Abstract Theory of Bogoliubov Linearizations with Application to Nonlinear Thermodynamic Formalism

J.-B. Bru W. de Siqueira Pedra A.O. Lopes

May 25, 2026

Abstract

Bogoliubov's 1947 approximation, originally developed in the microscopic theory of superfluidity, laid the foundation for solving previously intractable quantum models and later became part of "quantum mathematics". Regarding mathematically rigorous results, one of its most advanced forms – the only one that handles quantum equilibrium states – was published in the Memoirs of the AMS in 2013. Building on key results from convex analysis, the present work significantly extends it to obtain a general mathematical theory that enables nonlinear variational problems on convex compact spaces to be fully studied via a linearization process, referred to here as the "Bogoliubov linearization". This problem is particularly timely, given the current development of quantum algorithms and computers, which are inherently linear machines. A deep connection with the optimal transport is also proven. As a paradigmatic example of application, the approach proposed here is applied to the nonlinear thermodynamic formalism – an emerging field that can have important impacts on various fields of mathematics, such as ergodic transport, the fractals and multifractal formalism, discrete-time linear dynamics, C^*-algebras, etc. Notably, even in the case of finite alphabets the obtained results go beyond the scope of the existing literature in nonlinear thermodynamic formalism.

Keywords: Bogoliubov approximation, Ruelle operator, nonlinear thermodynamic formalism, optimal transportation, non-cooperative equilibria.

2020 AMS Subject Classification: 58E30, 37D35, 46N10, 46N55.

Contents

1 Introduction

1.1 Abstract theory of Bogoliubov linearizations

Nonlinear variational problems appear in such a wide variety of mathematical contexts that it would be difficult to provide an exhaustive list. Their interest lies in the wide range of applications in the real world, such as in physics, biology, engineering and economics, for systems governed by principles of equilibrium and optimality in the presence of nonlinear effects. From a mathematical perspective, nonlinear variational problems stimulate the development of sophisticated analytical techniques, such as compactness methods, variational inequalities and critical point theory. They also promote significant connections between functional and convex analysis, partial differential equations, geometry, as well as other areas of mathematics.

Although general approaches for studying them exist, such as the celebrated catastrophe theory, nonlinear variational problems are generally very difficult to analyze rigorously. In this paper, we propose a general method to study variational problems of the form $\sup \mathbb{F}(K)$, where $\mathbb{F} : K \to \mathbb{R}$ is a nonlinear function defined on a compact convex Hausdorff space K. This is done in the scope of convex analysis. Indeed, given two real normed spaces $\mathcal{X}_{\pm}$, the nonlinear functionals $\mathbb{F}$ we consider are of the form

$$\mathbb{F} \doteq f - g_{-} \circ \tau_{-} + g_{+} \circ \tau_{+} \ ,$$

where $f : K \to \{-\infty\} \cup \mathbb{R}$ is an upper semicontinuous concave function, $\tau_{\pm} : K \to \mathcal{X}_{\pm}$ are two continuous affine transformations and $g_{\pm} : \mathcal{X}_{\pm} \to \mathbb{R}$ are two lower semicontinuous and convex functions whose Legendre-Fenchel transforms, $g_{\pm}^{*}$, have a full domain[1] and grow sufficiently fast at large arguments. This situation is very general, as discussed in Remarks 3.5 and 3.6. For example, any C^1-functions on a compact subset of $\mathbb{R}^N$ ($N \in \mathbb{N}$) can be represented as the difference of convex and continuous functions. See Remark 3.6. In applications, f, $-g_{-} \circ \tau$ and $g_{+} \circ \tau_{+}$ typically refer to an entropy, a nonlinear attractive and a nonlinear repulsive interaction term, respectively.

[1] A slightly more general assumption for g_{-} can be used. See Condition B2 of Section 3.3.

Note that even if the function f is a priori only concave (rather than affine), our primary focus is on the nonlinearity introduced by the functions $g_{\pm}$. In fact, one might wonder why we do not include $-g_{-} \circ \tau_{-}$ in the term f, given that they are both upper semicontinuous and concave. The advantage of splitting $\mathbb{F}$ into three parts, f, $-g_{-} \circ \tau_{-}$ and $g_{+} \circ \tau_{+}$ becomes apparent when the Legendre-Fenchel transform of $g_{\pm}^{*}$ and certain associated linear variational problems can be easily controlled. This is the raison d'être of Bogoliubov linearizations. We demonstrate the effectiveness of this method by applying it to the *nonlinear thermodynamic formalism*. However, our new approach has a much wider range of applications.

In fact, it originates from a study of quantum lattice systems at equilibrium, which was published in the *Memoirs of the AMS* in 2013 [18] and has been extensively developed here. We prove in Theorem 3.8 that

$$\sup \mathbb{F}(K) = \sup_{y_{+} \in \mathcal{X}_{+}^{*}} \inf_{y_{-} \in \mathcal{X}_{-}^{*}} \{\sup \mathcal{G}_{y_{+},y_{-}}(K) + g_{-}^{*}(y_{-}) - g_{+}^{*}(y_{+})\} \doteq \mathrm{P}^{\flat} \in \mathbb{R} \,, \tag{1}$$

where $\mathcal{G}_{y_{+},y_{-}} \doteq f - y_{-} \circ \tau_{-} + y_{+} \circ \tau_{+}$, $y_{\pm} \in \mathcal{X}_{\pm}^{*}$, which are named the *Bogoliubov linearizations* of $\mathbb{F}$ here. Solutions to these variational problems are studied in detail, see again Theorem 3.8. For instance, we obtain the following unexpected result:

$$E_{\mathbb{F}} \doteq \left\{\mu \in K : \exists (\mu_j)_{j\in J} \subseteq K \text{ with } \lim_{J} \mu_j = \mu \text{ and } \lim_{J} \mathbb{F}(\mu_j) = \mathrm{P}^{\flat}\right\} = \left\{\mu \in K : \mathbb{F}(\mu) = \mathrm{P}^{\flat}\right\} \,,$$

keeping in mind that $\mathbb{F}$ is *generally not* upper semicontinuous. Additionally, this set is compact and its elements are *self-consistent* solutions to associated linear variational problems. Indeed, as can be seen from Equation (1), the nonlinear[2] problem can be studied through the family of *linear*[3] variational problems

$$P_{\mathrm{L}}(y_{+}, y_{-}) \doteq \sup \mathcal{G}_{y_{+},y_{-}}(K) \,, \qquad y_{\pm} \in \mathcal{X}_{\pm}^{*} \,. \tag{2}$$

This is achieved by determining beforehand the solutions $x_{\pm} \in \mathcal{X}_{\pm}^{*}$ to the $\sup_{y_{+} \in \mathcal{X}_{+}^{*}} \inf_{y_{-} \in \mathcal{X}_{-}^{*}}$ in (1). There is in particular a *canonical* two-person zero-sum game associated with the maximization of $\mathbb{F}$, the payoff function of which is given by the *nonlinear* approximating pressure

$$P_{\mathrm{NL}}(y_{+}, y_{-}) \doteq P_{\mathrm{L}}(y_{+}, y_{-}) + g_{-}^{*}(y_{-}) - g_{+}^{*}(y_{+}) \,.$$

This game is studied in detail in Section 3.4 because it itself has interesting applications, as shown in Section 2.7.4 in the context of the thermodynamic formalism or in [18, 21] for lattice quantum systems.

In Section 3.5, we explain how generalized nonlinear equilibrium states

$$\mu \in G_{\mathbb{F}} \doteq \overline{\mathrm{co}}(E_{\mathbb{F}}) \subseteq K$$

can lead to distributions (at equilibrium) of order parameters[4] for the system under consideration, which is an important concept in physics, related to phase transitions. In this context, we show that, for any given order parameter distribution at equilibrium, the nonlinear pressure $\sup \mathbb{F}(K)$ can be

[2] The nonlinearity here refers to the functions $g_{\pm}$.

[3] The function f is not necessarily linear. It is the functions $g_{\pm}$ that are linearized, being replaced by $y_{\pm} \circ \tau_{\pm}$ in $\mathcal{G}_{y_{+},y_{-}}$. We nonetheless refer to $\sup \mathcal{G}_{y_{+},y_{-}}(K)$ as a "linear" variational problem, because nonlinearity is represented here by $g_{\pm}$; furthermore, in statistical mechanics the function f represents an entropy functional (typically the entropy per unit volume of space-invariant states) that is usually both concave and convex in the thermodynamic limit, and is therefore affine. See, for example, the application of this method to nonlinear thermodynamic formalism, as explained in this paper.

[4] The term "order parameter" originates in physics and refers to a quantity that measures the degree of order in a system, distinguishing different phases of matter. It is expected to exhibit different behavior at phase transitions. In the context of mean-field theory, it refers to the c-number substitution that arises from the mean-field approximation. A typical example of an order parameter is the magnetization density in spin systems.

exactly recovered from a *Monge-Kantorovich (transportation) problem* [65, page 10] associated with these distributions, the cost function of which is nothing but the nonlinear approximating pressure P_{NL} defining the above thermodynamic game. This observation is very useful because of the celebrated Kantorovich dual problem [65], for which various high-performance numerical tools are available. Moreover, in usual applications, the distributions at equilibrium of order parameters are simpler mathematical objects than general equilibrium states. In fact, typically, the former are probability distributions in some compact subset of a finite-dimensional space, whereas the latter are positive functionals on some infinite-dimensional C^*-algebra.

Optimal transport is a fundamental theory that has numerous applications in fields such as economics, machine learning, fluid dynamics, signal processing, mechanics, etc. Therefore, in Theorem 3.16 we prove the Kantorovich dual problem for $\sup \mathbb{F}(K)$, thereby establishing a link between nonlinear equilibria and this theory. To our knowledge, this is a new result. For more details we recommend Section 3.5, which could even be a starting point for further developments in optimal transport theory. Indeed, our aim is not to present here a detailed analysis of the associated optimal transport problem, but rather to build a fundamental connection between the maximization of real-valued functions $\mathbb{F}$ on compact convex spaces and the dual Kantorovich problem, thereby opening the door to entirely new mathematical and numerical developments.

Moreover, as the frontiers of quantum computing and high-performance processing continue to expand, the ability to effectively linearize complex problems is becoming increasingly important. Our approach, based on Bogoliubov linearizations, offers an alternative to conventional tools used nowadays for dynamical systems, such as the Carleman or Koopman methods and perturbation / Taylor expansions [30, 34]. It thus unlocks a new framework for solving nonlinear variational problems of paramount importance at present.

1.2 The nonlinear thermodynamic formalism of dynamical systems

In this paper the paradigmatic example of application is the *nonlinear thermodynamic formalism*, a relatively new area of mathematical research: It appears in a series of seminal papers [45, 46, 7, 25, 42], which introduce a rigorous approach to studying several questions in mean-field theory of classical statistical mechanics, in particular in relation to the Curie-Weiss-Potts models.

The linear thermodynamic formalism, in the Bowen-Ruelle-Sinai sense (see, e.g., [49]), is a comparatively old but still extremely active area of mathematics. See [13, 54, 37, 40, 6, 58, 24, 8, 49, 67, 28]. As explained in [49], it has numerous ramifications towards other fields of mathematics, such as the ergodic transport, the fractals and multifractal formalism, discrete-time linear dynamics, C^*-algebras, Haar systems, groupoids and quasi-invariant probabilities for cocycles, mathematical statistics, etc.

In 2023, Buzzi, Kloeckner and Leplaideur [25] explained how the nonlinear thermodynamic formalism can be understood from the linear one. Our aim is the same, but we take a *completely different* approach with our general theory of Bogoliubov linearizations. Indeed, similar to Buzzi--Kloeckner-Leplaideur's method [25], the theory of Bogoliubov linearizations transforms the nonlinear problem into a family of linear ones. As explained above, this is achieved at the expense of some self-consistency condition, which can, however, be effectively studied in the linear thermodynamic formalism, as already demonstrated in [20] for some explicit examples.

This corresponds to Theorem 2.17, which allows us to determine equilibrium measures via the so-called thermodynamic game discussed above. In fact, this game is shown in Theorem 2.21 to have a direct and general interpretation in terms of (nonlinear) equilibrium measures. As explained in Section 2.3, the case of Hölder potentials in linear thermodynamic formalism remains a very general situation and has excellent mathematical properties. For example, in this case (linear) equilibrium measures are always unique and ergodic (see [50, 1]). We thus apply Theorem 2.17 to a nonlinear version of this situation and obtain a much stronger result in Corollary 2.18 (see also Corollary 2.22).

We study the nonlinear thermodynamic formalism here via its linear version, as Buzzi, Kloeckner and Leplaideur did in [25]; however, in our view, our method is simpler and more transparent. This allows us to reach the same conclusions under much more general assumptions, as well as derive new results, such as Theorems 2.17 and 2.21. See Section 2.6 which compares Buzzi-Kloeckner-Leplaideur's approach with ours. See also Conditions TF1–TF3 of Section 2.7.1.

Last but not least, with regard to the mathematical scope, it should be noted that the (symbolic) space considered in relation to the thermodynamic formalism is the set of all (infinite) sequences in an alphabet Ω, which is not necessarily finite but is only assumed to be compact with respect to some metric. Despite the great importance of this more general case in applications – such as in the classical XY model – the mathematical literature remains sparse. This has made it necessary to provide proofs for several fundamental results which, although anticipated, are essential to our analysis. In particular, we define the entropy via (transfer) Ruelle operators and make the connection with a thermodynamic limit of finite volume entropies, thanks to [1]. We also introduce new concepts, such as Δ-functionals. We believe that, in addition to establishing a very general mathematical foundation for the Bogoliubov linearization method, which stems from a long tradition in statistical physics, the present work provides a solid basis for developing the nonlinear version of the thermodynamic formalism for compact metric alphabets.

1.3 Structure of the paper

To summarize, Theorems 2.17, 2.21, 3.8 and 3.16, as well as Corollaries 2.18 and 2.22, are the main results of the paper, which is divided into two main parts:

- Section 2 explains the thermodynamic formalism, including Buzzi-Kloeckner-Leplaideur's seminal approach [25] to its nonlinear version, in Section 2.6. We begin with this example to directly illustrate a concrete application of our abstract theory of Bogoliubov linearizations. This example is indeed pedagogical and even paradigmatic, with a very general scope of application.

- Section 3 presents our general abstract theory of Bogoliubov linearizations, which could prove very useful in many other contexts. Note that in Section 3.2 we provide a more detailed explanation of why it was named after Bogoliubov.

Each of the two main sections ends with a more technical subsection and an appendix that compile mathematical assertions used in this work. The results presented in the two appendices are either standard, largely unknown (e.g., Theorem 2.40) or new. This makes the paper self-contained and accessible to a wide audience.

2 Paradigmatic Example: Nonlinear Thermodynamic Formalism of Dynamical Systems

The thermodynamic formalism is an old subject in strong connection with dynamical systems and ergodic theory. See, e.g., Ruelle's book [58] which first systematized this theory in 1978. It remains an extremely active and important area of mathematics. See, for example, the book [70] that emerged from a semester in 2019 at the Centre International de Rencontres Mathématiques on "Thermodynamic Formalism: Applications to Probability, Geometry and Fractals". It has indeed numerous ramifications across other areas of mathematics, including: the ergodic transport, the fractal and multifractal formalism, discrete-time linear dynamics, C^*-algebras, Haar systems, groupoids, quasi-invariant probabilities for cocycles, mathematical statistics, etc. For example, it has been known since the 1980s that the Hausdorff dimension of the support of a probability measure can, in certain cases, be determined through the derivative of a (linear) topological pressure. See, e.g., [47]. For various references and examples of works connected to the thermodynamic formalism, see [13, 54, 37, 40, 6, 58, 24, 8, 49, 67, 28].

In recent years, approaches to study questions in mean-field theory from the ergodic viewpoint were introduced, in particular for the Curie-Weiss type models. These results constitute the foundations of a new area called the *nonlinear thermodynamic formalism* [45, 46, 7, 25, 42]. As explained in [25], the nonlinear version of the thermodynamic formalism has been also considered in relation to the multifractal analysis [28].

As it turns out, there is an intricate relationship between the nonlinear formalism and the linear one. For compact metric alphabets, in the case of convex and concave nonlinearities, or a combination of both, we present a new method inspired by "quantum mathematics" to solve this problem effectively. In fact, the link between both formalisms is made via a min-max principle which leads to the concept of thermodynamic games, the cooperative equilibria of which provide a complete classification of the nonlinear equilibrium measures of a given nonlinear (topological) pressure as equilibrium measures of self-consistent (effective) linear pressures.

The general (abstract) version of this method is postponed to Section 3. To motivate it, we first present its application to the nonlinear version of the thermodynamic formalism as paradigmatic example. Although the nonlinear thermodynamic formalism is a known subject, our approach introduces new developments in the form of a novel yet natural variational problem on probability measures (Section 2.7.4). Even in the case of finite alphabets our results cover a more general class than the current literature in nonlinear thermodynamic formalism [45, 46, 7, 25, 42].

To make the paper self-contained, we begin with the mathematical framework of the symbolic dynamical systems formalism.

2.1 Mathematical framework

Alphabet. Let (Ω, d) be any compact metric space. It represents a general (possibly infinite) alphabet. For simplicity and without loss of generality, we assume that

$$\max d(\Omega \times \Omega) \doteq \max \{d(\omega_1, \omega_2) : \omega_1, \omega_2 \in \Omega\} = 1 , \tag{3}$$

i.e., Ω has normalized diameter. In addition, we set an arbitrary a priori probability measure m, which is fixed once and for all, on the Borel σ-algebra of the alphabet (Ω, d). Recall that such a measure is always regular, for the alphabet is a metric space. Below, it is important to assume that m has always support equal to Ω.

Infinite strings. From the compact metric space Ω we construct the compact[5] topological space $\Sigma \doteq \Omega^{\mathbb{N}}$ as the set of infinite strings in the alphabet Ω endowed with the product topology. A standard basis of the topology of Σ is given by the collection of *cylinders*, defined for any $N \in \mathbb{N}$ and finite sequence $O_1, \ldots, O_N$ of open sets of (Ω, d) by

$$[O_1, \ldots, O_N] \doteq \left\{\sigma = (\omega_n)_{n\in\mathbb{N}} \in \Sigma : \omega_1 \in O_1, \ldots, \omega_N \in O_N\right\} \subseteq \Sigma .$$

Fix once and for all throughout the paper a parameter $\eta \in (0, 1)$. Then, a metric generating the topology of Σ can be defined by

$$d_\eta(\sigma, \sigma') \doteq (1-\eta) \sum_{n\in\mathbb{N}} \eta^{n-1} d(\omega_n, \omega'_n) , \qquad \sigma = (\omega_n)_{n\in\mathbb{N}}, \sigma' = (\omega'_n)_{n\in\mathbb{N}} \in \Sigma . \tag{4}$$

[5] It is compact, by Tychonoff's theorem [57, Section A.3].

Note that the metric is defined to satisfy the normalization

$$\max d_\eta (\Sigma \times \Sigma) \doteq \max \{d_\eta (\sigma_1, \sigma_2) : \sigma_1, \sigma_2 \in \Sigma\} = 1,$$

given a fixed parameter $\eta \in (0,1)$. Observe that [54, Proposition 1.2] shows that results for equilibrium probabilities on $\Omega^{\mathbb{Z}}$ can be derived from the ones obtained in $\Omega^{\mathbb{N}}$, which are studied here.

Real-valued functions on strings. $C(\Sigma) \equiv C(\Sigma; \mathbb{R})$ is the Banach space of all real-valued continuous functions endowed with the supremum norm:

$$\|\varphi\|_\infty \equiv \|\varphi\|_{\infty,\Sigma} \doteq \sup |\varphi(\Sigma)| \doteq \sup \{|\varphi(\sigma)| : \sigma \in \Sigma\} \ , \qquad \varphi \in C(\Sigma) \ .$$

$C_{\mathrm{f}}(\Sigma) \equiv C_{\mathrm{f}}(\Sigma; \mathbb{R})$ denotes the subspace of all continuous functions $\Sigma \to \mathbb{R}$ that are supported in some cylinder $[O_1, \ldots, O_N]$. It is a dense subspace of $C(\Sigma)$.

For any $\alpha \in (0,1]$,

$$C^\alpha(\Sigma) \equiv C^\alpha(\Sigma; \mathbb{R}) \doteq \left\{\varphi \in C(\Sigma) : \exists C > 0 \text{ so that } |\varphi(\sigma) - \varphi(\sigma')| \leq C d_\eta (\sigma, \sigma')^\alpha \text{ for } \sigma, \sigma' \in \Sigma\right\}$$

is the space of all real-valued α-Hölder-continuous functions with respect to the metric d_η. We use in this space the so-called Hölder norm

$$\|\varphi\|_\alpha \doteq \|\varphi\|_\infty + \sup_{\sigma,\sigma' \in \Sigma,\, \sigma \neq \sigma'} \left\{d_\eta (\sigma, \sigma')^{-\alpha} |\varphi(\sigma) - \varphi(\sigma')|\right\} \ , \qquad \varphi \in C^\alpha(\Sigma) \ , \tag{5}$$

with respect to which $C^\alpha(\Sigma)$ is a Banach space. Note also that $C_{\mathrm{f}}(\Sigma) \subseteq C^\alpha(\Sigma)$. In particular, $C^\alpha(\Sigma)$ is a dense subspace of $C(\Sigma)$.

Some of our results require the use of the unit closed ball

$$S \doteq \{\varphi \in C(\Sigma) : \|\varphi\|_\infty \leq 1\} \subseteq C(\Sigma) \tag{6}$$

of the Banach space $C(\Sigma)$ of continuous functions. S is endowed with the uniform metric

$$d_S(\varphi, \varphi') \doteq \|\varphi - \varphi'\|_\infty \ , \qquad \varphi, \varphi' \in S \ .$$

Note that S is a separable metric space, $C(\Sigma)$ being a separable normed space. Then, $\mathcal{M}(S) \equiv \mathcal{M}(S; \mathbb{R})$ denotes the Banach space of bounded, real-valued Borel-measurable functions $S \to \mathbb{R}$ with the supremum norm:

$$\|f\|_\infty \equiv \|f\|_{\infty,S} \doteq \sup |f(S)| \doteq \sup \{|f(\varphi)| : \varphi \in S\} \ , \qquad f \in \mathcal{M}(S) \ .$$

Shift-invariant probability measures on strings. The so-called *shift mapping* $T : \Sigma \to \Sigma$ is defined by

$$T(\sigma)_n \doteq \omega_{n+1} \ , \qquad n \in \mathbb{N}, \ \sigma = (\omega_n)_{n \in \mathbb{N}} \in \Sigma \ . \tag{7}$$

Clearly, T is continuous and in particular Borel measurable. Then, a probability measure μ on the Borel σ-algebra of Σ is, by definition, T*-invariant* when $T_*(\mu) = \mu$, where $T_*(\mu)$ stands for the pushforward of the measure μ with respect to T, i.e., $T_*(\mu)(B) \doteq \mu(T^{-1}(B))$ for any Borel set $B \subseteq \Sigma$.

As Σ is a metrizable compact space, by the Riesz-Markov-Kakutani representation theorem, one can identify the space of Borel measures of finite variation on Σ with the (topological) dual space $C(\Sigma)^*$ of the Banach space $C(\Sigma)$. See, e.g., [3, 14.15 Corollary]. $\mathcal{P} \subseteq C(\Sigma)^*$ denotes the convex space of all Borel probability measures on Σ and the convex set of T-invariant ones is denoted

$$\mathcal{P}(T) \doteq \{\mu \in \mathcal{P} : T_*(\mu) = \mu\} \ . \tag{8}$$

It is a weak*-closed subset of $\mathcal{P}$ and, since $\mathcal{P}$ is weak*-compact, $\mathcal{P}(T)$ is also weak*-compact. Notice that the weak* topology of $\mathcal{P}$ and, consequently, that of $\mathcal{P}(T)$, is metrizable, because the Banach space $C(\Sigma)$ is separable, as Σ is a metrizable compact space.

Thanks to the Krein-Milman theorem [57, Theorem 3.23], $\mathcal{P}(T)$ is the weak*-closure of the convex hull of the (nonempty) set $\mathcal{P}_{\mathrm{erg}}(T)$ of its extreme points, i.e.,

$$\mathcal{P}(T) = \overline{\mathrm{co}}\, \mathcal{P}_{\mathrm{erg}}(T) \ .$$

By the metrizability of $\mathcal{P}(T)$ and Lemma 3.27, $\mathcal{P}_{\mathrm{erg}}(T)$ is a Borel set with respect to the weak* topology. Extreme T-invariant Borel probability measures on Σ, i.e., the elements of $\mathcal{P}_{\mathrm{erg}}(T)$, are the *ergodic* measures, because of Proposition 2.34. See also [66] for the case of a finite alphabet. In addition, by Proposition 2.36, the set $\mathcal{P}_{\mathrm{erg}}(T)$ of ergodic measures is a weak*-dense subset of $\mathcal{P}(T)$:

$$\mathcal{P}(T) = \overline{\mathcal{P}_{\mathrm{erg}}(T)} \ . \tag{9}$$

2.2 Entropy for compact metric alphabets

In the literature, entropy is usually only defined for finite alphabets Ω, but, motivated by possible applications to continuous-spin systems, following [50] (see also [1]), we allow the alphabet Ω to be an arbitrary compact metric space. Entropy is not as well understood in this broader context and much less literature is available.

To define it, recall that we fix in all the paper an arbitrary a priori probability measure m whose support is equal to Ω. It is used to introduce the (transfer) Ruelle operator $\mathcal{L}_0^{\mathrm{m}}$, which is the linear operator on $C(\Sigma)$ defined, for any $\varphi \in C(\Sigma)$ and $\sigma = (\omega_n)_{n\in\mathbb{N}} \in \Sigma$, by

$$\mathcal{L}_0^{\mathrm{m}}\varphi(\sigma) \doteq \int_\Omega \varphi(\omega_0\sigma)\, \mathrm{m}(\mathrm{d}\omega_0) \ , \qquad \omega_0\sigma \doteq (\omega_0, \omega_1, \ldots) \in \Sigma \ .$$

In the case the alphabet is the finite set $\{1, 2, ..., k\}$, it is natural to take the a priori probability measure m as the normalized counting measure, see [50] for details.

Let $\mathcal{C}^+$ be the cone of positive functions of $C(\Sigma)$, i.e., the set of continuous functions $\Sigma \to \mathbb{R}_0^+$. Then, the entropy for T-invariant probability measures is defined as follows:

Definition 2.1 (Entropy)
The entropy of any T-invariant probability measure is, by definition, equal to

$$h(\mu) \equiv h_{\mathrm{m}}(\mu) \doteq \inf_{\varphi\in\mathcal{C}^+} \left\{ \int_\Sigma \ln\left(\frac{\mathcal{L}_0^{\mathrm{m}}\varphi(\sigma)}{\varphi(\sigma)}\right) \mu(\mathrm{d}\sigma) \right\} \ , \qquad \mu \in \mathcal{P}(T) \ . \tag{10}$$

Remark that the infimum in (10) is not necessarily attained for some positive function $\varphi \in \mathcal{C}^+$. In addition, the entropy is a mapping $\mu \mapsto h(\mu)$ from $\mathcal{P}(T)$ to $[-\infty, 0]$. For instance, the T-invariant measure whose support is a fixed point for T is equal to $-\infty$, while

$$h(\mu) \leq \int_\Sigma \ln\left(\frac{\mathcal{L}_0^{\mathrm{m}}\mathbf{1}(\sigma)}{\mathbf{1}(\sigma)}\right) \mu(\mathrm{d}\sigma) = 0$$

for any T-invariant probability measure $\mu \in \mathcal{P}(T)$, where $\mathbf{1} \in \mathcal{C}^+$ is the (constant) function equal to 1.

A useful observation concerns an equivalent definition of the entropy defined above. To state it, we use a generalization of the Ruelle operator. For a given α-Hölder-continuous potential $f \in C^\alpha(\Sigma)$

($\alpha \in (0,1)$) and an a priori measure m on Ω whose support is equal to Ω, it is defined, for any $\varphi \in C^{\alpha}(\Sigma)$, by

$$\mathcal{L}_f^{\mathrm{m}}(\varphi)(\sigma) = \int_{\Omega} \exp(f(\omega_0\sigma))\varphi(\omega_0\sigma)\,\mathrm{m}(\mathrm{d}\omega_0) \ , \qquad \omega_0\sigma \doteq (\omega_0,\omega_1,\ldots) \in \Sigma \ .$$

This Ruelle operator is a linear operator on $C^{\alpha}(\Sigma)$. A normalized potential f is, by definition, a potential $f \in C^{\alpha}(\Sigma)$ ($\alpha \in (0,1)$) satisfying $\mathcal{L}_f^{\mathrm{m}}(\mathbf{1})(\sigma) = 1$ for all $\sigma \in \Sigma$. Then, we have the following fact:

Proposition 2.2 (Alternative definition of the entropy)
For the entropy of Definition 2.1, the following equality holds:

$$h(\mu) = \inf_{f \in C^{\alpha}(\Sigma)} \left\{ \log \lambda_f - \int_{\Sigma} f(\sigma)\,\mu(\mathrm{d}(\sigma)) \right\} \ , \tag{11}$$

where λ_f is the main (i.e., the largest) eigenvalue of the Ruelle operator $\mathcal{L}_f^{\mathrm{m}}$.

Proof. See in [33, Remark 7.8] and [1, Definition 2]. Note that the arguments in [33, Section 7] refer to a finite alphabet Ω, but a close analysis of the proofs shows that only the abstract properties of the Legendre-Fenchel transform are necessary. Therefore, everything can be done in the same way with a compact metric space Ω and, in particular, the first equation of [33, Section 7.1] applies to the case the alphabet is a compact metric space Ω. ■

By Definition 2.1, the mapping $\mu \mapsto h(\mu)$ from $\mathcal{P}(T)$ to $[-\infty, 0]$ is upper semicontinuous in the weak* topology, the entropy being defined as the infimum of weak*-continuous functions (indexed by $\varphi \in \mathcal{C}^+$). It is additionally affine, thanks to Proposition 2.38. That is, the entropy h is both concave and convex. See also [66] for the case of finite alphabets. This follows from the concept of specific entropy given in [1] for compact metric alphabets, along with the characterization of the entropy in terms of a thermodynamic limit, as given by [1, Theorems 3.1 and 3.4]. See Theorem 2.37. We elaborate on that in Section 2.9.2.

Remark 2.3 (Alternative formulations of entropy)
Non-dynamical entropies, such as the Gibbs-Boltzmann (or Shannon) entropy (density) can also be considered. Recent relevant examples include more general forms of entropy on the probability space, as presented in [51, 52], along with their relationship to the pushforward dynamics (the level-2 setting). There, the authors define an entropy that generalizes the one we use for a compact metric alphabet. [8] defines an abstract pressure and derives a corresponding entropy via the Legendre-Fenchel transform. In all these variants, the entropy functional remains concave and upper semicontinuous, ensuring that the results of Section 3 still apply.

2.3 Linear thermodynamic formalism

Given an a priori probability measure m whose support is equal to Ω, the so-called *topological pressure* (pressure for short) $\mathfrak{P}_L(\varphi)$ of a continuous function $\varphi \in C(\Sigma)$ is defined by

$$\mathfrak{P}_L(\varphi) \doteq \sup_{\mu \in \mathcal{P}(T)} \{ h(\mu) + \varphi(\mu) \} \ , \tag{12}$$

where

$$\varphi(\mu) \doteq \int_{\Sigma} \varphi(\sigma)\,\mu(\mathrm{d}\sigma) \ , \qquad \mu \in \mathcal{P}(T) \ .$$

In this context, the continuous function $\varphi \in C(\Sigma)$ is called *potential*. This pressure is nothing else than the Legendre-Fenchel transform of minus the entropy and, conversely, the entropy is minus the Legendre-Fenchel transform of the pressure (see [66, Theorem 9.12] for the case of a finite alphabet):

Proposition 2.4 (Pressure versus entropy)
Considering the entropy of Definition 2.1, one has

$$h(\mu) = \inf_{\varphi \in C(\Sigma)} \{\mathfrak{P}_L(\varphi) - \varphi(\mu)\}\ , \qquad \mu \in \mathcal{P}(T)\ .$$

Proof. By Definition 2.1, h is concave and weak*-upper semicontinuous, being the infimum of weak* continuous functions indexed by $\varphi \in \mathcal{C}^+$. So, the assertion is nothing else than Equation (122) for the concave and weak*-upper semicontinuous function g on the dual space $C(\Sigma)^*$ defined by $g = h$ on $\mathcal{P}(T)$ and $-\infty$ otherwise. ■

In the thermodynamic formalism (and statistical mechanics of one-dimensional lattices) one is interested in T-invariant probabilities μ_φ giving the exact pressure $\mathfrak{P}_L(\varphi)$, i.e., satisfying the equality

$$P_L(\mu_\varphi) = \sup P_L(\mathcal{P}(T))\ , \tag{13}$$

where

$$P_L(\mu) \doteq h(\mu) + \varphi(\mu)\ , \qquad \mu \in \mathcal{P}(T)\ .$$

(Remark that the potential φ considered here corresponds to minus the Hamiltonian in statistical physics.) Such probabilities refer to the following notion:

Definition 2.5 (Equilibrium measures – linear case)
Given a continuous function $\varphi \in C(\Sigma)$, a probability measure μ_φ satisfying (13) is called a linear equilibrium measure for φ.

The set

$$E_{P_L} \doteq \{\mu \in \mathcal{P}(T) : P_L(\mu) = \sup P_L(\mathcal{P}(T))\} \tag{14}$$

of linear equilibrium measures is nonempty, weak*-compact and convex, because the mapping $\mu \mapsto \varphi(\mu) + h(\mu)$ from $\mathcal{P}(T)$ to $\mathbb{R}$ is upper weak*-semicontinuous and affine, thanks to Proposition 2.38.

When φ is of Hölder class, i.e., $\varphi \in C^\alpha(\Sigma)$ for some $\alpha \in (0,1]$, the linear equilibrium measure μ_φ is unique and ergodic [50, 1]. The set of all linear equilibrium measures μ_φ for all possible Hölder potentials φ is even weak*-dense in $\mathcal{P}(T)$, i.e.,

$$\overline{\{\mu_\varphi : \exists \alpha \in (0,1] \text{ so that } \varphi \in C^\alpha(\Sigma)\}} = \mathcal{P}(T)\ .$$

(In particular one has that the ergodic probabilities are dense in $\mathcal{P}(T)$, see (9).) This follows from [33, Corollary 7.14]. It is easy to see that the proof of this result given in [50, 1] is general enough to include the case where the alphabet is a compact metric space.

2.4 Nonlinear thermodynamic formalism

A nonlinear version of the thermodynamic formalism is relatively recent. It appeared in a series of seminal papers [45, 46, 7, 25, 42], which contribute a rigorous approach to studying mean-field theory in classical statistical mechanics (cf. the Curie-Weiss-Potts models).

In previous works, the nonlinearity is encoded by a continuous function $F : \mathbb{R}^N \to \mathbb{R}$ ($N \in \mathbb{N}$), the arguments of which are the expectation values $\mu(\varphi_1), \ldots, \mu(\varphi_N)$ of N fixed α-Hölder-continuous functions $\varphi_1, \ldots, \varphi_N \in C^\alpha(\Sigma)$ with respect to a probability measure $\mu \in \mathcal{P}(T)$. This refers to Example 2.6 presented below. Here we consider a significant generalization of what has been done in the literature on nonlinear thermodynamic formalism by considering nonlinearities that, instead of depending on a finite vector $\varphi_1, \ldots, \varphi_N \in C^\alpha(\Sigma)$ of α-Hölder-continuous functions, now depend on a continuous function on the space $C^\alpha(\Sigma)$ of α-Hölder-continuous functions. This allows us, for instance, to include Example 2.7, which is the classical analogue of the quantum lattice systems considered in [18].

Intuitively, this allows us to use infinite, possibly uncountable, families $(\varphi_\alpha)_{\alpha\in I} \in C^\alpha(\Sigma)$ of Hölder potentials to formally define the nonlinear part as $F((\mu(\varphi_\alpha))_{\alpha\in I})$ for any probability measure $\mu \in \mathcal{P}(T)$. We restrict ourselves to uniformly bounded families $(\varphi_\alpha)_{\alpha\in I}$ to avoid innocuous technical complications and discussions. Since the function F will be completely arbitrary, we can take $(\varphi_\alpha)_{\alpha\in I}$ in the unit ball $S \subseteq C^\alpha(\Sigma)$ without loss of generality (for the uniformly bounded case).

Therefore, for any continuous linear functional $\mu : C(\Sigma) \to \mathbb{R}$, define the continuous bounded function $\mu_S \in C(S) \equiv C(S;\mathbb{R})$ by

$$\mu_S(\varphi) \doteq \mu(\varphi) , \qquad \varphi \in S . \tag{15}$$

In other words, μ_S is a kind of projectivization of the action of μ on $C(\Sigma)$. Then, our aim is to study the variational problem

$$\mathfrak{P}(F) \doteq \sup_{\mu\in\mathcal{P}(T)} \{h(\mu) + F(\mu_S)\} , \tag{16}$$

where $F : C(S) \to \mathbb{R}$ is some function satisfying

$$\sup_{\mu\in\mathcal{P}(T)} F(\mu_S) < \infty . \tag{17}$$

(For instance, F is bounded on the unit closed ball of $C(S)$. Note that μ_S is in this ball for all $\mu \in \mathcal{P}(T)$.) Recall that $h(\mu)$ stands for the entropy of the T-invariant measure $\mu \in \mathcal{P}(T)$. The quantity $\mathfrak{P}(F)$ is called here the nonlinear pressure of F.

As an example, if for some fixed α-Hölder-continuous potential $A \in C^\alpha(\Sigma)\backslash\{0\}$ we are interested in maximizing the quantity

$$\frac{1}{2}\mu(A)^2 + h(\mu)$$

over T-invariant measures $\mu \in \mathcal{P}(T)$, then we take

$$F(f) = \frac{1}{2}\left(\|A\|_\infty f\left(\frac{A}{\|A\|_\infty}\right)\right)^2 , \qquad f \in C(S) . \tag{18}$$

Moreover, since the metric d and the function F are arbitrary, our setting also includes the case studied by [25], which is described in detail in Section 2.6. The most important cases for applications in statistical mechanics are probably the following examples:

Example 2.6
Fix $N \in \mathbb{N}$, take any continuous function F on $\mathbb{R}^N$, along with (normalized) potentials $\varphi_1, \ldots, \varphi_N \in S$. Then

$$F^{(\varphi_1,\ldots,\varphi_N)}(f) \doteq F(f(\varphi_1), \ldots, f(\varphi_N)) , \qquad f \in \mathcal{M}(S) ,$$

defines a continuous function on $\mathcal{M}(S)$ that is bounded on bounded sets. This includes the following important particular choice:

$$F(\mu_S) = \mu(\varphi_1) + \lambda(\mu(\varphi_2))^2 , \qquad \mu \in \mathcal{P}(T) ,$$

for any continuous functions[6] *$\varphi_1, \varphi_2 \in C(\Sigma)$ and $\lambda \in \mathbb{R}$. In particular, the linear case $\lambda = 0$ is included.*

[6] A priori we should have $\varphi_1, \varphi_2 \in S$, but a simple change of the definition of the original function F gives $\mu(\varphi_1) + \lambda(\mu(\varphi_2))^2$ for arbitrary $\varphi_1, \varphi_2 \in C(\Sigma)$.

Example 2.7
Take any finite positive Borel measure $\mathfrak{a}$ on S, $\lambda \in \mathbb{R}$ and some continuous function $\varphi \in S$. Then,

$$F_{\pm}^{(\varphi,\mathfrak{a})}(f) \doteq \lambda f(\varphi) \pm \int_S f(\varphi')^2 \, \mathfrak{a}(\mathrm{d}\varphi') \ , \qquad f \in \mathcal{M}(S) \ ,$$

defines a continuous real-valued function on $\mathcal{M}(S)$ that is bounded on bounded sets.

Similar to Definition 2.5, we define equilibrium measures in the nonlinear thermodynamic formalism as follows:

Definition 2.8 (Equilibrium measures – nonlinear case)
A T-invariant probability measure maximizing (16) is called a nonlinear equilibrium measure.

We show, among other things, that the nonlinear equilibrium measures considered in the paper [25] fit (under its assumptions) within the above definition. However, our setting covers a much broader class of models, which are in line with important examples from mathematical physics. These are related to a convex nonlinearity F_+, representing long-range attractive forces, and a concave one F_-, associated with long-range repulsive forces.

