Third, given any $g \in C^r(V)$, the map $k \mapsto g + k$ is a self-homeomorphism of $C^r(V)$. Therefore there are arbitrarily small $k \in C^r(V)$ such that $g + k$ is nonresonant and Morse on K. Considering $\tilde{G}(z_0, \ldots, z_d) := G(z_0, \ldots, z_d) + k(z_1, \ldots, z_d)$ shows that $\mathcal{G}$ is dense in $C^r(V)$. □

Proof of Proposition 4.21. Let $f : \mathbb{R}^d \to [0, \infty)$ be a C^∞ function such that $E = \{z \in \mathbb{R}^d \mid f(z) = 0\}$ (such a function can be constructed as a convergent sum of functions that are each positive on one open balls, with the union of the balls equal to the complement of E). Let F be -1 outside $\operatorname{int} \rho(\vec{\varphi})$, coincide with $-f - \mathsf{h}$ on a compact subset of $\operatorname{int} \rho(\vec{\varphi})$ containing E in its interior, and be lesser than $-\mathsf{h}$ in between; such a function exists since E does not approach the boundary of the rotation set. Then maximizing $\mathsf{h}(z) + F(z)$ is the same as minimizing $f(z)$, and is achieved precisely on E. □

Since $\mathscr{Y} = -\nabla h(\mathscr{V})$ where $-\nabla h : \mathbb{R}^d \to \operatorname{int} \rho(\vec{\varphi})$ is a diffeomorphism, Proposition 4.21 gives:

Corollary 4.22. *Let $(T, \vec{\varphi})$ be a Legendre C^r regular system for some $2 \leq r \leq \infty$. For all compact sets $E \subset \mathbb{R}^d$, there exists a C^r nonlinearity F such that the set $\mathscr{Y}$ from Theorem C equals E. In particular the set of equilibrium measures can be infinite, even uncountable.*

5. Examples of phase transitions

This section is devoted to the application of the framework developed above to a few families of systems whose energy depends on a real multiplicative parameter (i.e., an inverse temperature) and exhibiting various behaviors when this parameter is modified: changes in the number of equilibrium measures, piecewise

analytic behavior with or without an affine piece. Most examples belong to the non-linear thermodynamical formalism, but even in the linear case we provide new insight thanks to the entropy-potential diagram $\mathcal{D}$, see Theorem 5.5.

5.1. **The Curie-Weiss Model - Symmetric case.** The *Curie-Weiss energy* for a potential φ is given by a quadratic nonlinearity, i.e., $\mathcal{E}(\mu) = \beta\mathcal{E}_1(\mu) = \frac{1}{2}\beta\mu(\varphi)^2$ where β is a parameter called the *inverse of temperature*. For this specific case, we shall first use our general machinery above to recover an example treated in [16], then provide a second example exhibiting a "metastable" phase transition.

We consider here the left shift T on $X := \{a, b\}^{\mathbb{N}}$, endowed for example with the distance

$$d(x, y) = 2^{-\inf\{i|x_i \neq y_i\}} \qquad \text{where } x = (x_i)_{i\in\mathbb{N}}, y = (y_i)_{i\in\mathbb{N}},$$

with the potential $\varphi : X \to \mathbb{R}$ defined by

$$\varphi(x) = \begin{cases} -1 & \text{if } x_0 = a \\ 1 & \text{if } x_0 = b \end{cases}$$

and the Curie-Weiss nonlinearity $F(z) = \beta F_1(z) := \frac{1}{2}\beta z^2$, with $\beta \geq 0$.

For any given $z \in [-1, 1]$, we consider the invariant measures $\mu \in \mathcal{M}(z)$, i.e., such that $\mu([b]) - \mu([a]) = z$ where $[i]$ is the cylinder of words starting with the letter i. Since these two cylinders form a partition of X, this equation rewrites as $\mu([a]) = \frac{1-z}{2}$ (and therefore $\mu([b]) = \frac{1+z}{2}$). Among invariant measures in $\mathcal{M}(z)$, the one of maximal entropy is the Bernoulli measure with weights $(\frac{1-z}{2}, \frac{1+z}{2})$, whose entropy is well-known:

$$\mathsf{h}(z) = -\frac{1-z}{2}\log\frac{1-z}{2} - \frac{1+z}{2}\log\frac{1+z}{2}$$

We thus are left with maximizing, given $\beta \geq 0$,

$$P_\beta(z) := \mathsf{h}(z) + \beta F(z) = -\frac{1-z}{2}\log\frac{1-z}{2} - \frac{1+z}{2}\log\frac{1+z}{2} + \frac{1}{2}\beta z^2.$$

A simple computation shows that there are two cases (see Figure 3):

- (*i*) For $0 \leq \beta \leq 1$, 0 is the unique critical point of $\mathbb{P}_\beta$ and is indeed a maximum. Thus, $\mathscr{V} = \{0\}$, there is a unique equilibrium state which is the Bernoulli measure of weights $(\frac{1}{2}, \frac{1}{2})$, and the nonlinear topological pressure is $\Pi_{\text{top}}^{\beta\mathcal{E}_1}(T) = \log 2$.
- (*ii*) For $\beta > 1$, there are three distinct critical points $\{-z_\beta, 0, z_\beta\}$ among which 0 is a local minimum and $-z_B < z_B$ are two global maxima. Hence, $\mathscr{V} = \{-z_\beta, z_\beta\}$ and there are two equilibrium measures, which are "symmetrical" Bernoulli measures, one with $\mu([a]) = \frac{1-z_\beta}{2}$ the other with $\mu([b]) = \frac{1-z_\beta}{2}$.