Note that the bare existence of a nonlinear equilibrium measure is unclear, depending on the choice of the function F in (16). Consequently, the existence of nonlinear equilibrium measures, and the entire scope of our study, require certain minimum conditions on the nonlinearity F. These are the following assumptions:

Condition 2.9
(i) *F is the restriction to $C(S) \subseteq \mathcal{M}(S)$ of the sum of two functions $F_\pm : \mathcal{M}(S) \to \mathbb{R}$ that are σ-normal, meaning here that, for any bounded sequence $(f_n)_{n\in\mathbb{N}} \subseteq \mathcal{M}(S)$ converging point-wise to $f \in \mathcal{M}(S)$, one has*

$$\lim_{n\to\infty} F_\pm(f_n) = F_\pm(f) \ .$$

(ii) *The functions F_+ and F_- are respectively convex and concave.*

Under these assumptions, the functions $F_\pm$ are necessarily continuous with respect to the uniform convergence in $\mathcal{M}(S)$ and they always satisfy (17): Observe that $\mathcal{P}(T)$ is weak*-compact and its weak* topology is metrizable, as $C(\Omega)$ is separable (with respect to the supremum norm, Ω being a compact metric space). In particular it is sequentially compact with respect to the weak* topology. Furthermore, the weak* convergence of μ in $\mathcal{P}(T)$ implies the point-wise convergence of μ_S in $\mathcal{M}(S)$. Consequently, the set $\{F(\mu_S) : \mu \in \mathcal{P}(T)\} \subseteq \mathbb{R}$ is (sequentially) compact and thus, bounded.

Typically, in statistical mechanics the convex function F_+ represents some mean-field attraction, while F_- represents a mean-field repulsion. This physical interpretation is similar to that given in [18] for quantum lattice systems. Here we simply refer to the concave and convex parts of the nonlinear pressure functional P, which is equal in this case to

$$P(\mu) = h(\mu) + F_-(\mu_S) + F_+(\mu_S) \ , \qquad \mu \in \mathcal{P}(T) \ . \tag{19}$$

Examples of such a σ-normal convex (concave) function F_+ (F_-) for applications in statistical mechanics are given by Examples 2.6 and 2.7[7] when the real-valued function F_+ (F_-) is additionally convex (concave) on $\mathbb{R}^N$. These situations are included in the following more general example:

[7]The σ-normality of $F_{\pm}^{(\varphi,\mathfrak{a})}$ is a direct consequence of Lebesgue's dominated convergence theorem.

Example 2.10
Let $\mathcal{X}_\pm$ be two real normed spaces and $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$ two continuous convex functions. Take two bounded linear transformations $\theta_\pm : \mathcal{M}(S) \to \mathcal{X}_\pm$ that are σ-normal, as explained in Condition 2.9 (i). Then define $F_\pm \doteq \pm g_\pm \circ \theta_\pm$, which are functions that are σ-normal. F_+ and F_- are obviously convex and concave, respectively.

The above examples $F_\pm^{(\varphi_1,\ldots,\varphi_N)}$ and $F_\pm^{(\varphi,\mathfrak{a})}$ refer to the (real normed) spaces $\mathcal{X}_\pm = \mathbb{R}^N$ and $\mathcal{X}_\pm = \mathbb{R} \times L^2(S,\mathfrak{a})$, respectively, with obvious choices for the convex functions $g_\pm$ and the linear transformations $\theta_\pm$.

Under Condition 2.9 the nonlinear pressure functional $P : \mathcal{P}(T) \to \mathbb{R}$ of a T-invariant measure is weak*-upper semicontinuous, observing that the mappings $\mu \mapsto F_\pm(\mu_S)$ from $\mathcal{P}(T)$ to $\mathbb{R}$ are weak*-continuous and the entropy h, weak*-upper semicontinuous. Therefore, the variational problem (16) has maximizers and the set

$$E_P \doteq \{\mu \in \mathcal{P}(T) : P(\mu) = \sup P(\mathcal{P}(T))\} \tag{20}$$

of nonlinear equilibrium measures forms a (nonempty) weak*-compact and metrizable set, thanks to the metrizability of the weak* topology in $\mathcal{P}(T)$. However, unlike the linear case (cf. (14)), this set is **not necessarily convex** if the functional F is not at least concave. This contrasts with the linear case, for which F is affine (i.e., both convex and concave).

2.5 Generalized equilibrium measures

The non-convexity of the set E_P (20) of nonlinear equilibrium measure can be highly problematic in physical applications, since, in this case, the (convex) mixture of two equilibrium states would no longer be a new equilibrium measure, as physically expected. The same issue occurs for quantum spin systems or fermions on lattices in the presence of mean-field interactions [18]. In other words, physically, one should not only consider equilibrium measures as being the solutions to the nonlinear pressure, but also any convex combination of them should be an equilibrium measure. This refers in the present work to *generalized* nonlinear equilibrium measures. See Section 2.9.3 and [18, Definition 2.15 and Theorem 2.21].

Definition 2.11 (Generalized nonlinear equilibrium measures)
A generalized nonlinear equilibrium measure is any element of the weak-closed convex hull $G_P \doteq \overline{\mathrm{co}}(E_P)$ of the set E_P of (usual) nonlinear equilibrium measures.*

G_P is in particular weak*-compact and convex. Thus, by the Milman theorem [55, Proposition 1.5], generalized nonlinear equilibrium measures that are extreme in G_P must belong to the weak*-compact set E_P. In other words, extreme generalized nonlinear equilibrium measures are always usual nonlinear equilibrium measures. Then, by the Choquet theorem (Theorem 3.28), for any $\mu \in G_P$, there is a probability measure m_μ on E_P such that

$$\mu = \int_{E_P} \nu m_\mu(\mathrm{d}\nu) \ ,$$

i.e., the generalized nonlinear equilibrium measures are nothing but the barycenters of probability measures on the set of (usual) nonlinear equilibrium measures. See Definition 3.25 and compare also with Corollary 2.30.

We use the terminology "generalized nonlinear equilibrium measure" because of the property stated in Theorem 2.21 (ii), which establishes a precise relationship between these measures and the maximization of an affine functional ($\mathfrak{F}^\flat$) that naturally appears in our setting for the nonlinear thermodynamic formalism. This approach is inspired by the study of quantum lattice systems done in [18], in which the affine functional is canonical from a thermodynamic point of view, being the function that determines the properties of such quantum systems at equilibrium.

To explain this in a simple way, we come back to Example 2.6 above. Let m be any fixed a priori probability measure on the character set (Ω, d), whose support is equal to Ω, as in Section 2.2, and $\mathrm{m}_{\otimes} \in \mathcal{P}(T)$ the corresponding product (probability) measure on the symbolic space $\Sigma \doteq \Omega^{\mathbb{N}}$. Notice that $\mathrm{m}_{\otimes}$ is even ergodic. Let further

$$\mathbb{E}_n [\varphi] \doteq n^{-1} \left(\varphi + \varphi \circ T + \cdots + \varphi \circ T^{n-1} \right) , \qquad n \in \mathbb{N},\ \varphi \in C(\Sigma) , \tag{21}$$

be the normalized Birkhoff sums of continuous potentials φ. Then, for any $n \in \mathbb{N}$, the corresponding Gibbs measure on $C(\Sigma)$ is defined for any $\varphi \in C(\Sigma)$ by

$$\mu_{\mathrm{Gibbs}}^{(n)} (\varphi) \doteq \frac{1}{Z_{\mathrm{Gibbs}}^{(n)}} \int_{\Sigma_n} \mathrm{e}^{F_+ (\mathbb{E}_n[\varphi_1](\sigma), \ldots, \mathbb{E}_n[\varphi_N](\sigma)) + F_- (\mathbb{E}_n[\varphi_1](\sigma), \ldots, \mathbb{E}_n[\varphi_N](\sigma))} \varphi(\sigma) \mathrm{m}_{\otimes} (\mathrm{d}\sigma) ,$$

where $\varphi_1, \ldots, \varphi_N$ are the potentials of Example 2.6 and

$$Z_{\mathrm{Gibbs}}^{(n)} \doteq \int_{\Sigma_n} \mathrm{e}^{F_+ (\mathbb{E}_n[\varphi_1](\sigma), \ldots, \mathbb{E}_n[\varphi_N](\sigma)) + F_- (\mathbb{E}_n[\varphi_1](\sigma), \ldots, \mathbb{E}_n[\varphi_N](\sigma))} \mathrm{m}_{\otimes} (\mathrm{d}\sigma) .$$

Then, using Varadhan's lemma one shows [20] that the following limit exists and is given by a variational problem on the set of T-invariant probability measures:

$$\lim_{n \to \infty} n^{-1} \ln Z_{\mathrm{Gibbs}}^{(n)} = \sup \mathfrak{F}^{\flat} (\mathcal{P}(T)) , \tag{22}$$

where $\mathfrak{F}^{\flat} : \mathcal{P}(T) \to \mathbb{R}$ is the affine, albeit **not** necessarily weak*-upper semicontinuous, functional defined below by (47) and named here the *affine nonlinear pressure*. Keeping in mind that $\mathcal{P}(T)$ is sequentially weak*-compact, the set of approximated maximizers of this variational problem is then defined by

$$E_{\mathfrak{F}^{\flat}} \doteq \left\{ \mu \in \mathcal{P}(T) : \exists \, (\mu_n)_{n \in \mathbb{N}} \subseteq \mathcal{P}(T) \text{ with } \lim_{n \to \infty} \mu_n = \mu \text{ and } \lim_{n \to \infty} \mathfrak{F}^{\flat} (\mu_n) = \sup \mathfrak{F}^{\flat} (\mathcal{P}(T)) \right\} , \tag{23}$$

where the limit of T-invariant probability measures refers to the weak* convergence. This set is nothing but the weak*-closed convex hull $G_P \doteq \overline{\mathrm{co}}(E_P)$ of E_P:

Proposition 2.12 (Generalized equilibrium measures)
Fix $N \in \mathbb{N}$, take any convex (concave) real-valued function F_+ (F_-) on $\mathbb{R}^N$, as well as norm-one potentials $\varphi_1, \ldots, \varphi_N \in S$, and define the continuous function $F_{\pm}^{(\varphi_1, \ldots, \varphi_N)}$ as in Example 2.6. Then, $E_{\mathfrak{F}^{\flat}} = \overline{\mathrm{co}}(E_P) \doteq G_P$.

Proof. The assertion is a direct consequence of Theorem 2.29. ■

Observe that weak*-accumulation points of probability measures $(\mu_{\mathrm{Gibbs}}^{(n)} \circ \mathbb{E}_n)_{n \in \mathbb{N}}$ always belong to $G_P = E_{\mathfrak{F}^{\flat}}$, but **not necessarily** to E_P. This last property is proven in [20] and we omit the details here because we are solely interested in the variational problems themselves, with no (extended) thermodynamic considerations other than those presented above for pedagogical reasons.

In fact, Theorem 2.29 is a pivotal result in this context, which can be used to generalize Proposition 2.12 beyond Example 2.6, i.e., for any σ-normal functions $F_{\pm} : \mathcal{M}(S) \to \mathbb{R}$, with F_+ and F_- being respectively convex and concave. This generalization is not considered in the present section to avoid repeating essentially the same arguments in a case that is more complicated to explain properly and, therefore, much less pedagogical. In fact, Proposition 2.12 already shows us that the notion of generalized nonlinear equilibrium states is the correct one for describing the thermodynamic limit of nonlinear systems at equilibrium.

Remark 2.13
Buzzi, Kloeckner and Leplaideur [25] already explain how to obtain nonlinear equilibrium measures from an affine pressure, but their approach ***is very different*** *from the one that leads to the affine pressure $\mathfrak{F}^{\flat}$ presented above. Compare (79) below with (25)–(26). In fact, the arguments used by these authors are reminiscent of the Bogoliubov linearization, which is explained below. However, they do not exploit it to the fullest extent, which is what we propose to do here.*

2.6 Buzzi-Kloeckner-Leplaideur's approach to nonlinear pressures

In 2019, Leplaideur and Watbled [25] pioneered the study of nonlinear topological pressure through an affine pressure. Considering a generalized Curie-Weiss model, they investigated the special case of quadratic non-linearity. In 2023, in the context of Example 2.6 Buzzi, Kloeckner and Leplaideur [25] show how nonlinear equilibrium measures can be obtained from an affine pressure. In fact, given $N \in \mathbb{N}$, $\varphi_1, \ldots, \varphi_N \in C(\Sigma)$ and a real-valued continuous function F on some set $U \subseteq \mathbb{R}^N$ (cf. Example 2.6) containing the compact convex set

$$\{(\mu(\varphi_1), \ldots, \mu(\varphi_N)) : \mu \in \mathcal{P}(T)\} \subseteq U \,, \tag{24}$$

observe that

$$\sup_{\mu \in \mathcal{P}(T)} P(\mu) \doteq \sup_{\mu \in \mathcal{P}(T)} \{F(\mu(\varphi_1), \ldots, \mu(\varphi_N)) + h(\mu)\} = \sup_{z \in U} \{F(z) + \mathrm{h}(z)\} \,, \tag{25}$$

where

$$\mathrm{h}(z) \doteq \sup \{h(\mu) : \mu \in \mathcal{P}(T) \text{ such that } z = (\mu(\varphi_1), \ldots, \mu(\varphi_N))\} \,.$$

The variational problem (25) (used in [25]) is a particular case of (16), in accordance with Example 2.6.

By [25, Proposition 3.3], solutions μ_z to this last variational problem are linear equilibrium measures ν for some linear combination $y_1\varphi_1 + \cdots + y_N\varphi_N$, i.e.,

$$p(\nu, y_1, \ldots, y_N) = \sup_{\mu \in \mathcal{P}(T)} p(\mu, y_1, \ldots, y_N) \doteq \mathrm{P}(y_1, \ldots, y_N) \,, \tag{26}$$

where

$$p(\mu, y_1, \ldots, y_N) \doteq h(\mu) + y_1\mu(\varphi_1) + \cdots + y_N\mu(\varphi_N) \,, \qquad y_1, \ldots, y_N \in \mathbb{R}^N, \ \mu \in \mathcal{P}(T) \,.$$

By [25, Theorem C], all the values of $y_1, \ldots, y_N$ to be taken to maximize the nonlinear pressure functional P in (25) are obtained from the maximizers in $U \subseteq \mathbb{R}^N$ of the variational problem on the right-hand side of Equation (25), in the following way: Define the gradient $\nabla \mathrm{P} : \mathbb{R}^N \to \mathbb{R}^N$ by

$$\nabla \mathrm{P}(y_1, \ldots, y_N) \doteq (\partial_{y_1}\mathrm{P}(y_1, \ldots, y_N), \ldots, \partial_{y_N}\mathrm{P}(y_1, \ldots, y_N)) \,.$$

Then, we must take all possible $(y_1, \ldots, y_N) \in \mathcal{Y} \doteq (\nabla \mathrm{P})^{-1}(\mathcal{V})$ in (26), where

$$\mathcal{V} \doteq \left\{ c \in \mathbb{R}^N : F(c) + \mathrm{h}(c) = \sup_{z \in \mathbb{R}^N} \{F(z) + \mathrm{h}(z)\} \right\} \,, \tag{27}$$

in order to obtain all nonlinear equilibrium measures, i.e., maximizers of P (see the left-hand side of the variational problem (25)), as linear equilibrium measures ν of (26). This is done for C^1-functions F.

The authors get interesting results, but there are however several drawbacks: First, the computation of the set $\mathcal{V}$ looks highly nontrivial for general C^1-functions F. It requires, in particular, a good control of the entropy function h, which is minus the Legendre-Fenchel transform of the pressure function P. Then, one has to be able to compute the preimage of the set $\mathcal{V}$ trough the function $\nabla\mathrm{P}$, which is again another nontrivial task, in general.

Below we present a more effective method related to the *Bogoliubov linearization*, which is explained in detail and great generality in Section 3. As applied to the setting of [25, Theorem C], this method allows us to study the variational problem (25), but the corresponding set $\mathcal{Y}$ is obtained much more straightforwardly than in [25, Theorem C], by means of a simple variational problem that takes advantage of the convexity or concavity of the function F. Summarizing, via the Buzzi-Kloeckner-Leplaideur's method,

$$\sup_{\mu\in\mathcal{P}(T)} P\left(\mu\right) = P\left(\mu_c\right)$$

for some T-invariant measure $\mu_c \in \mathcal{P}\left(T\right)$ satisfying so-called self-consistency equations $\mu_c(\varphi_j) = c_j$, $j \in \{1,\ldots,N\}$, and

$$\sup_{\mu\in\mathcal{P}(T)} p\left(\mu, y_1,\ldots,y_N\right) = p\left(\mu_c, y_1,\ldots,y_N\right) \ ,$$

for some $c \in \mathcal{V}$ and $y_1,\ldots,y_N \in \mathcal{Y} \doteq \left(\nabla\mathrm{P}\right)^{-1}\left(\mathcal{V}\right)$. Theorem 2.17 below yields essentially the same result, but it does not require explicit knowledge of the set $\mathcal{V}$, which avoids the need to study a variational problem involving the highly non-trivial entropy function $\mathrm{h}\left(z\right)$. Instead, the numbers $y_1,\ldots,y_N$ are computed from a variational problem involving only linear pressures and the Legendre-Fenchel transform of F.

To get a first intuitive idea, let us assume for simplicity that F is convex and satisfies

$$F\left(z\right) = \sup_{x\in\mathbb{R}^N}\left\{\langle z,x\rangle_{\mathbb{R}^d} - F^*\left(x\right)\right\} \qquad \text{with} \qquad F^*\left(x\right) \doteq \sup_{z\in\mathbb{R}^N}\left\{\langle x,z\rangle_{\mathbb{R}^d} - F\left(z\right)\right\} \ . \tag{28}$$

Then

$$\begin{aligned}\sup_{\mu\in\mathcal{P}(T)} P\left(\mu\right) &= \sup_{z\in\mathbb{R}^N}\left\{F\left(z\right)+\mathrm{h}\left(z\right)\right\} = \sup_{x\in\mathbb{R}^N}\sup_{z\in\mathbb{R}^N}\left\{\langle z,x\rangle_{\mathbb{R}^d}+\mathrm{h}\left(z\right)-F^*\left(x\right)\right\}\\ &= \sup_{x\in\mathbb{R}^N}\left\{\mathrm{P}\left(x\right)-F^*\left(x\right)\right\} = \sup_{x\in\mathbb{R}^N}\inf_{z\in\mathbb{R}^N}\left\{\mathrm{P}\left(x\right)-\langle x,z\rangle_{\mathbb{R}^d}+F\left(z\right)\right\} \ .\end{aligned}$$

In particular, if one can interchange the last $\sup$ over $x \in \mathbb{R}^N$ and $\inf$ over $z \in \mathbb{R}^N$, then

$$\begin{aligned}\sup_{\mu\in\mathcal{P}(T)} P\left(\mu\right) &= \inf_{z\in\mathbb{R}^N}\sup_{x\in\mathbb{R}^N}\left\{\mathrm{P}\left(x\right)-\langle x,z\rangle_{\mathbb{R}^d}+F\left(z\right)\right\} = \inf_{z\in\mathbb{R}^N}\left\{\mathrm{P}\left(y\right)-\langle y,z\rangle_{\mathbb{R}^d}+F\left(z\right)\right\}\\ &= \mathrm{P}\left(y\right)-\langle y,c\rangle_{\mathbb{R}^d}+F\left(c\right)\end{aligned}$$

with y being chosen so that

$$\mathrm{P}\left(y\right)-\langle y,z\rangle_{\mathbb{R}^d} = \sup_{x\in\mathbb{R}^N}\left\{\mathrm{P}\left(x\right)-\langle x,z\rangle_{\mathbb{R}^d}\right\} \ .$$

In the same way, c is chosen so that

$$\mathrm{P}\left(y\right)-\langle y,c\rangle_{\mathbb{R}^d}+F\left(c\right) = \inf_{z\in\mathbb{R}^N}\left\{\mathrm{P}\left(y\right)-\langle y,z\rangle_{\mathbb{R}^d}+F\left(z\right)\right\} \ .$$

In particular, $\nabla\mathrm{P}(y) = c$, i.e., $y = (y_1,\ldots,y_N) \in \left(\nabla\mathrm{P}\right)^{-1}\left(\{c\}\right)$. Compare with [25, Theorem C] and $y_1,\ldots,y_N \in \mathcal{Y} \doteq \left(\nabla\mathrm{P}\right)^{-1}\left(\mathcal{V}\right)$ above. Equation (28) refers to the Legendre-Fenchel transform of the convex functions F and F^*, see Section 3.1 below.

Our method makes this heuristics rigorous. In fact, the problematic entropy function $\mathrm{h}\left(z\right)$ is **not** used directly in **any** of our arguments. Instead, we use its Legendre-Fenchel transform, $\mathrm{P}\left(x\right)$, which is much more tractable in many important applications. Our method also avoids calculating any gradients, but requires the convex and concave parts $F_\pm$ of F to be separated and written as Legendre-Fenchel transforms. Since [25] only considers the compact convex set (24) and any C^1-functions on a compact subset of $\mathbb{R}^N$ ($N \in \mathbb{N}$) can be represented as the difference of convex and continuous functions (see Remark 3.6), our assumptions are much weaker than those of [25]. Moreover, we provide new results. See Conditions TF1–TF3 of Section 2.7.1, Theorems 2.17, 2.21 and Corollary 2.18. For example, we can handle not only the case of finitely many potentials $\varphi_1, \ldots, \varphi_N$, as in [25] or Example 2.6, but also the very general framework of Example 2.10, while making the approach transparent with regard to the main arguments used in proofs.

2.7 New approach to nonlinear thermodynamic formalism

The sets E_P and $G_P \doteq \overline{\mathrm{co}}(E_P)$ of equilibrium measures are pivotal; see Equation (20), Definitions 2.8 and 2.11. However, it is a priori not clear how useful the corresponding variational principles defining these sets are to study phase transitions. In fact, it turns out to be more convenient to reduce the nonlinear thermodynamic formalism to the linear one, for which concrete calculations can often be performed, similar to Buzzi-Kloeckner-Leplaideur's approach (Section 2.6).

Our abstract theory of *Bogoliubov linearizations*, which is explained in detail in Section 3, provides a systematic approach to this. It generalizes the study conducted in [18] on quantum lattice systems, which, in this special case, resolves an old open problem addressed by Ginibre in 1968 [32] concerning the exactness of Bogoliubov approximations of equilibrium states in the thermodynamic limit. Indeed, in the quantum framework previously studied [18] only the specific case of "Example 2.7[8]" is studied. Our framework here is much more general, encompassing Example 2.10 and thus going far beyond the quadratic case of Example 2.7.

2.7.1 Bogoliubov linearizations

General assumptions. We study the nonlinear pressure

$$P(\mu) \doteq h(\mu) + F_-(\mu_S) + F_+(\mu_S) \ , \qquad \mu \in \mathcal{P}(T) \ ,$$

for the general case given by Example 2.10. In fact, our abstract theory of Bogoliubov linearizations would allow us to extend this example even further to functions that are not necessarily of the $F_\pm = \pm g_\pm \circ \theta_\pm$, and even **non**-$\sigma$-normal ones. For simplicity, we refrain from making this further generalization here, bearing in mind that the objective of this section is to illustrate our approach using the nonlinear thermodynamic formalism as a paradigmatic example.

To present our general assumptions, we use the Legendre-Fenchel transform $g^* : \mathcal{X}^* \to (-\infty, \infty]$ of a function $g : \mathcal{X} \to \mathbb{R}$ defined for a given dual pair $(\mathcal{X}, \mathcal{X}^*)$ by

$$g^*(x) \doteq \sup_{y \in \mathcal{X}} \{x(y) - g(y)\} \ , \qquad x \in \mathcal{X}^* \ . \tag{29}$$

We say that this Legendre-Fenchel transform has *minimal linear growth* $\lambda \in \mathbb{R}^+$ if

$$\lim_{R \to \infty} \sup_{y \in \mathcal{X}^* \backslash B(0,R)} \left\{ \lambda \|y\|_{\mathrm{op}} - g^*(y) \right\} = -\infty \ , \tag{30}$$

where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm (see (126)) and $B(0, R)$ is the closed ball of radius $R \in \mathbb{R}^+$ and center $0 \in \mathcal{X}^*$ (see (125)). We recommend referring to Section 3.1 for more details. We consider the following assumptions on the nonlinear energies:

[8] Adapted appropriately to the fermionic/quantum spin situation.

TF1 $\mathcal{X}_\pm$ are two real normed spaces and $\theta_\pm : \mathcal{M}(S) \to \mathcal{X}_\pm$ are two linear transformations that are σ-normal, i.e., for any bounded sequence $(f_n)_{n\in\mathbb{N}} \subseteq \mathcal{M}(S)$ converging point-wise to $f \in \mathcal{M}(S)$, $(\theta_\pm(f_n))_{n\in\mathbb{N}}$ converges to $\theta_\pm(f)$ in $\mathcal{X}_\pm$.

TF2 $g_- : \mathcal{X}_- \to \mathbb{R}$ is a lower semicontinuous and convex function for which there is a positive radius $R_0 \in \mathbb{R}^+$ such that $\{x \in \mathcal{X}_- : \|x\|_{\mathrm{op}} < R,\ g_-^*(x) < \infty\}$ is nonempty and weak*-closed for all $R \in [R_0, \infty)$. Moreover, g_-^* has minimal linear growth $\|\theta_-\|_{\mathrm{op}}$ in the sense of Equation (30).

TF3 $g_+ : \mathcal{X}_+ \to \mathbb{R}$ is a lower semicontinuous, bounded and convex function for which $g_+^*(\mathcal{X}_+) \subseteq \mathbb{R}$ and g_+^* has minimal linear growth $\|\theta_+\|_{\mathrm{op}}$ in the sense of Equation (30).

Note that a σ-normal linear mapping on $\mathcal{M}(S)$ is always bounded. In particular, $\|\theta_\pm\|_{\mathrm{op}} < \infty$. Observe also that Conditions TF2–TF3 are trivially satisfied by quadratic functions $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$, as g_+^* and g_-^* are also quadratic in this case. This is the most relevant case in statistical physics. These assumptions are in fact very general. They are an adapted version of (the much more general) Conditions B1–B3 of Section 3.3. Their general nature is discussed in Remarks 3.5 and 3.6.

Bogoliubov linearizations. They refer to a μ-linearization method for the nonlinear pressure, which is now defined as follows:

$$P(\mu) \doteq h(\mu) - g_- \circ \theta_-(\mu_S) + g_+ \circ \theta_+(\mu_S) \ , \qquad \mu \in \mathcal{P}(T) \ , \tag{31}$$

under Conditions TF1–TF3. Since $g^{**} = g$ for any lower semicontinuous convex function (see (122)), we observe from Equation (31) that

$$P(\mu) = \sup_{y_+ \in \mathcal{X}_+^*} \inf_{y_- \in \mathcal{X}_-^*} \left\{ \mathrm{P}(y_+, y_-, \mu) + g_-^*(y_-) - g_+^*(y_+) \right\} \ , \qquad \mu \in \mathcal{P}(T) \ , \tag{32}$$

where, for any $y_\pm \in \mathcal{X}_\pm^*$ and $\mu \in \mathcal{P}(T)$,

$$\mathrm{P}(y_+, y_-, \mu) \doteq h(\mu) - y_- \circ \theta_-(\mu_S) + y_+ \circ \theta_+(\mu_S) \ , \tag{33}$$

is called here an approximating pressure of the T-invariant probability measure μ.

We call the new functionals $\mu \mapsto \mathrm{P}(y_+, y_-, \mu)$, $y_\pm \in \mathcal{X}_\pm^*$, *Bogoliubov linearizations* of P. Then, the method we present refers to the study the μ-linearized variational problem

$$\mathrm{P}^\flat \doteq \sup_{y_+ \in \mathcal{X}_+^*} \inf_{y_- \in \mathcal{X}_-^*} P_{\mathrm{NL}}(y_+, y_-) \ , \tag{34}$$

instead of $\sup P(\mathcal{P}(T))$, where, for any continuous linear functionals $y_\pm \in \mathcal{X}_\pm^*$,

$$P_{\mathrm{NL}}(y_+, y_-) \doteq P_{\mathrm{L}}(y_+, y_-) + g_-^*(y_-) - g_+^*(y_+) \ , \tag{35}$$

$$P_{\mathrm{L}}(y_+, y_-) \doteq \sup \mathrm{P}(y_+, y_-, \mathcal{P}(T)) \doteq \sup_{\mu \in \mathcal{P}(T)} \mathrm{P}(y_+, y_-, \mu) \ . \tag{36}$$

P_{NL} and P_{L} are called here the *nonlinear and linear approximating pressures*, respectively. Thus, proceeding in this manner, we first analyze the linear problem for T-invariant probability measures and then study the nonlinear part via the other two variational problems over $\mathcal{X}_\pm^*$.

The variational problem (34) is supposed to be much easier than $\sup P(\mathcal{P}(T))$ in many important cases because the linear thermodynamic formalism (Section 2.3), which is now very well developed, should give us good control over the linear approximating pressures $P_{\mathrm{L}}(y_+, y_-)$. We explain below how (34) can be used rigorously to study the original variational problem $\sup P(\mathcal{P}(T))$. When $g_- = 0$, note from (32) that (34) and $\sup P(\mathcal{P}(T))$ are trivially equivalent variational problems, because two suprema always commute with each other. The main difficulty is therefore to handle the case where $g_- \neq 0$. This is done in Sections 3.2–3.3 using the celebrated von Neumann minimax theorem (Theorem 3.24).

Bogoliubov linearizations – Hölder case. Remark that, for any continuous linear functional $y_\pm \in \mathcal{X}_\pm^*$, the new functional

$$\mu \mapsto y_+ \circ \theta_+ (\mu_S) - y_- \circ \theta_- (\mu_S)$$

from $\mathcal{P}$ to $\mathbb{R}$ in Equation (33) is affine and weak* continuous. Similar to [19, Proposition 3.9], for any $y_\pm \in \mathcal{X}_\pm^*$, there is a unique continuous function $\Theta_{y_-,y_+} \in C(\Sigma)$ such that

$$y_+ \circ \theta_+ (\mu_S) - y_- \circ \theta_- (\mu_S) = \mu \left(\Theta_{y_-,y_+}\right) , \qquad \mu \in \mathcal{P} . \tag{37}$$

We name the continuous function Θ_{y_-,y_+} the *approximating potential* associated with $y_-, y_+, \theta_-, \theta_+$.

Note that the approximating potentials are not necessarily Hölder continuous. In particular, the linear equilibrium measure, i.e., the T-invariant probability measure realizing the supremum in (36), is not necessarily unique. This property can nevertheless always be ensured for specific linear transformations $\theta_\pm$:

Definition 2.14 (Hölder-type linear functions)
The pair (θ_-, θ_+) of linear functions is Hölder-type if, for all $y_\pm \in \mathcal{X}_\pm^*$, their associated approximating potential (see (37)) is Hölder continuous, i.e., $\Theta_{y_-,y_+} \in C^\alpha(\Sigma)$ for some $\alpha = \alpha_{y_-,y_+} \in (0,1]$.

It corresponds in Example 2.6 to fix functions $\varphi_1, \ldots, \varphi_N$ only in $C^\alpha(\Sigma)$, instead of the whole space $C(\Sigma)$. Another general example of this situation can be given in the scope of Example 2.7 as follows: Fix $\alpha \in (0,1]$. Define the unit closed ball

$$S_\alpha \doteq \{\varphi \in C^\alpha (\Sigma) : \|\varphi\|_\alpha \leq 1\} \subseteq C^\alpha (\Sigma)$$

of the Banach space $C^\alpha(\Sigma)$ of α-Hölder continuous functions, endowed with the Hölder metric

$$d_{S_\alpha} (\varphi, \psi) \doteq \|\varphi - \psi\|_\alpha , \qquad \varphi, \psi \in S_\alpha .$$

See also Equation (5). Take any finite positive Borel measure $\mathfrak{a}_\alpha$ on S_α, $\lambda \in \mathbb{R}$ and some α-Hölder continuous function $\varphi \in S_\alpha$. Observe that the identity mapping $S_\alpha \to S$ is continuous. Thus, we can define a measure $\mathfrak{a}$ on S as the pushforward of $\mathfrak{a}_\alpha$ through the identity mapping. Then, with such a measure $\mathfrak{a}$ in Example 2.7 we again obtain Hölder-type functions (θ_-, θ_+), where $\mathcal{X}_\pm = \mathbb{R} \times L^2(S, \mathfrak{a})$ with obvious choices for the linear transformations $\theta_\pm$.

Hölder-type linear transformations $\theta_\pm$ are particularly useful since the linear approximating pressure (36) leads in this case to a unique and ergodic linear equilibrium measure for any $y_\pm \in \mathcal{X}_\pm^*$, as already explained in Section 2.3. See also [50, 1]. This is a very interesting situation for studying the nonlinear problem, as we shall see.

2.7.2 Thermodynamic game

The μ-linearized variational problem $\mathrm{P}^\flat$ refers to Equations (34)–(36). Observe in particular that the concave and convex parts of the nonlinear pressure do not have symmetric roles, since an infimum and a supremum do not commute in general. In special situations, this could be the case, but certainly not in the general case, as is explicitly shown in [18, Section 2.7] for lattices quantum systems. A sufficient condition for $\sup$ and $\inf$ to commute is given by Sion's minimax theorem [41].

The switching of $\sup$ and $\inf$ as it appears in $\mathrm{P}^\flat$ (34) leads to the min-max variational problem

$$\mathrm{P}^\sharp \doteq \inf_{y_- \in \mathcal{X}_-^*} \sup_{y_+ \in \mathcal{X}_+^*} \left\{ P_{\mathrm{L}} (y_+, y_-) + g_-^* (y_-) - g_+^* (y_+) \right\} \doteq \inf_{y_- \in \mathcal{X}_-^*} \sup_{y_+ \in \mathcal{X}_+^*} P_{\mathrm{NL}} (y_+, y_-) . \tag{38}$$

The max-min variational problem $\mathrm{P}^{\flat}$ and min-max variational problem $\mathrm{P}^{\sharp}$ have both a meaning in terms of equilibrium measures, as shown in Section 2.7.4. In the case of lattice quantum systems, this has already been observed in [18, Theorem 2.36]. The max-min variational problem also appears in the so-called Kac or van der Waals limit of lattice quantum models [21].

Similar to what is done in [18, Definition 2.35] for quantum lattice systems, the variational problems $\mathrm{P}^{\flat}$ (34) and $\mathrm{P}^{\sharp}$ (38) can be interpreted as the conservative values of a two-person zero-sum game whose payoff function is nothing but the nonlinear approximating pressure P_{NL} (35), called here *the thermodynamic game*. Note that we always consider here *non-zero* continuous functions $g_{\pm} : \mathcal{X}_{\pm} \to \mathbb{R}$, as otherwise there is no proper two-person game. The thermodynamic game is characterized by the following objects:

- For any continuous linear functionals $y_{\pm} \in \mathcal{X}_{\pm}^{*}$, we define the variational problems

$$P^{\flat}(y_{+}) \doteq \inf P_{\mathrm{NL}}\left(y_{+}, \mathcal{X}_{-}^{*}\right), \qquad y_{+} \in \mathcal{X}_{+}^{*}, \tag{39}$$

$$P^{\sharp}(y_{-}) \doteq \sup P_{\mathrm{NL}}\left(\mathcal{X}_{+}^{*}, y_{-}\right), \qquad y_{-} \in \mathcal{X}_{-}^{*}, \tag{40}$$

  as well as their set of minimizers:

$$M^{\flat}(y_{+}) \doteq \left\{x_{-} \in \mathcal{X}_{-}^{*} : P^{\flat}(y_{+}) = P_{\mathrm{NL}}(y_{+}, x_{-})\right\} \subseteq \mathcal{X}_{-}^{*}, \tag{41}$$

$$M^{\sharp}(y_{-}) \doteq \left\{x_{+} \in \mathcal{X}_{+}^{*} : P^{\sharp}(y_{-}) = P_{\mathrm{NL}}(x_{+}, y_{-})\right\} \subseteq \mathcal{X}_{+}^{*}. \tag{42}$$

- For the max-min variational problem $\mathrm{P}^{\flat}$ and min-max variational problem $\mathrm{P}^{\sharp}$ we consider the sets of optimizers defined by

$$\mathrm{M}^{\flat} \doteq \left\{x_{+} \in \mathcal{X}_{+}^{*} : P^{\flat}(x_{+}) = \sup P^{\flat}\left(\mathcal{X}_{+}^{*}\right) \doteq \mathrm{P}^{\flat}\right\} \subseteq \mathcal{X}_{+}^{*}, \tag{43}$$

$$\mathrm{M}^{\sharp} \doteq \left\{x_{-} \in \mathcal{X}_{-}^{*} : P^{\sharp}(x_{-}) = \sup P^{\sharp}\left(\mathcal{X}_{-}^{*}\right) \doteq \mathrm{P}^{\sharp}\right\} \subseteq \mathcal{X}_{-}^{*}. \tag{44}$$

As proven in Section 3.6, under Conditions TF1–TF3 these objects have the properties gathered in the proposition below. Let $\mathrm{dom}(g_{-}^{*})$ be the domain of the Legendre-Fenchel transform g_{-}^{*} of the function $g_{-} : \mathcal{X}_{-} \to \mathbb{R}$, i.e., the set of all $x \in \mathcal{X}_{-}^{*}$ such that $g_{-}^{*}(x) < \infty$. Since the function g_{-} never takes the value $-\infty$ (Condition TF2), observe that $\mathrm{dom}(g_{-}^{*}) \neq \emptyset$ (see (123)) and $\mathrm{dom}(g_{-}^{*})$ is a convex subset of $\mathcal{X}_{-}^{*}$, by convexity of g_{-}^{*}.

Proposition 2.15 (Properties of variational problems)
Assume Conditions TF1–TF3. Then, the following assertions hold:
(i) *For all $y_{+} \in \mathcal{X}_{+}^{*}$ and $y_{-} \in \mathrm{dom}(g_{-}^{*}) \subseteq \mathcal{X}_{-}^{*}$, $P^{\flat}(y_{+}), P^{\sharp}(y_{-}) \in \mathbb{R}$ and $M^{\flat}(y_{+}), M^{\sharp}(y_{-})$ are nonempty weak*-compact sets. For all $y_{+} \in \mathcal{X}_{+}^{*}$, $M^{\flat}(y_{+})$ is additionally convex.*
(ii) $\mathrm{P}^{\flat}, \mathrm{P}^{\sharp} \in \mathbb{R}$ *and* $\mathrm{M}^{\flat}, \mathrm{M}^{\sharp}$ *are nonempty, norm-bounded and weak*-compact.* $\mathrm{M}^{\sharp}$ *is additionally a convex subset of* $\mathrm{dom}(g_{-}^{*})$.

Proof. Conditions TF1–TF3 imply Conditions B1–B3 of Section 3.3 for $K = \mathcal{P}(T)$, $f = h$ and $\tau_{\pm} : \mathcal{P}(T) \to \mathcal{X}_{\pm}$ defined by $\tau_{\pm}(\mu) \doteq \theta_{\pm}(\mu_{S})$ for any $\mu \in \mathcal{P}(T)$. Indeed, $\mathcal{P}(T)$ is a compact convex Hausdorff space, h is affine and upper semicontinuous (Definition 2.1 and Proposition 2.38), and $\tau_{\pm}$ is affine and continuous in this case, because the weak* topology of $\mathcal{P}(T)$ is metrizable and the weak* convergence of μ in $\mathcal{P}(T)$ implies the point-wise convergence of μ_{S} in $\mathcal{M}(S)$. Note also that

$$\|\tau_{\pm}\|_{\infty} \doteq \sup\left\{\|\tau_{\pm}(\mu)\|_{\mathcal{X}} : \mu \in \mathcal{P}(T)\right\} \leq \|\theta_{\pm}\|_{\mathrm{op}} \sup_{\mu \in \mathcal{P}(T)} \|\mu_{S}\|_{\mathrm{op}} = \|\theta_{\pm}\|_{\mathrm{op}}.$$

So, it suffices to invoke Propositions 3.7 and 3.9 to get the assertion. ∎

Much more features of the thermodynamic game can be considered. This is done in Section 3.4. For example, we can define decision rules for the thermodynamic game (Definition 3.10) and find their particular properties. E.g., if we assume Conditions TF1–TF3 but also that g_{+}^{*} is continuous, $\mathrm{dom}(g_{-}^{*}) = \mathcal{X}_{-}^{*}$ and $g_{-}^{*} : \mathcal{X}_{-}^{*} \rightarrow \mathbb{R}$ is strictly convex, then we can deduce from Proposition 3.11 that, for all $x_{+} \in \mathrm{M}^{\flat}$, $M^{\flat}(x_{+})$ contains exactly one element $x_{-}(x_{+}) \in \mathcal{X}_{+}^{*}$. In this case, the function $x_{+} \mapsto x_{-}(y_{+})$ from $\mathrm{M}^{\flat}$ to $\mathcal{X}_{-}^{*}$ is continuous, both sets being endowed with the weak* topology. This means that there exists a unique $\flat$-decision rule, see Definition 3.10.

Remark 2.16
The simpler case where $g_{+} = 0$ or $g_{-} = 0$ can be studied in the same way (mutatis mutandis). For example, if $g_{+} = 0$ and $g_{-} \neq 0$, we consider $P^{\flat} \doteq \inf P_{\mathrm{NL}}\left(\mathcal{X}_{-}^{*}\right)$ and $M^{\flat}(0)$, since the set $\mathrm{M}^{\flat}$ does not make sense in this case. Similarly, if $g_{+} \neq 0$ and $g_{-} = 0$, we consider $\mathrm{P}^{\flat} \doteq \sup P^{\flat}\left(\mathcal{X}_{+}^{*}\right)$ and $\mathrm{M}^{\flat}$ with the new function $P^{\flat} \doteq P_{\mathrm{L}}(\cdot, 0) - g_{+}^{*}$. Then Proposition 2.15 holds true, mutatis mutandis. We refrain from going into further detail, as this special situation is already discussed in Section 3.

2.7.3 Validity of Bogoliubov linearizations

Under Condition 2.9 nonlinear equilibrium measures (Definition 2.8) form the weak*-compact set (20), that is,

$$E_{P} \doteq \{\mu \in \mathcal{P}(T) : P(\mu) = \sup P(\mathcal{P}(T))\} \ .$$

However, if we look at the same variational problem with the pressure P defined by (31) but under the *more general* Conditions TF1–TF3, it is not a priori clear whether the above set is nonempty, since P is not necessarily weak*-upper semicontinuous. So, in this case, it is instead convenient to use

$$E_{P} \doteq \left\{\mu \in \mathcal{P}(T) : \exists (\mu_{n})_{n\in\mathbb{N}} \subseteq \mathcal{P}(T) \text{ with } \lim_{n\rightarrow\infty} \mu_{n} = \mu \text{ and } \lim_{n\rightarrow\infty} P(\mu_{n}) = \sup P(\mathcal{P}(T))\right\} \ , \tag{45}$$

as the basic definition for the (nonempty) set of nonlinear equilibrium measures, similar to (23). Above, the limit $\mu_{n} \rightarrow \mu$ of T-invariant probability measures is taken with respect to the weak* topology. Surprisingly, this set is exactly what one would like from a conceptual point of view, the set

$$M_{P} \doteq \{\mu \in \mathcal{P}(T) : P(\mu) = \sup P(\mathcal{P}(T))\}$$

of the usual maximizers of P. This is proven in the next theorem.

In fact, the sets $\mathrm{M}^{\flat} \subseteq \mathcal{X}_{+}^{*}$ and $M^{\flat}(y_{+}) \subseteq \mathcal{X}_{-}^{*}$, $y_{+} \in \mathcal{X}_{+}^{*}$, of optimizers can be used to obtain all elements of E_{P} via the weak*-compact convex sets

$$E_{\mathrm{L}}(y_{+}, y_{-}) \doteq \{\nu \in \mathcal{P}(T) : \mathrm{P}(y_{+}, y_{-}, \nu) = \sup \mathrm{P}(y_{+}, y_{-}, \mathcal{P}(T)) \doteq P_{\mathrm{L}}(y_{+}, y_{-})\} \ , \qquad y_{\pm} \in \mathcal{X}_{\pm}^{*} \ ,$$

of approximating linear equilibrium measures. Note that the weak* compactness and convexity of $E_{\mathrm{L}}(y_{+}, y_{-})$ for all $y_{\pm} \in \mathcal{X}_{\pm}^{*}$ is a direct consequence of the affine property and weak*-upper semicontinuity of the approximating pressure $\mu \mapsto \mathrm{P}(y_{+}, y_{-}, \mu)$ defined by (33), thanks to Proposition 2.38 and the weak*-upper semicontinuity of the entropy (cf. Definition 2.1).

As is usual, here the so-called *subdifferential* of a convex function g at $z \in \mathrm{dom}(g)$ is denoted $\partial g(z)$, see Section 3.1 and in particular Equation (128) for more details. Recall that if g is Gateaux-differentiable at z then $\partial g(z)$ must be a singleton.

We are now in a position to show how the μ-linearized variational problem $\mathrm{P}^{\flat}$ (34)–(36) can be used to completely solve the original problem, including not only the calculation of the nonlinear pressure but also the nonlinear equilibrium measures:

Theorem 2.17 (Nonlinear pressure and nonlinear equilibrium measures)
Assume Conditions TF1–TF3. Let $\tau_{\pm}(\mu) \doteq \theta_{\pm}(\mu_S)$ for any T-invariant probability measure $\mu \in \mathcal{P}(T)$.
(i) *Nonlinear pressure:*

$$\sup P\left(\mathcal{P}\left(T\right)\right) = \mathrm{P}^{\flat} \doteq \sup_{y_{+}\in\mathcal{X}_{+}^{*}} \inf_{y_{-}\in\mathcal{X}_{-}^{*}} \left\{P_{\mathrm{L}}\left(y_{+},y_{-}\right) + g_{-}^{*}\left(y_{-}\right) - g_{+}^{*}\left(y_{+}\right)\right\} \ .$$

(ii) *Self-consistency conditions: For any $x_{+} \in \mathrm{M}^{\flat}$ and $x_{-} \in M^{\flat}\left(x_{+}\right)$, the set*

$$E_{\mathrm{L}}^{\mathrm{sc}}\left(x_{+},x_{-}\right) \doteq \left\{\mu \in E_{\mathrm{L}}\left(x_{+},x_{-}\right) : x_{-} \in \partial g_{-}\left(\tau_{-}\left(\mu\right)\right)\right\} \equiv E_{\mathrm{L}}^{\mathrm{sc}}\left(x_{+}\right)$$

of self-consistent equilibrium measures of Bogoliubov linearizations is nonempty, convex and weak*-compact, and does not depend upon the choice of $x_{-} \in M^{\flat}\left(x_{+}\right)$. Furthermore, for all $x_{+} \in \mathrm{M}^{\flat}$,

$$E_{\mathrm{L}}^{\mathrm{sc}}\left(x_{+}\right) \subseteq \left\{\mu \in \mathcal{P}\left(T\right) : x_{+} \in \partial g_{+}\left(\tau_{+}\left(\mu\right)\right)\right\} \ .$$

(iii) *Nonlinear equilibrium measures: $E_P = M_P$ is weak*-compact and corresponds to the union of all above sets of self-consistent equilibrium measures, that is,*

$$E_P = M_P = \bigcup_{x_{+}\in\mathrm{M}^{\flat}} E_{\mathrm{L}}^{\mathrm{sc}}\left(x_{+}\right) \ .$$

If the function $g_{+} : \mathcal{X}_{+} \to \mathbb{R}$ is additionally Gateaux-differentiable, then the above union is disjoint.

Proof. As explained in Proposition 2.15, Conditions TF1–TF3 imply Conditions B1–B3 of Section 3.3 for $K = \mathcal{P}(T)$, $f = h$ and $\tau_{\pm} : \mathcal{P}(T) \to \mathcal{X}_{\pm}$ defined by $\tau_{\pm}(\mu) \doteq \theta_{\pm}(\mu_S)$ for any $\mu \in \mathcal{P}(T)$. The assertions are therefore direct consequences of Theorem 3.8 (i)–(iii). ∎

Compared to [25, Theorem C], in particular Equations (25)–(27), the computation of the sets $\mathrm{M}^{\flat}$, $M^{\flat}\left(x_{+}\right)$, $x_{+} \in \mathrm{M}^{\flat}$, and $E_{\mathrm{L}}\left(x_{+},x_{-}\right)$ of optimizers is much simpler, while our setup allows for parameter sets in **infinite-dimensional** spaces $\mathcal{X}_{\pm}$. The cost is that we have to separate the convex and concave parts of the nonlinearity. This cannot be avoided and is not just a technical artefact. On the other hand, the C^1-character of the nonlinearity is not needed[9], although it can be very useful for solving the corresponding variational problems.

Note that the (nonempty) compact convex set $E_{\mathrm{L}}^{\mathrm{sc}}\left(x_{+}\right)$ is naturally called the set of self-consistent equilibrium measures of Bogoliubov linearizations. Indeed, take for instance $\mathcal{X}_{\pm} = \mathbb{R}$, $g_{-}(x) = g_{+}(x) = x^2/2$ and $\tau_{\pm}(\mu) \doteq \theta_{\pm}(\mu_S) = \mu(\varphi)$ for fixed Hölder potentials $\varphi_{\pm} \in C^{\alpha}(\Sigma)$ ($\alpha \in (0,1]$). In this case we canonically identify the dual spaces $\mathcal{X}_{\pm}^{*}$ with $\mathbb{R}$. Then, for all T-invariant probability measures $\mu \in \mathcal{P}\left(T\right)$,

$$\partial g_{\pm}\left(\tau\left(\mu\right)\right) = \left\{\mu\left(\varphi_{\pm}\right)\right\} \ .$$

Thus, $\mu \in E_{\mathrm{L}}^{\mathrm{sc}}\left(x_{+}\right)$ means that μ is the unique linear equilibrium probability measure of the Hölder potential $(x_{+}\varphi_{+} - x_{-}\varphi_{-})$ and satisfies $\mu(\varphi_{\pm}) = x_{\pm}$. By Theorem 2.17, it implies in particular that, for any fixed $\mu \in E_P = M_P$, $\mu(\varphi_{\pm}) = x_{\pm}$ for all $x_{+} \in \mathrm{M}^{\flat}$ and $x_{-} \in M^{\flat}\left(x_{+}\right)$, i.e., $\mathrm{M}^{\flat} = \{x_{+}\}$ and $M^{\flat}(x_{+}) = \{x_{-}\}$. Compare this with Corollary 2.18 below. More generally, similar results are true if the functions $g_{\pm}$ are Gateaux-differentiable and their gradient mapping is injective.

This last example refers to the Hölder case, which is very important in the linear thermodynamic formalism. Indeed, when a potential φ is of Hölder class, i.e. $\varphi \in C^{\alpha}(\Sigma)$ for a certain $\alpha \in (0,1]$, recall that the linear equilibrium measure μ_{φ} is unique and ergodic (see [50, 1]). In the nonlinear framework, this leads us to define the Hölder-type for pairs (θ_{-},θ_{+}) of linear functions, in Definition 2.14. In this particular case the nonlinear equilibrium measures have strong properties. Among other things, they are always ergodic.

[9] In contrast to the corresponding results of [25].

Corollary 2.18 (Nonlinear equilibrium measures – Hölder case)
Assume Conditions TF1–TF3 and that the pair (θ_-, θ_+) of linear functions is Hölder-type.
(i) *Ergodic nonlinear equilibrium measures: For any $x_+ \in \mathrm{M}^\flat$ and $x_- \in M^\flat(x_+)$,*

$$E_{\mathrm{L}}^{\mathrm{sc}}(x_+) = E_{\mathrm{L}}^{\mathrm{sc}}(x_+, x_-) = E_{\mathrm{L}}(x_+, x_-) = \left\{\mu_{x_+}\right\} \subseteq \mathcal{P}_{\mathrm{erg}}(T)$$

and $E_P = M_P$ is a nonempty weak-compact set of ergodic probability measures:*

$$E_P = M_P = \left\{\mu_{x_+} : x_+ \in \mathrm{M}^\flat\right\} \subseteq \mathcal{P}_{\mathrm{erg}}(T) \ .$$

(ii) *If $g_+ : \mathcal{X}_+^* \to \mathbb{R}$ is Gateaux-differentiable then the mapping $x_+ \mapsto \mu_{x_+}$ from the weak*-compact $\mathrm{M}^\flat \subseteq \mathcal{X}_+^*$ to $E_P \subseteq \mathcal{P}_{\mathrm{erg}}(T)$ is a homeomorphism with respect to the weak* topology of $\mathrm{M}^\flat$ and E_P.*

Proof. Use Theorem 2.17, Proposition 3.14 (iii) and the fact that the nonlinear equilibrium measure for Hölder potentials is unique and ergodic (see [50, 1]). ■

The conditions of Corollary 2.18 are satisfied in most cases of interest. For instance, consider Example 2.6 with $\varphi_1, \ldots, \varphi_N \in C^\alpha(\Sigma)$ for some $\alpha \in (0,1]$ and a lower semicontinuous convex function $F = F_+$ on $\mathbb{R}^N$ for which the Legendre-Fenchel transform is strictly convex (e.g., $N = 1$ and $F(x) = x^2/2$). We refrain to discuss in more detail such explicit examples here, since this is already done in [20].

Recall that $G_P \doteq \overline{\mathrm{co}}(E_P)$ is the set of so-called generalized nonlinear equilibrium measures, which can be equivalently defined in a more fundamental way by a property stated below in Theorem 2.21 (ii). See Definition 2.11 and discussions thereafter. Under the conditions of Corollary 2.18, by the Milman theorem [55, Proposition 1.5] and observing that ergodic measures are extreme in the convex set $\mathcal{P}(T)$ of T-invariant measures, the set G_P of generalized nonlinear equilibrium measures is a *face*[10] of $\mathcal{P}(T)$, the extreme boundary of which is precisely the set E_P of (simple) nonlinear equilibrium measures. In particular, as $\mathcal{P}(T)$ is a Choquet simplex (Proposition 2.35), G_P is in this case a *Bauer simplex*, i.e., a Choquet simplex whose extreme boundary is compact.

Remark 2.19
Theorem 2.17 and Corollary 2.18 also hold when $g_+ = 0$ or $g_- = 0$ with obvious modifications. In fact, this case is even simpler. For more details, see Remark 2.16 as well as the discussions following Theorem 3.8.

Remark 2.20
Theorem 2.17 and Corollary 2.18 are, of course, only relevant when the functions $g_\pm$ are not affine, because otherwise we are dealing with the linear thermodynamic formalism.

2.7.4 Validity of the thermodynamic game

The conservative values $\mathrm{P}^\flat$ (34) and $\mathrm{P}^\sharp$ (38) of the thermodynamic game are related to two variational problems on T-invariant measures respectively setup from the functionals $\mathfrak{F}^\sharp : \mathcal{P}(T) \to \mathbb{R}$ and $\mathfrak{F}^\flat : \mathcal{P}(T) \to \mathbb{R}$ defined, for any T-invariant measure $\mu \in \mathcal{P}(T)$, by

$$\mathfrak{F}^\sharp(\mu) \doteq \Delta^{g_+ \circ \theta_+}(\mu) - g_- \circ \theta_-(\mu_S) + h(\mu) \ , \qquad (46)$$

$$\mathfrak{F}^\flat(\mu) \doteq \Delta^{g_+ \circ \theta_+}(\mu) - \Delta^{g_- \circ \theta_-}(\mu) + h(\mu) \ , \qquad (47)$$

where $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$ are *continuous* convex functions and $\theta_\pm : \mathcal{M}(S) \to \mathcal{X}_\pm$ are two linear transformations that are σ-normal. Here,

$$\Delta^F(\mu) \doteq \int_{\mathcal{P}_{\mathrm{erg}}(T)} F(\nu_S)\, \xi_\mu(\mathrm{d}\nu) \ , \qquad \mu \in \mathcal{P}(T) \ , \qquad (48)$$

[10] A face F of a convex set K is a subset of K with the property that, if $\rho = \lambda_1\rho_1 + \cdots + \lambda_n\rho_n \in F$ with $\rho_1, \ldots, \rho_n \in K$, $\lambda_1, \ldots, \lambda_n \in (0,1)$ and $\lambda_1 + \cdots + \lambda_n = 1$, then $\rho_1, \ldots, \rho_n \in F$.

for any convex and σ-normal function $F : \mathcal{M}(S) \to \mathbb{R}$, where ξ_μ is the (unique) Choquet measure associated with μ. To understand why the functional Δ^F in both (46) and (47) is important and natural, we recommend Proposition 2.25 below. See meanwhile Definition 2.26.

Recall that the entropy h is affine weak*-upper semicontinuous (Definition 2.1 and Proposition 2.38). We can thus apply [18, Lemma 10.17] to h in order to get – via Equations (47)–(48) – the ergodic decomposition of the functional $\mathfrak{F}^\flat$:

$$\mathfrak{F}^\flat (\mu) = \int_{\mathcal{P}_{\mathrm{erg}}(T)} P(\nu) \, \xi_\mu (\mathrm{d}\nu) \ , \qquad \mu \in \mathcal{P}(T) \ , \tag{49}$$

where ξ_μ is the unique Choquet measure associated to μ (see Proposition 2.35). It is an elementary result which, whilst certainly not surprising given that the functional $\mathfrak{F}^\flat$ is affine, is highlighted here because it is a very useful observation, enabling the functional $\mathfrak{F}^\flat$ to be calculated in practice via the nonlinear pressure functional P.

In this subsection, we only consider *continuous* functions g_+ and g_- in order to highlight the singular nature of Δ-functionals, which even in this simpler context are discontinuous on a dense set (cf. Theorem 2.27). The aim is not to obtain the most general framework possible, but rather to explain the associated variational problems in an accessible manner, given that they are new in the thermodynamic formalism.

The functional $\mathfrak{F}^\flat$ is exactly the one obtained from the limit (22). It is therefore quite intuitive. The functional $\mathfrak{F}^\sharp$ turns out to be very useful in the proof of Proposition 2.28, in which **no** Bogoliubov linearization is considered. See Equation (83). In fact, these seemingly incidental functionals are very natural and the theorem below shows their importance for the study of nonlinear equilibrium measures.

By Definition 2.1, Theorem 2.27 and Proposition 2.38, for any (weak*-)lower semicontinuous and convex function g_- and σ-normal linear transformations $\theta_\pm : \mathcal{M}(S) \to \mathcal{X}_\pm$, the functional $\mathfrak{F}^\sharp$ is concave and weak*-upper semicontinuous. Notice also that Definition 2.1 implies the weak*-upper semicontinuity of the entropy. Thus, the set

$$E_{\mathfrak{F}^\sharp} = \left\{ \mu \in \mathcal{P}(T) : \mathfrak{F}^\sharp(\mu) = \sup \mathfrak{F}^\sharp (\mathcal{P}(T)) \right\}$$

of maximizers of $\mathfrak{F}^\sharp$ is convex and weak*-compact. By contrast, the functional $\mathfrak{F}^\flat$ is affine but (generally) **not** weak*-upper semicontinuous, which therefore forces us to take sequences of approximating maximizers (similar to (23) and (45)):

$$E_{\mathfrak{F}^\flat} \doteq \left\{ \mu \in \mathcal{P}(T) : \exists (\mu_n)_{n\in\mathbb{N}} \subseteq \mathcal{P}(T) \text{ with } \lim_{n\to\infty} \mu_n = \mu \text{ and } \lim_{n\to\infty} \mathfrak{F}^\flat(\mu_n) = \sup \mathfrak{F}^\flat(\mathcal{P}(T)) \right\} , \tag{50}$$

where the first limit $\mu_n \to \mu$ refers to the weak* topology. See again Theorem 2.27 for the continuity properties of Δ-functionals.

Now we are in a position to associate the conservative values $\mathrm{P}^\flat$ (34) and $\mathrm{P}^\sharp$ (38) of the thermodynamic game with two different variational problems on T-invariant measures:

Theorem 2.21 (Thermodynamic game and nonlinear equilibrium measures)
Assume Conditions TF1–TF3 with continuous functions $g_\pm$. Let $\tau_\pm(\mu) \doteq \theta_\pm(\mu_S)$ for any T-invariant probability measure $\mu \in \mathcal{P}(T)$.
(i) *Nonlinear pressures:*

$$\begin{aligned} \sup \mathfrak{F}^\flat (\mathcal{P}(T)) &= \mathrm{P}^\flat \doteq \sup_{y_+ \in \mathcal{X}_+^*} \inf_{y_- \in \mathcal{X}_-^*} \left\{ P_{\mathrm{L}}(y_+, y_-) + g_-^*(y_-) - g_+^*(y_+) \right\} \ , \\ \sup \mathfrak{F}^\sharp (\mathcal{P}(T)) &= \mathrm{P}^\sharp \doteq \inf_{y_- \in \mathcal{X}_-^*} \sup_{y_+ \in \mathcal{X}_+^*} \left\{ P_{\mathrm{L}}(y_+, y_-) + g_-^*(y_-) - g_+^*(y_+) \right\} \ . \end{aligned}$$

(ii) *Generalized nonlinear equilibrium measures:*

$$\begin{aligned} E_{\mathfrak{F}^\flat} &= \overline{\mathrm{co}}(E_P) \doteq G_P \ , \\ E_{\mathfrak{F}^\sharp} &= \left\{ \mu \in \overline{\mathrm{co}}\left(E_{\mathrm{NL}}\left(x_-\right)\right) : x_- \in \mathrm{M}^\sharp \cap \partial g_-\left(\tau_-\left(\mu\right)\right) \right\} \ , \end{aligned}$$

where $E_P = M_P$ is given by Theorem 2.17 (ii)–(iii), while

$$E_{\mathrm{NL}}\left(x_-\right) \doteq \bigcup_{x_+ \in M^\sharp(x_-)} E_{\mathrm{L}}\left(x_+, x_-\right) \ .$$

Proof. Assertions (i)–(ii) for the case ($\flat$) are direct consequences of Theorems 2.17 and 2.29. See also Definition 2.11. Assertions (i)–(ii) for the case ($\sharp$) are proven as follows:

Step 1: The functional $f : \mathcal{P}(T) \to \mathbb{R}$, as defined by

$$f = \Delta^{g_+ \circ \theta_+} + h \ , \tag{51}$$

is affine (in particular concave) and weak* upper semicontinuous, thanks to Definition 2.1, Proposition 2.38 and Theorem 2.27 (i). Thus, using Proposition 3.2 for $K = \mathcal{P}\left(T\right)$, $\mathcal{X} = \mathcal{X}_-$, $g = g_-$, $\tau(\mu) = \theta_-(\mu_S)$ and f defined by (51), we obtain that

$$\sup \mathfrak{F}^\sharp\left(\mathcal{P}\left(T\right)\right) = \inf_{y_- \in \mathcal{X}_-^*} \left\{ \sup \mathcal{G}_{y_-}\left(\mathcal{P}\left(T\right)\right) + g_-^*\left(y_-\right) \right\} \tag{52}$$

with

$$\mathcal{G}_{y_-}\left(\mu\right) \doteq \Delta^{g_+ \circ \theta_+}\left(\mu\right) + h\left(\mu\right) - y_- \circ \theta_-(\mu_S) \ , \qquad \mu \in \mathcal{P}\left(T\right) \ .$$

In addition, Corollary 3.4 also yields

$$E_{\mathfrak{F}^\sharp} = \left\{ \mu \in E_{\mathcal{G}_{x_-}} : x_- \in \mathrm{M}^\sharp \cap \partial g_-\left(\tau_-\left(\mu\right)\right) \right\} \ , \tag{53}$$

where

$$E_{\mathcal{G}_{y_-}} \doteq \left\{ \mu \in \mathcal{P}\left(T\right) : \mathcal{G}_{y_-}\left(\mu\right) = \sup \mathcal{G}_{y_-}\left(\mathcal{P}\left(T\right)\right) \right\}$$

for any continuous linear functional $y_- \in \mathcal{X}_-^*$.

Step 2: Note that, with obvious adaptations in the proofs, Theorem 2.29 and Corollary 2.30 also hold if we add any arbitrary affine and weak* continuous function to the entropy h. It follows that

$$\sup \mathcal{G}_{y_-}\left(\mathcal{P}\left(T\right)\right) = \sup P_{y_-}\left(\mathcal{P}\left(T\right)\right) \tag{54}$$

with

$$P_{y_-}\left(\mu\right) \doteq g_+ \circ \theta_+\left(\mu_S\right) + h\left(\mu\right) - y_- \circ \theta_-(\mu_S) \ , \qquad \mu \in \mathcal{P}\left(T\right) \ ,$$

while (the extended version of) Corollary 2.30 yields

$$E_{\mathcal{G}_{y_-}} = \overline{\mathrm{co}}(E_{P_{y_-}}) \ , \qquad y_- \in \mathcal{X}_-^* \ , \tag{55}$$

with

$$E_{P_{y_-}} = \left\{ \mu \in \mathcal{P}\left(T\right) : P_{y_-}\left(\mu\right) = \sup P_{y_-}\left(\mathcal{P}\left(T\right)\right) \right\}$$

for any continuous linear functional $y_- \in \mathcal{X}_-^*$.

Step 3: Now, we apply the simplified version of Theorem 3.8 (see the discussions following Theorem 3.8) with $K = \mathcal{P}\left(T\right)$, $g_- = 0$, and

$$\tau_+(\mu) = \theta_+(\mu_S) \ , \quad f\left(\mu\right) = h\left(\mu\right) - y_- \circ \theta_-(\mu_S) \ , \qquad \mu \in \mathcal{P}\left(T\right) \ ,$$

observing that g_+ is by assumption a continuous and convex function, that the functional f defined as above is affine and weak*-upper semicontinuous (cf. Definition 2.1 and Proposition 2.38) and that τ_+ is a continuous linear transformation, θ_+ being a σ-normal linear transformation. Doing so, for any continuous linear functional $y_- \in \mathcal{X}_-^*$, we get

$$\sup P_{y_-}(\mathcal{P}(T)) = \sup_{y_+ \in \mathcal{X}_+^*} \left\{ P_{\mathrm{L}}(y_+, y_-) - g_+^*(y_+) \right\} \tag{56}$$

and

$$E_{P_{y_-}} = \bigcup_{y_+ \in M^\sharp(y_-)} E_{\mathrm{L}}(y_+, y_-) \ . \tag{57}$$

It remains now to combine the three steps.

Step 4: We combine Equations (52), (54) and (56) to get Assertion (i) for the case ($\sharp$). Assertion (ii) for the case ($\sharp$) is a consequence of Equations (53), (55) and (57). ■

By Theorem 2.21, the thermodynamic game has a direct interpretation in terms of equilibrium measures. This is a very useful information that can be employed, for example, to show that generally $\mathrm{P}^\sharp > \mathrm{P}^\flat$, that is, the sup and inf in $\mathrm{P}^\flat$ (34) or $\mathrm{P}^\sharp$ (38) do not commute, as is done in the quantum case in [18, Discussions after Theorem 2.36]. The interpretation of the conservative values $\mathrm{P}^\flat$ (34) and $\mathrm{P}^\sharp$ (38) as variational problems on T-invariant measures given by Theorem 2.21 also turns out to be crucial for the Kac, or van der Waals, limit, very well-known in statistical mechanics. This is exploited in [21] in the context of quantum lattice models, but the arguments used, in the light of the results presented here, can clearly be adapted to produce a version of [21] for the nonlinear thermodynamic formalism. We therefore believe that the above results on the thermodynamic game can be useful for further developments of the nonlinear version of the thermodynamic formalism. Notice additionally that Theorem 2.21 follows relatively easily from our general approach, which is presented in Section 3.

We conclude this subsection by examining whether generalized equilibrium states can be true maximizers of the affine pressure $\mathfrak{F}^\flat$. So, consider the set

$$M_{\mathfrak{F}^\flat} \doteq \left\{ \mu \in \mathcal{P}(T) : \mathfrak{F}^\flat(\mu) = \sup \mathfrak{F}^\flat(\mathcal{P}(T)) \right\} \subseteq E_{\mathfrak{F}^\flat}$$

of strict maximizers of $\mathfrak{F}^\flat$. Compare this definition with (50). By the affine property of $\mathfrak{F}^\flat$, $M_{\mathfrak{F}^\flat}$ is convex. One might wonder whether the equality $M_{\mathfrak{F}^\flat} = E_{\mathfrak{F}^\flat}$ holds true even if $\mathfrak{F}^\flat$ is not necessarily upper semicontinuous. In fact, in general, one only has the strict inclusion $M_{\mathfrak{F}^\flat} \subsetneq E_{\mathfrak{F}^\flat}$. The reason is that, because of the affine property of $\mathfrak{F}^\flat$, $M_{\mathfrak{F}^\flat}$ must be a (convex) face of $\mathcal{P}(T)$, but this is generally not the case for $E_{\mathfrak{F}^\flat}$. We proved this in the context of quantum lattice systems, but our arguments can easily be adapted to the nonlinear thermodynamic formalism, see [18, Lemma 9.8]. However, if the linear transformations $\theta_\pm$ are Hölder-type then the equality $M_{\mathfrak{F}^\flat} = E_{\mathfrak{F}^\flat}$ does indeed hold. This is proven in the next corollary, deduced from Theorem 2.21 (i)–(ii) and Equation (49).

Corollary 2.22 (Generalized equilibrium measures and maximizers of the affine pressure) *Assume Conditions TF1–TF3 with continuous functions $g_\pm$. Then,*

$$\mathrm{co}\left(E_P \cap \mathcal{P}_{\mathrm{erg}}(T)\right) \subseteq M_{\mathfrak{F}^\flat} \subseteq \overline{\mathrm{co}}\left(E_P\right) = E_{\mathfrak{F}^\flat} \tag{58}$$

and if $\theta_\pm$ are additionally Hölder-type then $E_{\mathfrak{F}^\flat} = M_{\mathfrak{F}^\flat}$.

Proof. Clearly, $M_{\mathfrak{F}^\flat} \subseteq E_{\mathfrak{F}^\flat}$ and thus, by Theorem 2.21 (ii) for the case ($\flat$), $M_{\mathfrak{F}^\flat} \subseteq \overline{\mathrm{co}}(E_P)$. Note that $\mathfrak{F}^\flat = P$ on ergodic measures, by the definition of Δ-functionals. Hence, by combining Theorem 2.17 (i) with Theorem 2.21 (i) for the case ($\flat$), we deduce that $E_P \cap \mathcal{P}_{\mathrm{erg}}(T) \subseteq M_{\mathfrak{F}^\flat}$ and (58) follows from the affine property of $\mathfrak{F}^\flat$. If $\theta_\pm$ are additionally Hölder-type then, by Corollary 2.18 (i), E_P

is a weak*-compact set of ergodic measures only. Hence, the Choquet measure in $\overline{\mathrm{co}}(E_P)$ of any $\mu \in \overline{\mathrm{co}}(E_P)$, as given by Corollary 2.30 ($E_{\mathfrak{F}^\flat} = \overline{\mathrm{co}}(E_P)$), coincides with its unique Choquet measure in $\mathcal{P}(T)$, as given by Proposition 2.35. As explained in Corollary 2.30 (cf. the Milman theorem [55, Proposition 1.5]), this measure must be supported in E_P. Therefore, we can combine Theorem 2.21 (ii) with Equation (49) to get $\overline{\mathrm{co}}(E_P) = M_{\mathfrak{F}^\flat}$ when $\theta_\pm$ are additionally Hölder-type. ■

As a consequence, in the Hölder case, generalized nonlinear equilibrium measures are precisely all (strict) maximizers of the affine pressure functional $\mathfrak{F}^\flat$, which is a very satisfactory characterization of such T-invariant probability measures.

2.8 Technical Results: Δ-functionals and nonlinear pressures

We address the maximization of the *nonlinear* pressure functional $P : \mathcal{P}(T) \to \mathbb{R}$ defined on T-invariant measures by (19). Although P is generally neither convex nor concave, this variational problem can be studied via another real-valued functional on T-invariant measures, which is *affine*, i.e., both concave and convex. This fact may seem curious, or even strange, at a first sight. To properly understand this, we must first precisely define the affine functional, which requires a few preliminary definitions and observations.

Recall that $\mathbb{E}_n[\varphi]$ stands for Birkhoff sums (21) on $C(\Sigma)$. It defines a linear contraction $\mathbb{E}_n$ mapping S to itself. Now, we use the following observation:

Lemma 2.23 (Properties of ergodic measures)
For any ergodic measure $\mu \in \mathcal{P}_{\mathrm{erg}}(T)$ and for μ-a.s. $\sigma \in \Sigma$,

$$\lim_{n\to\infty} \mathbb{E}_n[\varphi](\sigma) = \mu(\varphi) \doteq \mu_S(\varphi) , \qquad \varphi \in S .$$

Proof. Recall that any countable intersection of measurable sets of full measure has full measure. Therefore, we infer from Proposition 2.33 that, for any countable subset $\tilde{S} \subseteq S$ and for μ-a.s. $\sigma \in \Sigma$,

$$\lim_{n\to\infty} \mathbb{E}_n[\varphi] = \mu(\varphi) , \qquad \varphi \in \tilde{S} .$$

Since $S \subseteq C(\Sigma)$ is separable (with respect to the supremum norm) and for any $n \in \mathbb{N}$, $\sigma \in \Sigma$ and $\varphi_1, \varphi_2 \in C(\Sigma)$,

$$\max\{|\mathbb{E}_n[\varphi_1](\sigma) - \mathbb{E}_n[\varphi_2](\sigma)|, |\mu(\varphi_1) - \mu(\varphi_2)|\} \le \|\varphi_1 - \varphi_2\|_\infty ,$$

the lemma directly follows. ■

Remark 2.24
Ergodic probability measures are mutually singular. See [66, Theorem 6.10 (iv)]. Note also that we restrict Lemma 2.23 to functions $\varphi \in S$ in view of our applications. By linearity, the same assertion holds true for all $\varphi \in C(\Sigma)$.

Recall that a function $F : \mathcal{M}(S) \to \mathbb{R}$ is σ-normal when, for any bounded sequence $(f_n)_{n\in\mathbb{N}} \subseteq \mathcal{M}(S)$ converging point-wise to $f \in \mathcal{M}(S)$, one has

$$\lim_{n\to\infty} F(f_n) = F(f) .$$

We also remind that $\mathcal{P}_{\mathrm{erg}}(T)$ denotes the set of ergodic (or extreme) measures of the space $\mathcal{P}(T)$ of T-invariant probability measures. Note also that, given $\sigma \in \Sigma$ and $n \in \mathbb{N}$, Equation (21) defines a continuous and bounded function $\mathbb{E}_{n,\sigma} : S \to \mathbb{R}$ by

$$\mathbb{E}_{n,\sigma}[\varphi] \doteq \mathbb{E}_n[\varphi](\sigma) , \qquad \varphi \in S . \tag{59}$$

In particular, $\mathbb{E}_{n,\sigma} \in \mathcal{M}(S)$. Clearly, for any $\sigma \in \Sigma$ and $n \in \mathbb{N}$, $\mathbb{E}_{n,\sigma'}$ converges point-wise to $\mathbb{E}_{n,\sigma}$, as $\sigma' \to \sigma$. Therefore, for any σ-normal function $F : \mathcal{M}(S) \to \mathbb{R}$, the mapping $\sigma \mapsto F(\mathbb{E}_{n,\sigma})$ is a continuous function from Σ to $\mathbb{R}$. By a slight abuse of notation, we denote this function by "$F \circ \mathbb{E}_n$".

We can now show an important property that leads to a convenient definition of the affine pressure functional on T-invariant measures.

Proposition 2.25
Let $F : \mathcal{M}(S) \to \mathbb{R}$ be any convex and σ-normal function. Then, for any T-invariant measure $\mu \in \mathcal{P}(T)$,

$$\lim_{n\to\infty} \mu\left(F \circ \mathbb{E}_n\right) = \inf_{n\in\mathbb{N}} \mu\left(F \circ \mathbb{E}_n\right) = \int_{\mathcal{P}_{\mathrm{erg}}(T)} F\left(\nu_S\right) \xi_\mu\left(\mathrm{d}\nu\right) \ ,$$

where ξ_μ is the Choquet measure associated with μ (Proposition 2.35).

Proof. Fix $\mu \in \mathcal{P}(T)$ and denote by ξ_μ the associated Choquet measure. Since F is σ-normal, remark that the mapping $\nu \mapsto F\left(\nu_S\right)$ from $\mathcal{P}(T)$ to $\mathbb{R}$ is weak*-continuous and, thus, integrable with respect to the Choquet measure ξ_μ. We divide the proof into two steps:

Step 1: Since F is a convex continuous function on the Banach space $\mathcal{M}(S)$, it has continuous tangents at any $f \in \mathcal{M}(S)$. That is, for any fixed $f \in \mathcal{M}(S)$, there is a continuous linear functional $\vartheta \in \mathcal{M}(S)^*$ such that, for all $u \in \mathcal{M}(S)$,

$$F\left(u\right) - F\left(f\right) \geq \vartheta\left(u\right) - \vartheta\left(f\right) \ . \tag{60}$$

Such a linear functional ϑ is also σ-normal. Indeed, take any bounded sequence $(f_n)_{n\in\mathbb{N}} \subseteq \mathcal{M}(S)$ converging pointwise to f. By Inequality (60) for $u = f_n$ and the σ-normality of F, we deduce that

$$\vartheta\left(f\right) \geq \limsup_{n\to\infty} \vartheta\left(f_n\right) \ . \tag{61}$$

Similarly, from Inequality (60) for $u = 2f - f_n$,

$$\liminf_{n\to\infty} \vartheta\left(f_n\right) \geq \vartheta\left(f\right) \ . \tag{62}$$

Altogether, Inequalities (61) and (62) yield

$$\lim_{n\to\infty} \vartheta\left(f_n\right) = \vartheta\left(f\right) \tag{63}$$

for any bounded sequence $(f_n)_{n\in\mathbb{N}} \subseteq \mathcal{M}(S)$ converging pointwise to f. By linearity, it follows that ϑ is σ-normal. Since (Ω, d) is a compact metric space, Σ is also a compact metric space (see, e.g., (4)). In particular, $C(\Sigma)$ is a separable Banach space, which implies that the weak* topology of $\mathcal{P}$, the space of all (i.e., not necessarily T-invariant) probability measures on Σ, is metrizable. From this fact and the σ-normality of ϑ it follows that the mapping $\nu \mapsto \vartheta(\nu_S)$ from $\mathcal{P}$ to $\mathbb{R}$ is weak*-continuous. Note that, for any $\sigma \in \Sigma$,

$$\delta_\sigma(\vartheta \circ \mathbb{E}_n) = \vartheta \circ \mathbb{E}_n(\sigma) = \vartheta((\delta_\sigma \circ \mathbb{E}_n)_S) \ ,$$

where $\vartheta \circ \mathbb{E}_n$ stands for the continuous function $\sigma \mapsto \vartheta\left(\mathbb{E}_{n,\sigma}\right)$ and δ_σ is the δ-probability measure at $\sigma \in \Sigma$, that is, $\delta_\sigma(\{\sigma\}) = 1$. Since Σ is a separable (being compact) metric space, any probability measure on Σ is the weak* limit of a sequence of convex combinations of δ-probability measures. See, for instance, [53, Chapter II, Theorem 7.1]. Using this fact, along with the weak* continuity of the mapping $\nu \mapsto \vartheta(\nu_S)$ and the linearity of ϑ, one arrives at

$$\nu(\vartheta \circ \mathbb{E}_n) = \vartheta\left((\nu \circ \mathbb{E}_n)_S\right)$$

for all $\nu \in \mathcal{P}$ and $n \in \mathbb{N}$. In particular, if ν is T-invariant then

$$\nu(\vartheta \circ \mathbb{E}_n) = \vartheta(\nu_S) \ .$$

We then conclude from Inequality (60) applied to $f = \nu_S$ and $u = \mathbb{E}_{n,\sigma}$ that

$$\nu(F \circ \mathbb{E}_n) - F(\nu_S) \geq 0 \ , \tag{64}$$

for any T-invariant probability measure $\nu \in \mathcal{P}(T)$, which is nothing but a version of Jensen's inequality. For a more general statement, see [18, Lemma 10.33]. Hence, by convexity of F and (the above version of) Jensen's inequality,

$$\mu(F \circ \mathbb{E}_n) = \int_{\mathcal{P}_{\mathrm{erg}}(T)} \nu(F \circ \mathbb{E}_n)\, \xi_\mu(\mathrm{d}\nu) \geq \int_{\mathcal{P}_{\mathrm{erg}}(T)} F(\nu_S)\xi_\mu(\mathrm{d}\nu) \ . \tag{65}$$

Step 2: Keeping in mind (65), it suffices now to show that

$$\lim_{n\to\infty} \int_{\mathcal{P}_{\mathrm{erg}}(T)} \nu(F \circ \mathbb{E}_n)\, \xi_\mu(\mathrm{d}\nu) = \int_{\mathcal{P}_{\mathrm{erg}}(T)} F(\nu_S)\xi_\mu(\mathrm{d}\nu) \tag{66}$$

to get the assertion. Using the σ-normality of F and Lemma 2.23, for any ergodic measure $\nu \in \mathcal{P}_{\mathrm{erg}}(T)$ and for ν-a.s. $\sigma \in \Sigma$,

$$\lim_{n\to\infty} F(\mathbb{E}_{n,\sigma}) = F(\nu_S) \ . \tag{67}$$

The continuous functions

$$\sigma \mapsto F \circ \mathbb{E}_n(\sigma) \doteq F(\mathbb{E}_{n,\sigma})$$

are uniformly bounded for all $n \in \mathbb{N}$, i.e.,

$$\sup_{n\in\mathbb{N}} \|F \circ \mathbb{E}_n\|_\infty = \sup_{\sigma\in\Sigma}\sup_{n\in\mathbb{N}} |F(\mathbb{E}_{n,\sigma})| \leq \sup\{|F(f)| : f \in \mathcal{M}(S) \text{ with } \|f\|_\infty < 1\} < \infty \ ,$$

because

$$\sup_{n\in\mathbb{N}} \|\mathbb{E}_{n,\sigma}\|_\infty = \sup\{|\mathbb{E}_{n,\sigma}[\varphi]| : n \in \mathbb{N},\ \varphi \in S \text{ with } \|\varphi\|_\infty = 1\} \leq 1$$

and any σ-normal function F satisfies (17), as explained after Condition 2.9. So, we can use Lebesgue's dominated convergence to infer from (67) that, for any ergodic measure $\nu \in \mathcal{P}_{\mathrm{erg}}(T)$,

$$\lim_{n\to\infty} \nu(F \circ \mathbb{E}_n) = F(\nu_S) \ .$$

Applying again Lebesgue's dominated convergence, we obtain the limit (66). ■

Proposition 2.25 motivates the following definition of a real-valued functional on T-invariant probability measures:

Definition 2.26 (Δ-Functionals)
Given a convex and σ-normal function $F : \mathcal{M}(S) \to \mathbb{R}$, Δ^F denotes the mapping $\mathcal{P}(T) \to \mathbb{R}$ defined by

$$\Delta^F(\mu) \doteq \int_{\mathcal{P}_{\mathrm{erg}}(T)} F(\nu_S)\, \xi_\mu(\mathrm{d}\nu) \ , \qquad \mu \in \mathcal{P}(T) \ ,$$

where ξ_μ is the Choquet measure associated with μ (Proposition 2.35).