We have recovered the result of [16] that the nonlinear equilibrium measure is unique for $0 \leq \beta \leq 1$ but that there are two of them for $\beta > 1$, in line with the physical model.

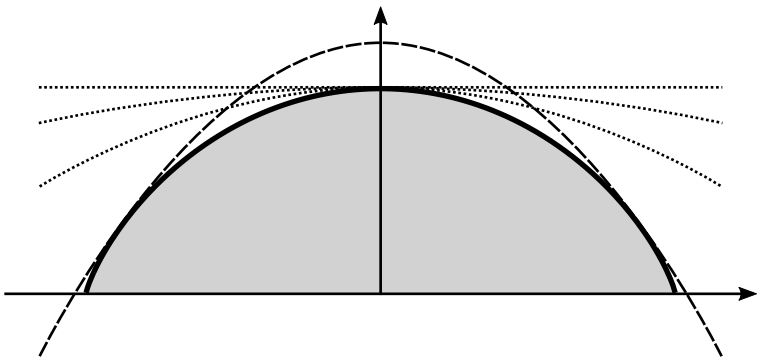

FIGURE 3. The symmetric Curie-Weiss example: graph of h (solid line), highest translates of the graph of $-\beta F$ touching it (dotted line: $\beta < 1$; dashed line: $\beta > 1$).

Note that any C^2 Legendre system (T, φ) with an entropy-potential diagram that is symmetric with respect to the vertical axis will provide a similar example. Indeed the symmetry ensures that for all β, 0 is a critical point; and as long as $\beta < \mathsf{h}''(0)$, the graph of h being more concave at 0 than the graph of $-\beta F$, 0 will be a local maximum. It will then be a global maximum at least when β is close enough to 0. For $\beta > \mathsf{h}''(0)$, 0 will be a local minimum and one will get (at least) two non-zero symmetric equilibrium values.

5.2. **An asymmetric Curie-Weiss model.** Consider now the space of three-letter words $X = \{a, b, c\}^{\mathbb{N}}$ and let T be the left shift on X. We will again consider the Curie-Weiss nonlinearities $F(z) = \beta F_1(z) = \beta \frac{z^2}{2}$ where $\beta \in [0, +\infty)$ is the inverse of the temperature, but with a potential exhibiting a specific asymmetry:

$$\varphi(x) = \begin{cases} -2 & \text{when } x_0 = a \text{ or } x_0 = b, \\ 3 & \text{when } x_0 = c \end{cases}$$

Here $\rho(\varphi) = [-2, 3]$ and a measure maximizing entropy under the constraint $\mu(\varphi) = z$ must, as above, be a Bernoulli measure. If we write $(p, q, 1 - (p + q))$ for its weights, the constraint translates as

$$p + q = \frac{3 - z}{5}. \tag{5.1}$$

Given this constraint, it is easily checked that entropy is maximized when $p = q$. Setting $p(z) = (3 - z)/10$, we get that the measure in $\mathcal{M}(z)$ maximizing entropy is the Bernoulli measure with weights $(p(z), p(z), 1 - 2p(z))$ and we obtain

$$\begin{aligned} \mathsf{h}(z) &= -2p(z) \log p(z) - (1 - 2p(z)) \log(1 - 2p(z)) \\ &= \frac{z - 3}{5} \log \frac{3 - z}{10} - \frac{2 + z}{5} \log \frac{2 + z}{5}. \end{aligned}$$

We are left with maximizing

$$P_\beta(z) := \mathsf{h}(z) + \beta F_1(z) = \frac{z - 3}{5} \log \frac{3 - z}{10} - \frac{2 + z}{5} \log \frac{2 + z}{5} + \frac{1}{2} \beta z^2$$

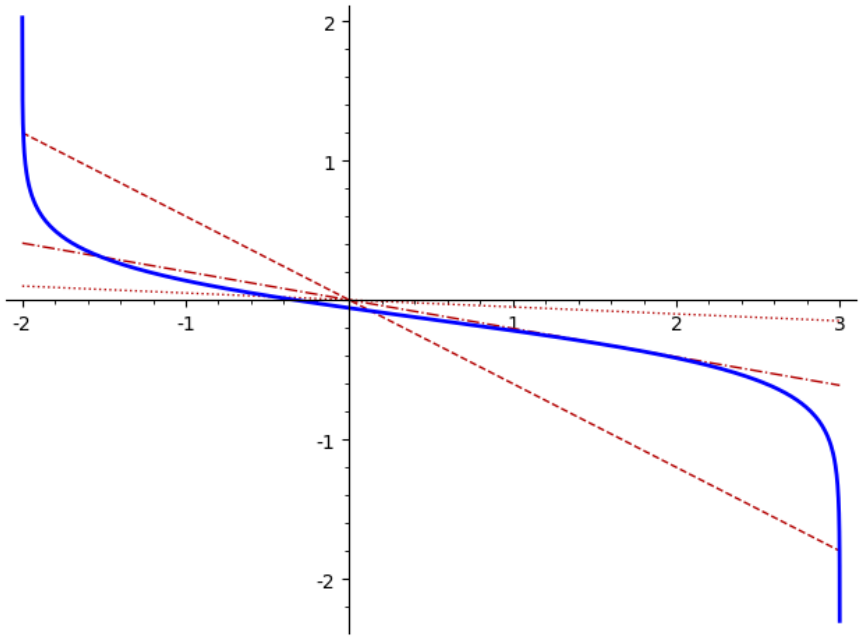

FIGURE 4. The graph of h' and ℓ_β for three values of β: $\beta < \beta_1$ (dotted), $\beta = \beta_1$ (dot dash), $\beta > \beta_1$ (dashed).