As far as we know, the use of such a functional is entirely new to thermodynamic formalism. Additionally, mutatis mutandis, it is a broad extension of the space-averaging functional introduced in [18, Section 1.3] for quantum lattice systems. Important properties of this functional are gathered in the next theorem. But before stating them, we recall that the Γ-regularization of a functional $f : \mathcal{P}(T) \to (-\infty, \infty]$ is the convex functional $\Gamma(f) : \mathcal{P}(T) \to (-\infty, \infty]$ defined by the supremum over all affine and continuous minorants of f, i.e., for all $\mu \in \mathcal{P}(T)$,

$$\Gamma(f)(\mu) \doteq \sup\left\{m(\mu) : m \in \mathrm{A} \text{ and } m|_{\mathcal{P}(T)} \leq f\right\} , \tag{68}$$

where A denotes the set of all affine and weak*-continuous functionals on the dual space $C(\Sigma)^*$. It is the largest weak*-lower semicontinuous convex function below f, see [17, Corollary 3.2]. Cf. Section 2.9.3, in particular Equation (120). It is also convenient to define the concave counterpart of Γ-regularizations: For any functional $f : \mathcal{P}(T) \to [-\infty, \infty)$ and all $\mu \in \mathcal{P}(T)$,

$$\Gamma_-(f)(\mu) \doteq -\Gamma(-f)(\mu) = \inf\left\{-m(\mu) : m \in \mathrm{A} \text{ and } -m|_{\mathcal{P}(T)} \geq f\right\} , \tag{69}$$

which is the smallest weak*-upper semicontinuous concave function above f. We call $\Gamma_-(f)$ the upper Γ-regularization of f.

Note that taking the Γ-regularization is equivalent to performing the Legendre-Fenchel transform twice, i.e., it is equal to the biconjugate of a function. See Sections 3.1 and 2.9.3. However, the concept of Γ-regularization is more convenient, in the sense that we do not need to talk about dual pairs $(\mathcal{X}, \mathcal{X}^*)$ to introduce it. Furthermore, its definition via affine weak*-continuous functionals is also technically very useful. In particular, as shown in [17, Lemma 3.4], it leads to an extension of the Bauer maximum principle.

Theorem 2.27 (Properties of the functional Δ^F)
Let $F : \mathcal{M}(S) \to \mathbb{R}$ be a convex and σ-normal function.
(i) *Δ^F is a weak*-upper semicontinuous affine functional.*
(ii) *Δ^F is weak*-continuous iff the mapping $\mu \mapsto F(\mu_S)$ is affine on $\mathcal{P}(T)$. In this case, there is $f \in C(\Sigma)$ such that*

$$\Delta^F(\mu) = \mu(f) , \qquad \mu \in \mathcal{P}(T) . \tag{70}$$

(iii) *If the mapping $\mu \mapsto F(\mu_S)$ is strictly convex then Δ^F is weak*-discontinuous on a weak*-dense subset of $\mathcal{P}(T)$.*
(iv) *Δ^F is continuous on the G_δ weak*-dense subset $\mathcal{P}_{\mathrm{erg}}(T)$ of ergodic measures. In particular, the set of all T-invariant measures on which Δ^F is weak*-discontinuous is weak*-meager.*
(v) *Its Γ-regularization $\Gamma(\Delta^F)$ is the weak*-continuous convex mapping $\mu \mapsto F(\mu_S)$ on $\mathcal{P}(T)$.*

Proof. We take a function F as specified in the theorem and divide the proof into several steps. Note that this proof strongly relies on the one of [18, Theorems 1.18-1.19], which regards quantum lattice systems.

Step 1: By Definition 2.26, Δ^F is clearly affine. By Proposition 2.25,

$$\Delta^F(\mu) = \lim_{n\to\infty} \mu(F \circ \mathbb{E}_n) = \inf_{n\in\mathbb{N}} \mu(F \circ \mathbb{E}_n) , \qquad \mu \in \mathcal{P}(T) . \tag{71}$$

As a consequence, Δ^F is an infimum over weak*-continuous functionals $\mu \mapsto \mu(F \circ \mathbb{E}_n)$ and is therefore weak*-upper semicontinuous. (i) is thus proven. From Jensen's inequality (see (64)), for any $n \in \mathbb{N}$,

$$\mu(F \circ \mathbb{E}_n) \geq F(\mu_S) .$$

This last inequality combined with Definition 2.26 and (71) leads to

$$\Delta^F(\mu) \geq F(\mu_S) , \qquad \mu \in \mathcal{P}(T) . \tag{72}$$

Note that $\Delta^F(\mu) = F(\mu_S)$ for all $\mu \in \mathcal{P}_{\text{erg}}(T)$, by Definition 2.26.

Step 2: Assume that the mapping $\mu \mapsto F(\mu_S)$ from $\mathcal{P}(T)$ to $\mathbb{C}$ is affine. We infer from Definition 2.26 and the σ-normality of F that

$$\Delta^F(\mu) = F(\mu_S) \ , \qquad \mu \in \mathcal{P}(T) \ .$$

By the σ-normality of F, the mapping $\mu \mapsto F(\mu_S)$ is affine and weak*-continuous on the whole (topological) dual space $C(\Sigma)^*$. In particular, Δ^F is continuous. Similar to [19, Proposition 3.9], there is $f \in C(\Sigma)$ such that

$$F(\mu_S) = \mu(f) \ , \qquad \mu \in \mathcal{P}(T) \ ,$$

which in turn implies (70).

Step 3: Assume now that $\mu \mapsto F(\mu_S)$ is not affine. Then, there are at least two T-invariant probability measures $\nu, \zeta \in \mathcal{P}(T)$ and $\lambda \in (0,1)$ such that, for $\mu = \lambda\nu + (1-\lambda)\zeta$,

$$F(\mu_S) < \lambda F(\nu_S) + (1-\lambda) F(\zeta_S) \ , \tag{73}$$

the function F being convex. By Proposition 2.36, the set $\mathcal{P}_{\text{erg}}(T)$ of ergodic probability measures is weak*-dense in $\mathcal{P}(T)$. So, there is a sequence $(\nu^{(n)})_{n\in\mathbb{N}} \subseteq \mathcal{P}_{\text{erg}}(T)$ of ergodic measures converging with respect to the weak* topology to the above probability measure μ. Then, by Definition 2.26 as well as Equations (72) and (73),

$$\lim_{n\to\infty} \Delta^F(\nu^{(n)}) = \lim_{n\to\infty} F(\nu_S^{(n)}) = F(\mu_S) < \lambda F(\nu_S) + (1-\lambda) F(\zeta_S) \leq \Delta^F(\mu) \ , \tag{74}$$

because F is σ-normal, $\lambda \in (0,1)$ and Δ^F is affine. Consequently, Δ^F is discontinuous on the probability measure $\mu = \lambda\nu + (1-\lambda)\zeta$. Steps 2 and 3 yield Assertion (ii).

Step 4: Assume that the mapping $\mu \mapsto F(\mu_S)$ is strictly convex. In particular, it is not the constant mapping. Therefore, for any $\nu \in \mathcal{P}(T)$, there is at least one T-invariant probability measure $\zeta^{(\nu)} \in \mathcal{P}(T)$ such that $F(\nu_S) \neq F(\zeta_S^{(\nu)})$. For all $\nu \in \mathcal{P}(T)$, we define the subset

$$\mathrm{I}(\nu) \doteq \left\{ \lambda\nu + (1-\lambda)\zeta^{(\nu)} \text{ for any } \lambda \in (0,1) \right\} \subseteq \mathcal{P}(T) \ .$$

Finally, let us consider the subset

$$\mathrm{D} \doteq \bigcup_{\nu\in\mathcal{P}(T)} \mathrm{I}(\nu) \subseteq \mathcal{P}(T) \backslash \mathcal{P}_{\text{erg}}(T) \ .$$

By continuity of the mapping $\lambda \mapsto \lambda\nu$ for $\lambda \in \mathbb{C}$ and $\nu \in \mathcal{P}(T)$, the set D is weak*-dense in $\mathcal{P}(T)$. Any $\mu \in \mathrm{D}$ is clearly of the form

$$\mu = \lambda\nu + (1-\lambda)\zeta \tag{75}$$

for some $\lambda \in (0,1)$ and $\nu, \zeta \in \mathcal{P}(T)$ with $F(\nu) \neq F(\zeta)$. By Proposition 2.36, for any $\mu \in \mathrm{D}$, there is a sequence $(\nu^{(n)})_{n\in\mathbb{N}} \subseteq \mathcal{P}_{\text{erg}}(T)$ of ergodic measures converging with respect to the weak* topology to μ. Then, by Definition 2.26 and Equation (72), for any $\mu \in \mathrm{D}$ of the form (75), Equation (74) holds true because F is σ-normal and strictly convex, $\lambda \in (0,1)$ and Δ^F is affine. As a consequence, the mapping Δ^F is weak*-discontinuous at any $\mu \in \mathrm{D}$ and hence on a weak*-dense subset of $\mathcal{P}(T)$. In other words, (iii) is proven.

Step 5: We show now that Δ^F is weak*-continuous for any ergodic probability measure, which yields (iv), thanks to Proposition 2.36. Take $\nu \in \mathcal{P}_{\text{erg}}(T)$ and consider any sequence $(\nu^{(n)})_{n\in\mathbb{N}} \subseteq \mathcal{P}(T)$

of probability measures converging with respect to the weak* topology to ν. By Definition 2.26, Assertion (i) and Equation (72),

$$F(\nu_S) = \Delta^F(\nu) \geq \limsup_{n\to\infty} \Delta^F(\nu^{(n)}) \geq \liminf_{n\to\infty} \Delta^F(\nu^{(n)}) \geq \lim_{n\to\infty} F(\nu_S^{(n)}) = F(\nu_S)$$

because F is σ-normal. In other words, the functional Δ^F is weak*-continuous on $\mathcal{P}_{\text{erg}}(T)$. Notice that the weak* topology of $\mathcal{P}(T)$ is metrizable. (So, there is no need to use nets here to check continuity on $\mathcal{P}(T)$.)

Step 6: Recall that the Γ-regularization $\Gamma(f)$ of functionals f, as defined by Equation (68), is the largest weak*-lower semicontinuous convex function below f. See, e.g., [17, Corollary 3.2]. By Inequality (72), it follows that

$$F(\mu_S) \leq \Gamma\left(\Delta^F\right)(\mu) \leq \Delta^F(\mu) \ , \qquad \mu \in \mathcal{P}(T) \ , \tag{76}$$

because the mapping $\mu \mapsto F(\mu_S)$ is weak*-continuous and convex on $\mathcal{P}(T)$, thanks the convexity and σ-normality of F. In addition, for any ergodic probability measure $\nu \in \mathcal{P}_{\text{erg}}(T)$,

$$F(\nu_S) = \Gamma\left(\Delta^F\right)(\nu) = \Delta^F(\nu) \ . \tag{77}$$

Therefore, using the weak*-density of $\mathcal{P}_{\text{erg}}(T)$ in $\mathcal{P}(T)$ (Proposition 2.36), the weak*-continuity of $\mu \mapsto F(\mu_S)$ and the weak*-lower semicontinuity of $\Gamma(\Delta^F)$, we deduce from (76)–(77) that

$$\Gamma\left(\Delta^F\right)(\mu) = F(\mu_S) \ , \qquad \mu \in \mathcal{P}(T) \ ,$$

that is, (v) follows. This concludes the proof of the theorem. ■

Thanks to Theorem 2.27, under Condition 2.9 the *nonlinear* pressure functional

$$P(\mu) \doteq F(\mu_S) + h(\mu) = F_+(\mu_S) + F_-(\mu_S) + h(\mu) \ , \qquad \mu \in \mathcal{P}(T) \ , \tag{78}$$

on T-invariant measures has a direct relation to the functional $\mathfrak{F}^\flat : \mathcal{P}(T) \to \mathbb{R}$ defined by (47), that is,

$$\mathfrak{F}^\flat \doteq \Delta^{F_+} - \Delta^{-F_-} + h \ . \tag{79}$$

In contrast to the nonlinear pressure functional $P : \mathcal{P}(T) \to \mathbb{R}$ (78) which is not necessarily affine but weak*-upper semicontinuous (for σ-normal $F_\pm$) (see Proposition 2.38), the functional $\mathfrak{F}^\flat$ is affine but not necessarily weak*-upper semicontinuous, because of Theorem 2.27 and Proposition 2.38. However, the variational problems $\sup P(\mathcal{P}(T))$ and $\sup \mathfrak{F}^\flat(\mathcal{P}(T))$ can be related, as well as can their corresponding maximizers. This is a consequence of the next observation. Recall that $\mathcal{P}_{\text{erg}}(T)$ stands for the (dense in $\mathcal{P}(T)$) set of ergodic probability measures on Σ.

Proposition 2.28
Let $F_\pm : \mathcal{M}(S) \to \mathbb{R}$ be two σ-normal functions and assume that F_+ and F_- are respectively convex and concave.
(i) *We have*

$$\sup_{\mu\in\mathcal{P}(T)} \mathfrak{F}^\flat(\mu) = \sup_{\nu\in\mathcal{P}_{\text{erg}}(T)} \mathfrak{F}^\flat(\nu) = \sup_{\nu\in\mathcal{P}_{\text{erg}}(T)} P(\nu) = \sup_{\mu\in\mathcal{P}(T)} P(\mu) < \infty \ . \tag{80}$$

(ii) $\Gamma_-(\mathfrak{F}^\flat) = \Gamma_-(P)$ *on the whole space* $\mathcal{P}(T)$ *and* $\Gamma_-(P) = P$ *on* $\mathcal{P}_{\text{erg}}(T)$. *Furthermore,*

$$\sup_{\mu\in\mathcal{P}(T)} \mathfrak{F}^\flat(\mu) = \sup_{\mu\in\mathcal{P}(T)} \Gamma_-\left(\mathfrak{F}^\flat\right)(\mu) = \sup_{\mu\in\mathcal{P}(T)} \Gamma_-(P)(\mu) = \sup_{\mu\in\mathcal{P}(T)} P(\mu) < \infty \ . \tag{81}$$

Proof. Note that all suprema in (i) are finite, because any σ-normal function F such as $F_\pm$ satisfies (17), as explained after Condition 2.9, and the entropy functional is uniformly bounded from above. Assertion (i) is proven in the same way as [18, Lemma 2.9], while the proof of (ii) uses the same arguments as the first part of the proof of [18, Theorem 2.21]. We reproduce these proofs below, adapting them to the present situation.

Step 1: The Bauer maximum principle [2, Theorem I.5.3] says that an upper semicontinuous convex real-valued function f over a compact convex subset K, such as the weak*-compact convex set $\mathcal{P}(T)$, attains its maximum at an extreme point of K. However, by Theorem 2.27, $\mathfrak{F}^\flat$ is the sum of a convex weak*-lower semicontinuous functional and a convex weak*-upper semicontinuous functional, both on $\mathcal{P}(T)$. We thus need the extension of the Bauer maximum principle given by [17, Lemma 3.4], which says that the supremum of the values of a sum $f \doteq f_- + f_+$ of a lower and upper semicontinuous affine functionals f_- and f_+ on a compact convex subset K can be restricted to the set $\mathcal{E}(K)$ of extreme points of K (see Section 3.7.2), i.e.,

$$\sup\, f\left(K\right) = \sup\, f\left(\mathcal{E}(K)\right)\ .$$

Applying this result to $f = \mathfrak{F}^\flat$ and $K = \mathcal{P}(T)$ while using additionally that $\mathfrak{F}^\flat = P$ on $\mathcal{P}_{\mathrm{erg}}(T)$ (see Definition 2.26), we thus obtain the equality

$$\sup_{\mu\in\mathcal{P}(T)} \mathfrak{F}^\flat(\mu) = \sup_{\nu\in\mathcal{P}_{\mathrm{erg}}(T)} \mathfrak{F}^\flat\left(\nu\right) = \sup_{\nu\in\mathcal{P}_{\mathrm{erg}}(T)} P\left(\nu\right)\ . \tag{82}$$

Since P is weak*-upper semicontinuous, it has a maximizer μ_0 over $\mathcal{P}(T)$. Thus, by Corollary 2.39, there is a sequence $(\nu_n)_{n\in\mathbb{N}} \subseteq \mathcal{P}_{\mathrm{erg}}(T)$ of ergodic probability measures converging in the weak* topology to μ_0 with the property that $P(\nu_n)$ converges to $P(\mu_0)$ as $n \to \infty$. In other words, maximizing P over $\mathcal{P}(T)$ results the same as maximizing P over $\mathcal{P}_{\mathrm{erg}}(T)$. Equation (80) then follows. Observe that Equation (81) is a direct consequence of Equation (80) and Theorem 2.40. We now prove the equality $\Gamma_-(\mathfrak{F}^\flat) = \Gamma_-(P)$ on the whole space $\mathcal{P}\left(T\right)$ and $\Gamma_-(P) = P$ on the dense subset $\mathcal{P}_{\mathrm{erg}}(T)$ of ergodic measures.

Step 2: We start with a partial upper Γ-regularization of $\mathfrak{F}^\flat$ by defining the new functional $\mathfrak{F}^\sharp : \mathcal{P}(T) \to \mathbb{R}$ as follows:

$$\mathfrak{F}^\sharp\left(\mu\right) \doteq \Delta^{F_+}\left(\mu\right) + F_-\left(\mu_S\right) + h\left(\mu\right)\ , \qquad \mu \in \mathcal{P}\left(T\right)\ . \tag{83}$$

By Theorem 2.27, $\mathfrak{F}^\sharp$ is concave, weak*-upper semicontinuous and bounded from below by $\mathfrak{F}^\flat$. Recall that the Γ-regularization of a functional f is the largest weak*-lower semicontinuous convex function below f, see [17, Corollary 3.2] or Section 2.9.3. I.e., (the upper Γ-regularization) $\Gamma_-\left(f\right)$ is the smallest weak*-upper semicontinuous concave function above f. Using this property and Equation (69),

$$\mathfrak{F}^\flat \leq \Gamma_-\left(\mathfrak{F}^\flat\right) \leq \mathfrak{F}^\sharp\ .$$

Hence, for any ergodic probability measure $\nu \in \mathcal{P}_{\mathrm{erg}}(T)$,

$$P\left(\nu\right) = \mathfrak{F}^\sharp\left(\nu\right) = \Gamma_-\left(\mathfrak{F}^\flat\right)\left(\nu\right) = \mathfrak{F}^\flat\left(\nu\right)\ . \tag{84}$$

Step 3: We show now that $\Gamma_-\left(\mathfrak{F}^\flat\right)$ is an upper bound for $\Gamma_-(P)$. By Corollary 2.39 and the σ-normality of $F_\pm$, any T-invariant probability measure $\mu \in \mathcal{P}(T)$ is the limit of some sequence $(\nu_n)_{n\in\mathbb{N}} \subseteq \mathcal{P}_{\mathrm{erg}}(T)$ of ergodic probability measures converging in the weak* topology to μ and such that $P(\nu_n)$ converges to $P(\mu)$. By (84), $\Gamma_-(\mathfrak{F}^\flat)(\nu_n)$ also converges to $P(\mu)$. Since $\Gamma_-(\mathfrak{F}^\flat)$ is, by construction, weak*-upper semicontinuous on $\mathcal{P}(T)$, we thus conclude that, for any T-invariant probability measure $\mu \in \mathcal{P}(T)$,

$$P\left(\mu\right) = \lim_{n\to\infty} \Gamma_-\left(\mathfrak{F}^\flat\right)\left(\nu_n\right) \leq \Gamma_-\left(\mathfrak{F}^\flat\right)\left(\mu\right)\ , \tag{85}$$

which, combined with the fact that $\Gamma_-(P)$ is the smallest weak*-upper semicontinuous concave function above P (see [17, Corollary 3.2]), yields

$$\Gamma_-(P) \leq \Gamma_-\left(\mathfrak{F}^\flat\right) . \tag{86}$$

We show next the converse inequality.

Step 4: We use now the concavity of $\Gamma_-(P)$, the (unique) Choquet measure ξ_μ associated with an arbitrary T-invariant probability measure $\mu \in \mathcal{P}(T)$, along with Jensen's inequality applied to ξ_μ and $\Gamma_-(P)$ (see [18, Lemma 10.33] for its proof in this context) in order to bound $\Gamma_-(P)$ from below by

$$\Gamma_-(P)(\mu) \geq \int_{\mathcal{P}_{\mathrm{erg}}(T)} \Gamma_-(P)(\nu)\,\xi_\mu(\mathrm{d}\nu) \geq \int_{\mathcal{P}_{\mathrm{erg}}(T)} P(\nu)\,\xi_\mu(\mathrm{d}\nu) \tag{87}$$

for all $\mu \in \mathcal{P}(T)$. Now, by applying [18, Lemma 10.17] to the entropy h (which is an affine weak*-upper semicontinuous functional) and using Definition 2.26, recall that one gets an ergodic decomposition of $\mathfrak{F}^\flat$ in the following sense:

$$\int_{\mathcal{P}_{\mathrm{erg}}(T)} P(\nu)\,\xi_\mu(\mathrm{d}\nu) = \mathfrak{F}^\flat(\mu) , \qquad \mu \in \mathcal{P}(T) ,$$

see Equation (49). By Inequality (87), It follows that

$$\Gamma_-(P) \geq \Gamma_-\left(\mathfrak{F}^\flat\right) , \tag{88}$$

thanks to Equation (69) and the fact that $\Gamma_-\left(\mathfrak{F}^\flat\right)$ is the smallest weak*-upper semicontinuous concave function above $\mathfrak{F}^\flat$ (see [17, Corollary 3.2]). Consequently, $\Gamma_-(P) = \Gamma_-(\mathfrak{F}^\flat)$, which, combined with (84) also yield $\Gamma_-(P) = \Gamma_-(\mathfrak{F}^\flat) = P$ on $\mathcal{P}_{\mathrm{erg}}(T)$. ■

Proposition 2.28 combined with Theorem 2.40 not only implies that the corresponding variational problems for all functionals $\mathfrak{F}^\flat$, $\Gamma_-(\mathfrak{F}^\flat)$, P and $\Gamma_-(P)$ give the same numerical value but also that the corresponding (possibly generalized) maximizers are directly related to each other. To explain this, similar to Equations (23) and (50), given a bounded functional $f : \mathcal{P}(T) \to \mathbb{R}$, we define the (nonempty) set

$$E_f \doteq \left\{ \mu \in \mathcal{P}(T) : \exists (\mu_n)_{n\in\mathbb{N}} \subseteq \mathcal{P}(T) \text{ with } \lim_{n\to\infty} \mu_n = \mu \text{ and } \lim_{n\to\infty} f(\mu_n) = \sup f(\mathcal{P}(T)) \right\} , \tag{89}$$

of all weak*-limits of approximating maximizers, i.e., generalized maximizers. E_f differs a priori from the set

$$M_f \doteq \{\mu \in \mathcal{P}(T) : f(\mu) = \sup f(\mathcal{P}(T))\}$$

of (strict) maximizers of f. One has a priori only the inclusion $M_f \subseteq E_f$. If the functional f is weak*-upper semicontinuous then $E_f = M_f$, that is, E_f is nothing but the (weak*-compact) set of (strict) maximizers of f over $\mathcal{P}(T)$.

Theorem 2.29 (Sets of generalized maximizers)
Let $F_\pm : \mathcal{M}(S) \to \mathbb{R}$ be two σ-normal functions, with F_+ and F_- being respectively convex and concave. Then,

$$E_{\mathfrak{F}^\flat} = E_{\Gamma_-(\mathfrak{F}^\flat)} = M_{\Gamma_-(\mathfrak{F}^\flat)} = M_{\Gamma_-(P)} = E_{\Gamma_-(P)} = \overline{\mathrm{co}}(E_P) ,$$

that is, all sets of (possibly generalized) maximizers are nothing but the weak-closed convex hull of the weak*-compact set $E_P = M_P$ of strict maximizers of the (weak*-upper semicontinuous) functional P.*

Proof. As upper Γ-regularizations are always upper semicontinuous, note that $E_{\Gamma_-(\mathfrak{F}^\flat)} = M_{\Gamma_-(\mathfrak{F}^\flat)}$ and $E_{\Gamma_-(P)} = M_{\Gamma_-(P)}$. The equality $E_{\Gamma_-(\mathfrak{F}^\flat)} = E_{\Gamma_-(P)}$ is an obvious consequence of Proposition 2.28 (ii) and we infer from Theorem 2.40 and (69) that $E_{\Gamma_-(\mathfrak{F}^\flat)} = \overline{\mathrm{co}}(E_{\mathfrak{F}^\flat})$. On the other hand, $E_{\mathfrak{F}^\flat}$ is a convex set, by the affine property of $\mathfrak{F}^\flat$ (cf. Theorem 2.27 and Proposition 2.38). By [18, Lemma 10.36], it is also weak*-compact because $\mathcal{P}(T)$ is weak*-compact and the weak*-topology is metrizable on $\mathcal{P}(T)$. As a consequence, $E_{\Gamma_-(\mathfrak{F}^\flat)} = \overline{\mathrm{co}}(E_{\mathfrak{F}^\flat}) = E_{\mathfrak{F}^\flat}$. Applying again Theorem 2.40 and (69), we obtain meanwhile $E_{\Gamma_-(P)} = \overline{\mathrm{co}}(E_P)$, which leads to $E_{\mathfrak{F}^\flat} = E_{\Gamma_-(\mathfrak{F}^\flat)} = E_{\Gamma_-(P)} = \overline{\mathrm{co}}(E_P)$. Finally, E_P is the weak*-compact set of strict maximizers of the functional P as a consequence of the weak*-upper semicontinuity of this functional together with the weak*-compactness of $\mathcal{P}(T)$. ∎

Corollary 2.30 (Choquet decomposition)
Let $F_\pm : \mathcal{M}(S) \to \mathbb{R}$ be two σ-normal functions with F_+ and F_- being respectively convex and concave. Then, equilibrium measures $\nu \in E_{\mathfrak{F}^\flat}$ that are extreme in $E_{\mathfrak{F}^\flat}$ belong to E_P and for any $\mu \in E_{\mathfrak{F}^\flat}$, there is a probability measure m_μ on E_P such that

$$m_\mu(E_P) = 1 \quad \textit{and} \quad \mu = \int_{E_P} \nu m_\mu (\mathrm{d}\nu) \ . \tag{90}$$

Proof. By the Milman theorem [55, Proposition 1.5] and Theorem 2.29, equilibrium measures that are extreme in $E_{\mathfrak{F}^\flat}$ must belong to E_P. Then, using this property and the Choquet theorem (Theorem 3.28), we derive (90), because the set $E_{\mathfrak{F}^\flat}$ is convex, weak*-compact and metrizable, by Theorem 2.29 and the metrizability of the weak* topology in $\mathcal{P}(T)$. ∎

2.9 Appendix

This appendix contains a selection of useful results that are difficult to find within the mathematical framework we use. Indeed, in most previous works the alphabet Ω is assumed to be a finite set, whereas we only assume it to be a compact metric space, which, of course, may be infinite. Therefore, in Sections 2.9.1–2.9.2 we re-examine known facts within this broader context.

One thing that is not very well-known is the properties of the entropy functional for infinite alphabets. Our definition (Definition 2.1) is based on a variational problem involving the (transfer) Ruelle operator. This definition is simple and elegant, but it poses certain difficulties in proving the convexity of this functional. To this end, we use recent results [1] to write the entropy functional for compact alphabets as a thermodynamic limit of finite-volume entropies (Theorem 2.37). Based on this observation, we are able to prove the affine property[11] of the entropy functional of Definition 2.1. See Proposition 2.38. Another consequence of this result is the so-called ergodic abundance of the entropy functional, as given by Corollary 2.39, which is also a fundamental property for our proofs, along with the weak* density of ergodic measures (Proposition 2.36).

2.9.1 Ergodic measures

Let $(X, \mathcal{A})$ be any measurable space. I.e., X is any nonempty set and $\mathcal{A}$ some σ-algebra on X. In ergodic theory, the existence of limits of the form

$$\tilde{\varphi}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} \varphi \circ f^m (x) \ , \qquad x \in X \ , \tag{91}$$

is studied for any measurable function $f : X \to X$ and appropriate functions $\varphi : X \to \mathbb{R}$. Such a question led to two important (types of) ergodic theorems, which were proven for the first time by

[11] I.e. convex and concave. The concavity of the entropy functional is a direct consequence of its definition.

von Neumann and Birkhoff, respectively. See [64, Theorems 3.1.6 and 3.2.3]. The most interesting one here is Birkhoff's ergodic theorem [64, Theorem 3.2.3], which ensures the existence of the limit (91) almost surely for integrable functions φ:

Theorem 2.31 (Birkhoff)
Let $(X, \mathcal{A})$ be any measurable space, $f : X \to X$ a measurable mapping and μ a probability measure that is invariant[12] *with respect to f. For any μ-integrable function $\varphi : X \to \mathbb{R}$, the limit (91) exists μ-almost surely and defines another μ-integrable function $\tilde{\varphi} : X \to \mathbb{R}$ (μ-almost everywhere) satisfying*

$$\int_X \tilde{\varphi}(x)\, \mu(\mathrm{d}x) = \int_X \varphi(x) \mu(\mathrm{d}x) \ .$$

The Birkhoff ergodic theorem stated above is proven, for instance, in [64, Section 3.2.2], see [64, Theorem 3.2.3].

The μ-integrable function $\tilde{\varphi} : X \to \mathbb{R}$, uniquely defined μ-almost everywhere, is called a *time average* of the (μ-integrable) function $\varphi : X \to \mathbb{R}$. For any given measurable set $A \subseteq X$ (i.e., $A \in \mathcal{A}$), τ_A denotes the time average of the corresponding characteristic function χ_A. It is called the *mean sojourn time* associated with $A \in \mathcal{A}$. These objects allow us to explain the notion of ergodicity:

Definition 2.32 (Ergodicity)
Under the conditions of Theorem 2.31, the pair (f, μ) is ergodic[13] *if, for all $A \in \mathcal{A}$, $\tau_A = \mu(A)$, i.e., the constant function $\mu(A)$ on X is a mean sojourn time for A.*

By definition, ergodicity of a probability measure μ thus refers to the fact that, for all Borel sets $A \in \mathcal{A}$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} \chi_A \circ f^m = \int_X \chi_A(x)\, \mu(\mathrm{d}x) \ , \qquad \mu\text{-a.s.}$$

See Theorem 2.31. This property is equivalent to the corresponding one with the characteristic functions χ_A, $A \in \mathcal{A}$, replaced by general μ-integrable functions. See, for instance, [64, Proposition 4.1.3]. The ergodicity is also equivalent to a clustering property [64, Proposition 4.1.4]. All this leads to the following proposition:

Proposition 2.33 (Properties equivalent to ergodicity)
Under the assumptions of Theorem 2.31, the following conditions are equivalent:
(i) *(f, μ) is ergodic.*
(ii) *For any μ-integrable function $\varphi : X \to \mathbb{R}$,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{n} \varphi \circ f^m = \int_X \varphi(x) \mu(\mathrm{d}x) \ , \qquad \mu\text{-a.s.}$$

(iii) *For all $p, q \in (1, \infty)$ with $1/p + 1/q = 1$ and all functions $\varphi \in \mathcal{L}^p(\mu)$, $\psi \in \mathcal{L}^q(\mu)$,*

$$\lim_{k \to \infty} \frac{1}{k} \sum_{n=0}^{k-1} \int_X \psi(x)\, \varphi \circ f^n(x)\, \mu(\mathrm{d}x) = \left(\int_X \psi(x)\, \mu(\mathrm{d}x) \right) \left(\int_X \varphi(x)\, \mu(\mathrm{d}x) \right) \ .$$

[12] I.e., $f_*(\mu) = \mu$, where $f_*(\mu)(A) = \mu(f(A))$ for all Borel sets $A \in \mathcal{A}$.

[13] I.e., the transformation f is ergodic with respect to the measure μ or the measure μ is ergodic with respect to the transformation f.

Proof. By [64, Proposition 4.1.3, cf. (i) and (iii)], (i) holds true iff (ii) is satisfied, while [64, Proposition 4.1.4, cf. (i) and (iii)] proves that (i) is equivalent to the clustering properties (iii). ■

It is well known that the ergodicity property precisely characterizes the extreme invariant measures. In fact, under the conditions of Theorem 2.31, let

$$\mathcal{P}(f) \doteq \{\mu \in \mathcal{P} : f_*(\mu) = \mu\} \tag{92}$$

be the set of all probability measures on $\mathcal{A}$ that are invariant with respect to f, where $\mathcal{P}$ is the set of all (i.e., not necessarily invariant) probability measures on $\mathcal{A}$. It is a convex set and $\mathcal{P}_{\mathrm{erg}}(f)$ denotes the subset of its extreme points, which turns out to be nothing but the set of ergodic measures with respect to the measurable transformation f:

Proposition 2.34 (Ergodicity and extremality)
Under the conditions of Theorem 2.31, the invariant measure $\mu \in \mathcal{P}(f)$ is ergodic iff $\mu \in \mathcal{P}_{\mathrm{erg}}(f)$.

See, for instance, [64, Proposition 4.3.2] or [66, Theorem 6.10] for more details, including the proof.

Of course, all these results can be applied to the case where X is the compact topological space $\Sigma \doteq \Omega^{\mathbb{N}}$, the set of infinite strings on the alphabet Ω endowed with the product topology, and f is the transformation $T : \Sigma \to \Sigma$ defined by (7). In this particular case, we can identify $\mathcal{P}(T)$, defined above by (92), with the weak* compact and convex space of all positive normalized linear functionals on $C(\Sigma)$, thanks to the Riesz-Markov theorem [23, Theorem 4.68]. In particular, $\mathcal{P}(T) \subseteq C(\Sigma)^*$. By the Krein-Milman theorem [57, Theorems 3.4 (b) and 3.21], the subset $\mathcal{P}_{\mathrm{erg}}(T)$ of its extreme points is nonempty and

$$\mathcal{P}(T) = \overline{\mathrm{co}}\, \mathcal{P}_{\mathrm{erg}}(T) \ .$$

Therefore, by Proposition 2.34, there are ergodic measures in $\mathcal{P}(T)$ and any T-invariant probability measure $\mu \in \mathcal{P}(T)$ is the weak* limit of convex combinations of ergodic probability measures.

Since Σ is a metrizable compact space, $C(\Sigma)$ is also separable with respect to the topology of uniform convergence and, hence, the weak* topology of $\mathcal{P}(T) \subseteq C(\Sigma)^*$ is also metrizable. In particular, by Lemma 3.27, $\mathcal{P}(T)$ is a Borel set for the weak* topology and using the Choquet theorem (Theorem 3.28), one checks that $\mathcal{P}(T)$ is a (Choquet) simplex:

Proposition 2.35 ($\mathcal{P}(T)$ as a Choquet simplex)
For any T-invariant measure $\mu \in \mathcal{P}(T)$ there is a unique (Choquet) probability measure ξ_μ on the Borel σ-algebra associated with the weak* topology of $\mathcal{P}(T)$, that is supported in $\mathcal{P}_{\mathrm{erg}}(T)$, the barycenter of which is μ. In particular,

$$\xi_\mu(\mathcal{P}(T)\backslash\mathcal{P}_{\mathrm{erg}}(T)) = 0 \ . \tag{93}$$

Proof. The existence of a (Choquet) probability measure ξ_μ with the asserted property is a direct consequence of Theorem 3.28, since $\mathcal{P}(T)$ is a metrizable, weak*-compact and convex set, as already explained. It only remains to show its uniqueness. So fix $\mu \in \mathcal{P}(T)$ and let ξ_μ be any Choquet measure on $\mathcal{P}(T)$ whose barycenter is μ, as given by Theorem 3.28. Using Proposition 2.33, Equation (93), Lebesgue's dominated convergence theorem as well as the notation $\mathbb{E}_n[\cdot]$ for the Birkhoff sum (21), we deduce that, for any $\varphi_1, \varphi_2 \in C(\Sigma)$,

$$\begin{aligned}
\int_{\mathcal{P}(T)} \hat{\mu}(\varphi_2)\, \hat{\mu}(\varphi_1)\, \xi_\mu(\mathrm{d}\hat{\mu}) &= \int_{\mathcal{P}_{\mathrm{erg}}(T)} \lim_{n\to\infty} \hat{\mu}(\varphi_2 \mathbb{E}_n[\varphi_1])\, \xi_\mu(\mathrm{d}\hat{\mu}) \\
&= \lim_{n\to\infty} \int_{\mathcal{P}_{\mathrm{erg}}(T)} \hat{\mu}(\varphi_2 \mathbb{E}_n[\varphi_1])\, \xi_\mu(\mathrm{d}\hat{\mu}) \\
&= \lim_{n\to\infty} \mu(\varphi_2 \mathbb{E}_n[\varphi_1]) \ .
\end{aligned}$$

The second equality follows from Lebesgue's dominated convergence theorem because

$$|\nu(\varphi_2 \mathbb{E}_n[\varphi_1])| \leq \|\varphi_1\|_\infty \|\varphi_2\|_\infty , \qquad \nu \in \mathcal{P}(T), \; n \in \mathbb{N} .$$

More generally, by the same arguments, for any $k \in \{2, \ldots, \infty\}$ and $\varphi_1, \ldots, \varphi_k \in C(\Sigma)$,

$$\int_{\mathcal{P}(T)} \hat{\mu}(\varphi_k) \cdots \hat{\mu}(\varphi_1) \, \xi_\mu(\mathrm{d}\hat{\mu}) = \lim_{n_{k-1} \to \infty} \cdots \lim_{n_1 \to \infty} \mu\left(\varphi_k \mathbb{E}_{n_{k-1}}\left[\varphi_{k-1}\right] \cdots \mathbb{E}_{n_1}[\varphi_1]\right) . \tag{94}$$

Let

$$\mathfrak{X} \doteq \mathrm{span}\{\hat{\varphi}_1 \cdots \hat{\varphi}_k : k \in \mathbb{N}_0, \; \varphi_1, \ldots, \varphi_k \in C(\Sigma)\} \subseteq C(\mathcal{P}(T); \mathbb{R}) ,$$

where $\hat{\varphi}_1 \cdots \hat{\varphi}_k \doteq 1$ (i.e., the constant function 1) when $k = 0$, $\hat{\varphi} : \mathcal{P}(T) \to \mathbb{R}$ is defined for any $\varphi \in C(\Sigma)$ by

$$\hat{\varphi}(\nu) = \nu(\varphi) , \qquad \nu \in \mathcal{P}(T) , \tag{95}$$

and $C(\mathcal{P}(T); \mathbb{R})$ stands for the space of weak*-continuous functions $\mathcal{P}(T) \to \mathbb{R}$. By Equation (94), the quantities

$$\int_{\mathcal{P}(T)} \hat{\mu}(\varphi_k) \cdots \hat{\mu}(\varphi_1) \, \xi_\mu(\mathrm{d}\hat{\mu}) , \qquad k \in \{2, \ldots, \infty\}, \; \varphi_1, \ldots, \varphi_k \in C(\Sigma) ,$$

depend only on the probability measure μ, but not on the particular choice of the Choquet measure ξ_μ representing μ. As a consequence, the quantities

$$\xi_\mu(f) \equiv \int_{\mathcal{P}(T)} f(\hat{\mu}) \, \xi_\mu(\mathrm{d}\hat{\mu}) , \qquad f \in \mathfrak{X} , \tag{96}$$

also depend only on the probability measure μ. Observe that the elements of $C(\Sigma)$, seen as weak*-continuous linear functionals[14] on the dual of the Banach space $(C(\Sigma), \|\cdot\|_\infty)$, (trivially) separate the points of this dual space. As $\mathcal{P}(T)$ is weak*-compact, we can invoke the Stone-Weierstrass theorem [23, Theorem 7.191] to deduce that $\mathfrak{X}$ is a uniformly dense subalgebra of $C(\mathcal{P}(T); \mathbb{R})$. Thus, the quantities (96) extended to all continuous functions $f \in C(\mathcal{P}(T); \mathbb{R})$ only depend on μ. Therefore, the Choquet measure ξ_μ, which is a probability measure on $\mathcal{P}(T)$, must be unique. ■

Proposition 2.35 means that $\mathcal{P}(T)$ is a so-called *Choquet simplex*. In fact, $\mathcal{P}(T)$ is even the so-called Poulsen simplex (which is unique [44, Theorem 2.3.] up to an affine homeomorphism), as a consequence of the weak* density of its extreme points:

Proposition 2.36 (Weak* density of ergodic measures)
The set $\mathcal{P}_{\mathrm{erg}}(T)$ of ergodic measures is a G_δ weak-dense subset of $\mathcal{P}(T)$.*

Proof. Note that the weak* topology of $\mathcal{P}(T)$ is metrizable and thus, the set $\mathcal{P}_{\mathrm{erg}}(T)$ of extreme points of $\mathcal{P}(T)$ is a G_δ set, thanks to Lemma 3.27. We now prove the weak*-density of $\mathcal{P}_{\mathrm{erg}}(T)$ in several steps. We start with a general observation on the Banach space $C(\Sigma)$ of all continuous functions on character strings $\Sigma \doteq \Omega^{\mathbb{N}}$.