for $z \in [-2, 3]$. The critical points of P_β are given by the intersections of the graph of h' with the line $\ell_\beta = \{(z_0, z) \mid z_0 = -\beta z\}$. We have

$$\mathsf{h}'(z) = \frac{1}{5}\log\left(\frac{3-z}{4+2z}\right)$$
$$\mathsf{h}''(z) = -\frac{1}{(2+z)(3-z)} = \frac{1}{5}\left(\frac{1}{z-3} - \frac{1}{z+2}\right)$$
$$\mathsf{h}'''(z) = \frac{1}{5}\left(\frac{1}{(z+2)^2} - \frac{1}{(z-3)^2}\right)$$

so that h' is strictly decreasing, from $+\infty$ when $z \to -2$ to $-\infty$ when $z \to 3$; it has a single inflection point at $z = \frac{1}{2}$, is convex on $(-2, \frac{1}{2}]$ and concave on $[\frac{1}{2}, 3)$ (see its graph in Figure 4).

It follows that for $\beta \geq 0$ small enough, P_β has only one critical point, which must be a maximum; in this regime, there is only one equilibrium state, with equilibrium value $z < 0$, and the pressure varies analytically.

Increasing β, at some value β_1 the line ℓ_β touches the graph of h' on the right, and a second critical point appears. However, at this moment there is still only one equilibrium measure: P_β is unimodal, decreasing around the second critical point. Increasing β any further makes P_β bimodal, with three critical points: one local minimum located between two local maximums $z_1(\beta) < z_2(\beta)$.

At first, $z_1(\beta)$ is the unique global maximum, but it ultimately gets surpassed by $P_\beta(z_2(\beta))$, precisely at the inverse temperature β_0 when the vertical translate of the graph of $-\frac{\beta}{2}z^2$ touching the graph of h does so at two points. The choice of

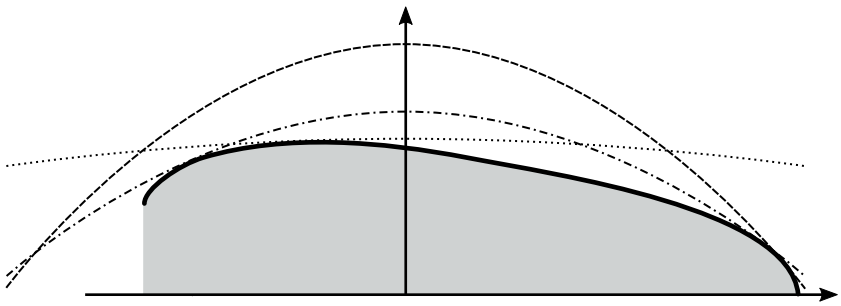

FIGURE 5. A metastable phase transition: graph of h (solid line, graph modified for readability), highest translates of the graph of $-\beta F_1$ touching it (dotted: $\beta < \beta_0$; dashed: $\beta > \beta_0$; dot-dashed: $\beta = \beta_0$).

φ has been made to ensure this happens, by giving the entropy-potential diagram a larger overhang to the right than to the left (see Figure 5): as $\beta \to \infty$, the highest translate of the graph of $-\beta F$ that touches the graph of h converges to the two vertical lines of equations $(z = 3)$ and $(z = -3)$. The latter of these vertical lines is far from the entropy-potential diagram since $\rho(\varphi) = [-2, 3]$, and for large enough β the unique global maximum of P_β must be attained at $z_2(\beta) \to 3$.

Again the pressure is analytic for $\beta > \beta_0$, but we have a phase transition at β_0: the pressure is $\beta \mapsto \max(P_\beta(z_1(\beta)), P_\beta(z_2(\beta)))$ and cannot be analytical at the point where the arguments of the max cross each other. Observe that the value β_1 $(< \beta_0)$ does not correspond to a phase transition: pressure is analytic in the vicinity of β_1.

This example motivates the following definition.

Definition 5.2. *A system $(T, \mathcal{E}_1)$ is said to exhibit a* metastable phase transition *at inverse temperature $\beta_0 > 0$ when there are two curves of invariant probability measures (μ_β), (ν_β) defined on a neighborhood I of β_0 with $\beta \mapsto \Pi^{\beta\mathcal{E}_1}(\mu_\beta)$ and $\beta \mapsto \Pi^{\beta\mathcal{E}_1}(\nu_\beta)$ both C^ω, such that:*

(i) for all $\beta \in I$, (μ_β), (ν_β) are local maximums of $\Pi^{\beta\mathcal{E}}$,
(ii) for $\beta < \beta_0$, μ_β is an equilibrium measure of $\beta\mathcal{E}$ but ν_β is not, and for $\beta > \beta_0$, ν_β is an equilibrium measure but μ_β is not.

Observe that the pressure function $\beta \mapsto \Pi^{\beta\mathcal{E}_1}$ is not analytic at β_0, for otherwise $\Pi^{\beta\mathcal{E}_1}(\mu_\beta)$ and $\Pi^{\beta\mathcal{E}_1}(\nu_\beta)$ would have to coincide and both ν_β and μ_β would be equilibrium measures throughout I.