Step 1: For any $m \in \mathbb{N}$, we define the subspace

$$C_{\mathrm{cyl}}(\Sigma) \doteq \bigcup_{m \in \mathbb{N}} C(\Sigma) \cap \mathcal{J}_m$$

of so-called cylinder functions, where

$$\mathcal{J}_m \doteq \left\{\varphi : \Sigma \to \mathbb{R} : \exists f : \mathbb{R}^m \to \mathbb{R} \text{ such that } \varphi(\sigma) = f(\omega_1, \ldots, \omega_m) \text{ for all } \sigma = (\omega_n)_{n \in \mathbb{N}} \in \Sigma\right\} .$$

[14] I.e., for any $\varphi \in C(\Sigma; \mathbb{R})$, $\hat{\varphi}(\nu) = \nu(\varphi)$ for all $\nu \in C(\Sigma; \mathbb{R})^*$. Cf. Equation (95).

The set $C_{\mathrm{cyl}}(\Sigma)$ is a dense subspace of $C(\Sigma)$, by compactness of Σ (cf. Tychonoff's theorem [57, Section A.3]) and the equicontinuity of well-chosen families of continuous functions on compacts (via Ascoli's theorem [57, Section A.5]). Indeed, fix some $s = (s_n)_{n\in\mathbb{N}} \in \Sigma$. For any $\varphi \in C(\Sigma)$ and $m \in \mathbb{N}$, we define the continuous function $\varphi_m \in C(\Sigma) \cap \mathcal{J}_m$ by

$$\varphi_m(\sigma) \doteq \varphi(\sigma_m) \ , \qquad \sigma = (\omega_n)_{n\in\mathbb{N}} \in \Sigma \ ,$$

where

$$\sigma_m = (\omega_1, \ldots, \omega_m, s_{m+1}, s_{m+2}, \ldots) \in \Sigma \ , \qquad \sigma = (\omega_n)_{n\in\mathbb{N}} \in \Sigma \ .$$

Recall that the metric (4) for $\eta = 1/2$, that is,

$$d_{1/2}(\sigma, \sigma') \doteq \sum_{n\in\mathbb{N}} 2^{-n} d(\omega_n, \omega'_n) \ , \qquad \sigma = (\omega_n)_{n\in\mathbb{N}}, \sigma' = (\omega'_n)_{n\in\mathbb{N}} \in \Sigma \ ,$$

generates the topology of Σ. From (3) we observe that

$$d_{1/2}(\sigma, \sigma_m) = \sum_{n=m}^{\infty} 2^{-n} d(\omega_n, s_n) \leq 2^{1-m} \ , \qquad \sigma = (\omega_n)_{n\in\mathbb{N}} \in \Sigma, \ m \in \mathbb{N} \ . \tag{97}$$

This implies that the family $\{\varphi_m\}_{n\in\mathbb{N}}$ of functions on the compact Σ is equicontinuous. By compactness of Σ and continuity of φ, the family is also uniformly bounded. Again by (97), φ_m converges pointwise to φ, as $m \to \infty$. Now, using Ascoli's theorem [57, Section A.5] we deduce that, at least along some subsequence, the convergence is uniform, that is, for any $\varphi \in C(\Sigma)$,

$$\liminf_{m\to\infty} \|\varphi - \varphi_m\|_\infty = \liminf_{m\to\infty} \sup_{\sigma\in\Sigma} |\varphi(\sigma) - \varphi(\sigma_m)| = 0 \ .$$

In other words, $C_{\mathrm{cyl}}(\Sigma)$ is dense in $C(\Sigma)$.

Step 2: For any $m \in \mathbb{N}$ and $k \in \mathbb{N}$, let

$$C_{m,k}(\Sigma) \doteq \left\{\varphi \circ T^{(k-1)m} : \varphi \in C(\Sigma) \cap \mathcal{J}_m\right\} \ .$$

In fact, $C_{m,k}(\Sigma)$ is nothing but the space of continuous functions $(\omega_n)_{n\in\mathbb{N}} \mapsto \varphi((\omega_n)_{n\in\mathbb{N}})$ that only depend on $\omega_{(k-1)m+1}, \ldots, \omega_{km}$. These closed subalgebras of $C(\Sigma)$ are all naturally isomorphic to $C(\Omega^m)$. Given any positive normalized functionals (i.e., probability measures on Ω^m) $\nu_{m,k}$ on the subalgebras $C_{m,k}(\Sigma) \subseteq C(\Sigma)$, $m, k \in \mathbb{N}$, there is a unique positive normalized functional $\bigotimes_{k\in\mathbb{N}} \nu_{m,k}$ on $C(\Sigma)$, for which, for all $l \in \mathbb{N}$, $k_1, \ldots, k_l \in \mathbb{N}$, $k_1 < \cdots < k_l$, and $\varphi_1 \in C_{m,k_1}(\Sigma), \ldots, \varphi_l \in C_{m,k_l}(\Sigma)$,

$$\bigotimes_{k\in\mathbb{N}} \nu_{m,k}(\varphi_1 \cdots \varphi_l) = \nu_{m,k_1}(\varphi_1) \cdots \nu_{m,k_l}(\varphi_l) \ .$$

In other words, identifying positive normalized functionals on functions on compacts with probability measures, $\bigotimes_{k\in\mathbb{N}} \nu_{m,k}$ is nothing but the product probability measure of the probability measures $\nu_{m,k}$, $m, k \in \mathbb{N}$. Fix now any T-invariant measure $\mu \in \mathcal{P}(T) \subseteq C(\Sigma)^*$ and, for all $m, k \in \mathbb{N}$, define the positive normalized functional $\mu_{m,k} \in C_{m,k}(\Sigma)^*$ as being the restriction to $C_{m,k}(\Sigma)$ of μ, i.e.,

$$\mu_{m,k}(\varphi) \doteq \mu(\varphi) \ , \qquad \varphi \in C(\Sigma) \cap \mathcal{J}_m \ . \tag{98}$$

Then, for all $m \in \mathbb{N}$, let the probability measure μ_m on Σ be defined by

$$\mu_m^{\otimes} \doteq \bigotimes_{k\in\mathbb{N}} \mu_{m,k} \in C(\Sigma)^* \ . \tag{99}$$

Observe that, by construction, $T_*^m(\mu_m^\otimes) = \mu_m^\otimes$, where $T_*(\nu)$ stands for the pushforward of an arbitrary probability measure $\nu \in \mathcal{P}$ with respect to T, i.e., $T_*(\nu)(A) \doteq \nu(T^{-1}(A))$ for any Borel set $A \subseteq \Sigma$. That is, $\mu_m^\otimes$ is a m-periodic probability measure. We can now define a T-invariant measure by averaging out $\mu_m^\otimes$ within a period:

$$\mu_m \doteq \frac{1}{m}\sum_{n=0}^{m-1} T_*^n(\mu_m^\otimes) \in \mathcal{P}(T) \ . \tag{100}$$

By construction we also have

$$\lim_{m\to\infty} \mu_m(\varphi) = \mu(\varphi) \ , \qquad \varphi \in C_{\mathrm{cyl}}(\Sigma) \ .$$

Since $C_{\mathrm{cyl}}(\Sigma)$ is a dense subset of $C(\Sigma)$ (Step 1), we deduce that the sequence $(\mu_m)_{m\in\mathbb{N}}$ converges to μ, in the weak* topology. It remains to show that μ_m is an ergodic measure for each $m \in \mathbb{N}$.

Step 3: Fix again a T-invariant measure $\mu \in \mathcal{P}(T)$. Let $j_1, j_2 \in \mathbb{N}$ and $p, q \in (1,\infty)$ with $1/p+1/q = 1$. Take $\varphi \in C(\Sigma) \cap \mathcal{J}_{j_1}$ and $\psi \in C(\Sigma) \cap \mathcal{J}_{j_2}$. Then, for any $k \in \mathbb{N}$, $k \geq j_2 + m$, using Equations (99)–(100), we compute that

$$\begin{aligned}\frac{1}{k}\sum_{n=0}^{k}\int_\Sigma \psi(\sigma)\left(\varphi\circ T^k\right)(\sigma)\,\mu_m(\mathrm{d}\sigma) &= \frac{1}{k}\sum_{n=0}^{j_2+m}\int_\Sigma \psi(\sigma)\left(\varphi\circ T^k\right)(\sigma)\,\mu_m(\mathrm{d}\sigma)\\ &\quad + \left(1-\frac{(j_2+m)}{k}\right)\int_\Sigma \psi(\sigma)\,\mu_m(\mathrm{d}\sigma)\int_\Sigma \varphi\circ T^k(\sigma)\,\mu_m(\mathrm{d}\sigma)\,,\end{aligned}$$

which, combined with Hölder's inequality and the T-invariance of μ_m, yields the bound

$$\begin{aligned}&\left|\frac{1}{k}\sum_{n=0}^{k}\int_\Sigma \psi(\sigma)\left(\varphi\circ T^k\right)(\sigma)\,\mu_m(\mathrm{d}\sigma) - \int_\Sigma \psi(\sigma)\,\mu_m(\mathrm{d}\sigma)\int_\Sigma \varphi(\sigma)\,\mu_m(\mathrm{d}\sigma)\right|\\ \leq\ & \frac{2(j_2+m)}{k}\left(\left(\int_\Sigma |\psi(\sigma)|^q\,\mu_m(\mathrm{d}\sigma)\right)^{1/q}\left(\int_\Sigma |\varphi(\sigma)|^p\,\mu_m(\mathrm{d}\sigma)\right)^{1/p}\right)\end{aligned} \tag{101}$$

It follows that

$$\lim_{k\to\infty}\frac{1}{k}\sum_{n=0}^{k}\int_\Sigma \psi(\sigma)\left(\varphi\circ T^k\right)(\sigma)\,\mu_m(\mathrm{d}\sigma) = \int_\Sigma \psi(\sigma)\,\mu_m(\mathrm{d}\sigma)\int_\Sigma \varphi(\sigma)\,\mu_m(\mathrm{d}\sigma) \ .$$

By Step 1, $C_{\mathrm{cyl}}(\Sigma)$ is dense in $C(\Sigma)$. Therefore, we can extend the last limit to all $\varphi \in C(\Sigma)$ and $\psi \in C(\Sigma)$. In addition, since $C(\Sigma)$ is dense in $\mathcal{L}^p(\mu_m)$ for any $p \in (1,\infty)$, using Hölder's inequality as done in (101), we can then extend the last limit to all $\varphi \in \mathcal{L}^p(\mu_m)$ and $\psi \in \mathcal{L}^q(\mu_m)$, $p, q \in (1,\infty)$ with $1/p + 1/q = 1$. Now, by Proposition 2.33, μ_m is an ergodic measure for all $m \in \mathbb{N}$. ∎

2.9.2 The entropy functional as a thermodynamic limit

Our initial goal in this subsection is to show that the entropy is affine in the case where the alphabet is a general (i.e., not necessarily finite) compact metric space (see Proposition 2.38), what is well-known for finite alphabets [66, Theorem 8.1]. We summarize the main steps of the proof of this property, which stems from [1]. Then, we give in Corollary 2.39 a proof of ergodic abundance for alphabets that are general compact metric spaces; this result was already known in the case of a finite alphabet (see [48] and also [29]).

First, just like in [1, Section 2], we define the relative entropy (also known as Kullback-Leibler divergence) of a probability measure $\mu \in \mathcal{P}$ with respect to second one ν, both on the measurable space $(\Sigma, \mathfrak{S})$: For any sub-σ-algebra $\mathcal{A} \subseteq \mathfrak{S}$,

$$\mathcal{H}_{\mathcal{A}}(\mu|\nu) \doteq \begin{cases} \int_{\Sigma} f_{\mu}(\sigma) \ln f_{\mu}(\sigma) \ \nu(\mathrm{d}\sigma) \geq 0 \ , & f_{\mu} = \frac{d\mu|_{\mathcal{A}}}{d\nu|_{\mathcal{A}}} \text{ if } \mu|_{\mathcal{A}} \ll \nu|_{\mathcal{A}} \\ \infty \ , & \text{otherwise} \ , \end{cases} \tag{102}$$

where f_{μ} is the Radon-Nikodym derivative of the restrictions to $\mathcal{A}$, $\mu|_{\mathcal{A}}$ and $\nu|_{\mathcal{A}}$, of the measures μ and ν. Here, $\mu|_{\mathcal{A}} \ll \nu|_{\mathcal{A}}$ means that $\mu|_{\mathcal{A}}$ is absolutely continuous with respect to $\nu|_{\mathcal{A}}$ and we use the convention $f_{\mu}(\sigma) \ln f_{\mu}(\sigma) \doteq 0$ when $f_{\mu}(\sigma) = 0$, as is usual. As it is well-known, the relative entropy is monotonically decreasing with respect to "coarse graining", that is, for any σ-subalgebra $\mathcal{B} \subseteq \mathcal{A}$, we have

$$\mathcal{H}_{\mathcal{B}}(\mu|\nu) \leq \mathcal{H}_{\mathcal{A}}(\mu|\nu) \ . \tag{103}$$

For any finite subset $\Lambda \subseteq \mathbb{N}$, we apply this definition to the initial σ-algebra $\mathfrak{S}_{\Lambda}$ of the canonical projections[15] $\Sigma \to \Omega^{\Lambda}$. For simplicity we use the notation

$$\mathcal{H}_{\Lambda}(\mu|\nu) \equiv \mathcal{H}_{\mathfrak{S}_{\Lambda}}(\mu|\nu) \tag{104}$$

for any two probability measures μ, ν on the measurable space $(\Sigma, \mathfrak{S})$ and finite set $\Lambda \subseteq \mathbb{N}$. We call this quantity the relative entropy of μ in Λ with respect to ν. Note from Inequality (103) that

$$\mathcal{H}_{\Lambda'}(\mu|\nu) \leq \mathcal{H}_{\Lambda}(\mu|\nu) \ , \qquad \Lambda' \subseteq \Lambda \ . \tag{105}$$

We fix an *a priori* probability measure m on Ω and, for any probability measure $\mu \in \mathcal{P}$ and finite subset $\Lambda \subseteq \mathbb{N}$, we define $\mathcal{H}_{\Lambda}(\mu)$, the so-called specific entropy in Λ of μ. It is nothing but *minus* the relative entropy of μ in Λ with respect to the product measure on Σ constructed from m, i.e.,

$$\mathrm{m}_{\otimes} \doteq \bigotimes_{k \in \mathbb{N}} \mathrm{m} \in \mathcal{P}(T) \ . \tag{106}$$

In other words, for any probability measure μ and finite subset $\Lambda \subseteq \mathbb{N}$,

$$\mathcal{H}_{\Lambda}(\mu) \doteq -\mathcal{H}_{\Lambda}(\mu|\mathrm{m}_{\otimes}) \ . \tag{107}$$

In the limit $n \to \infty$ of finite subsets $\{1, \ldots, n\}$ it gives the so-called specific entropy per site of any *T-invariant* probability measure $\mu \in \mathcal{P}(T)$, which is well-defined in this case by the limit

$$\mathfrak{h}(\mu) \doteq \lim_{n \to \infty} n^{-1} \mathcal{H}_{\{1,\ldots,n\}}(\mu) \in [-\infty, 0] \ . \tag{108}$$

The functional $\mathfrak{h} : \mathcal{P}(T) \to [-\infty, 0]$ is both concave and weak* upper semicontinuous. See [31] for more details. Note in particular that [31, Definition 15.13, page 315] essentially correspond to this claim, but the lattice is $\mathbb{Z}$ there, instead of $\mathbb{N}$.

A thermodynamic limit similar to (108) can also be obtained by using other reference measures, not just the product measure (106), like in [31, Theorem 15.30]. Observe that the product measure is nothing but the Gibbs measure associated with a constant potential. Indeed, more generally, we can take the (unique) Gibbs measure associated with any Hölder potential f as reference measure. This is done as follows:

Step 1: Given any Hölder potential $f \in C^{\alpha}(\Sigma)$ ($\alpha \in (0,1)$), we define an appropriate normalized potential $\bar{f}$ and consider the associated Ruelle operator $\mathcal{L}^{\mathrm{m}}_{\bar{f}}$. In particular, we should have $\mathcal{L}^{\mathrm{m}}_{\bar{f}}(\mathbf{1})(\sigma) = 1$

[15] $(\sigma_i)_{i \in \mathbb{N}} \mapsto (\sigma_i)_{i \in \Lambda}$

for all $\sigma \in \Sigma$. This is done in the following way: As already explained in [1], one can use the generalization of the Ruelle-Perron-Frobenius theorem given in [50] to define a cohomologous normalized potential defined by

$$\bar{f} \doteq f + \ln h_f - \ln(h_f \circ T) - \ln \lambda_f \ ,$$

for any $f \in C^\alpha(\Sigma)$, where λ_f is a maximal eigenvalue of $\mathcal{L}_f^{\mathrm{m}}$ having $h_f \in C^\alpha(\Sigma) \cap \mathcal{C}^+$ as an eigenfunction.

Step 2: The relation between f and $\bar{f}$, or more explicitly $\mathcal{L}_f^{\mathrm{m}}$ and $\mathcal{L}_{\bar{f}}^{\mathrm{m}}$, is explained in [50] and refers to the following observations: Since λ_f is a maximal eigenvalue of $\mathcal{L}_f^{\mathrm{m}}$ there is an eigenprobability ν_f such that

$$\left(\mathcal{L}_f^{\mathrm{m}}\right)^* (\nu_f) = \lambda_f \nu_f$$

with $(\mathcal{L}_f^{\mathrm{m}})^*$ being the usual dual operator acting on probability measures. It turns out that the product $h_f \nu_f$, properly normalized, is an equilibrium measure $\mu_f \in \mathcal{P}(T)$ for f, in the sense of Definition 2.5. In other words, μ_f maximizes the pressure functional $\mathfrak{P}_L(f)$, as defined by (12) with $\varphi = f$. Moreover, $\log \lambda_f = P_L(f)$ and the set

$$\left\{ \nu \in \mathcal{P}(T) : (\mathcal{L}_{\bar{f}}^{\mathrm{m}})^* \nu = \nu \right\} \tag{109}$$

has only one element, which is equal to $\mu_f = \mu_{\bar{f}} \in \mathcal{P}(T)$ (notice that the normalized potential $\bar{f}$ is of Hölder class, as f is a Hölder potential), called[16] the Gibbs measure associated with the potential f.

Having now the Gibbs measure μ_f at our disposal, we study the thermodynamic limit of the relative entropy of a T-invariant measure with respect to μ_f. This refers to [1, Theorem 3.1], which states that, for any probability measure m on Ω having full support, each α-Hölder-continuous function $f \in C^\alpha(\Sigma)$ ($\alpha \in (0,1)$) and any T-invariant probability measure $\nu \in \mathcal{P}(T)$, the following limit exists:

$$\mathfrak{h}\left(\nu|\mu_f\right) \doteq \lim_{n\to\infty} n^{-1} \mathcal{H}_{\{1,\ldots,n\}}\left(\nu|\mu_f\right) = \ln \lambda_f - \int_\Sigma f(\sigma)\,\nu(\mathrm{d}\sigma) - \mathfrak{h}(\nu) \in [0,\infty] \ , \tag{110}$$

where $\mu_f \in \mathcal{P}(T)$ is the Gibbs measure associated with f, which is the unique element of the set (109). From [1, Theorem 3.1] together with the positivity of the relative entropy it also follows that, for any $\alpha \in (0,1)$,

$$\mathfrak{h}(\mu) \leq \ln \lambda_f - \int_\Sigma f(\sigma)\,\mu(\mathrm{d}\sigma) \ , \qquad \mu \in \mathcal{P}(T)\,,\ f \in C^\alpha(\Sigma) \ ,$$

provided m is a probability measure on Ω having full support. By Proposition 2.2, we thus deduce that

$$\mathfrak{h}(\mu) \leq h(\mu) = \inf_{f \in C^\alpha(\Sigma)} \left\{ \ln \lambda_f - \int_\Sigma f(\sigma)\,\mu(\mathrm{d}(\sigma)) \right\} \ , \qquad \mu \in \mathcal{P}(T) \ ,$$

when m is a probability measure on Ω having full support. Recall that $h : \mathcal{P}(T) \to \mathbb{R}$ is the entropy of Definition 2.1. In fact, the functionals $\mathfrak{h}$ and h are equal to each other. This claim refers to [1, Theorem 3.4]. Combined with Proposition 2.2 and Equation (108), we thus arrive at the Aguiar-Cioletti-Ruviaro theorem:

Theorem 2.37 (The entropy as a thermodynamic limit)
Assume that m *is a probability measure on* Ω *having full support. Then, for any* $\mu \in \mathcal{P}(T)$*,*

$$h(\mu) = \inf_{f \in C^\alpha(\Sigma)} \left\{ \ln \lambda_f - \int_\Sigma f(\sigma)\,\mu(\mathrm{d}(\sigma)) \right\} = \lim_{n\to\infty} n^{-1} \mathcal{H}_{\{1,\ldots,n\}}(\mu) \ .$$

[16] It is sometimes named the normalized DLR probability for $\bar{f}$; for the exact definition and equivalences, see [27].

The representation of entropy as a thermodynamic limit is very useful because, among other things, it allows us to derive the affine property of the entropy as a functional on $\mathcal{P}(T)$:

Proposition 2.38 (Affine property of the entropy)
Assume that m *is a probability measure on* Ω *having full support. Then, the mapping* $\mu \mapsto h(\mu)$ *from* $\mathcal{P}(T)$ *to* $[-\infty, 0]$ *is affine.*

Proof. We start with the concavity of the entropy h, which is easy to prove directly from its definition (Definition 2.1): For any $\lambda \in (0,1)$, consider the convex combination

$$\mu_3 = \lambda \mu_1 + (1-\lambda)\,\mu_2 \tag{111}$$

of two arbitrary T-invariant probability measures $\mu_1, \mu_2 \in \mathcal{P}(T)$. For the T-invariant probability measure μ_3 and an arbitrary strictly positive parameter $\epsilon \in \mathbb{R}^+$, take an approximating minimizer of (10), i.e., a positive function $f \in \mathcal{C}^+$ such that

$$h(\mu_3) + \epsilon \geq \int_\Sigma \ln\left(\frac{\mathcal{L}_0^{\mathrm{m}} f(\sigma)}{f(\sigma)}\right) \mu_3(\mathrm{d}\sigma) \ .$$

It follows from Equation (111) that

$$\begin{aligned} h(\mu_3) + \epsilon &\geq \lambda \int_\Sigma \ln\left(\frac{\int_\Omega f(\omega_0\sigma)\,\mathrm{m}(\mathrm{d}\omega_0)}{f(\sigma)}\right) \mu_1(\mathrm{d}\sigma) \\ &\quad + (1-\lambda) \int_\Sigma \ln\left(\frac{\int_\Omega f(\omega_0\sigma)\,\mathrm{m}(\mathrm{d}\omega_0)}{f(\sigma)}\right) \mu_2(\mathrm{d}\sigma) \\ &\geq \lambda h(\mu_1) + (1-\lambda)\,h(\mu_2) \ . \end{aligned}$$

Since $\epsilon \in \mathbb{R}^+$ is arbitrary,

$$h(\mu_3) \geq \lambda h(\mu_1) + (1-\lambda)\,h(\mu_2) \tag{112}$$

for any $\lambda \in (0,1)$ and $\mu_1, \mu_2 \in \mathcal{P}(T)$.

We now prove that the convexity of the entropy h, which has become relatively easy to prove, thanks to Theorem 2.37: Take again arbitrary $\lambda \in (0,1)$ and $\mu_1, \mu_2 \in \mathcal{P}(T)$. Fix also some $n \in \mathbb{N}$. Regarding the σ-algebra $\mathfrak{S}_{\{1,\ldots,n\}}$, (the convex combination (111)) μ_3 has Radon-Nikodym derivative equal to

$$f_{\mu_3} = \frac{d\mu_3|_{\mathfrak{S}_{\{1,\ldots,n\}}}}{d\mathrm{m}_\otimes|_{\mathfrak{S}_{\{1,\ldots,n\}}}} = \lambda f_{\mu_1} + (1-\lambda)\,f_{\mu_2} \ , \tag{113}$$

provided f_{μ_1} and f_{μ_2}, the corresponding Radon-Nikodym derivatives of μ_1 and μ_2, exist. In this case,

$$\mathcal{H}_{\{1,\ldots,n\}}(\mu_3) = -\mathcal{H}_{\{1,\ldots,n\}}(\mu_3|\mathrm{m}_\otimes) = -\int_\Sigma f_{\mu_3}(\sigma) \ln f_{\mu_3}(\sigma)\ \mathrm{m}_\otimes(\mathrm{d}\sigma) \ ,$$

see Equations (102)–(107). If one of the two Radon-Nikodym derivatives f_{μ_1} and f_{μ_2} does not exist, the Radon-Nikodym derivative f_{μ_3} does not exist either and, for any $\lambda \in (0,1)$, we trivially have

$$\mathcal{H}_{\{1,\ldots,n\}}(\mu_3) \leq \lambda \mathcal{H}_{\{1,\ldots,n\}}(\mu_1) + (1-\lambda)\,\mathcal{H}_{\{1,\ldots,n\}}(\mu_2) \ , \tag{114}$$

both sides of the inequality being equal to ∞. Thus, assume that both f_{μ_1} and f_{μ_2} exist. Then, since the Radon-Nikodym derivatives are $\mathrm{m}_\otimes$-almost everywhere positive and, by (113) and the fact that the function $\ln(\cdot)$ is monotonically increasing,

$$\ln\left(f_{\mu_3}(\sigma)\right) = \ln\left(\lambda f_{\mu_1}(\sigma) + (1-\lambda)\,f_{\mu_2}(\sigma)\right) \geq \ln\left(\lambda f_{\mu_1}(\sigma)\right) \ ,$$

($\mathrm{m}_\otimes$-almost everywhere), we arrive at

$$\begin{aligned}\lambda \int_\Sigma f_{\mu_1} \ln \left(f_{\mu_3}\right) \mathrm{m}_\otimes (\mathrm{d}\sigma) &\geq \lambda \int_\Sigma f_{\mu_1} \ln \left(\lambda f_{\mu_1}\right) \mathrm{m}_\otimes (\mathrm{d}\sigma) \\ &= \lambda \int_\Sigma f_{\mu_1} \ln \left(f_{\mu_1}\right) \mathrm{m}_\otimes (\mathrm{d}\sigma) + \lambda \ln \lambda \int_\Sigma f_{\mu_1} \mathrm{m}_\otimes (\mathrm{d}\sigma) \\ &= \lambda \int_\Sigma f_{\mu_1} \ln \left(f_{\mu_1}\right) \mathrm{m}_\otimes (\mathrm{d}\sigma) + \lambda \ln \lambda \, . \end{aligned} \tag{115}$$

In the same way, we obtain that

$$(1-\lambda) \int_\Sigma f_{\mu_2} \ln \left(f_{\mu_3}\right) \mathrm{m}_\otimes (\mathrm{d}\sigma) \geq (1-\lambda) \int_\Sigma f_{\mu_2} \ln \left(f_{\mu_2}\right) \mathrm{m}_\otimes (\mathrm{d}\sigma) + (1-\lambda) \ln (1-\lambda) \ . \tag{116}$$

By adding (115) and (116), we get the upper bound

$$\mathcal{H}_{\{1,\ldots,n\}} (\mu_3) \leq \lambda \mathcal{H}_{\{1,\ldots,n\}} (\mu_1) + (1-\lambda) \mathcal{H}_{\{1,\ldots,n\}} (\mu_2) - \lambda \ln \lambda - (1-\lambda) \ln (1-\lambda) \tag{117}$$

for any $\lambda \in (0,1)$, $\mu_1, \mu_2 \in \mathcal{P}(T)$ and $n \in \mathbb{N}$ such that f_{μ_1} and f_{μ_2} both exist. By Theorem 2.37 combined with (114) and (117), it follows that

$$h(\mu_3) \leq \lambda h(\mu_1) + (1-\lambda) h(\mu_2) \tag{118}$$

for any $\lambda \in (0,1)$ and $\mu_1, \mu_2 \in \mathcal{P}(T)$. Thus, h is both concave and convex, i.e., affine. ■

Corollary 2.39 (Ergodic abundance)
Assume that m *is a probability measure on* Ω *having full support. For any* $\mu \in \mathcal{P}(T)$*, there exists a sequence* $(\mu_m)_{m\in\mathbb{N}}$ *of ergodic measures converging in the weak* topology to* μ *such that*

$$\lim_{m\to\infty} h(\mu_m) = h(\mu) \ .$$

Proof. Let $\mu \in \mathcal{P}(T)$ and define the sequence $(\mu_m)_{m\in\mathbb{N}}$ by (98)–(100). Indeed, by the proof of Proposition 2.36 (cf. Steps 2 and 3), $(\mu_m)_{m\in\mathbb{N}}$ converges to μ in the weak* topology while μ_m is an ergodic measure for each $m \in \mathbb{N}$. By Theorem 2.37,

$$h(\mu_m) = \lim_{n\to\infty} n^{-1} \mathcal{H}_{\{1,\ldots,n\}} (\mu_m) = \lim_{n\to\infty} n^{-1} \mathcal{H}_{\{1,\ldots,n\}} \left(\frac{1}{m} \sum_{k=0}^{m-1} T_*^k \left(\mu_m^\otimes\right) \right) \ .$$

By convexity of the relative entropy, $\mathcal{H}_{\{1,\ldots,n\}}$ is concave. In particular,

$$\mathcal{H}_{\{1,\ldots,n\}} \left(\frac{1}{m} \sum_{k=0}^{m-1} \mu_{m,k}^\otimes \right) \geq \frac{1}{m} \sum_{k=0}^{m-1} \mathcal{H}_{\{1,\ldots,n\}} \left(\mu_{m,k}^\otimes\right) \ ,$$

where $\mu_{m,k}^\otimes \doteq T_*^k(\mu_m^\otimes)$. Since the Radon-Nikodym derivatives are $\mathrm{m}_\otimes$-almost everywhere positive and the function $\ln(\cdot)$ is monotonically increasing, we have that

$$\begin{aligned}\left(\frac{1}{m} \sum_{k=0}^{m-1} f_{\mu_{m,k}^\otimes} \right) \ln \left(\frac{1}{m} \sum_{k=0}^{m-1} f_{\mu_{m,k}^\otimes} \right) &\geq \frac{1}{m} \sum_{k=0}^{m-1} f_{\mu_{m,k}^\otimes} \ln \left(\frac{1}{m} f_{\mu_{m,k}^\otimes} \right) \\ &= -\left(\frac{1}{m} \sum_{k=0}^{m-1} f_{\mu_{m,k}^\otimes} \right) \ln m + \frac{1}{m} \sum_{k=0}^{m-1} f_{\mu_{m,k}^\otimes} \ln \left(f_{\mu_{m,k}^\otimes} \right)\end{aligned}$$

($\mathrm{m}_{\otimes}$-almost everywhere). Therefore,

$$\frac{1}{m}\sum_{k=0}^{m-1}\mathcal{H}_{\{1,\ldots,n\}}\left(\mu_{m,k}^{\otimes}\right)\leq\mathcal{H}_{\{1,\ldots,n\}}\left(\frac{1}{m}\sum_{k=0}^{m-1}\mu_{m,n}^{\otimes}\right)\leq\ln m+\frac{1}{m}\sum_{k=0}^{m-1}\mathcal{H}_{\{1,\ldots,n\}}\left(\mu_{m,k}^{\otimes}\right)\ .$$

(This is essentially the same argument used in the proof of Proposition 2.38 to prove the affine property of the entropy.) It follows that

$$\begin{aligned}h\left(\mu_{m}\right)&=\lim_{n\to\infty}\frac{1}{mn}\sum_{k=0}^{m-1}\mathcal{H}_{\{1,\ldots,n\}}\left(\mu_{m,k}^{\otimes}\right)=\lim_{n\to\infty}\frac{1}{m^{2}n}\sum_{k=0}^{m-1}\mathcal{H}_{\{1,\ldots,mn\}}\left(\mu_{m,k}^{\otimes}\right)\\&=\lim_{n\to\infty}\frac{1}{m^{2}n}\sum_{k=0}^{m-1}\mathcal{H}_{\{1+k,\ldots,mn+k\}}\left(\mu_{m}^{\otimes}\right)\\&=\lim_{n\to\infty}\frac{1}{m^{2}n}\Bigg(n\mathcal{H}_{\{1,\ldots,m\}}\left(\mu\right)+\sum_{k=1}^{m-1}\mathcal{H}_{\{1,\ldots,k\}}\left(\mu\right)\\&\qquad\qquad+\sum_{k=1}^{m-1}\left((n-1)\mathcal{H}_{\{1,\ldots,m\}}\left(\mu\right)+\mathcal{H}_{\{k+1,\ldots,m\}}\left(\mu\right)\right)\Bigg)\\&=\frac{1}{m}\mathcal{H}_{\{1,\ldots,m\}}\left(\mu\right)\ .\end{aligned}\tag{119}$$

To get the third equality we use the additivity of the relative entropy with respect to product probability measures, along with the T-invariance of μ. Notice also, that by Inequality (105) and the T-invariance of μ, the terms

$$\mathcal{H}_{\{1,\ldots,k\}}\left(\mu\right)\qquad\text{and}\qquad\mathcal{H}_{\{k+1,\ldots,m\}}\left(\mu\right)=\mathcal{H}_{\{1,\ldots,m-k\}}\left(\mu\right)$$

for $k\in\{1,\ldots,m-1\}$ have to be finite (i.e., not $-\infty$) whenever $\mathcal{H}_{\{1,\ldots,m\}}\left(\mu\right)>-\infty$. In turn, $\mathcal{H}_{\{1,\ldots,m\}}\left(\mu\right)>-\infty$ for all $m\in\mathbb{N}$ when the entropy $h\left(\mu\right)$ is finite, i.e., $h\left(\mu\right)>-\infty$), by Theorem 2.37 and Inequality (105). In particular $h\left(\mu_{m}\right)$ is finite for all $m\in\mathbb{N}$ when $h\left(\mu\right)$ is finite. Now, we deduce once again from Theorem 2.37 and (119) that $h\left(\mu_{m}\right)$ tends to $h\left(\mu\right)$, as $m\to\infty$. ■

2.9.3 Minimization of real-valued functions via Γ-regularization

In 2012 we proved [17] that, in great generality, the minimization problem of any non-convex and non-lower semicontinuous function that is bounded from below and defined on a compact convex subset of a locally convex real topological vector space can be analyzed through an associated convex and lower semicontinuous function, which is the so-called Γ-regularization of the original function. We present this result because it is important in Sections 2.7.4 and 2.8, but is not common knowledge (at least in the thermodynamic formalism community).

Fix once and for all in all the subsection a (nonempty) compact convex subset K of a locally convex real (topological vector) space[17] $\mathcal{X}$. The aim of [17] is to characterize of the set of *generalized* minimizers of real-valued functions on K. Given an (extended) function $f:K\to(-\infty,\infty]$ the set of its generalized minimizers is, by definition, the closure of the nonempty[18] set

$$\varOmega\left(f,K\right)\doteq\left\{x\in K:\exists(x_{i})_{i\in I}\subseteq K\ \text{ with } x_{i}\to x\text{ and }\lim_{I}f(x_{i})=\inf\ f(K)\right\}$$

of all limit points of approximating minimizers of f. Note that the function f is here completely general. In particular, it is **not** necessarily convex or lower semicontinuous.

[17] Topological vector spaces are here Hausdorff spaces, i.e., singletons in those spaces are closed sets.

[18] $\varOmega\left(h,K\right)$ is non-empty because any net $(x_{i})_{i\in I}\subseteq K$ converges along a subnet, K being compact.

It turns out that the study of the above set of generalized minimizers can be performed via Γ-regularization, which is defined as follows [2, Eq. (1.3) in Chapter I]: For any extended real-valued function $f : K \to [\mathrm{k}, \infty]$ with $\mathrm{k} \in \mathbb{R}$, its Γ-regularization (on K) is the function defined as the supremum over all affine and continuous minorants $m : \mathcal{X} \to \mathbb{R}$ of f, i.e., for all $x \in K$,

$$\Gamma(f)(x) \doteq \sup \{m(x) : m \in \mathrm{A}(\mathcal{X}) \text{ and } m|_K \leq f\} \ , \tag{120}$$

where $\mathrm{A}(\mathcal{X})$ is the space of all affine continuous real-valued functions on $\mathcal{X}$. Since the Γ-regularization $\Gamma(f)$ is a supremum over continuous functions, $\Gamma(f)$ is a convex and lower semicontinuous function on K. In fact, it is the largest weak*-lower semicontinuous convex function below f, by [17, Corollary 3.2]. See, e.g., [17] or [18, Section 10.5] for a short review on Γ-regularizations.

Theorem 2.40 (Minimization of real-valued functions)
For any function $f : K \to [\mathrm{k}, \infty]$ *with* $\mathrm{k} \in \mathbb{R}$*, the following assertions hold true:*
(i)

$$\inf f(K) = \inf \Gamma(f)(K) \ .$$

(ii) *The set* M *of minimizers of* $\Gamma(f)$ *over* K *is the closed convex hull of the set* $\Omega(f, K)$ *of generalized minimizers of* f *over* K*, i.e.,*

$$M = \overline{\mathrm{co}}(\Omega(f, K)) \ .$$

(iii) *All extreme points of the compact convex set* M *belong to the set of generalized minimizers of* f*, i.e.,* $\mathcal{E}(M) \subseteq \overline{\Omega(f, K)}$.

This theorem correspond to [17, Theorems 1.4–1.5] and implies, among other things, a generalization to locally convex real spaces of the Lanford III-Robinson theorem, proven for separable real Banach spaces, which characterizes the subdifferentials of continuous convex functions. See [17, Theorem 1.8].

3 Abstract Theory of Bogoliubov Linearizations

The abstract structure of the arguments used above for the nonlinear thermodynamic formalism of dynamical systems can be extracted, revealing a much broader spectrum of applications than that considered in Section 2. In fact, all one needs is a generic compact convex Hausdorff space K, instead of the particular case of the metrizable weak*-compact and convex space $\mathcal{P}(T)$ of T-invariant measures of Section 2, or the metrizable weak*-compact and convex space of states on the CAR C^*-algebra used in [18]. The considered nonlinear functions can also be very general, a sum of a concave upper semicontinuous and a convex lower semicontinuous function. Despite this being a very general framework, explicit results can be obtained via self-consistency when the associated linear variational problems are well understood. This transition between nonlinear variational problems and linear ones is what we call here *Bogoliubov linearization*, in honour of Bogoliubov's seminal approach used in his famous microscopic theory of He^4 superfluidity, proposed in 1947. That is what is presented in this section, forming a general theoretical framework that we believe is both useful and elegant.

For the reader's convenience, we begin by recalling basic definitions and properties of the Legendre-Fenchel transform and subdifferentials (Section 3.1). Next, we study the purely concave case (Section 3.2), which is then used as a springboard to the general case (Section 3.3). In Section 3.4, this general study leads to decision rules of a canonical two-person zero-sum game, referred to here as thermodynamic game, which is the terminology proposed by [18]. In Section 3.5, we conclude by linking this formalism to Monge-Kantorovich's optimal transport and duality formulas. This opens up our approach to a broader field of research.

3.1 The Legendre-Fenchel transform and subdifferentials

Let $\mathcal{X}$ be any real normed space. Its (topological) dual space is the set of all continuous linear functionals on $\mathcal{X}$ and is denoted $\mathcal{X}^*$, as is usual. The dual space $\mathcal{X}^*$ is then endowed with the weak* topology. Recall that it is a locally convex space. By [57, Theorem 3.10], $(\mathcal{X}, \mathcal{X}^*)$ forms a so-called dual pair, meaning that, for all $x \in \mathcal{X}$, the functional $y \mapsto y(x) \in \mathbb{C}$ on $\mathcal{X}^*$ is continuous with respect to the weak* topology, and all linear functionals which are continuous with respect to the weak* topology have this form.

Given such a dual pair $(\mathcal{X}, \mathcal{X}^*)$, the Legendre-Fenchel transform[19] g^* of a function $g : \mathcal{X} \to (-\infty, \infty]$ (that is, g is allowed to take the value ∞) with nonempty domain

$$\mathrm{dom}\,(g) \doteq \{x \in \mathcal{X} : g\,(x) < \infty\} \neq \emptyset$$

is the convex lower semicontinuous functional from $\mathcal{X}^*$ to $(-\infty, \infty]$ defined by

$$g^*\,(x) \doteq \sup_{y \in \mathcal{X}} \{x\,(y) - g\,(y)\}\ , \qquad x \in \mathcal{X}^*\ . \tag{121}$$

Note that the conditions imposed on g means that we only consider here so-called *proper* functions, as defined in [59, Definition 1.3.1]. In this context, the double Legendre-Fenchel transform g^{**}, also called the biconjugate of g, is nothing but the so-called Γ-regularization of g (see (120)), that is, the largest lower semicontinuous convex function below g. In particular, if g is a lower semicontinuous convex function then

$$g\,(x) = \sup_{y \in \mathcal{X}^*} \{y\,(x) - g^*\,(y)\}\ , \qquad x \in \mathcal{X}\ . \tag{122}$$

Recall that the Legendre-Fenchel transform is always weak*-lower semicontinuous, being the supremum over weak*-continuous (and affine) functionals. See [59, Theorem 2.2.4] and [17, Corollary 3.2]. Observe from (122) that, for lower semicontinuous convex functions g, one must have

$$\mathrm{dom}\,(g^*) \doteq \{x \in \mathcal{X} : g^*\,(x) < \infty\} \neq \emptyset\ , \tag{123}$$

since the function g never takes the value $-\infty$ (it is even proper, by assumption). In this case, the nonempty sets $\mathrm{dom}(g)$ and $\mathrm{dom}(g^*)$ are always convex, by convexity of Legendre-Fenchel transforms.

This essential property (that is, (122)) is crucial for allowing us to implement the *Bogoliubov linearization* in a mathematically rigorous manner.

Below, we focus on functions whose Legendre-Fenchel transform grows faster than linearly. More precisely, we consider functions with large enough minimal linear growth. In fact, given a dual pair $(\mathcal{X}, \mathcal{X}^*)$, we say that the Legendre-Fenchel transform g^* of a function $g : \mathcal{X} \to (-\infty, \infty]$ with nonempty domain $\mathrm{dom}(g) \neq \emptyset$ has *minimal linear growth* $\lambda \in \mathbb{R}^+$ if

$$\lim_{R \to \infty} \sup_{y \in \mathcal{X}^* \backslash B(0,R)} \left\{\lambda \left\|y\right\|_{\mathrm{op}} - g^*\,(y)\right\} = -\infty\ , \tag{124}$$

where

$$B(0, R) \doteq \left\{y \in \mathcal{X}^* : \left\|y\right\|_{\mathrm{op}} \leq R\right\} \subseteq \mathcal{X}^* \tag{125}$$

is the closed ball of radius $R \in \mathbb{R}^+$ and center $0 \in \mathcal{X}^*$, while, as is usual,

$$\left\|y\right\|_{\mathrm{op}} \doteq \sup\{|y\,(x)| : x \in \mathcal{X},\ \|x\|_{\mathcal{X}} \leq 1\}\ , \qquad y \in \mathcal{X}^*\ . \tag{126}$$

[19] It is also called the conjugate of g.

If such a function g is additionally convex and lower semicontinuous, then we deduce from (122) and (124) the existence of a strictly positive radius $R \in \mathbb{R}^+$ such that, for any $x \in \mathcal{X}$ satisfying $\|x\|_{\mathcal{X}} \leq \lambda$,

$$g(x) = \sup_{y \in \mathcal{X}^*} \{y(x) - g^*(y)\} = \sup_{y \in B(0,R)} \{y(x) - g^*(y)\} \ . \tag{127}$$

This property is important to define below "equilibrium states".

Given again a dual pair $(\mathcal{X}, \mathcal{X}^*)$ and a function $g : \mathcal{X} \to (-\infty, \infty]$ with nonempty domain $\mathrm{dom}(g) \neq \emptyset$, the corresponding Legendre-Fenchel transform is directly related to the so-called subdifferential of g at $z \in \mathrm{dom}(g)$, which is defined by

$$\partial g(z) \doteq \{x \in \mathcal{X}^* : \forall y \in \mathcal{X}, \quad x(y - z) \leq g(y) - g(z)\} \ . \tag{128}$$

In particular, by [59, Proposition 4.4.1], when g is also convex, $z \in \mathrm{dom}(g)$ is solution to the variational problem (121) if and only if $x \in \partial g(z)$. When g is a lower semicontinuous convex function, we can also infer from [59, Proposition 4.4.1] that, for any $x \in \mathrm{dom}(g)$, $z \in \mathcal{X}^*$ is solution to (122) if and only if $z \in \partial g(x) \subseteq \mathcal{X}^*$. Note indeed that $z \in \mathcal{X}^*$ is solution to (122) if and only if

$$g(x) + g^*(z) = z(x) \ .$$

3.2 Bogoliubov linearizations for concave interactions

General assumptions. In all this subsection, we consider the following assumptions:

A1 $\mathcal{X}$ is a real normed space, K is a compact convex Hausdorff space, $\tau : K \to \mathcal{X}$ is a continuous affine transformation and $f : K \to \{-\infty\} \cup \mathbb{R}$ is an upper semicontinuous concave function (which is of course not identically equal to $-\infty$).

A2 $g : \mathcal{X} \to \mathbb{R}$ is a lower semicontinuous and convex function for which there is a radius $R_0 \in \mathbb{R}^+$ such that $B(0,R) \cap \mathrm{dom}(g^*) \subseteq \mathcal{X}^*$ is nonempty and weak*-closed for all $R \in [R_0, \infty)$. Moreover, the Legendre-Fenchel transforms g^* has minimal linear growth $\|\tau\|_\infty \in \mathbb{R}^+$ in the sense of Equation (124).

Above, recall that $B(0,R) \subseteq \mathcal{X}^*$ is the weak*-closed ball of radius $R \in \mathbb{R}^+$ and center $0 \in \mathcal{X}^*$, see (125)–(126). We use the standard notation

$$\|\tau\|_\infty \doteq \sup \{\|\tau(\mu)\|_{\mathcal{X}} : \mu \in K\} \tag{129}$$

for the uniform norm (or sup norm) of τ. Recall that any continuous function on a compact is uniformly bounded. In particular, in Condition A2 one always has that $\|\tau\|_\infty < \infty$. Remark also from Condition A2 that g has obviously full domain, i.e., $\mathrm{dom}(g) = \mathcal{X}$ (it is also proper), and the convex set $\mathrm{dom}(g^*)$ is nonempty, thanks to (123).

We call the elements of K *(generalized) states*. In this subsection we study the set

$$E_{\mathcal{F}} = \{\mu \in K : \mathcal{F}(\mu) = \sup \mathcal{F}(K)\}$$

of *nonlinear equilibrium states*, where the function $\mathcal{F} : K \to \{-\infty\} \cup \mathbb{R}$ is defined by

$$\mathcal{F} \doteq f - g \circ \tau \ . \tag{130}$$

Note from Conditions A1–A2 that $\mathcal{F}$ is upper semicontinuous and concave, and so, $E_{\mathcal{F}}$ is a (nonempty) compact convex set.

Remark 3.1
Clearly, $\mathcal{F} : K \rightarrow \{-\infty\} \cup \mathbb{R}$ is an upper semicontinuous concave function like f in Condition A1. The advantage of splitting $\mathcal{F}$ into two parts f and $-g \circ \tau$ arises when the Legendre-Fenchel transform of g^* and the maximization of $f - y \circ \tau$, $y \in \mathcal{X}^*$, can be easily controlled. This is the raison d'être of Bogoliubov linearizations.

Bogoliubov linearizations. We associate to the function (130) the family of functions

$$\mathcal{G}_y \doteq f - y \circ \tau , \qquad y \in \mathcal{X}^* . \tag{131}$$

This leads in particular to the (nonempty) compact convex sets

$$E_{\mathcal{G}_y} \doteq \{\mu \in K : \mathcal{G}_y(\mu) = \sup \mathcal{G}_y(K)\} , \qquad y \in \mathcal{X}^* , \tag{132}$$

of associated *linear* equilibrium states. We call these new functions $\mathcal{G}_y$, $y \in \mathcal{X}^*$, *Bogoliubov linearizations* of $\mathcal{F}$.

The name "Bogoliubov linearization" used here is reminiscent to what is known in Quantum Statistical Mechanics as the Bogoliubov approximation or, sometimes, the approximating Hamiltonian method. It is a generic method that has been first used in theoretical physics, starting with Bogoliubov's microscopic superfluidity theory in 1947 [9]. It can be made mathematically rigorous, as shown in [10, 14, 15, 11, 12, 16, 18, 69, 43] for many cases in quantum physics, ranging from Bose gases to quantum-spin and lattice-fermion systems.

However, apart from [18], all the mathematical results obtained beyond perturbative arguments (with in particular (often dilute) limits of infinite particle numbers) on the Bogoliubov approximation concern only the computation of pressures (or free-energy densities), i.e., the calculation of infinite-volume limits of the logarithm of partition functions. As far as we know, the monograph [18] is the only study of the Bogoliubov approximation for equilibrium states. It thus provides results regarding all correlation functions (or all N-body density matrices) and solves in the case of lattice-fermion systems an old open problem that was first raised by Ginibre in 1968 [32]. This approach [18] has been significantly generalized here. Indeed, in the quantum framework, previously considered in [18], only the specific case of "Example 2.7[20]" is studied.

Exactness of Bogoliubov linearizations at the variational problem level. The relationship between the maximization of the function $\mathcal{F} : K \rightarrow \{-\infty\} \cup \mathbb{R}$ and its Bogoliubov linearizations $\mathcal{G}_y : K \rightarrow \{-\infty\} \cup \mathbb{R}$ is established in the sequel using the celebrated von Neumann minimax theorem (Theorem 3.24). In fact, they are related via the variational problem

$$\mathrm{P} \doteq \inf_{y \in \mathcal{X}^*} P_{\mathrm{NL}}(y)$$

and its set

$$\mathrm{M} \doteq \{y \in \mathcal{X}^* : P_{\mathrm{NL}}(y) = \mathrm{P}\} \tag{133}$$

of minimizers, where

$$P_{\mathrm{NL}}(y) \doteq P_{\mathrm{L}}(y) + g^*(y) , \qquad P_{\mathrm{L}}(y) \doteq \sup \mathcal{G}_y(K) , \qquad y \in \mathcal{X}^* .$$

We call the quantities $P_{\mathrm{L}}(y)$ and $P_{\mathrm{NL}}(y)$, $y \in \mathcal{X}^*$, respectively the linear and nonlinear approximating pressures associated with the Bogoliubov linearization $\mathcal{G}_y$. As $P_{\mathrm{L}}(y) \in \mathbb{R}$ for all $y \in \mathcal{X}^*$, P_{NL} defines a real-valued function on the nonempty convex set $\mathrm{dom}(g^*)$ and $P_{\mathrm{NL}}(x) = \infty$ when $x \notin \mathrm{dom}(g^*)$.

Note that $\mathrm{P} < \infty$, because $\mathrm{dom}(g^*) \neq \emptyset$. From Lemma 3.21 ($\flat$) (trivially adapted to the case $g_+^* = 0$) we can infer that Conditions A1–A2 imply $\mathrm{P} \in \mathbb{R}$, i.e., $\mathrm{P} > -\infty$. Under the same conditions, we meanwhile have $\sup \mathcal{F}(K) \in \mathbb{R}$, because $\mathcal{F}$ (130) is bounded from above, as it is upper semicontinuous, $\mathcal{F}(\mu) < \infty$ for all $\mu \in K$, and K is compact. In fact, the supremum of the function $\mathcal{F}$ (130) turns out to be nothing but the variational problem P:

[20] Adapted appropriately to the fermionic situation.

Proposition 3.2 (Exactness of Bogoliubov linearizations)
Under Conditions A1–A2,

$$\sup \mathcal{F}(K) = \mathrm{P} \doteq \inf_{y\in\mathcal{X}^*} P_{\mathrm{NL}}(y) \in \mathbb{R}$$

and there are $x \in \mathrm{M}$ *and* $\nu \in E_{\mathcal{G}_x}$ *such that* $x \in \partial g(\tau(\nu))$, $\nu \in E_{\mathcal{F}}$ *and*

$$x \circ \tau(\nu) - g^*(x) = \sup_{y\in\mathcal{X}^*} \{y \circ \tau(\nu) - g^*(y)\} \ . \tag{134}$$

Proof. We divide the proof into three steps:

Step 1: Recall that there is a radius $R_0 \in \mathbb{R}^+$ such that the (convex) set $B(0,R) \cap \mathrm{dom}(g^*) \subseteq \mathcal{X}^*$ is nonempty and weak*-closed for all $R \in [R_0, \infty)$ (see Condition A2). On the one hand, by Lemma 3.20 ($\flat$), there is $R_1 \in [R_0, \infty)$ such that, for all $R \in [R_1, \infty)$,

$$\mathrm{P} = \inf_{y\in B(0,R)\cap\mathrm{dom}(g^*)} P_{\mathrm{NL}}(y) < \inf_{y\in\mathcal{X}^*\backslash B(0,R)} P_{\mathrm{NL}}(y) \ . \tag{135}$$

On the other hand, by Equations (122) and (130) together with Condition A2 (cf. (127)) there is also $R_2 \in [R_0, \infty)$ such that, for all $R \in [R_2, \infty)$,

$$\sup \mathcal{F}(K) = \sup_{\mu\in K} \inf_{y\in B(0,R)\cap\mathrm{dom}(g^*)} \{f(\mu) - y \circ \tau(\mu) + g^*(y)\} \ . \tag{136}$$

Therefore, there is $R \in \mathbb{R}^+$ ($R \geq R_j$, $j = 0, 1, 2$) such that Equations (135)–(136) hold true and the (convex) set $B(0,R) \cap \mathrm{dom}(g^*) \subseteq \mathcal{X}^*$ is nonempty and weak*-closed.

Step 2: The space K is by assumption compact, convex and Hausdorff (see Condition A1). Observe that the set

$$N_0 \doteq \{\mu \in K : f(\mu) > -\infty\}$$

is weak*-closed and convex because the function f is by assumption upper semicontinuous and concave (see again Condition A1). In particular, N_0 is a compact convex Hausdorff subspace of K. Meanwhile, $B(0,R)$ is a (Hausdorff) weak*-compact convex subset dual vector space $\mathcal{X}^*$, thanks to the Banach-Alaoglu theorem [57, Theorem 3.15]. By taking a radius $R \in \mathbb{R}^+$ as in Step 1 we ensure that (135)–(136) hold true and the nonempty convex set

$$M_0 \doteq B(0,R) \cap \mathrm{dom}(g^*)$$

is weak*-closed, and therefore weak*-compact.

Step 3: Define the real-valued function $\mathfrak{f} : M_0 \times N_0 \to \mathbb{R}$ by

$$\mathfrak{f}(y,\mu) \doteq f(\mu) - y \circ \tau(\mu) + g^*(y) \in \mathbb{R} \ , \qquad y \in M_0,\ \mu \in N_0 \ .$$

For all $\mu \in N_0$, the mapping $y \mapsto \mathfrak{f}(y,\mu)$ is convex and weak*-lower semicontinuous, whereas, for all $y \in M_0$, the mapping $\mu \mapsto \mathfrak{f}(y,\mu)$ is concave and upper semicontinuous, since f is by assumption concave and upper semicontinuous and $\tau : K \to \mathcal{X}$ is continuous and affine (see Condition A1). By applying the von Neumann minimax theorem (Theorem 3.24), we conclude the existence of a saddle point $(x,\nu) \in M_0 \times N_0$ of the function $\mathfrak{f}$. Since the radius $R \in \mathbb{R}^+$ is taken so that (135)–(136) hold true, we infer from (136) that

$$\sup \mathcal{F}(K) = \sup_{\mu\in N_0} \inf_{y\in M_0} \mathfrak{f}(y,\mu) = \inf_{y\in M_0} \sup_{\mu\in N_0} \mathfrak{f}(y,\mu) = \inf_{y\in B(0,R)} P_{\mathrm{NL}}(y) \ .$$

By the definition of a saddle point, the pair (x,ν) satisfies $\nu \in E_{\mathcal{G}_x} \cap E_{\mathcal{F}}$, $x \in \mathrm{M}$ and

$$g \circ \tau(\nu) = \sup_{y\in\mathcal{X}^*} \{y \circ \tau(\nu) - g^*(y)\} = x \circ \tau(\nu) - g^*(x) \ .$$

In particular, since the function g never takes an infinite value (see Condition A2), we deduce from [59, Proposition 4.4.1] that $x \in \partial g(\tau(\nu)) \subseteq \mathcal{X}^*$. ■

Non-exactness of Bogoliubov linearizations at the level of equilibrium states. Under Conditions A1–A2, the function $\mathcal{G}_y$ is upper semicontinuous for all $y \in \mathcal{X}^*$ and therefore, there is $\nu_y \in K$ such that

$$\sup \mathcal{G}_y(K) = \mathcal{G}_y(\nu_y) \ . \tag{137}$$

In other words, $E_{\mathcal{G}_y} \neq \emptyset$ for all $y \in \mathcal{X}^*$. In particular, each minimizer $x \in \mathrm{M}$ (see (133)) leads to a nonempty set $E_{\mathcal{G}_x}$ of linear equilibrium states of the corresponding Bogoliubov linearization. These sets are expected to provide good candidates for the nonlinear equilibrium states of the function $\mathcal{F}$ (130). However, as shown in [18, Section 9.2] for fermion-lattice systems where K is the (metrizable) weak*-compact convex space of translation-invariant states on the CAR C^*-algebra of the infinite lattice, elements of the sets $E_{\mathcal{G}_x}$, $x \in \mathrm{M}$, are not necessarily nonlinear equilibrium states of the function (130), i.e., elements of $E_{\mathcal{F}}$. In particular, we cannot generally expect the inclusion $E_{\mathcal{G}_x} \subseteq E_{\mathcal{F}}$ for $x \in \mathrm{M}$. Instead, we only have the opposite inclusion, as the following proposition demonstrates:

Proposition 3.3 (Bogoliubov linearizations – Equilibrium states)
Under Conditions A1–A2,

$$E_{\mathcal{F}} \subseteq \bigcap_{x \in \mathrm{M}} E_{\mathcal{G}_x} \qquad \textit{and} \qquad \emptyset \neq \mathrm{M} \subseteq \bigcap_{\nu \in E_{\mathcal{F}}} \partial g(\tau(\nu)) .$$

Proof. We divide the proof into two steps:

Step 1: By Proposition 3.2, there is a continuous linear functional $x \in \mathrm{M}$, along with an equilibrium state $\nu \in E_{\mathcal{G}_x} \cap E_{\mathcal{F}}$, satisfying (134). Using (122) we thus observe that

$$\begin{aligned} \sup \mathcal{G}_x(K) &= f(\nu) - x \circ \tau(\nu) \qquad (138) \\ &= f(\nu) + \inf_{y \in \mathcal{X}^*} \{-y \circ \tau(\nu) + g^*(y)\} - g^*(x) \\ &= f(\nu) - g \circ \tau(\nu) - g^*(x) \\ &= \sup \mathcal{F}(K) - g^*(x) \ . \qquad (139) \end{aligned}$$

Take now any nonlinear equilibrium state $\tilde{\nu} \in E_{\mathcal{F}}$. Then, going backwards from (139) to (138) we obtain that

$$\mathcal{F}(\tilde{\nu}) - g^*(x) = \sup \mathcal{G}_x(K) \geq \mathcal{G}_x(\tilde{\nu}) \ .$$

Keeping in mind (122), (130) and (131), we deduce from the last inequality that

$$g \circ \tau(\tilde{\nu}) = \sup_{y \in \mathcal{X}^*} \{y \circ \tau(\tilde{\nu}) - g^*(y)\} \leq x \circ \tau(\tilde{\nu}) - g^*(x) \ .$$

It follows that the continuous linear functional $x \in \mathrm{M}$ is always solution to

$$g \circ \tau(\tilde{\nu}) = \sup_{y \in \mathcal{X}^*} \{y \circ \tau(\tilde{\nu}) - g^*(y)\} = x \circ \tau(\tilde{\nu}) - g^*(x) \tag{140}$$

for all nonlinear equilibrium states $\tilde{\nu} \in E_{\mathcal{F}}$. By going again backwards from (139) to (138) with $\tilde{\nu}$ in place of ν, along with (140), we conclude that, for any $\tilde{\nu} \in E_{\mathcal{F}}$,

$$\sup \mathcal{G}_x(K) = f(\tilde{\nu}) - x \circ \tau(\tilde{\nu})$$

and, hence, $E_{\mathcal{F}} \subseteq E_{\mathcal{G}_x}$.

Step 2: Take now any continuous linear functional $\tilde{x} \in \mathrm{M}$. By Proposition 3.2, for any nonlinear equilibrium state $\tilde{\nu} \in E_{\mathcal{F}}$,

$$\mathcal{F}(\tilde{\nu}) = \sup \mathcal{F}(K) = \sup \mathcal{G}_{\tilde{x}}(K) + g^*(\tilde{x}) \geq \mathcal{G}_{\tilde{x}}(\tilde{\nu}) + g^*(\tilde{x}) \ .$$

Via Equations (122), (130) and (131), we thus get that, for all $\tilde{\nu} \in E_{\mathcal{F}}$,

$$g \circ \tau(\tilde{\nu}) = \sup_{y \in \mathcal{X}^*} \{y \circ \tau(\tilde{\nu}) - g^*(y)\} = \tilde{x} \circ \tau(\tilde{\nu}) - g^*(\tilde{x}) \ .$$

Since the function g never takes an infinite value (cf. Condition A2), we deduce from [59, Proposition 4.4.1] and the last equation that $\mathrm{M} \subseteq \partial g(\tau(\tilde{\nu}))$ for all $\tilde{\nu} \in E_{\mathcal{F}}$. From the above equalities in this step, we also get that, for all $\tilde{\nu} \in E_{\mathcal{F}}$,

$$\mathcal{F}(\tilde{\nu}) = \sup \mathcal{G}_{\tilde{x}}(K) + g^*(\tilde{x}) = \sup \mathcal{G}_{\tilde{x}}(K) - g \circ \tau(\tilde{\nu}) + \tilde{x} \circ \tau(\tilde{\nu}) \ ,$$

from which we conclude that

$$\sup \mathcal{G}_{\tilde{x}}(K) = \mathcal{F}(\tilde{\nu}) + g \circ \tau(\tilde{\nu}) - \tilde{x} \circ \tau(\tilde{\nu}) = f(\tilde{\nu}) - \tilde{x} \circ \tau(\tilde{\nu}) \ .$$

In other words, $E_{\mathcal{F}} \subseteq E_{\mathcal{G}_{\tilde{x}}}$ for all $\tilde{x} \in \mathrm{M}$. ■

The next corollary shows that the elements of $E_{\mathcal{F}}$ are in fact elements of the sets $E_{\mathcal{G}_x}$, $x \in \mathrm{M}$, that satisfy a self-consistency condition:

Corollary 3.4 (Bogoliubov linearizations and self-consistency)
Under Conditions A1–A2,

$$E_{\mathcal{F}} = \{\nu \in E_{\mathcal{G}_x} : x \in \partial g(\tau(\nu))\} \doteq E_{\mathcal{G}_x}^{\mathrm{sc}} \ , \qquad x \in \mathrm{M} \ .$$

In particular, $E_{\mathcal{G}_x}^{\mathrm{sc}} \equiv E_{\mathcal{G}.}^{\mathrm{sc}}$ is a nonempty, compact, convex set which does not depend upon the choice of $x \in \mathrm{M}$. (For this reason the notation $E_{\mathcal{G}.}^{\mathrm{sc}}$ is used.)

Proof. Under Condition A1, recall that the function $\mathcal{G}_y$ is upper semicontinuous for all $y \in \mathcal{X}^*$. So, for all $x \in \mathrm{M}$, there is a solution to the variational problem (137) for $y = x$. So take an arbitrary solution $\nu_x \in E_{\mathcal{G}_x}$ to (137). By Proposition 3.2 and Equation (122) it follows that, for all $x \in \mathrm{M}$,

$$\begin{aligned} f(\nu_x) - x \circ \tau(\nu_x) + g^*(x) &= \sup \mathcal{F}(K) \\ &\geq f(\nu_x) - g \circ \tau(\nu_x) = f(\nu_x) + \inf_{y \in \mathcal{X}^*} \{-y \circ \tau(\nu_x) + g^*(y)\} \ . \end{aligned}$$

The above inequality is satisfied with equality iff

$$x \circ \tau(\nu_x) - g^*(x) = \sup_{y \in \mathcal{X}^*} \{y \circ \tau(\nu_x) - g^*(y)\} < \infty \ ,$$

meaning that $\tau(\nu_x) \in \mathrm{dom}(g)$ and $x \in \partial g(\tau(\nu_x))$, thanks to [59, Proposition 4.4.1]. Note that $g^*(x)$ is a real number because $x \in \mathrm{M}$. Thus, for all $x \in \mathrm{M}$, $E_{\mathcal{G}_x}^{\mathrm{sc}} \subseteq E_{\mathcal{F}}$. Now, by Proposition 3.3, for any $\nu \in E_{\mathcal{F}}$ and $x \in \mathrm{M}$, one has that $\nu \in E_{\mathcal{G}_x}$ and, thus, thanks again to Proposition 3.2 and Equation (122),

$$f(\nu) - x \circ \tau(\nu) + g^*(x) = f(\nu) - g \circ \tau(\nu) = f(\nu) + \inf_{y \in \mathcal{X}^*} \{-y \circ \tau(\nu) + g^*(y)\} \ .$$

Hence,

$$x \circ \tau(\nu) - g^*(x) = \sup_{y \in \mathcal{X}^*} \{y \circ \tau(\nu) - g^*(y)\} < \infty \ ,$$

and, by [59, Proposition 4.4.1], $x \in \partial g(\tau(\nu))$. Note in particular that, for all $x \in \mathrm{M}$, $E_{\mathcal{G}_x}^{\mathrm{sc}}$ is nonempty, as $E_{\mathcal{F}}$ is nonempty and $E_{\mathcal{G}_x}^{\mathrm{sc}} = E_{\mathcal{F}}$. ■

The (nonempty) compact convex set $E_{\mathcal{G}.}^{\mathrm{sc}}$ is called here the set of *self-consistent equilibrium states* of Bogoliubov linearizations $\mathcal{G}_x$, $x \in \mathrm{M}$. This terminology refers to the paradigmatic example $K = \mathcal{P}(T)$. See the discussions following Theorem 2.17.

3.3 Bogoliubov linearizations for interactions with concave and convex components

General assumptions. In all this subsection, we consider the following assumptions:

B1 $\mathcal{X}_\pm$ are two real normed spaces, K is a compact convex Hausdorff space, $\tau_\pm : K \to \mathcal{X}_\pm$ are two continuous affine transformations and $f : K \to \{-\infty\} \cup \mathbb{R}$ is an upper semicontinuous concave function (which is of course not identically equal to $-\infty$).

B2 $g_- : \mathcal{X}_- \to \mathbb{R}$ is a lower semicontinuous and convex function for which there is a positive radius $R_0 \in \mathbb{R}^+$ such that $B_-(0,R) \cap \mathrm{dom}\left(g_-^*\right)$ is nonempty and weak*-closed for all $R \in [R_0, \infty)$. Moreover, g_-^* has minimal linear growth $\|\tau_-\|_\infty$ in the sense of Equation (124).

B3 $g_+ : \mathcal{X}_+ \to \mathbb{R}$ is a lower semicontinuous, bounded and convex function for which the Legendre-Fenchel transform g_+^* has full domain, i.e., $\mathrm{dom}(g_+^*) = \mathcal{X}_+^*$, and has minimal linear growth $\|\tau_+\|_\infty$ in the sense of Equation (124).

Above, $B_\pm(0,R) \subseteq \mathcal{X}_\pm^*$ are the weak*-closed ball of radius $R \in \mathbb{R}^+$ and center $0 \in \mathcal{X}_\pm^*$. Note that the rather technical condition $\mathrm{dom}(g_+^*) = \mathcal{X}_+^*$ in Condition B3 is taken to ensure that the difference $g_-^* \circ \tau_-(\mu) - g_+^* \circ \tau_+(\mu)$, $\mu \in K$, which is an important object here, is always well-defined as an element of $\mathbb{R} \cup \{\infty\}$. It essentially says that $g_+(x)$ grows faster than the norm $\|x\|_{\mathcal{X}}$, as $\|x\|_{\mathcal{X}} \to \infty$.

Again, we call the elements of K *(generalized) states*. In this subsection we study the sets

$$M_{\mathbb{F}} \doteq \{\mu \in K : \mathbb{F}(\mu) = \sup \mathbb{F}(K)\} \tag{141}$$

and

$$E_{\mathbb{F}} \doteq \left\{\mu \in K : \exists (\mu_j)_{j\in J} \subseteq K \text{ with } \lim_J \mu_j = \mu \text{ and } \lim_J \mathbb{F}(\mu_j) = \sup \mathbb{F}(K)\right\} , \tag{142}$$

of *nonlinear equilibrium states* of the (Borel-measurable) function $\mathbb{F} : K \to \{-\infty\} \cup \mathbb{R}$ defined by

$$\mathbb{F} \doteq f - g_- \circ \tau_- + g_+ \circ \tau_+ . \tag{143}$$

Note from Conditions B1–B3 that $\mathbb{F}$ is bounded from above but **neither upper semicontinuous nor concave**, and so, in contrast to the (purely) concave case ($E_{\mathcal{F}}$), the set $M_{\mathbb{F}}$ is not necessarily convex anymore. Furthermore, the set $M_{\mathbb{F}}$ could a priori be empty. By contrast, the set $E_{\mathbb{F}} \supseteq M_{\mathbb{F}}$ is necessarily nonempty.

All elements of $E_{\mathbb{F}}$, including those of $M_{\mathbb{F}}$ if they exist, are called *nonlinear equilibrium states*. In fact, as we prove below, the two sets turn out to be equal. This may seem very surprising, given that we are maximizing a function $\mathbb{F}$ that is not upper semicontinuous.

Remark 3.5 (Decomposition as a difference of semicontinuous functions)
The decomposition property of a function as the difference of lower semicontinuous ones refers to the set of point-wise limits of a sequence $(f_n)_{n\in\mathbb{N}}$ of continuous functions $K \to \mathbb{R}$ such that

$$\sum_{n\in\mathbb{N}} |f_{n+1}(\mu) - f_n(\mu)| < \infty , \qquad \mu \in K .$$

If this inequality is uniform with respect to $\mu \in K$, the resulting lower semicontinuous functions are bounded. Such classes of functions, defined as differences of semicontinuous ones, are for example discussed in [35] for compact metric spaces K. However, this set of functions does not generally exhaust all point-wise limits of continuous functions $K \to \mathbb{R}$, which, for separable and metrizable K, form the Baire-1 class, by the Lebesgue-Hausdorff-Banach theorem [39, (24.10)]; see more generally [39, Chapter 24], in particular [39, (24.3)] for (possibly non separable) metrizable K.

Remark 3.6 (Decomposition as a difference of convex functions)
In various situations, very general functions can be represented as the difference of convex functions. For example, any Gateaux-differentiable function with domain in a Hilbert space, whose gradient mapping is Lipschitz-continuous, can be written[21] *as a difference of two convex Gateaux-differentiable functions. This is in particular true for C^1-functions on any compact subset of $\mathbb{R}^N$ ($N \in \mathbb{N}$). Compare with [25] and Section 2.6.*

Bogoliubov linearizations. Similar to the concave case, we associate to the function (143) the family of functions

$$\mathcal{G}_{y_+,y_-} \doteq f - y_- \circ \tau_- + y_+ \circ \tau_+, \qquad y_\pm \in \mathcal{X}_\pm^* \, . \tag{144}$$

This object generalizes the functions $\mathcal{G}_y$ defined by Equation (131) and leads to the (nonempty) compact convex set

$$E_{\mathcal{G}_{y_+,y_-}} \doteq \left\{ \mu \in K : \mathcal{G}_{y_+,y_-}(\mu) = \sup \mathcal{G}_{y_+,y_-}(K) \right\}$$

of *linear* equilibrium states of $\mathcal{G}_{y_+,y_-}$ for all $y_\pm \in \mathcal{X}_\pm^*$. The functions $\mathcal{G}_{y_+,y_-}$, $y_\pm \in \mathcal{X}_\pm^*$, are again called *Bogoliubov linearizations* of $\mathbb{F}$.

Exactness of Bogoliubov linearizations. Similar to the concave case, the maximization of the function $\mathbb{F}$ is related to Bogoliubov linearizations via the variational problem

$$\mathrm{P}^\flat \doteq \sup_{y_+ \in \mathcal{X}_+^*} \inf_{y_- \in \mathcal{X}_-^*} P_{\mathrm{NL}}(y_+, y_-) \ ,$$

where, for any continuous linear functionals $y_\pm \in \mathcal{X}_\pm^*$,

$$P_{\mathrm{NL}}(y_+, y_-) \doteq P_{\mathrm{L}}(y_+, y_-) + g_-^*(y_-) - g_+^*(y_+) \ , \tag{145}$$

$$P_{\mathrm{L}}(y_+, y_-) \doteq \sup \mathcal{G}_{y_+,y_-}(K) \ . \tag{146}$$

As in the previous case, given $y_\pm \in \mathcal{X}_\pm^*$ we call the quantities $P_{\mathrm{L}}(y_+, y_-)$ and $P_{\mathrm{NL}}(y_+, y_-)$ respectively the linear and nonlinear approximating pressures associated with the Bogoliubov linearization $\mathcal{G}_{y_+,y_-}$. Conditions B1–B3 imply that $\mathcal{X}_+^* \times \mathrm{dom}(g_-^*)$ is a nonempty convex subset of $\mathcal{X}_+^* \times \mathcal{X}_-^*$ and

$$P_{\mathrm{NL}}\left(\mathcal{X}_+^* \times \mathrm{dom}(g_-^*)\right) \subseteq \mathbb{R} \, ,$$

while $P_{\mathrm{NL}}(y_+, y_-) = \infty$ when $y_- \notin \mathrm{dom}(g_-^*)$.

It is convenient to define the function $P^\flat : \mathcal{X}_+^* \to \{-\infty\} \cup \mathbb{R}$ by

$$P^\flat(y_+) \doteq \inf P_{\mathrm{NL}}\left(y_+, \mathcal{X}_-^*\right) \ , \qquad y_+ \in \mathcal{X}_+^* \, , \tag{147}$$

along with the following sets of optimizers:

$$\mathrm{M}^\flat \doteq \left\{ x_+ \in \mathcal{X}_+^* : P^\flat(x_+) = \sup P^\flat\left(\mathcal{X}_+^*\right) \doteq \mathrm{P}^\flat \right\} \subseteq \mathcal{X}_+^* \tag{148}$$

and, for all fixed continuous linear functionals $y_+ \in \mathcal{X}_+^*$,

$$M^\flat(y_+) \doteq \left\{ x_- \in \mathcal{X}_-^* : P_{\mathrm{NL}}(y_+, x_-) = P^\flat(y_+) \right\} \subseteq \mathcal{X}_-^* \, , \tag{149}$$

which are associated with the solutions to the variational problems $\mathrm{P}^\flat$ and $P^\flat(y_+)$, respectively. The study of variational problems $P^\flat(y_+)$, $y_+ \in \mathcal{X}_+^*$, and $\mathrm{P}^\flat$, including the sets $M^\flat(y_+)$, $y_+ \in \mathcal{X}_+^*$, and $\mathrm{M}^\flat$ of optimizers is postponed to Section 3.6 in order to reduce the technical arguments and focus on the main result of this section. We therefore only give the essential properties of these variational problems in the next proposition:

[21] For such a function f, write $f(x) = (f(x) + C\|x\|^2) - C\|x\|^2$ for a sufficiently large constant $C \in \mathbb{R}^+$.