The "metastable" terminology is suggested by the analogy with the physical phenomenon of the same name. A simple example of it is that of water remaining liquid below the freezing point in some circumstances. This is modeled by the liquid state (described by μ_β) admitting a continuation to $\beta > \beta_0$ as a local maximum and the global maximal, the solid state (described by ν_β), being too far from μ_β to allow the water to easily reorganize itself from one state to the other.

What we have proven can be summarized as follows.

Theorem F. *There exists a locally constant potential φ on a full shift X such that the Curie-Weiss energy $\mathcal{E}_1(\mu) = \frac{1}{2}\mu(\varphi)^2$ exhibits a metastable phase transition.*

This gives another concrete example of multiple nonlinear equilibrium measures in a context where the linear thermodynamical formalism is long known to be flawless (analytic pressure, etc.)

5.3. **The mean-field Potts model.** The mean-field Potts model is given by the full shift (X, T) over a finite alphabet $\{\theta_1, \dots, \theta_n\}^{\mathbb{N}}$ or $\{\theta_1, \dots, \theta_n\}^{\mathbb{Z}}$, with $n \geq 3$. The potential is $\vec{\varphi} := (\mathbb{1}_{\theta_1}, \dots, \mathbb{1}_{\theta_n})$ and the nonlinearity $F(z) = \beta F_1(z) = \frac{\beta}{2}|\vec{z}|^2$ where $|\cdot|$ is the usual Euclidean norm. The energy is thus given by

$$\mathcal{E}(\mu = \beta\mathcal{E}_1(\mu) = \frac{\beta}{2}\Big|\int \vec{\varphi}\,\mathrm{d}\mu\Big|^2 = \frac{\beta}{2}\sum_i \mu([\theta_i])^2$$

where, as above, $[\theta_i]$ is a cylinder, the set of words having the letter θ_i in zeroth position.

The framework developed above seems not to apply since the potentials are not linearly independent up to (coboundaries and) constants: $\sum_i \mathbb{1}_{\theta_i} \equiv 1$, and the rotation set has empty interior. Let us take this as an opportunity to explain how this hypothesis is easily recovered: one simply extract a maximal independent subfamily of potentials, here $\vec{\varphi}_\circ = (\mathbb{1}_{\theta_1}, \dots, \mathbb{1}_{\theta_{n-1}})$, and adjusts the nonlinearity to ensure $F_\circ(\mu(\vec{\varphi}_\circ)) = F(\mu(\vec{\varphi}))$ for all $\mu \in \mathscr{P}(T)$, here

$$F_\circ(z_1, \dots, z_{n-1}) = \frac{\beta}{2}\Big(z_1^2 + \dots + z_{n-1}^2 + \big(1 - \sum_{i<n} z_i\big)^2\Big).$$

It is always possible to construct such an $F_\circ$, since by maximality each the potentials that are present in $\vec{\varphi}$ can be expressed as linear combination of the potentials in $\vec{\varphi}_\circ$ up to a coboundary and a constant, and a coboundary $g - g \circ T$ can be neglected since $\mu(g - g \circ T) = 0$ for all invariant measures μ.

Now $(T, \vec{\varphi}_\circ)$ is C^ω Legendre and we can apply Theorems B and C (recall that moreover $(T, \vec{\varphi}_\circ)$ has unique linear equilibrium measures, hence each $z \in \mathscr{V}$ yields a unique nonlinear equilibrium measure), and these results translate to the original system $(T, \vec{\varphi})$ with the nonlinearity F: accumulation points of Gibbs ensembles are convex combinations of the nonlinear equilibrium measures, each of which coincides with a linear equilibrium measure for some linear combination of the (φ_i); however, due to the lack of independence, several different linear combinations lead to the same equilibrium state.

In the specific case of the mean-field Potts model one can work out the equilibrium measures by (nontrivial) direct computations. Given a vector $z := (z_1, \dots, z_n)$ in the rotation set

$$\rho(\vec{\varphi}) := \Big\{\int \vec{\varphi}\,d\mu,\ \mu \in \mathscr{P}(T)\Big\} = \Big\{(z_1, \dots z_n) \in [0,1]^n \colon \sum_i z_i = 1\Big\},$$

the maximal entropy among invariant measures μ satisfying $\mu(\vec{\varphi}) = z$ is $\mathsf{h}(z) = -\sum_i z_i \log z_i$ with the convention $0 \log 0 = 0$. It is achieved by a unique measure, the Bernoulli measure giving each cylinder $[\theta_i]$ the mass z_i.

For $\beta \geq 0$, the nonlinear pressure is

$$\Pi_{\text{top}}^{\beta\mathcal{E}_1} = \max_{\vec{z}} -\sum_i z_i \log z_i + \frac{\beta}{2}\sum_i z_i^2.$$

We now summarize results from [9]. For $0 \leq \beta < \beta_c := 2\dfrac{n-1}{n-2}\log(n-1)$, $\Pi_{\text{top}}^{\beta\mathcal{E}_1}$ is reached for $z = (\frac{1}{n}, \ldots, \frac{1}{n})$. The value is $\frac{\beta}{2n} + \log n$ and is achieved by a unique measure.

For $\beta > \beta_c$, $\Pi_{\text{top}}^{\beta\mathcal{E}_1}$ is given by an implicit equation. It is realized by z equal to any permutation of $\widetilde{z}$ defined by

$$\widetilde{z}_1 = \frac{1+(n-1)s}{n},\ \widetilde{z}_i = \frac{1-s}{n},\ 2 \leq i \leq n$$

where s is the biggest solution for

$$s = \frac{1-e^{-\beta s}}{1+(n-1)e^{-\beta s}}.$$

Each permutation of $\widetilde{z}$ gives a distinct equilibrium measure. Thus we get exactly n equilibrium measures.