Proposition 3.7 (Properties of variational problems ($\flat$))
Under Conditions B1–B3 the following assertions hold:
(i) *For all $y_+ \in \mathcal{X}_+^*$, $P^\flat(y_+) \in \mathbb{R}$ and $M^\flat(y_+)$ is nonempty, convex and weak*-compact. There is $R \in [R_0, \infty)$ such that*

$$M^\flat(y_+) \subseteq B_-(0,R) \cap \mathrm{dom}(g_-^*) , \qquad y_+ \in \mathcal{X}_+^* .$$

(ii) $\mathrm{P}^\flat \in \mathbb{R}$ *and* $\mathrm{M}^\flat \subseteq \mathcal{X}_+^*$ *is nonempty, norm-bounded and weak*-compact.*

Proof. Assertion (i) refers to Lemmata 3.21 ($\flat$) and 3.22 ($\flat$). Assertion (ii) corresponds to Lemma 3.23 ($\flat$). ■

We are now in a position to examine the exactness of Bogoliubov linearizations both at the level of the variational problem and at the level of the equilibrium states. This brings us to our next theorem, which is the central result of the present paper and can be applied to very general situations, such as quantum lattice systems or classical dynamical systems:

Theorem 3.8 (Exactness of Bogoliubov linearizations)
Under Conditions B1–B3 the following assertions hold:
(i) *Nonlinear pressure:*

$$\sup \mathbb{F}(K) = \mathrm{P}^\flat \doteq \sup_{y_+ \in \mathcal{X}_+^*} \inf_{y_- \in \mathcal{X}_-^*} P_{\mathrm{NL}}(y_+, y_-) \in \mathbb{R} .$$

(ii) *Self-consistency conditions: For any $x_+ \in \mathrm{M}^\flat$ and $x_- \in M^\flat(x_+)$, the set*

$$E_{\mathcal{G}_{x_+,x_-}}^{\mathrm{sc}} \doteq \left\{ \nu \in E_{\mathcal{G}_{x_+,x_-}} : x_- \in \partial g_-(\tau_-(\nu)) \right\} \equiv E_{\mathcal{G}_{x_+,\cdot}}^{\mathrm{sc}}$$

of self-consistent equilibrium states of the Bogoliubov linearization $\mathcal{G}_{x_+,x_-}$ is nonempty, convex and compact, and does not depend upon the choice of $x_- \in M^\flat(x_+)$. (That is the raison for the notation $E_{\mathcal{G}_{x_+,\cdot}}^{\mathrm{sc}}$.) Furthermore, for all $x_+ \in \mathrm{M}^\flat$,

$$E_{\mathcal{G}_{x_+,\cdot}}^{\mathrm{sc}} \subseteq \{\nu \in K : x_+ \in \partial g_+(\tau_+(\nu))\} .$$

(iii) *Nonlinear equilibrium states: $E_{\mathbb{F}} = M_{\mathbb{F}}$ is compact and the union of all above sets of self-consistent equilibrium states, that is,*

$$E_{\mathbb{F}} = M_{\mathbb{F}} = \bigcup_{x_+ \in \mathrm{M}^\flat} E_{\mathcal{G}_{x_+,\cdot}}^{\mathrm{sc}} .$$

If the function $g_+ : \mathcal{X}_+ \to \mathbb{R}$ is additionally Gateaux-differentiable, then the above union is disjoint.

(iv) *If the (concave) function $f : K \to \mathbb{R}$ defining $\mathbb{F}$ (see Equation (143)) is affine and all Bogoliubov linearizations $\mathcal{G}_{x_+,x_-}$ have one single linear equilibrium state, then $E_{\mathbb{F}} = M_{\mathbb{F}} \subseteq K$ is a subset of extreme points of the compact convex Hausdorff space K.*

Proof. Many assertions need to be proven. So, we divide our arguments into several steps:

Step 1: By Equation (122), observe that

$$\begin{aligned} \sup \mathbb{F}(K) &= \sup_{\mu \in K} \left\{ f(\mu) - g_- \circ \tau_-(\mu) + \sup_{y_+ \in \mathcal{X}_+^*} \{ y_+ \circ \tau_+(\mu) - g_+^*(y_+) \} \right\} \\ &= \sup_{y_+ \in \mathcal{X}_+^*} \left\{ \sup_{\mu \in K} \{ f(\mu) + y_+ \circ \tau_+(\mu) - g_- \circ \tau_-(\mu) \} - g_+^*(y_+) \right\} , \end{aligned} \tag{150}$$

and a solution in $\mathcal{X}_+^* \times K$ to the variational problems (if it exists) does not depend on the order in which the suprema are taken. Note that $f + y_+ \circ \tau_+$ is a concave and upper semicontinuous function on K for any $y_+ \in \mathcal{X}_+^*$. Hence, we can apply Proposition 3.2 for $f + y_+ \circ \tau_+$ in place of f to deduce from (150) Assertion (i).

Step 2: Similarly, we can apply Corollary 3.4 to arrive at Assertion (ii), except for the property $x_+ \in \partial g_+ (\tau_+ (\nu))$ for all $x_+ \in \mathrm{M}^\flat$ and $\nu \in E^{\mathrm{sc}}_{\mathcal{G}_{x_+,\cdot}}$, which is proven as follows: Take any $x_+ \in \mathrm{M}^\flat$ and $\nu \in E^{\mathrm{sc}}_{\mathcal{G}_{x_+,\cdot}}$. Then, using Assertion (i), along with Corollary 3.4 for $f + x_+ \circ \tau_+$, we deduce that $\nu \in E_{\mathbb{F}}$ and

$$f(\nu) - g_- \circ \tau_- (\nu) + g_+ \circ \tau_+ (\nu) = f(\nu) - g_- \circ \tau_- (\nu) + x_+ \circ \tau_+ (\nu) - g_+^* (x_+) \ .$$

In other words, by Equation (122), one arrives at the equalities

$$g_+ \circ \tau_+ (\nu) = \sup_{y_+ \in \mathcal{X}_+^*} \left\{ y_+ \circ \tau_+ (\nu) - g_+^* (y_+) \right\} = x_+ \circ \tau_+ (\nu) - g_+^* (x_+) \ .$$

Using again [59, Proposition 4.4.1] we conclude that $x_+ \in \partial g_+ (\tau_+ (\nu))$.

Step 3: By Proposition 3.7 (ii), the supremum in (150) with respect to $y_+ \in \mathcal{X}_+^*$ can be restricted to a closed ball $B_+(0, R)$, which is always weak*-compact (cf. the Banach-Alaoglu theorem [57, Theorem 3.15]). Then, by Conditions B1–B3, observe that the mapping

$$(y_+, \mu) \mapsto f(\mu) + y_+ \circ \tau_+ (\mu) - g_- \circ \tau_- (\mu) - g_+^* (y_+)$$

defined on $B_+(0, R) \times K$ is upper semicontinuous with respect to the product topology when $B_+(0, R)$ is equipped with the weak* topology. Therefore, there are solutions in $B_+(0, R) \times K$ to the variational problem (150) and the set of all such solutions is compact, because K and $B_+(0, R)$ are both compact. In particular, $M_{\mathbb{F}}$ is nonempty. Furthermore, for any $\mu \in M_{\mathbb{F}}$, there is $x_+ \in B_+(0, R)$ such that (x_+, μ) is solution to (150) and all solutions to (150) have this form. As the projection $(x_+, \mu) \mapsto \mu$ is continuous, we conclude that $M_{\mathbb{F}}$ is compact.

Step 4: Take any net $(\mu_j)_{j \in J} \subseteq K$ converging to $\mu \in E_{\mathbb{F}}$ such that

$$\lim_J \mathbb{F} (\mu_j) = \sup \mathbb{F} (K) \ .$$

Because of Condition B3, there is some fixed radius $R \in \mathbb{R}^+$ such that, for any $j \in J$, we have a solution $x_{j,+} \in B_+(0, R)$ in this ball such that

$$\sup_{y_+ \in \mathcal{X}_+^*} \left\{ y_+ \circ \tau_+ (\mu_j) - g_+^* (y_+) \right\} = x_{j,+} \circ \tau_+ (\mu_j) - g_+^* (x_{j,+}) \ .$$

By the weak* compactness of $B_+(0, R)$, we can assume that $(x_{j,+})_{j \in J}$ converges to some $x_+ \in B_+(0, R)$. In particular, via Equation (150), Conditions B1–B2 and the weak*-lower semicontinuity of g_+^*, we conclude that

$$\sup \mathbb{F} (K) = \lim_J \mathbb{F} (\mu_j) \leq f(\mu) - g_- \circ \tau_- (\mu) + x_+ \circ \tau_+ (\mu) - g_+^* (x_+)$$

and (x_+, μ) is a solution to (150), which yields $\mu \in M_{\mathbb{F}}$. In other words, $E_{\mathbb{F}} \subseteq M_{\mathbb{F}} \subseteq E_{\mathbb{F}}$.

Step 5: We already prove in Steps 3 and 4 that $E_{\mathbb{F}} = M_{\mathbb{F}}$ is compact. Therefore, having (150) in mind (cf. Steps 1, 3 and 4), we can again apply Corollary 3.4 to arrive at Assertion (iii), except for the disjointness of the sets $E^{\mathrm{sc}}_{\mathcal{G}_{x_+,\cdot}}$, $x_+ \in \mathrm{M}^\flat$, when the convex function g_+ is additionally Gateaux--differentiable. This last property is proven as follows: If g_+ is Gateaux-differentiable then the subdifferential $\partial g_+ (\tau_+ (\nu))$, $\nu \in E_{\mathbb{F}}$, contains at most one point. Hence, in this case, for $x_+, x'_+ \in \mathrm{M}^\flat$,

$x_+ \neq x'_+$, and $\nu \in E^{\mathrm{sc}}_{\mathcal{G}_{x_+,\cdot}}$, $\nu' \in E^{\mathrm{sc}}_{\mathcal{G}_{x'_+,\cdot}}$, one necessarily has that $\nu \neq \nu'$, that is, $E^{\mathrm{sc}}_{\mathcal{G}_{x_+,\cdot}}$ and $E^{\mathrm{sc}}_{\mathcal{G}_{x'_+,\cdot}}$ are disjoint.

Step 6: If f is affine then all Bogoliubov linearizations $\mathcal{G}_{x_+,x_-}$ are affine, see Equation (144). So, if they all have a single linear equilibrium state, then these states have to be extreme points of K. As $E^{\mathrm{sc}}_{\mathcal{G}_{x_+,\cdot}} \subseteq E_{\mathcal{G}_{x_+,x_-}}$, $x_- \in M^\flat(x_+)$, it follows in this case that $E_{\mathbb{F}} \subseteq K$ is a subset of extreme points of the convex set K. In other words, we obtain Assertion (iv). ■

The above assumptions on g_+ and g_- in Theorem 3.8 exclude the cases where one of these functions is zero. However, if one of them is zero, or both, then Theorem 3.8 still holds true, mutatis mutandis:

- If both g_- and g_+ are zero then $\mathbb{F} = f$, that is, $E_{\mathbb{F}}$ is nothing but the set of maximizers of f and Theorem 3.8 is useless in this case.

- If $g_+ = 0$ and $g_- \neq 0$ then one has the (purely) concave case and, therefore, $E_{\mathbb{F}} = M_{\mathbb{F}} = E_{\mathcal{F}}$ for $g = g_-$ and $\tau = \tau_-$ in (130), which defines the function $\mathcal{F}$. Thus, in this case, Theorem 3.8 (i) and (ii)–(iii) correspond to Proposition 3.2 and Corollary 3.4, respectively. If $f : K \to \mathbb{R}$ is affine and $E_{\mathcal{G}_x}$, $x \in \mathrm{M}$, (see (132) and (133)) are singletons, then $E_{\mathbb{F}} = M_{\mathbb{F}} = E_{\mathcal{F}}$ is a set of extreme points of K.

- If $g_+ \neq 0$ and $g_- = 0$, there is no infimum, only the two suprema, in Theorem 3.8 (i), that is,

$$\sup \mathbb{F}(K) = \mathrm{P}^\flat \doteq \sup_{y_+ \in \mathcal{X}_+^*} \left\{ P_{\mathrm{L}}(y_+, 0) - g_+^*(y_+) \right\} \in \mathbb{R} .$$

  In this case, Condition B2 is irrelevant. As for (ii) one defines the function $P^\flat : \mathcal{X}_+^* \to \{-\infty\} \cup \mathbb{R}$ now by

$$P^\flat(y_+) \doteq \sup \mathcal{G}_{y_+,0}(K) - g_+^*(y_+) , \qquad y_+ \in \mathcal{X}_+^* ,$$

  along with the corresponding set of maximizers:

$$\mathrm{M}^\flat \doteq \left\{ x \in \mathcal{X}_+^* : P^\flat(x) = \sup P^\flat\left(\mathcal{X}_+^*\right) \doteq \mathrm{Q}^\flat \right\} \subseteq \mathcal{X}_+^* .$$

  Then, for all $x_+ \in \mathrm{M}^\flat$ and $\nu \in E_{\mathcal{G}_{x_+,0}}$, $x_+ \in \partial g_+(\tau_+(\nu))$. In other words, *all* linear equilibrium states of the Bogoliubov linearizations $\mathcal{G}_{x_+,0}$, $x_+ \in \mathrm{M}^\flat$, are self-consistent in this case. Again, $E_{\mathbb{F}} = M_{\mathbb{F}}$ is compact and

$$E_{\mathbb{F}} = M_{\mathbb{F}} = \bigcup_{x_+ \in \mathrm{M}^\flat} E_{\mathcal{G}_{x_+,0}} .$$

  For Gateaux-differentiable g_+, the above union is disjoint. If $f : K \to \mathbb{R}$ is affine and $E_{\mathcal{G}_{x_+,0}}$, $x_+ \in \mathrm{M}^\flat$, are singletons, then $E_{\mathbb{F}} = M_{\mathbb{F}}$ is a set of extreme points of K.

3.4 Decision rules of the thermodynamic game

The nonlinear approximating pressure P_{NL} of Equation (145) can be seen as the payoff function of a two-person zero-sum game. The conservative values of such a game are then the real quantities

$$\mathrm{P}^\flat \doteq \sup_{y_+ \in \mathcal{X}_+^*} \inf_{y_- \in \mathcal{X}_-^*} P_{\mathrm{NL}}(y_+, y_-) \quad \text{and} \quad \mathrm{P}^\sharp \doteq \inf_{y_- \in \mathcal{X}_-^*} \sup_{y_+ \in \mathcal{X}_+^*} P_{\mathrm{NL}}(y_+, y_-) . \tag{151}$$

This game is named here the *thermodynamic game* associated with f, $g_\pm$ and $\tau_\pm$. In all this subsection, we only consider non-zero functions $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$. If one of them is zero, we have essentially the same results (mutatis mutandis). See, e.g., the discussions at the end of Section 3.3.

Having in mind the thermodynamic game, in particular the conservative values of Equation (151), it is natural to consider the function

$$P^{\sharp}\left(y_{-}\right) \doteq \sup P_{\mathrm{NL}}\left(\mathcal{X}_{+}^{*}, y_{-}\right) , \qquad y_{-} \in \mathcal{X}_{-}^{*} , \tag{152}$$

along with $P^{\flat} : \mathcal{X}_{+}^{*} \to \{-\infty\} \cup \mathbb{R}$ already defined by (147), that is,

$$P^{\flat}\left(y_{+}\right) \doteq \inf P_{\mathrm{NL}}\left(y_{+}, \mathcal{X}_{-}^{*}\right) , \qquad y_{+} \in \mathcal{X}_{+}^{*} . \tag{153}$$

The subsets of solutions to the variational problems $P^{\flat}(y_{+})$ and $P^{\sharp}(y_{-})$ for all $y_{\pm} \in \mathcal{X}_{\pm}^{*}$, i.e.,

$$M^{\flat}\left(y_{+}\right) \doteq \left\{x_{-} \in \mathcal{X}_{-}^{*} : P^{\flat}\left(y_{+}\right) = P_{\mathrm{NL}}\left(y_{+}, x_{-}\right)\right\} \subseteq \mathcal{X}_{-}^{*} , \tag{154}$$

$$M^{\sharp}\left(y_{-}\right) \doteq \left\{x_{+} \in \mathcal{X}_{+}^{*} : P^{\sharp}\left(y_{-}\right) = P_{\mathrm{NL}}\left(x_{+}, y_{-}\right)\right\} \subseteq \mathcal{X}_{+}^{*} , \tag{155}$$

are important, along with the sets

$$\mathrm{M}^{\flat} \doteq \left\{x_{+} \in \mathcal{X}_{+}^{*} : \mathrm{P}^{\flat} = P^{\flat}\left(x_{+}\right)\right\} \subseteq \mathcal{X}_{+}^{*} , \tag{156}$$

$$\mathrm{M}^{\sharp} \doteq \left\{x_{-} \in \mathcal{X}_{-}^{*} : \mathrm{P}^{\sharp} \doteq P^{\sharp}\left(x_{-}\right)\right\} \subseteq \mathcal{X}_{-}^{*} , \tag{157}$$

which refer to optimal strategies for the thermodynamic game. The sets $\mathrm{M}^{\flat}$ and $M^{\flat}(y_{+})$, $y_{+} \in \mathcal{X}_{+}^{*}$, are already defined in (148)–(149) and their precise definition are recalled here for convenience.

The study of the variational problems $P^{\flat}(y_{+})$ and $P^{\sharp}(y_{-})$ ($y_{\pm} \in \mathcal{X}_{\pm}^{*}$), as well as $\mathrm{P}^{\flat}$ and $\mathrm{P}^{\sharp}$, including also the sets $M^{\flat}(y_{+})$, $M^{\sharp}(y_{-})$, $\mathrm{M}^{\flat}$ and $\mathrm{M}^{\sharp}$ of optimizers, is carried out in Section 3.6. The essential properties of case ($\flat$) are summarized in Proposition 3.7 above: $P^{\flat}(y_{+}) \in \mathbb{R}$ and $M^{\flat}(y_{+})$ is nonempty, convex and weak*-compact and uniformly norm bounded for all $y_{+} \in \mathcal{X}_{+}^{*}$; $\mathrm{P}^{\flat} \in \mathbb{R}$ and $\mathrm{M}^{\flat} \subseteq \mathcal{X}_{+}^{*}$ is also nonempty, norm-bounded and weak*-compact. We now give a similar statement for the case ($\sharp$):

Proposition 3.9 (Properties of variational problems ($\sharp$))
Under Conditions B1–B3 the following assertions hold:
(i) *For all $y_{-} \in \mathrm{dom}(g_{-}^{*}) \subseteq \mathcal{X}_{-}^{*}$, $P^{\sharp}(y_{-}) \in \mathbb{R}$ and the set $M^{\sharp}(y_{-})$ is nonempty and weak*-compact. There is $R \in \mathbb{R}^{+}$ such that*

$$M^{\sharp}(y_{-}) \subseteq B_{+}(0, R) , \qquad y_{-} \in \mathrm{dom}(g_{-}^{*}) \in \mathcal{X}_{-}^{*} .$$

(ii) $\mathrm{P}^{\sharp} \in \mathbb{R}$ *and the set* $\mathrm{M}^{\sharp} \subseteq \mathcal{X}_{-}^{*}$ *is a nonempty, convex, norm-bounded and weak*-compact subset of* $\mathrm{dom}(g_{-}^{*})$.

Proof. Assertion (i) refers to Lemmata 3.21 ($\sharp$) and 3.22 ($\sharp$). Assertion (ii) corresponds to Lemma 3.23 ($\sharp$). ■

We are now able to define the decision rules for the thermodynamic game, which is the main topic of this subsection.

Definition 3.10 (Decision rules of the thermodynamic game)
We call all weak*-to-weak*[22] continuous mappings

$$\mathfrak{d}^{\flat} : \mathrm{M}^{\flat} \to \mathcal{X}_{-}^{*} \qquad \text{and} \qquad \mathfrak{d}^{\sharp} : \mathrm{M}^{\sharp} \to \mathcal{X}_{+}^{*}$$

such that $\mathfrak{d}^{\flat}(x_{+}) \in M^{\flat}(x_{+})$ and $\mathfrak{d}^{\sharp}(x_{-}) \in M^{\sharp}(x_{-})$ $\flat$-decision rules and $\sharp$-decision rules, respectively. The two sets of $\flat$-decision and $\sharp$-decision rules are respectively denoted by $\mathfrak{D}^{\flat}$ and $\mathfrak{D}^{\sharp}$.

[22]That is, continuity refers to the weak* topology for both the domain and codomain of the mapping.

As $M^\flat(x_+)$ is convex for all $x_+ \in \mathrm{M}^\flat$ (see Proposition 3.7), $\mathfrak{D}^\flat$ is a convex subset the space $C_{\mathrm{w}^*}(\mathrm{M}^\flat;\mathcal{X}_-^*)$ of all weak*-to-weak* continuous functions $\mathrm{M}^\flat \to \mathcal{X}_-^*$, with the usual point-wise vector space operations.

Observe that

$$\inf_{\mathfrak{f}\in C_{\mathrm{w}^*}(\mathrm{M}^\flat;\mathcal{X}_-^*)} \sup_{y_+\in \mathrm{M}^\flat} P_{\mathrm{NL}}\left(y_+,\mathfrak{f}\left(y_+\right)\right) \geq \sup_{y_+\in \mathrm{M}^\flat} \inf_{\mathfrak{f}\in C_{\mathrm{w}^*}(\mathrm{M}^\flat;\mathcal{X}_-^*)} P_{\mathrm{NL}}\left(y_+,\mathfrak{f}\left(y_+\right)\right) \geq \mathrm{P}^\flat .$$

Therefore, if the set $\mathfrak{D}^\flat$ of $\flat$-decision rules is **nonempty** then, for all $\mathfrak{d}^\flat \in \mathfrak{D}^\flat$ and $x_+ \in \mathrm{M}^\flat$, one clearly has the following equalities:

$$\begin{aligned}\mathrm{P}^\flat &= \inf_{\mathfrak{f}\in C_{\mathrm{w}^*}(\mathrm{M}^\flat;\mathcal{X}_-^*)} \sup_{y_+\in \mathrm{M}^\flat} P_{\mathrm{NL}}\left(y_+,\mathfrak{f}\left(y_+\right)\right) = \sup_{y_+\in \mathrm{M}^\flat} \inf_{\mathfrak{f}\in C_{\mathrm{w}^*}(\mathrm{M}^\flat;\mathcal{X}_-^*)} P_{\mathrm{NL}}\left(y_+,\mathfrak{f}\left(y_+\right)\right) \\ &= \inf_{\mathfrak{f}\in C_{\mathrm{w}^*}(\mathrm{M}^\flat;\mathcal{X}_-^*)} P_{\mathrm{NL}}\left(x_+,\mathfrak{f}\left(x_+\right)\right) = \sup_{y_+\in \mathrm{M}^\flat} P_{\mathrm{NL}}\left(y_+,\mathfrak{d}^\flat\left(y_+\right)\right) = P_{\mathrm{NL}}\left(x_+,\mathfrak{d}^\flat\left(x_+\right)\right) .\end{aligned}$$

In particular the two-person zero-sum game on $\mathrm{M}^\flat \times C_{\mathrm{w}^*}(\mathrm{M}^\flat;\mathcal{X}_-^*)$ whose payoff function is

$$(y_+,\mathfrak{f}) \mapsto P_{\mathrm{NL}}\left(y_+,\mathfrak{f}\left(y_+\right)\right)$$

has a non-cooperative equilibrium (i.e., a saddle-point) and its unique conservative value is nothing but $\mathrm{P}^\flat$, which is equal to $\sup \mathbb{F}(K)$, thanks to Theorem 3.8 (i). This new two-person zero-sum game is known as "extended game with information transfer". Such a game is defined for instance in [5, Ch. 7, Section 7.2]. This is related to a very general result of game theory: Lasry's theorem [4, Theorem 8.4].

The same is true for $\mathrm{P}^\sharp$ and $\mathfrak{D}^\sharp$, mutatis mutandis. Note however that $\mathfrak{D}^\sharp$ is not studied in further detail, as it is a subject of lesser interest, even though this study would be no more difficult than the one of $\mathfrak{D}^\flat$. By contrast, as discussed above, $\mathfrak{D}^\flat$ is related to the variational problem $\sup \mathbb{F}(K)$ on the state space, which is our main object of study.

We give in the sequel some general condition ensuring that $\mathfrak{D}^\flat$ is nonempty. We start with additional conditions on the Legendre-Fenchel transforms g_-^* and g_+^*. It turns out that the strict convexity and the finiteness of g_-^* together with the continuity of g_+^* imply that $\mathfrak{D}^\flat$ is a singleton. In particular, it is nonempty. We then show that the uniqueness of the linear equilibrium states of the Bogoliubov linearization also implies the existence of a $\flat$-decision rule (but not necessarily its uniqueness). This last condition is very useful in our paradigmatic example, because it is fulfilled when the functions (θ_-,θ_+) are of Hölder type (Definition 2.14).

Proposition 3.11 (Existence of $\flat$-decision rules for strictly convex g_-^* and continuous g_+^*) *Assume Conditions B1–B3. If, additionally, g_+^* is weak*-continuous,* $\mathrm{dom}(g_-^*) = \mathcal{X}_-^*$ *and $g_-^* : \mathcal{X}_-^* \to \mathbb{R}$ is strictly convex, then $\mathfrak{D}^\flat$ is a singleton.*

Proof. If g_-^* is strictly convex then, for any fixed $y_+ \in \mathcal{X}_+^*$, the mapping $y_- \mapsto P_{\mathrm{NL}}(y_+,y_-)$, defined by (145)–(146), is strictly convex (as $y_- \mapsto P_{\mathrm{L}}(y_+,y_-)$ is convex, by Lemma 3.17) and thus, $M^\flat(y_+) = \{y_-(y_+)\}$ is a singleton. Therefore, $\mathfrak{D}^\flat$ has at most one element. We now show that the mapping $x_+ \mapsto x_-(x_+)$ from $\mathrm{M}^\flat$ to $\mathcal{X}_-^*$ is weak*-to-weak* continuous. Take any net $(x_+^{(j)})_{j\in J}$ in $\mathrm{M}^\flat$ converging in the weak* topology to some $x_+ \in \mathrm{M}^\flat$. By Proposition 3.7 (i) or Lemma 3.22 ($\flat$), for some $R \in \mathbb{R}^+$, $M^\flat(x_+) \subseteq B_-(0,R)$ for all $x_+ \in \mathrm{M}^\flat$, where $B_-(0,R) \subseteq \mathcal{X}_-^*$ is the closed ball of radius $R \in \mathbb{R}^+$ and center $0 \in \mathcal{X}_-^*$. In particular, $x_-(x_+^{(j)}) \in B_-(0,R)$ for all $j \in J$. Thus, as $B_-(0,R)$ is weak*-compact (cf. the Banach-Alaoglu theorem [57, Theorem 3.15]), we can assume without loss of generality that $(x_-(x_+^{(j)}))_{j\in J}$ converges to some $x_- \in B_-(0,R)$. For all $j \in J$ and $y_- \in \mathcal{X}_-^*$,

$$P_{\mathrm{NL}}(x_+^{(j)},x_-(x_+^{(j)})) = P_{\mathrm{L}}(x_+^{(j)},x_-(x_+^{(j)})) + g_-^*(x_-(x_+^{(j)})) - g_+^*(x_+^{(j)}) \leq P_{\mathrm{NL}}(x_+^{(j)},y_-) .$$

As $\mathrm{M}^\flat$ is norm bounded (Proposition 3.7 (ii) or Lemma 3.23 ($\flat$)), g_-^* is always weak*-lower semicontinuous and g_+^* is by assumption weak*-continuous, by taking the j-limit and using Lemma 3.18, we conclude that

$$P_{\mathrm{NL}}\left(x_+, x_-\right) = P_{\mathrm{L}}(x_+, x_-) + g_-^*(x_-) - g_+^*(x_+) \le P_{\mathrm{NL}}\left(x_+, y_-\right) \ , \qquad y_- \in \mathcal{X}_-^* \ .$$

In other words, $x_- = x_-(x_+)$ and so, the net $(x_-(x_+^{(j)}))_{j\in J}$ has to converge to $x_-(x_+)$ in the weak* topology, from which it follows that the mapping $x_+ \mapsto x_-(x_+)$ is weak*-to-weak* continuous. ■

We now provide a method for constructing decision rules, based on equilibrium states. To this end, we recall some general notions of convex analysis, more specifically the notion of "selection" [56, Definition 2.7]: Given a real normed space $\mathcal{X}$, a subset $\Omega \subseteq \mathcal{X}$ and a continuous convex function $g : \mathcal{X} \to \mathbb{R}$, a function $s : \Omega \to \mathcal{X}^*$ is a *selection* for g if $s(x) \in \partial g(x)$ for all $x \in \Omega$. Interestingly, for Banach spaces $\mathcal{X}$, g is Gateaux-differentiable (Fréchet-differentiable) at $x \in \mathcal{X}$ iff there is a selection $\xi_x : \Omega_x \to \mathcal{X}^*$ for g that is norm-to-weak* continuous[23] (norm-to-norm continuous[24]) at x, where Ω_x is some neighborhood of x. See [56, Proposition 2.8]. In particular, if g is Gateaux-differentiable (Fréchet-differentiable) on the whole space $\mathcal{X}$ then the unique selection $s_g : \mathcal{X} \to \mathcal{X}^*$ for g is norm-to-weak* continuous (norm-to-norm continuous).

Inspired by the general concept of *selection*, we introduce a similar notion in relation to equilibrium states:

Definition 3.12 (Equilibrium state selection)
A mapping $\xi : \mathrm{M}^\flat \to E_{\mathbb{F}} \subseteq K$ is a equilibrium state selection if it is weak* continuous and $\xi(x_+) \in E_{\mathcal{G}_{x_+,\cdot}}^{\mathrm{sc}}$ for all $x_+ \in \mathrm{M}^\flat$. See Theorem 3.8 for the definition of $E_{\mathcal{G}_{x_+,\cdot}}^{\mathrm{sc}} \subseteq E_{\mathbb{F}}$.

Because of [56, Proposition 2.8] and the self-consistency of equilibrium states, if the function $g_- : \mathcal{X}_- \to \mathbb{R}$ is Gateaux-differentiable then any equilibrium state selection naturally induces a $\flat$-decision rule $\mathfrak{d}_\xi^\flat : \mathrm{M}^\flat \to \mathcal{X}_-^*$:

Proposition 3.13 (Construction of decision rules from equilibrium state selections)
Assume Conditions B1–B3. If the function $g_- : \mathcal{X}_- \to \mathbb{R}$ is Gateaux-differentiable and $\xi : \mathrm{M}^\flat \to E_{\mathbb{F}}$ is an equilibrium state selection, then the mapping

$$\mathfrak{d}_\xi^\flat \doteq s_{g_-} \circ \tau_- \circ \xi : \mathrm{M}^\flat \to \mathcal{X}_-^*$$

is a decision rule, where $s_{g_-} : \mathcal{X}_- \to \mathcal{X}_-^*$ is the unique selection for g_-. It is weak*-to-norm[25] continuous when g_- is Fréchet-differentiable.

Proof. By [56, Proposition 2.8], $\mathfrak{d}_\xi^\flat$ is weak*-to-weak* continuous and it is weak*-to-norm continuous when g_- is Fréchet-differentiable. Take now any optimizer $x_+ \in \mathrm{M}^\flat$. Then, as ξ is an equilibrium state selection, $\xi(x_+) \in E_{\mathcal{G}_{x_+,\cdot}}^{\mathrm{sc}}$. By Theorem 3.8 (ii), for any $x_- \in M^\flat(x_+)$, $x_- \in \partial g_-(\tau_- \circ \xi(x_+))$. Since

$$\partial g_-\left(\tau_- \circ \xi\left(x_+\right)\right) = \left\{s_{g_-} \circ \tau_- \circ \xi\left(x_+\right)\right\}$$

is a singleton (g_- being Gateaux-differentiable), it follows that $\mathfrak{d}_\xi^\flat(x_+) \in M^\flat(x_+)$ for all $x_+ \in \mathrm{M}^\flat$. In other words, $\mathfrak{d}_\xi^\flat$ is a $\flat$-decision rule. ■

We now conclude our study of decision rules by giving sufficient conditions for the existence of an equilibrium state selection as well as its properties. All of this is summarized in the following proposition, which concludes our subsection:

[23] That is, the continuity is considered with respect to the norm topology in the domain and the weak* topology in the codomain of the mapping.

[24] That is, the continuity is considered with respect to the norm topology for both the domain and codomain of the mapping.

[25] That is, the continuity is considered with respect to the weak* topology in the domain and the norm topology in the codomain of the mapping.

Proposition 3.14 (Properties of equilibrium state selections)
Under Conditions B1–B3 the following assertions hold:
(i) *If the function $g_+ : \mathcal{X}_+ \to \mathbb{R}$ is Gateaux-differentiable then any equilibrium state selection is injective.*
(ii) *If, for all $x_+ \in \mathrm{M}^\flat$, the set $E^{\mathrm{sc}}_{\mathcal{G}_{x_+,\cdot}}$ of self-consistent equilibrium states is a singleton, then there is a unique equilibrium state selection $\mathrm{M}^\flat \to E_\mathbb{F}$. Additionally, in this case the unique equilibrium state selection is surjective. This occurs in particular when $g_- \circ \tau_- : K \to \mathbb{R}$ is strictly convex, or if the Bogoliubov linearization $\mathcal{G}_{y_+,y_-}$ has a unique linear equilibrium state for all $y_\pm \in \mathcal{X}^*_\pm$.*
(iii) *If $g_+ : \mathcal{X}^*_+ \to \mathbb{R}$ is Gateaux-differentiable and, for all $x_+ \in \mathrm{M}^\flat$, the set $E^{\mathrm{sc}}_{\mathcal{G}_{x_+,\cdot}}$ is a singleton, then the unique equilibrium state selection $\mathrm{M}^\flat \to E_\mathbb{F}$ is a homeomorphism with respect to the weak* topology of $\mathrm{M}^\flat \subseteq \mathcal{X}^*_+$.*

Proof. If the function $g_+ : \mathcal{X}^*_+ \to \mathbb{R}$ is Gateaux-differentiable then, for any $x_+, x'_+ \in \mathrm{M}^\flat$ such that $x_+ \neq x'_+$, we know from Theorem 3.8 (iii) that $E_{\mathcal{G}^{\mathrm{sc}}_{x_+,\cdot}} \cap E_{\mathcal{G}^{\mathrm{sc}}_{x'_+,\cdot}} = \emptyset$. Thus, any equilibrium state selection is necessarily injective in this case. By the proof of Theorem 3.8, for any $x_+ \in \mathrm{M}^\flat$,

$$E_{\mathcal{G}^{\mathrm{sc}}_{x_+,\cdot}} = E_{\mathcal{F}_{x_+}} \subseteq E_{\mathcal{G}_{x_+,x_-}} \ , \tag{158}$$

where $\mathcal{F}_{x_+} : K \to \{-\infty\} \cup \mathbb{R}$ is the concave and upper semicontinuous function defined by

$$\mathcal{F}_{x_+}(\mu) \doteq f(\mu) - g_- \circ \tau_-(\mu) + x_+ \circ \tau_+(\mu) \ , \qquad \mu \in K \ .$$

Clearly,

$$E_{\mathcal{G}^{\mathrm{sc}}_{x_+,\cdot}} = \{\mu(x_+)\} = E_{\mathcal{F}_{x_+}} \ , \qquad x_+ \in \mathrm{M}^\flat \ , \tag{159}$$

is a singleton under the assumptions of Assertion (ii). Thus, the mapping $\xi : \mathrm{M}^\flat \to E_\mathbb{F}$ defined by $\xi(x_+) \doteq \mu(x_+)$ is the unique possible equilibrium state selection in this case. It remains to prove that it is weak* continuous. With this aim, take any net $(x^{(j)}_+)_{j\in J}$ in $\mathrm{M}^\flat$ that converges in the weak* topology to some $x_+ \in \mathrm{M}^\flat$. By the weak*-compactness of $E_\mathbb{F}$ and Theorem 3.8 (iii), we can assume without loss of generality that $\mu(x^{(j)}_+) \subseteq E_\mathbb{F}$ converges to some $\mu \in E_\mathbb{F}$ in the weak* topology. By Equation (159), for all $j \in J$ and $\nu \in K$,

$$f(\nu) - g_- \circ \tau_-(\nu) + x^{(j)}_+ \circ \tau_+(\nu) \leq f(\mu(x^{(j)}_+)) - g_- \circ \tau_-(\mu(x^{(j)}_+)) + x^{(j)}_+ \circ \tau_+(\mu(x^{(j)}_+)) \ .$$

Taking the j-limit at fixed $\nu \in K$, we arrive at

$$f(\nu) - g_- \circ \tau_-(\nu) + x_+ \circ \tau_+(\nu) \leq f(\mu) - g_- \circ \tau_-(\mu) + x_+ \circ \tau_+(\mu) \ , \qquad \nu \in K \ ,$$

using that $\mathrm{M}^\flat$ is norm bounded (Proposition 3.7 (ii) or Lemma 3.23 ($\flat$)). In other words, $\mu \in E_{\mathcal{F}_{x_+}} = E_{\mathcal{G}^{\mathrm{sc}}_{x_+,\cdot}}$. As $E_{\mathcal{G}^{\mathrm{sc}}_{x_+,\cdot}} = \{\mu(x_+)\}$, it follows that $\mu(x^{(j)}_+)$ converges to $\mu(x_+)$ in the weak* topology and ξ is thus weak* continuous. If $E_{\mathcal{G}^{\mathrm{sc}}_{\cdot,x_+}}$ is a singleton for all $x_+ \in \mathrm{M}^\flat$ then ξ is surjective, because of Theorem 3.8 (iii). Finally, under the assumption of Assertion (iii), the unique equilibrium state selection $\xi : \mathrm{M}^\flat \to E_\mathbb{F}$ is weak* continuous and bijective, thanks to Assertions (i)–(ii). Since $\mathrm{M}^\flat$ is weak*-compact (Proposition 3.7 (ii) or Lemma 3.23 ($\flat$)) and $E_\mathbb{F} \subseteq K$ with K being a Hausdorff space, ξ is in this case a homeomorphism. ∎

3.5 Generalized equilibria and optimal transport

Generalized equilibria and order parameters. Physically, the set of all equilibrium states of a system is expected to be a convex set, in order to allow for phase mixtures. By contrast, the set $E_\mathbb{F} =$

$M_{\mathbb{F}}$ of nonlinear equilibrium states defined in Section 3.3 (Equations (141)–(142)) is not necessarily convex, because the function $\mathbb{F}$ is generally not concave. That is why, in the scope of our paradigmatic example, we introduced the notion of generalized nonlinear equilibrium measures in Section 2.5 (Definition 2.11) by considering the closed convex hull $G_P \doteq \overline{\mathrm{co}}(E_P)$ of the set E_P of nonlinear equilibrium measures. Exactly the same can be done in our abstract version of the theory of nonlinear equilibria.

We thus define the compact convex set

$$G_{\mathbb{F}} \doteq \overline{\mathrm{co}}(E_{\mathbb{F}}) \subseteq K \ . \tag{160}$$

We call its elements *generalized nonlinear equilibrium states*. Physical equilibria should be seen as elements of this set. If K is a compact convex set in some topological vector space (which is the case in virtually all important cases, like our paradigmatic example, the nonlinear thermodynamic formalism of classical dynamical systems, or quantum lattice systems), note from the Milman theorem [55, Proposition 1.5] that the extreme points of the convex set $G_{\mathbb{F}}$ are all in $E_{\mathbb{F}}$. Thus, the previous set $E_{\mathbb{F}}$ of nonlinear equilibrium states is rather related to what are called the *pure equilibrium states*, or *pure phases*, in Physics.

One also expects that equilibrium states of a physical system are exactly those that maximize a pressure function associated with the system under consideration. It turns out that $G_{\mathbb{F}}$ can also be characterized in this way. It is the set of maximizers of the *upper*[26] Γ-regularization of $\mathbb{F}$, which is the smallest concave and upper semicontinuous function above $\mathbb{F}$. For more details, see Section 2.9.3, in particular Theorem 2.40.

However, an upper Γ-regularization is usually only concave, which is not as good as being affine, a more desirable property for a pressure function from a physical point of view. In fact, in Section 2.7.4 we show for our paradigmatic example that generalized equilibrium states (measures in this case) are indeed the limits of (weak*-)convergent sequences of approximating maximizers of an affine pressure function. See Equation (50). This function turns out to be not semicontinuous in general (it is only Borel-measurable), but it has the same upper Γ-regularization as the original nonlinear pressure. See Theorem 2.40 in this context. Moreover, it is maximized by any equilibrium state $\rho \in E_{\mathbb{F}} \cap \mathcal{E}(K)$ that is extreme in the convex set K. As the function is affine, it is also maximized by any convex combination $\rho \in \mathrm{co}(E_{\mathbb{F}} \cap \mathcal{E}(K)) \subseteq \overline{\mathrm{co}}(E_{\mathbb{F}}) \doteq G_{\mathbb{F}}$ of such equilibrium states, but not necessarily by any generalized equilibrium state $\rho \in G_{\mathbb{F}}$, as it is not necessarily upper semicontinuous. However, in the Hölder case (which is the relevant one in this section) the generalized equilibrium states turn out to be precisely the strict maximizers of the affine pressure function. See Corollary 2.22. In fact, this is not a particular property of our paradigmatic example, i.e., the thermodynamic formalism of (classical) dynamical systems. In [18] we made the same observation in the case of quantum lattice systems (quantum spins or fermionic systems on lattices) when the functions $g_{\pm}$ defining the nonlinear pressure function $\mathbb{F}$ are quadratic.