For $\beta = \beta_c$, the maximal value is simultaneously realized by $(\frac{1}{n}, \ldots, \frac{1}{n})$ and by the n distinct permutations of $\tilde{z}$. Thus we get exactly $n+1$ equilibrium measures. In this case, the convergence of Gibbs measures to a convex combination of these equilibrium measures was previously shown in [17].

5.4. **Freezing phase transitions.** Let us explain how the entropy-potential diagram can be used to visualize "freezing phase transitions", i.e., situations where for some β_0, the set of equilibrium measures of the energy $\beta\mathcal{E}_1$ is constant for $\beta > \beta_0$. These measures are called the *ground states*. The physical interpretation is that once the temperature goes below some positive value $1/\beta_0$, the system freezes in a macroscopic state corresponding to zero temperature, described by (one of) the ground states. In the linear thermodynamical formalism, the first freezing phase transition was exhibited by Hofbauer [11], motivated by giving examples with multiple equilibrium states (this is sometimes achieved at $\beta = \beta_0$). Concretely, the typical examples are for the shift T on $X = \{a,b\}^{\mathbb{N}}$ or $X = \{a,b\}^{\mathbb{Z}}$ with potentials

$$\varphi(x) = -\frac{1}{k(x)^\alpha}, \qquad k(x) := \min\{|k| \colon x_k \neq a\}$$

with $\alpha \in (0,1]$, and the freezing equilibrium measure is $\mu_0 = \delta_{aaa\ldots}$. It has more recently been shown by Bruin and Leplaideur [2, 3] that one can produce in a similar way a freezing phase transition with more interesting ground states, supported on some uniquely ergodic, zero-entropy compact subsets of X such as given by the Thue-Morse or the Fibonacci substitutions.

Let us interpret in the entropy-potential diagram $\mathcal{D}$ such a freezing phase transition, with potential φ being maximized by some invariant measure μ_0, say with $\mu_0(\varphi) = 0$ for normalization. By definition, for $\beta \geq \beta_0$ the pressure is affine and

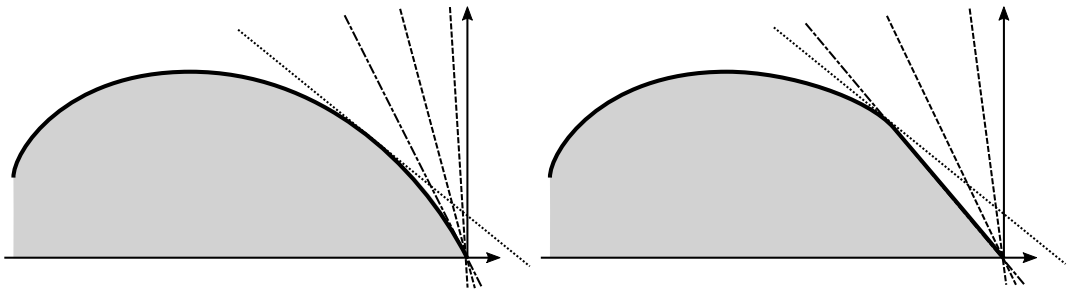

FIGURE 6. Freezing phase transitions in the linear thermodynamical formalism: for $\beta > \beta_0$, all support lines are concurrent, and $\mathcal{D}$ must exhibit an acute corner at its right end. Left: h is strictly concave, there might be a unique equilibrium measure throughout (case $\alpha = 1$ in Hofbauer's example). Right: h has a flat part, at β_0 there are (at least) two ergodic equilibrium measure, one at each end of the flat edge (case $\alpha < 1$ in Hofbauer's example).

achieved at μ_0, meaning that all lines of slope $< -\beta_0$ touching $\mathcal{D}$ do it at the same point (see Figure 6).

This observation immediately implies a characterization of (linear) freezing phase transition by a linear inequality between the entropy and the integral of the potential.

Proposition 5.3. *Let $T : X \to X$ be a measurable map, $\varphi : X \to \mathbb{R}$ be a potential whose rotation set has the form $[r, 0]$ for some $r \in (-\infty, 0)$, such that there is an invariant measure μ_0 realizing $\mu_0(\varphi) = 0$ and maximizing entropy among such measures: $h(T, \mu_0) = \max\{h(T, \mu) \colon \mu \in \mathscr{P}(T), \mu(\varphi) = 0\}$. The following are equivalent:*

(i) the linear thermodynamical formalism for the system (T, φ) exhibits a freezing phase transition, i.e., for some $\beta_0 > 0$ and all $\beta > \beta_0$, the set of equilibrium measures is non-empty and independent of β,
(ii) there is some finite β such that μ_0 is an equilibrium measure for $\beta\varphi$,
(iii) the topological pressure function

$$\begin{aligned} \mathsf{P} \colon \mathbb{R} &\to \mathbb{R} \\ \beta &\mapsto \sup\{h(T, \mu) + \beta\mu(\varphi) \colon \mu \in \mathscr{P}(T)\} \end{aligned}$$

is affine on some interval $[\beta_0, +\infty)$,
(iv) there exists $C > 0$ such that $h(T, \mu) \leq h(T, \mu_0) - C\mu(\varphi)$ for all $\mu \in \mathscr{P}(T)$.

When these conditions are realized, the critical inverse temperature, i.e., the least possible value of β_0, is the least possible C in the entropy-potential inequality (iv). The intercept of the affine part of the graph of P is then the entropy of equilibrium measures after the freezing phase transition, and its slope is their energy $\mu(\varphi)$ (here 0 is given by the chosen normalization of the rotation set).

Proof. The main novelty here is the observation that (*iv*) characterizes Freezing Phase Transitions, but for the sake of completeness we prove all the equivalences, through the cycle *(i)* $\implies$ *(iii)* $\implies$ *(iv)* $\implies$ *(ii)* $\implies$ *(i)*.

Assume *(i)* and let μ_1 be any equilibrium measure for any $\beta > \beta_0$. For all $\beta > \beta_0$ we get $\mathsf{P}(\beta) = h(T, \mu_1) + \beta\mu_1(\varphi)$, an affine expression.

Convex duality translates angular points to flat regions and vice-versa; that P is affine on an interval means that the entropy-potential diagram has an angular point with a supporting line of slope $-\beta$ for each β in the interval. Let us explain this, a simple case of what we left hidden behind the appeal to Legendre duality above. Using the notation $\mathsf{h}(z) = \sup\{h(T, \mu) \colon \mu(\varphi) = z\}$ for all $z \in [r, 0]$, h is concave thus continuous on $(r, 0)$, and has a continuous extension $\bar{\mathsf{h}}$ on $[r, 0]$. We can the rewrite $\mathsf{P}(\beta) = \max_z \bar{\mathsf{h}}(z) + \beta z$. Denoting by z_β an abscissa realizing $\mathsf{P}(\beta)$, observe that for all $\varepsilon > 0$, $\mathsf{P}(\beta + \varepsilon) \geq \bar{\mathsf{h}}(z_\beta) + (\beta + \varepsilon)z_\beta \geq \mathsf{P}(\beta) + \varepsilon z_\beta$ so that the right derivative of P is at least z_β. Similarly, $\mathsf{P}(\beta - \varepsilon) \geq \mathsf{P}(\beta) - \varepsilon z_\beta$ shows that the left derivative is at most z_β. Whenever P is differentiable, $\mathsf{P}'(\beta) = z_\beta$. On an affine part, the derivative exists and is constant, therefore z_β is (locally) constant and h has an angular point. Moreover the abscissa of the angular point is the slope of the line extending the affine part of the graph of P, while the ordinate of that point is the intercept of that line.

Item *(iii)* thus implies that the entropy-potential diagram has an angular point with supporting lines of slope $-\beta$ for all $\beta \geq \beta_0$. Since slopes are arbitrarily high in magnitude, the abscissa of this angular point must be the supremum of the rotation set, i.e., 0. It must then have ordinate equal to the supremum of the realizable entropies for this energy, i.e., $h(T, \mu_0)$. In particular, the entropy-potential diagram is constrained under a line of equation $(h(T, \mu) = h(T, \mu_0) - \beta_0\mu(\varphi))$, which is *(iv)*.

Assume *(iv)* and take any $\beta \geq C$. For all $\mu \in \mathscr{P}(T)$,

$$h(T, \mu) + \beta\mu(\varphi) \leq h(T, \mu_0) + (\beta - C)\mu(\varphi) \leq h(T, \mu_0) = h(T, \mu_0) + \beta\mu_0(\varphi)$$

so that μ_0 is an equilibrium measure for such β, proving *(ii)*.

Assume *(ii)*, let β_1 be such that μ_0 is an equilibrium measure for $\beta_1\varphi$ and $\beta > \beta_1$. For all $\mu \in \mathscr{P}(T)$, since $\mu(\varphi) \leq 0$ and $\mu_0(\varphi) = 0$,

$$h(T, \mu) + \beta\mu(\varphi) \leq h(T, \mu) + \beta_1\mu(\varphi) \leq h(T, \mu_0) + \beta_1\mu_0(\varphi) = h(T, \mu_0) + \beta\mu_0(\varphi)$$

and μ_0 is an equilibrium measure for $\beta\varphi$. It follows that the set of β's such μ_0 is an equilibrium measure for $\beta\varphi$ is an interval $\beta_0, +\infty)$. The above computation shows that for all $\beta > \beta_0$, the set of equilibrium measure is $\{\mu \in \mathscr{P}(T) \colon \mu(\varphi) = 0, h(T, \mu) = h(T, \mu_0)\}$, and is thus independent of β. □

Remark 5.4. *If we consider several potentials $\varphi_1, \ldots, \varphi_d$, the condition in Legendre regularity that $|\nabla \mathsf{h}|$ goes to $+\infty$ as one approaches the boundary is violated exactly when some linear combination of the (φ_k) exhibit a (linear) freezing phase transition.*

The entropy-potential diagram makes it clear how to prove existence of freezing phase transition in both the linear and nonlinear settings. We divide Theorem E of the introduction in two parts.

Theorem 5.5. *Let $T : X \to X$ be a continuous map of finite, positive topological entropy such that $\mu \mapsto h(T, \mu)$ is upper semi-continuous. Consider $\mu_0 \in \mathscr{P}_{\mathrm{erg}}(T)$ with zero entropy. Then there exists a continuous potential $\varphi : X \to \mathbb{R}$ such that the linear thermodynamical formalism of (T, φ) exhibits a freezing phase transition with ground state μ_0. Moreover we can ensure that μ_0 is the unique ground state, and that at the critical inverse temperature β_0 there are exactly two equilibrium states.*

In particular, if K is a compact T-invariant set with zero topological entropy, then we can find a potential exhibiting a freezing phase transition supported on K. This broadly extends [2, 3] by proving existence of freezing phase transitions for all zero-entropy subshifts, instead of very specific ones; but it is not constructive, since the potential φ is ultimately obtained through the Hahn-Banach theorem.

Proof. According to a Theorem of Jenkinson [12], there exists a continuous potential $\tilde{\varphi} : X \to \mathbb{R}$ such that μ_0 is the unique equilibrium state of $\tilde{\varphi}$, i.e., the unique maximizer of $h(T, \mu) + \beta\mu(\tilde{\varphi})$ for $\beta = 1$. Since $h(T, \mu_0) = 0$ is minimal, μ_0 must be a maximizing measure for $\tilde{\varphi}$. The conclusion then follows from Proposition 5.3 applied to the adjusted potential $\varphi = \tilde{\varphi} - \mu_0(\tilde{\varphi})$.

To have a second equilibrium state at the critical inverse temperature, it suffices to consider an arbitrary ergodic measure μ_1 of positive entropy: Jenkinson's theorem provides a continuous potential whose only ergodic equilibrium states (at $\beta = 1$) are μ_0 and μ_1. This also fixes the critical inverse temperature at $\beta_0 = 1$. □

Theorem 5.6. *Let $T : X \to X$ be a continuous dynamical system of finite, positive topological entropy such that $\mu \mapsto h(T, \mu)$ is upper semi-continuous. Let $\varphi : X \to (-\infty, 0]$ be a continuous potential such that $K = \varphi^{-1}(0)$ is T-invariant and has zero topological entropy.*

Then there exists a continuous nonlinearity $F_1 : (-\infty, 0] \to (-\infty, 0]$ with $F(0) = 0$ such that the energy $\mathcal{E}_1(\mu) = F_1(\mu(\varphi))$ exhibits a "strong freezing phase transition" in the following sense. There is a $\beta_0 > 0$ such that:

- *for each $\beta < \beta_0$ the energy $\beta\mathcal{E}_1$ has at least one equilibrium measure, and none of them are supported on K,*
- *at $\beta = \beta_0$ there are several equilibrium measures, at least one supported on K and one not supported on K,*
- *for each $\beta > \beta_0$ the equilibrium measures are exactly the K-supported, T-invariant measures and the topological pressure function $\beta \mapsto \Pi_{\mathrm{top}}^{\beta\mathcal{E}_1}(T)$ is affine.*

Observe that here F_1 will only be continuous at 0; we can extend it continuously to $\mathbb{R}$, but we cannot make F_1 differentiable in a neighborhood of 0.

Proof. Take for F_1 any increasing convex continuous function $(-\infty, 0] \to (-\infty, 0]$ such that $\mathsf{h}(z) = o(-F_1(z))$ as $z \to 0$. Theorem A ensures that equilibrium measures are found by optimizing $\mathsf{h}(z) + \beta F_1(z)$ and then maximizing entropy in $\mathcal{M}(z)$, as in Section 4 (we did not assume Legendre regularity, but we assumed enough to ensure that each optimal z comes with at least one equilibrium measure).

Since h is bounded by $h_{\text{top}}(T)$, for β large enough the graph of $-\beta F_1$ is above the graph of h except at 0 where they meet. This means that for these βs, $\mathsf{h}(z) + \beta F_1(z)$ is non positive and always negative for $z < 0$, i.e., the unique optimal z is 0.

Let β_0 the least β such that $\mathsf{h}(z) \leq -\beta F_1(z)$ for all z. Since $\mathsf{h}(z) = o(-\beta_0 F_1(z))$ as $z \to 0$, there must be a touching point distinct from 0, and we get two optimal values $z = 0$ and $z = z_0 < 0$, and at least two equilibrium measures. For $\beta < \beta_0$, $z = 0$ cannot be optimal anymore since z_0 is strictly better. The conclusion follows. □

A simple example can be worked out in the case of the shift over $X = \{a, b\}^{\mathbb{N}}$ and the potential φ taking the values 0 on the cylinder $[a]$ and -1 on the cylinder $[b]$. We have $\mathsf{h}(z) \sim z \log(-z)$ at zero, so that we can take $F_1(z) = -(-z)^\alpha$ with any $\alpha \in (0, 1)$: the nonlinear thermodynamical formalism associated with the energy

$$\mu \mapsto -|\mu(\varphi)|^\alpha$$

exhibits a strong freezing phase transition with ground state $\mu_0 = \delta_{aaa\ldots}$.

References

[1] V. Baladi. *Positive transfer operators and decay of correlations.* Advanced Series in Nonlinear Dynamics, 16. World Scientific Publishing Co., 2000. 1.3

[2] H. Bruin & R. Leplaideur Renormalization, thermodynamic formalism and quasicrystals in subshifts. *Comm. Math. Phy.*, 321(1), 209-247. 5.4, 5.4

[3] H. Bruin & R. Leplaideur Renormalization, freezing phase transitions and Fibonacci quasicrystals. *Ann. Sci. Éc. Norm. Supér.*, 48 (2015), no. 3, 739–763. 5.4, 5.4

[4] R. Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms.* Lecture Notes in Mathematics, Vol. 470. Springer-Verlag, Berlin, 1975. 2nd ed. - 2008 by JR Chazottes. 1

[5] R. Bowen. Some systems with unique equilibrium states. *Math. Systems Theory* 8, (1974/75), no. 3, 193–202.

[6] M. Brin & A. Katok *On Local Entropy* Lecture Notes in Mathematics, vol. 1007, Springer, Berlin, 1983, pp. 30–38 2.2.1

[7] J. Buzzi. Intrinsic ergodicity of smooth interval maps. *Israel J. Math.* 100 (1997), 125–161. 1

[8] R. S. Ellis. *Entropy, large deviations, and statistical mechanics.* Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1985 original.

[9] Richard S. Ellis and Kongming Wang. Limit theorems for the empirical vector of the Curie-Weiss-Potts model. *Stochastic Processes Appl.*, 35(1):59–79, 1990. 5.3

[10] P. Giulietti, B. R. Kloeckner, A. O. Lopes, & D. Marcon Farias. *The calculus of thermodynamical formalism.* arXiv:1508.01297, *J. Eur. Math. Soc.* 20 (2018), no. 10, pp. 2357–2412. 4.4

[11] F. Hofbauer. Examples for the nonuniqueness of the equilibrium state. *Trans. Amer. Math. Soc. 228 (1977), 223–241* 5.4

[12] O. Jenkinson. Every ergodic measure is uniquely maximizing. *Discrete Contin. Dyn. Syst. 16 (2006), no. 2, 383–392* 5.4

[13] A. Katok & B. Hasselblatt. *Introduction to the modern theory of dynamical systems* Encyclopedia of Mathematics and Its Applications 54. Cambridge Univ. Press, 1995. xviii, 802 p. ISBN: 0-521-34187-6 1.3

[14] S. Krantz, H. Parks, *A primer on real analytic functions,* Springer, 2002. 4.6

[15] R. Leplaideur. Chaos: butterflies also generate phase transitions. *J. Stat. Phys.*, 161 (1), 2015, 151–170. 1.5

[16] R. Leplaideur & F. Watbled, Generalized Curie-Weiss model and quadratic pressure in Ergodic Theory. *Bull. Soc. Math. Fr.* 147 (2), 2019, p. 197–219. 1, 1.3, 1.3, 1.4.2, 5.1, 5.1

[17] R. Leplaideur & F. Watbled, Curie–Weiss type models for general spin spaces and quadratic pressure in ergodic theory *J. Stat. Phys.* 181 (1), 2020, 263–292. 1.4.4, 5.3

[18] C. Liverani, Decay of correlations for piecewise expanding maps. *J. Stat. Phys.* 78 (3-4), 1995, 1111–1129.

[19] I. Melbourne & M. Nicol, Almost sure invariance principle for nonuniformly hyperbolic systems. *Comm. Math. Phys. 260 (2005), no. 1, 131–146.* 1.3

[20] M. Misiurewicz. A short proof of the variational principle for a $\mathbb{Z}_+^N$ action on a compact space. International Conference on Dynamical Systems in Mathematical Physics (Rennes, 1975), pp. 147–157. Astérisque, No. 40, Soc. Math. France, Paris, 1976. 2.6

[21] Boris S. Mityagin. The zero set of a real analytic function. arxiv:1512.07276. 4.6

[22] Liviu Nicolaescu. *An Invitation to Morse Theory.* Universitext. Springer, 2011. 4.6

[23] K. Petersen. *Ergodic theory.* Cambridge Studies in Advanced Mathematics, 2. Cambridge University Press, Cambridge, 1983. xii+329 pp. ISBN: 0-521-23632-0.

[24] R. Phelps *Lectures on Choquet's Theorem.* Lecture Notes in Mathematics 1757. Springer-Verlag Berlin Heidelberg, 2001. vi+122 pp. ISBN: 3-540-41834-2. 1.3

[25] R. Rockafellar. *Convex analysis.* Princeton Mathematical Series, 28. Princeton university press, Princeton, New Jersey, 1970. xviii+450 pp. ISBN: 0-691-08069-0. 4.3, 8, 4.3, 4.3

[26] D. Ruelle. *Thermodynamic formalism.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, second edition, 2004. The mathematical structures of equilibrium statistical mechanics. 1, 4.4

[27] O. Sarig. Phase Transitions for Countable Topological Markov Shifts. *Commun. Math. Phys.* **217** (2001), 555–577. 1.5

[28] Y. Sinai. Gibbs measures in ergodic theory,. *Uspehi Mat. Nauk*, 27(4(166)):21–64, 1972. 1

[29] M. Thaler. Estimates of the invariant densities of endomorphisms with indifferent fixed points. *Israel J. Math.*, 37(4):303–314, 1980. 1.5

[30] C. Villani Optimal transport, old and new. Grundlehren der Mathematischen Wissenschaften 338 Springer-Verlag, Berlin, 2009. xxii+973 pp. ISBN: 978-3-540-71049-3 2.1

[31] P. Walters. An introduction to ergodic theory. Graduate Texts in Mathematics, 79. Springer-Verlag, New York-Berlin, 1982. ix+250 pp. ISBN: 0-387-90599-5 1.1, 1

[32] L.-S. Young, Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math. (2) 147 (1998), no. 3, 585–650.*

Laboratoire de Mathématiques d'Orsay - CNRS & Université Paris-Sud
Email address: jerome.buzzi@universite-paris-saclay.fr

Univ Paris Est Creteil, CNRS, LAMA, F-94010 Creteil, France; Univ Gustave Eiffel, LAMA, F-77447 Marne-la-Vallée, France
Email address: benoit.kloeckner@u-pec.fr

ISEA, Université de la Nouvelle-Calédonie & LMBA UMR6205
Email address: renaud.leplaideur@unc.nc