Remark from [55, Proposition 1.2] that if A is a compact subset of a locally convex (Hausdorff) space $\mathcal{Y}$, then $y \in \overline{\mathrm{co}}(A) \subseteq \mathcal{Y}$ iff y is the barycenter of a probability measure $\mathfrak{m}_y$ supported in A (see Definition 3.25). If $\overline{\mathrm{co}}(A)$ is a compact Choquet simplex and A only contains extreme points of $\overline{\mathrm{co}}(A)$ then the probability measure $\mathfrak{m}_y$ is always unique and the mapping $y \mapsto \mathfrak{m}_y$ from $\overline{\mathrm{co}}(A)$ to the set of probability measures on A is an affine one-to-one correspondence. For more details see [55].

By Theorem 3.8 (iii)–(iv), $E_{\mathbb{F}}$ is compact and, if the equilibrium states of the associated Bogoliubov linearizations are unique and f is affine, then $E_{\mathbb{F}}$ is even a compact set of extreme points of K and $G_{\mathbb{F}} \doteq \overline{\mathrm{co}}(E_{\mathbb{F}}) \subseteq K$ is, in particular, a face of the convex set K. So, in this case, provided K is a compact convex subset of a locally convex space, $\mu \in G_{\mathbb{F}}$ iff μ is the barycenter of a probability measure $\mathfrak{m}_\mu$ supported in $E_{\mathbb{F}}$ (see Definition 3.25). As illustrated in the nonlinear thermodynamic

[26] It corresponds to the function $-\Gamma(-\mathbb{F})$ with $\Gamma(f)$ defined by Equation (120).

formalism (see Proposition 2.35) or for quantum lattice systems [18, Theorem 1.9], $G_{\mathbb{F}}$ is usually even a compact Choquet simplex and $\mathfrak{m}_\mu$ is uniquely defined.

If the functions $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$ are Gateaux-differentiable then the corresponding gradient mappings

$$\nabla g_\pm : \mathcal{X}_\pm \to \mathcal{X}_\pm^*$$

are norm-to-weak* continuous. They are even norm-to-norm continuous when the functions $g_\pm$ are Fréchet-differentiable. See [56, Proposition 2.8]. Recall that the subdifferentials $\partial g_\pm(x_\pm) = \{\nabla g_\pm(x_\pm)\}$ are singletons for any $x_\pm \in \mathcal{X}_\pm$, $g_\pm$ being Gateaux-differentiable. In particular, in this case, to any $\mu \in G_{\mathbb{F}}$ we naturally associate measures $\mathfrak{y}_\pm(\mu)$ on $\mathcal{X}_\pm^*$, by considering the pushforward of a probability measure $\mathfrak{m}_\mu$ representing $\mu \in G_{\mathbb{F}}$ (in $G_{\mathbb{F}}$) through the mapping

$$\nabla g_\pm \circ \tau_\pm : K \to \mathcal{X}_\pm^* \ .$$

As $\nabla g_\pm$ is always norm-to-weak* continuous and $\tau_\pm : K \to \mathcal{X}_\pm$ are by assumption continuous (Condition B2), the measures $\mathfrak{y}_\pm(\mu)$ are supported on the weak*-compact sets $\nabla g_\pm \circ \tau_\pm(E_{\mathbb{F}}) \subseteq \mathcal{X}_\pm^*$. In fact, by the self-consistency of nonlinear equilibrium states (Theorem 3.8 (ii)–(iii)), the probability measure $\mathfrak{y}_+(\mu)$ is supported in the weak*-compact set $\mathrm{M}^\flat \subseteq \mathcal{X}_+^*$, while $\mathfrak{y}_-(\mu)$ is supported in the union

$$\mathrm{M}_-^\flat \doteq \bigcup_{x_+ \in \mathrm{M}^\flat} M^\flat(x_+) \subseteq \mathcal{X}_-^* \ . \tag{161}$$

If $g_\pm$ are Fréchet-differentiable then the sets $\nabla g_\pm \circ \tau_\pm(E_{\mathbb{F}})$ are even norm-compact, as the gradient mapping is norm-to-norm continuous in this case, thanks again to [56, Proposition 2.8].

Physically, the elements of the dual spaces $\mathcal{X}_\pm^*$ refer to so-called *order parameters* of the system under consideration. For example, if we consider a spin system, one typical order parameter of interest is the macroscopic magnetization density. Similarly, if one considers models for superconductors, then one typical order parameter would be the macroscopic density of Cooper pairs. Thus, given a generalized nonlinear equilibrium state $\mu \in G_{\mathbb{F}}$, the measures $\mathfrak{y}_\pm(\mu)$ give the corresponding distributions of order parameters. The order parameter distribution is well-defined in virtually all cases of interest, because $G_{\mathbb{F}}$ is usually a Choquet simplex in these cases. In a so-called pure phase ($\mu \in E_{\mathbb{F}}$) such distributions are delta distributions, that is, they are concentrated in single points and thus have zero variance. If these distributions have non-vanishing variance then we have a so-called phase mixture, which is the general situation at a (first order) phase transition. Of course, the order parameter distributions $\mathfrak{y}_\pm(\mu)$ also make sense for states $\mu \in K$ that are more general, that is, not necessarily elements of $G_{\mathbb{F}}$, as far as the compact convex space K is a Choquet simplex (and $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$ stay Gateaux-differentiable of course). For simplicity we restrict ourselves to the equilibrium case.

One question that arises is how to calculate, from possible experiments, including numerical ones, the order parameter distributions $\mathfrak{y}_\pm(\mu)$ at least for generalized nonlinear equilibrium measures $\mu \in G_{\mathbb{F}}$. In our paradigmatic example (Section 2) we can give a more precise empirical interpretation to these order parameter distributions, which can thus be recovered from possible experiments:

Paradigmatic example – Nonlinear Thermodynamic Formalism of Dynamical Systems. In this example, $K = \mathcal{P}(T)$ is the set (8) of T-invariant probability measures on the Banach space $C(\Sigma)^*$ of continuous functions $\Sigma \to \mathbb{R}$, $f = h$ is the entropy (Definition 2.1) and $\tau_\pm : \mathcal{P}(T) \to \mathcal{X}_\pm$ are defined by $\tau_\pm(\mu) \doteq \theta_\pm(\mu_S)$ for any $\mu \in \mathcal{P}(T)$. In this particular case $\mathbb{F} = P$ is defined by Equation (31) and we assume in Section 2.7.1 Conditions TF1–TF3, which are nothing more than Conditions B1–B3 of Section 3.3 applied to our paradigmatic example, as explained in the proof of Proposition 2.15.

Observe first that $K = \mathcal{P}(T)$ is a weak*-compact Choquet simplex (Proposition 2.35). It is a subset of the dual space $C(\Sigma)^*$, which is a locally convex (Hausdorff) space when endowed with the weak* topology. Since the entropy $f = h$ is affine (Proposition 2.38), G_P is also a Choquet simplex and one can naturally identify generalized nonlinear equilibrium states $\mu \in G_P$ with probability measures $\mathfrak{m}_\mu$ on the weak*-compact set E_P of (simple) nonlinear equilibrium states.

Recall that S denotes the unit closed ball in $C(\Sigma)$ and $\mathcal{M}(S)$ is the Banach space of bounded, real-valued Borel-measurable functions $S \to \mathbb{R}$ with the supremum norm. See Section 2.1, in particular Equation (6). For any $\sigma \in \Sigma$ and $n \in \mathbb{N}$, let $\mathbb{E}_{n,\sigma} : S \to \mathbb{R}$ be the continuous function defined by Equation (59), that is,

$$\mathbb{E}_{n,\sigma}[\varphi] \doteq \mathbb{E}_n[\varphi](\sigma) \doteq n^{-1}\left(\varphi(\sigma) + \varphi \circ T(\sigma) + \cdots + \varphi \circ T^{n-1}(\sigma)\right) , \qquad \varphi \in S .$$

Observe that, at any fixed $n \in \mathbb{N}$, $\mathbb{E}_{n,\sigma'}$ converges point-wise to $\mathbb{E}_{n,\sigma}$, as $\sigma' \to \sigma$. Thus, if the linear mappings $\theta_\pm : \mathcal{M}(S) \to \mathcal{X}_\pm$ are σ-normal (see Condition TF1 of Section 2.7.1) then $\theta_\pm(\mathbb{E}_{n,\sigma'})$ converges in norm to $\theta_\pm(\mathbb{E}_{n,\sigma'})$, as $\sigma' \to \sigma$. That is, for any $n \in \mathbb{N}$, the mappings $\sigma \mapsto \theta_\pm(\mathbb{E}_{n,\sigma})$ from Σ to $\mathcal{X}_\pm$ are continuous. Recall that $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$ are assumed here to be Gateaux-differentiable and by continuity of the gradient mappings $\nabla g_\pm$ the functions $Y_{\pm,n} : \Sigma \to \mathcal{X}_\pm^*$ defined by

$$\sigma \mapsto \nabla g_\pm \circ \theta_\pm(\mathbb{E}_{n,\sigma}) , \tag{162}$$

are also continuous. In particular, they are Borel-measurable and thus define real-valued random variables

$$\begin{array}{rcll} Y_{\pm,n}(x_\pm) & : & \Sigma & \to \mathbb{R} \\ & & \sigma & \mapsto \nabla g_\pm \circ \theta_\pm(\mathbb{E}_{n,\sigma})(x_\pm) \end{array}$$

for any $x_\pm \in \mathcal{X}_\pm$, the distribution law of which is thus the pushforward of some generalized nonlinear equilibrium measure $\mu \in G_P$ through the functions (162), whose values are linear functionals, applied to vectors $x_\pm \in \mathcal{X}_\pm$. ($Y_{\pm,n}$ are also random variables taking values in $\mathcal{X}_\pm^*$.) The distribution law of $Y_{\pm,n}(x_\pm)$ is denoted below by $\mathfrak{y}_\pm^{(n)}(\mu)$.

If the pair (θ_-, θ_+) is a Hölder-type linear functions (Definition 2.14) then all the Bogoliubov linearizations have *unique* equilibrium measures, which are all ergodic, thanks to Corollary 2.18. Then, E_P is in this case a compact set of extreme points of $\mathcal{P}(T)$ and $G_P \doteq \overline{\mathrm{co}}(E_P) \subseteq \mathcal{P}(T)$ is, in particular, a face of the convex set $\mathcal{P}(T)$ and a weak*-compact Choquet simplex. Therefore, as already explained, $\mu \in G_P$ iff μ is the barycenter of a *unique* probability measure $\mathfrak{m}_\mu$ supported in $E_P \subseteq \mathcal{P}_{\mathrm{erg}}(T)$ (see Definition 3.25). Furthermore, in the Hölder case, the mapping $\mu \mapsto \mathfrak{m}_\mu$ from G_P to the set of probability measures on E_P is an affine one-to-one correspondence. For more details see again [55]. In this situation, we can prove that, given a (generalized nonlinear equilibrium) T-invariant probability measure $\mu \in G_P$, the characteristic functions of $\mathfrak{y}_\pm^{(n)}(\mu)$ point-wise converge to the characteristic functions of the order parameter distributions $\mathfrak{y}_\pm(\mu)$, as $n \to \infty$:

Proposition 3.15 (Order parameter distributions from Birkhoff sums – Hölder case)
Assume Conditions TF1–TF3 of Section 2.7.1 and that the pair (θ_-, θ_+) of linear functions is Hölder-type and $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$ are Gateaux-differentiable. Let $\mathfrak{y}_\pm \equiv \mathfrak{y}_\pm(\mu)$ for some $\mu \in G_P$. Then, for all $x_\pm \in \mathcal{X}_\pm$,

$$\lim_{n\to\infty} \int_{\mathcal{X}_\pm^*} \exp\left(iy_\pm(x_\pm)\right) \mathfrak{y}_\pm^{(n)}(\mathrm{d}y_\pm) = \int_{\mathcal{X}_\pm^*} \exp\left(iy_\pm(x_\pm)\right) \mathfrak{y}_\pm(\mathrm{d}y_\pm) .$$

Proof. Fix $\mu \in G_P$. For all $n \in \mathbb{N}$ and $x_\pm \in \mathcal{X}_\pm$,

$$\int_{\mathcal{X}_\pm^*} \exp\left(iy_\pm(x_\pm)\right) \mathfrak{y}_\pm^{(n)}(\mathrm{d}y_\pm) = \int_\Sigma \exp\left(i\left[\nabla g_\pm \circ \theta_\pm(\mathbb{E}_{n,\sigma})\right](x_\pm)\right) \mu(\mathrm{d}\sigma) .$$

Note that the mappings

$$\sigma \mapsto \left[\nabla g_\pm \circ \theta_\pm(\mathbb{E}_{n,\sigma})\right](x_\pm)$$

from Σ to $\mathbb{R}$ are continuous. Further, since $\mathcal{P}(T)$ is a weak*-compact Choquet simplex, μ is the barycenter of a (unique) probability measure $\mathfrak{m}_\mu$ on E_P and therefore,

$$\int_\Sigma \exp\left(i\left[\nabla g_\pm \circ \theta_\pm\left(\mathbb{E}_{n,\sigma}\right)\right](x_\pm)\right)\mu\left(\mathrm{d}\sigma\right) = \int_K \left\{\int_\Sigma \exp\left(i\left[\nabla g_\pm \circ \theta_\pm\left(\mathbb{E}_{n,\sigma}\right)\right](x_\pm)\right)\nu\left(\mathrm{d}\sigma\right)\right\}\mathfrak{m}_\mu\left(\mathrm{d}\nu\right)\ .$$

By Corollary 2.18 (i), all $\nu \in E_P$ are ergodic measures and we can infer from Lemma 2.23 and Lebesgue's dominated convergence theorem that, for all $\nu \in E_P$,

$$\lim_{n\to\infty}\int_\Sigma \exp\left(i\left[\nabla g_\pm \circ \theta_\pm\left(\mathbb{E}_{n,\sigma}\right)\right](x_\pm)\right)\nu\left(\mathrm{d}\sigma\right) = \exp\left(i\left[\nabla g_\pm \circ \theta_\pm\left(\nu_S\right)\right](x_\pm)\right)\ .$$

Hence, again by Lebesgue's dominated convergence theorem,

$$\begin{aligned}\lim_{n\to\infty}\int_\Sigma \exp\left(i\left[\nabla g_\pm \circ \theta_\pm\left(\mathbb{E}_{n,\sigma}\right)\right](x_\pm)\right)\mu\left(\mathrm{d}\sigma\right) &= \int_K \exp\left(i\left[\nabla g_\pm \circ \theta_\pm\left(\nu_S\right)\right](x_\pm)\right)\mathfrak{m}_\mu\left(\mathrm{d}\nu\right)\\ &= \int_{\mathcal{X}_\pm^*}\exp\left(iy_\pm\left(x_\pm\right)\right)\mathfrak{y}_\pm\left(\mathrm{d}y_\pm\right)\ .\end{aligned}$$

Indeed, recall that the measures $\mathfrak{y}_\pm(\mu)$ on $\mathcal{X}_\pm^*$ are the pushforward of $\mathfrak{m}_\mu$ through the mappings $\nabla g_\pm \circ \tau_\pm$, where $\tau_\pm(\nu) \doteq \theta_\pm(\nu_S)$ for any $\nu \in \mathcal{P}(T)$. Using well-known properties of the characteristic functions of real-valued random variables (cf. Lévy's continuity theorem [36, Theorem 5.3]), we conclude from the above proposition that, for all $x_\pm \in \mathcal{X}_\pm$, the distribution laws $\mathfrak{y}_\pm^{(n)}(\mu)$ of the real-valued random variables $Y_{\pm,n}(x_\pm)$ weak* converge, as $n \to \infty$, to the order parameter distributions $\mathfrak{y}_\pm(\mu)$. ∎

It means that the order parameter distributions can be recovered from experiments, possibly numerical ones. The order parameter distributions are also related to optimal transport [65], for which various high-performance numerical tools are available. This link is explained below for the general (abstract) case.

Optimal transport and distributions of order parameters. We show here that, for any given order parameter distribution at equilibrium, that is, $\mathfrak{y}_\pm \equiv \mathfrak{y}_\pm(\mu)$ for some $\mu \in G_{\mathbb{F}}$, which is not necessarily explicitly known, the nonlinear pressure $\sup \mathbb{F}(K)$ can be exactly recovered from an optimal-transport problem associated with these order parameter distributions, the cost function of which is nothing but the nonlinear approximating pressure P_{NL} (145) (similar to the thermodynamic game).

Recall from Theorem 3.8 (i) that the nonlinear pressure $\sup \mathbb{F}(K)$ satisfies

$$\sup \mathbb{F}\left(K\right) = \mathrm{P}^\flat \doteq \sup_{y_+\in\mathcal{X}_+^*}\ \inf_{y_-\in\mathcal{X}_-^*} P_{\mathrm{NL}}\left(y_+, y_-\right)$$

and consider the continuous mapping

$$\begin{array}{rcl} E_{\mathbb{F}} & \to & \mathcal{X}_+^* \times \mathcal{X}_-^* \\ \nu & \mapsto & \left(\nabla g_+ \circ \tau_+\left(\nu\right), \nabla g_- \circ \tau_-\left(\nu\right)\right)\end{array}$$

for Gateaux-differentiable functions $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$. Fix now any $\mu \in G_{\mathbb{F}}$ and let $\mathfrak{n}_\mu$ be the pushforward of a probability measure $\mathfrak{m}_\mu$ representing μ in $G_{\mathbb{F}}$ through the above continuous mapping. By construction, the marginal distributions on $\mathcal{X}_+^*$ and $\mathcal{X}_-^*$ are $\mathfrak{y}_+(\mu)$ and $\mathfrak{y}_-(\mu)$, respectively. Moreover, by Theorem 3.8,

$$\begin{aligned}\int_{\mathcal{X}_+^*\times\mathcal{X}_-^*} P_{\mathrm{NL}}\left(y_+, y_-\right)\mathfrak{n}_\mu\left(\mathrm{d}\left(y_+, y_-\right)\right) &= \int_K P_{\mathrm{NL}}\left(\nabla g_+ \circ \tau_+\left(\nu\right), \nabla g_- \circ \tau_-\left(\nu\right)\right)\mathfrak{m}_\mu\left(\mathrm{d}\nu\right)\\ &= \int_K \sup \mathbb{F}\left(K\right)\mathfrak{m}_\mu\left(\mathrm{d}\nu\right) = \sup \mathbb{F}\left(K\right)\ . \end{aligned} \tag{163}$$

On the other hand, for any probability measure $\mathfrak{n}$ on $\mathcal{X}_{+}^{*} \times \mathcal{X}_{-}^{*}$ whose marginal distributions are $\mathfrak{y}_{+} \equiv \mathfrak{y}_{+}(\mu)$ and $\mathfrak{y}_{-} \equiv \mathfrak{y}_{-}(\mu)$, one has

$$\int_{\mathcal{X}_{+}^{*} \times \mathcal{X}_{-}^{*}} P_{\mathrm{NL}}\left(y_{+}, y_{-}\right) \mathfrak{n}\left(\mathrm{d}\left(y_{+}, y_{-}\right)\right) \geq \int_{\mathcal{X}_{+}^{*} \times \mathcal{X}_{-}^{*}} P^{\flat}\left(y_{+}\right) \mathfrak{n}\left(\mathrm{d}\left(y_{+}, y_{-}\right)\right) = \int_{\mathcal{X}_{+}^{*}} P^{\flat}\left(y_{+}\right) \mathfrak{y}_{+}\left(\mathrm{d} y_{+}\right) , \tag{164}$$

where we recall that $P^{\flat}(y_{+})$ is defined by (147), i.e.,

$$P^{\flat}\left(y_{+}\right) \doteq \inf P_{\mathrm{NL}}\left(y_{+}, \mathcal{X}_{-}^{*}\right) , \qquad y_{+} \in \mathcal{X}_{+}^{*} .$$

Note that the function $y_{+} \mapsto P^{\flat}\left(y_{+}\right)$ is Borel-measurable, because it is upper semicontinuous, by Lemma 3.21 ($\flat$). As $\mathfrak{y}_{+}$ is supported in $\mathrm{M}^{\flat}$ (see Theorem 3.8), one conclude that

$$\int_{\mathcal{X}_{+}^{*} \times \mathcal{X}_{-}^{*}} P_{\mathrm{NL}}\left(y_{+}, y_{-}\right) \mathfrak{n}\left(\mathrm{d}\left(y_{+}, y_{-}\right)\right) \geq \int_{\mathcal{X}_{+}^{*}} \sup \mathbb{F}\left(K\right) \mathfrak{y}_{+}\left(\mathrm{d} y_{+}\right) = \sup \mathbb{F}\left(K\right) . \tag{165}$$

In other words, by (163) and (165), the nonlinear pressure can be recovered from the following *Monge-Kantorovich (transportation) problem* [65, page 10]:

$$\sup \mathbb{F}\left(K\right) = \inf_{\mathfrak{n} \in \mathfrak{d}(\mathfrak{y}_{+}, \mathfrak{y}_{-})} \int_{\mathcal{X}_{+}^{*} \times \mathcal{X}_{-}^{*}} P_{\mathrm{NL}}\left(y_{+}, y_{-}\right) \mathfrak{n}\left(\mathrm{d}\left(y_{+}, y_{-}\right)\right) , \tag{166}$$

where $\mathfrak{d}(\mathfrak{y}_{+}, \mathfrak{y}_{-})$ is the set of all probability measures on $\mathcal{X}_{+}^{*} \times \mathcal{X}_{-}^{*}$ whose marginal distributions are respectively $\mathfrak{y}_{+} \equiv \mathfrak{y}_{+}(\mu)$ and $\mathfrak{y}_{-} \equiv \mathfrak{y}_{-}(\mu)$ for arbitrary fixed equilibrium order parameter distributions associated with some $\mu \in G_{\mathbb{F}}$. As a consequence, the dual Kantorovich problem [65, Equation (5.3)], along with appropriate assumptions, leads to the following representation of the nonlinear pressure:

Theorem 3.16 (Nonlinear pressure and the dual Kantorovich problem)
Assume Conditions B1–B3. Suppose additionally that $\mathcal{X}_{\pm}$ are separable normed spaces, $g_{\pm} : \mathcal{X}_{\pm} \rightarrow \mathbb{R}$ are Gateaux-differentiable, $g_{\pm}^{*} : \mathcal{X}_{\pm}^{*} \rightarrow \mathbb{R}$ are continuous with g_{-}^{*} being strictly convex. Let $\mathfrak{y}_{\pm} \equiv \mathfrak{y}_{\pm}(\mu)$ for some $\mu \in G_{\mathbb{F}}$. Then,

$$\sup \mathbb{F}\left(K\right) = \sup \left\{ \int_{\mathcal{X}_{+}^{*}} P_{\mathrm{NL}}^{+}\left(y_{+}\right) \mathfrak{y}_{+}\left(\mathrm{d} y_{+}\right) - \int_{\mathcal{X}_{-}^{*}} P_{\mathrm{NL}}^{-}\left(y_{-}\right) \mathfrak{y}_{-}\left(\mathrm{d} y_{-}\right) \right\} ,$$

where the supremum is taken with respect to all possible choices of continuous functions $P_{\mathrm{NL}}^{+} : \mathcal{X}_{+}^{*} \rightarrow \mathbb{R}$ and $P_{\mathrm{NL}}^{-} : \mathcal{X}_{-}^{*} \rightarrow \mathbb{R}$ such that

$$P_{\mathrm{NL}}^{+}\left(y_{+}\right) - P_{\mathrm{NL}}^{-}\left(y_{-}\right) \leq P_{\mathrm{NL}}\left(y_{+}, y_{-}\right) , \qquad y_{\pm} \in \mathcal{X}_{\pm}^{*} .$$

Proof. On the one hand, observe from Theorem 3.8 that $\mathfrak{y}_{+}$ is always supported in $\mathrm{M}^{\flat} \equiv \mathrm{M}_{+}^{\flat}$, while $\mathfrak{y}_{-}$ is always supported on the set $\mathrm{M}_{-}^{\flat}$ defined by (161). As a consequence, Equation (166) can be rewritten as

$$\sup \mathbb{F}\left(K\right) = \inf_{\mathfrak{n} \in \mathfrak{d}(\mathfrak{y}_{+}, \mathfrak{y}_{-})} \int_{\mathrm{M}_{+}^{\flat} \times \mathrm{M}_{-}^{\flat}} P_{\mathrm{NL}}\left(y_{+}, y_{-}\right) \mathfrak{n}\left(\mathrm{d}\left(y_{+}, y_{-}\right)\right) . \tag{167}$$

By Proposition 3.7 (ii) (or Lemma 3.23 ($\flat$)), $\mathrm{M}^{\flat} \equiv \mathrm{M}_{+}^{\flat}$ is a (nonempty) weak*-compact subset of $\mathcal{X}_{+}^{*}$. Since, by assumption, g_{+}^{*} is continuous, $\mathrm{dom}(g_{-}^{*}) = \mathcal{X}_{-}^{*}$ and $g_{-}^{*} : \mathcal{X}_{-}^{*} \rightarrow \mathbb{R}$ is strictly convex, we can deduce from Propositions 3.7 and 3.11 that the nonempty set $\mathrm{M}_{-}^{\flat}$ is also a norm-bounded weak*-compact subset of $\mathcal{X}_{-}^{*}$. Since $\mathcal{X}_{\pm}$ are assumed to be separable normed spaces, the weak* topology is metrizable on the weak*-compacts sets $\mathrm{M}_{\pm}^{\flat} \subseteq \mathcal{X}_{\pm}^{*}$, which can thus be considered as metric spaces. Because compact metric spaces are separable and complete, it follows that $(\mathrm{M}_{\pm}^{\flat}, \mathfrak{y}_{\pm})$ are

Polish probability spaces. As $\mathrm{M}_{\pm}^{\flat}$ are norm-bounded, by Lemma 3.18 and continuity of $g_{\pm}^{*}:\mathcal{X}_{\pm}^{*}\rightarrow\mathbb{R}$, the nonlinear approximating pressure

$$P_{\mathrm{NL}}:\mathrm{M}_{+}^{\flat}\times\mathrm{M}_{-}^{\flat}\rightarrow\mathbb{R}$$

is weak* continuous. Moreover, for any $y_{\pm}\in\mathcal{X}_{\pm}^{*}$,

$$P_{\mathrm{NL}}\left(y_{+},y_{-}\right)\doteq P_{\mathrm{L}}\left(y_{+},y_{-}\right)+g_{-}^{*}\left(y_{-}\right)-g_{+}^{*}\left(y_{+}\right)\geq P^{\flat}\left(y_{+}\right)+g_{-}^{*}\left(y_{-}\right)-g_{+}^{*}\left(y_{+}\right)$$

with $P^{\flat}(y_{+})$ defined by (147). Since $g_{\pm}^{*}:\mathcal{X}_{\pm}^{*}\rightarrow\mathbb{R}$ are continuous and $\mathrm{M}_{\pm}^{\flat}$ are weak*-compact,

$$\int_{\mathrm{M}_{\pm}^{\flat}}g_{\pm}^{*}\left(y_{\pm}\right)\mathfrak{y}_{\pm}\left(\mathrm{d}y_{\pm}\right)<\infty\qquad\text{and}\qquad\int_{\mathrm{M}_{+}^{\flat}}P^{\flat}\left(y_{+}\right)\mathfrak{y}_{+}\left(\mathrm{d}y_{+}\right)=\sup\mathbb{F}\left(K\right)<\infty\ ,$$

as already explained in (164)–(165). Observe again that the function $y_{+}\mapsto P^{\flat}\left(y_{+}\right)$ is upper semi-continuous, by Lemma 3.21 ($\flat$). All the conditions of the Kantorovich duality theorem [65, Theorem 5.10] are thus satisfied. So, the assertion follows from Equation (167) combined with [65, Theorem 5.10 (i)]. ■

Solutions (or at least approximate solutions $\pm P_{\mathrm{NL}}^{\pm}$) to the dual Kantorovich problem expressed in Theorem 3.16 should be physically interpreted as being effective pressure functions on the spaces of order parameters.

Note that the assumptions of Theorem 3.16 are trivially satisfied for physically relevant choices of convex functions $g_{\pm}$. Take for example $\beta_{\pm}\in\mathbb{R}^{+}$ and

$$g_{\pm}(x)=\beta_{\pm}x^{2}/2\qquad\text{or}\qquad g_{\pm}(x)=2\cosh x$$

with $\mathcal{X}_{\pm}=\mathbb{R}$ and $\mathcal{X}_{\pm}^{*}\equiv\mathbb{R}$, and remark that the corresponding Legendre-Fenchel transforms,

$$g_{\pm}^{*}(s)=s^{2}/(2\beta_{\pm})\qquad\text{or}\qquad g_{\pm}^{*}(s)=s\sinh^{-1}(s/2)-\sqrt{4+s^{4}}\ ,$$

are smooth and strictly convex. See also Remarks 3.5 and 3.6. Observe additionally that the strict convexity of the function g_{-}^{*} assumed in Theorem 3.16 is a sufficient condition only used for simplicity in order to ensure the weak* compactness of the set $\mathrm{M}_{-}^{\flat}$ (161), but this should not, of course, be essential.

In fact, in the proof of Theorem 3.16, we only use the first assertion (i) of the Kantorovich duality theorem [65, Theorem 5.10]. There are many other results in [65, Theorem 5.10] that could have been invoked, but we refrain from mentioning them here, as our goal is not to provide a detailed analysis of the corresponding optimal transport under the more general conditions, but to establish a fruitful bridge between the maximization of real-valued functions $\mathbb{F}$ on compact spaces and the dual Kantorovich problem, thus paving the way for entirely new mathematical and numerical developments.

3.6 Technical results on the thermodynamic game

Within the general setting of the previous sections we study here the linear and nonlinear approximating pressures associated with Bogoliubov linearizations, that is, the quantities

$$P_{\mathrm{L}}\left(y_{+},y_{-}\right)\doteq\sup\mathcal{G}_{y_{+},y_{-}}\left(K\right)\quad\text{and}\quad P_{\mathrm{NL}}\left(y_{+},y_{-}\right)\doteq P_{\mathrm{L}}\left(y_{+},y_{-}\right)+g_{-}^{*}\left(y_{-}\right)-g_{+}^{*}\left(y_{+}\right)\tag{168}$$

for any $y_{\pm}\in\mathcal{X}_{\pm}^{*}$, previously defined by (145)–(146). We also analyze in this context the variational problems associated with the two-person zero-sum game (Section 3.4) whose payoff function is the nonlinear approximating pressure P_{NL}. Note that we always consider here *non-zero* functions $g_{\pm}$. If one of them is zero, we have essentially the same results (mutatis mutandis). See, e.g., the discussions at the end of Section 3.3.

Before we start our technical study, we go over a few of the notation: $B_{\pm}(0,R) \subseteq \mathcal{X}_{\pm}^{*}$ are closed balls (125) of radius $R \in \mathbb{R}^{+}$ and center $0 \in \mathcal{X}_{\pm}^{*}$ and $\|\cdot\|_{\mathrm{op}}$ is the usual operator norm, defined by (126) for any continuous linear functional, while $\|\cdot\|_{\infty}$ is the uniform norm (or sup norm), defined by (129). Recall that any continuous function defined on a compact is uniformly bounded and the corresponding uniform norm is thus finite.

We start with two simple continuity properties of the linear approximating pressure P_{L}, seen as a function

$$(y_{+},y_{-}) \mapsto P_{\mathrm{L}}\left(y_{+},y_{-}\right)$$

from $\mathcal{X}_{+}^{*} \times \mathcal{X}_{-}^{*}$ to $\mathbb{R}$ and defined by Equation (168) (see also (146)). Some arguments are elementary and are only given below for completeness. Recall that Condition B1 is stated in Section 3.3.

Lemma 3.17 (Norm continuity of linear approximating pressures)
Assume Condition B1. Then, $P_{\mathrm{L}} : \mathcal{X}_{+}^{*} \times \mathcal{X}_{-}^{*} \to \mathbb{R}$ is convex and Lipschitz-norm continuous:

$$\left|P_{\mathrm{L}}\left(x_{+},x_{-}\right) - P_{\mathrm{L}}\left(y_{+},y_{-}\right)\right| \leq \|\tau_{+}\|_{\infty} \left\|x_{+}-y_{+}\right\|_{\mathrm{op}} + \|\tau_{-}\|_{\infty} \left\|x_{-}-y_{-}\right\|_{\mathrm{op}} , \qquad x_{\pm},y_{\pm} \in \mathcal{X}_{\pm}^{*} .$$

Proof. Observe that, for any $x_{\pm},y_{\pm} \in \mathcal{X}_{\pm}^{*}$ and $\nu \in K$,

$$\left|x_{\pm} \circ \tau_{\pm}\left(\nu\right) - y_{\pm} \circ \tau_{\pm}\left(\nu\right)\right| \leq \left\|x_{\pm}-y_{\pm}\right\|_{\mathrm{op}} \|\tau_{\pm}\|_{\infty}$$

and the inequality of the lemma directly follows from the definition

$$\mathcal{G}_{y_{+},y_{-}} \doteq f - y_{-} \circ \tau_{-} + y_{+} \circ \tau_{+} , \qquad y_{\pm} \in \mathcal{X}_{\pm}^{*} , \tag{169}$$

of Bogoliubov linearizations, see Equation (144). P_{L} is convex because it is the supremum of affine (real-valued) mappings $(y_{+},y_{-}) \mapsto \mathcal{G}_{y_{+},y_{-}}(\mu)$, $\mu \in K$. ■

Lemma 3.18 (Weak* continuity of linear approximating pressures)
Assume Condition B1. Then, P_{L} is weak* continuous on $B_{+}(0,R) \times B_{-}(0,R)$ for any fixed $R \in \mathbb{R}^{+}$. If $\mathcal{X}_{\pm}$ are Banach spaces then P_{L} is weak* continuous on the whole space $\mathcal{X}_{+}^{*} \times \mathcal{X}_{-}^{*}$.

Proof. Fix $R \in \mathbb{R}^{+}$. Take any nets $(y_{\pm}^{(j)})_{j \in J} \subseteq B_{\pm}(0,R)$ weak* converging to arbitrary points $y_{\pm} \in B_{\pm}(0,R)$, respectively. By Condition B1, K is (Hausdorff) compact and $\tau_{\pm}$ are continuous. As a consequence, $\tau_{\pm}(K) \subseteq \mathcal{X}_{\pm}$ are compact (in the norm topology of $\mathcal{X}_{\pm}$). Thus, for any $\varepsilon \in \mathbb{R}^{+}$, there are finite sets $M_{\pm,\varepsilon} \subseteq K$ such that

$$\mathrm{dist}\left(\tau_{\pm}\left(M_{\pm,\varepsilon}\right), \tau_{\pm}\left(K\right)\right) = \sup_{\mu \in K} \inf_{\nu \in M_{\pm,\varepsilon}} \left\|\tau_{\pm}\left(\mu\right) - \tau_{\pm}\left(\nu\right)\right\|_{\mathcal{X}_{\pm}} \leq \frac{\varepsilon}{4R} . \tag{170}$$

Remark that the above quantity is the so-called Hausdorff distance between the compact sets $\tau_{\pm}\left(M_{\pm,\varepsilon}\right)$ and $\tau_{\pm}\left(K\right)$ in $\mathcal{X}_{\pm}$. Then, by the triangle inequality and Inequality (170), for any $\mu \in K$ and $\nu \in M_{\pm,\varepsilon}$,

$$\left|y_{\pm} \circ \tau_{\pm}\left(\mu\right) - y_{\pm}^{(j)} \circ \tau_{\pm}\left(\mu\right)\right| \leq \inf_{\nu \in M_{\pm,\varepsilon}} \left|y_{\pm} \circ \tau_{\pm}\left(\nu\right) - y_{\pm}^{(j)} \circ \tau_{\pm}\left(\nu\right)\right| + \frac{\varepsilon}{2} .$$

Hence, there is $j_{\varepsilon} \in J$ such that, for all $\mu \in K$ and $j \in J$ satisfying $j \geq j_{\varepsilon}$,

$$\left|y_{\pm} \circ \tau_{\pm}\left(\mu\right) - y_{\pm}^{(j)} \circ \tau_{\pm}\left(\mu\right)\right| \leq \varepsilon ,$$

because $M_{\pm,\varepsilon}$ is finite and $(y_{\pm}^{(j)})_{j \in J}$ weak* converges to $y_{\pm}$. Now, as the last bound is uniform with respect to $\mu \in K$, from the definition (169) of $\mathcal{G}_{y_{+},y_{-}}$, it follows that the mapping $(y_{+},y_{-}) \mapsto$

$P_{\mathrm{L}}(y_+, y_-)$ from $B_+(0,R) \times B_-(0,R)$ to $\mathbb{R}$ is weak* continuous. If the normed spaces $\mathcal{X}_\pm$ are complete, then any weak* convergent net in the dual spaces $\mathcal{X}_\pm^*$ is norm bounded, thanks to the Uniform Boundedness Principle (the Banach-Steinhaus theorem [57, Theorem 2.5]). Consequently, in the Banach case, P_{L} is weak* continuous on $\mathcal{X}_+^* \times \mathcal{X}_-^*$. ■

After briefly studying linear pressures, we now provide a series of lemmata regarding important properties of the nonlinear approximating pressure P_{NL}, seen as a mapping

$$(y_+, y_-) \mapsto P_{\mathrm{NL}}(y_+, y_-)$$

from $\mathcal{X}_+^* \times \mathcal{X}_-^*$ to $\mathbb{R} \cup \{\infty\}$ and defined by Equation (168) (see also (145)).

Lemma 3.19 (Properties of the nonlinear approximating pressure)
Assume Condition B1. Let $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$ be non-zero lower semicontinuous and convex functions with $\mathrm{dom}(g_+^*) = \mathcal{X}_+^*$.
$(+)$ *For $R \in \mathbb{R}^+$ and $y_- \in \mathrm{dom}(g_-^*)$, the mapping $y_+ \mapsto P_{\mathrm{NL}}(y_+, y_-)$ from $B_+(0,R)$ to $\mathbb{R}$ is weak*-upper semicontinuous. If $\mathcal{X}_+$ is a Banach space then it is weak*-upper semicontinuous on $\mathcal{X}_+^*$.*
$(-)$ *For $R \in \mathbb{R}^+$ and $y_+ \in \mathcal{X}_+^*$, the mapping $y_- \mapsto P_{\mathrm{NL}}(y_+, y_-)$ from $B_-(0,R)$ to $\mathbb{R} \cup \{\infty\}$ is convex and weak*-lower semicontinuous. If $\mathcal{X}_-$ is a Banach space then it is weak*-lower semicontinuous on $\mathcal{X}_-^*$.*

Proof. As Legendre-Fenchel transforms, the functions $g_+^* : \mathcal{X}_+^* \to \mathbb{R}$ and $g_-^* : \mathcal{X}_-^* \to \mathbb{R} \cup \{\infty\}$ are always weak*-lower semicontinuous. Note also that the convex set $\mathrm{dom}(g_-^*)$ is nonempty, thanks to (123). Therefore, by Lemma 3.18, for any fixed $R \in \mathbb{R}^+$ and $y_- \in \mathrm{dom}(g_-^*)$, the mapping $y_+ \mapsto P_{\mathrm{NL}}(y_+, y_-)$ from $B_+(0,R)$ to $\mathbb{R}$ is weak*-upper semicontinuous. The Banach case is also a direct consequence of Lemma 3.18. Mutatis mutandis for $(-)$. Notice that

$$y_- \mapsto P_{\mathrm{NL}}(y_+, y_-) \doteq P_{\mathrm{L}}(y_+, y_-) + g_-^*(y_-) - g_+^*(y_+)$$

is convex, because Legendre-Fenchel transforms are always convex and $y_- \mapsto P_{\mathrm{L}}(y_+, y_-)$ is also convex, thanks to Lemma 3.17. ■

We next undertake the study of the two functions

$$P^\flat : \mathcal{X}_+^* \to \{-\infty\} \cup \mathbb{R} \qquad \text{and} \qquad P^\sharp : \mathcal{X}_-^* \to \mathbb{R} \cup \{\infty\}$$

defined by (152)–(153), that is,

$$P^\flat(y_+) \doteq \inf P_{\mathrm{NL}}\left(y_+, \mathcal{X}_-^*\right) \qquad \text{and} \qquad P^\sharp(y_-) \doteq \sup P_{\mathrm{NL}}\left(\mathcal{X}_+^*, y_-\right) \tag{171}$$

for any $y_\pm \in \mathcal{X}_\pm^*$. We assume that Legendre-Fenchel transforms g^* have minimal linear growth $\lambda \in \mathbb{R}^+$ in the sense of (124), that is,

$$\lim_{R\to\infty} \sup_{y \in \mathcal{X}^* \backslash B(0,R)} \left\{ \lambda \|y\|_{\mathrm{op}} - g^*(y) \right\} = -\infty \,. \tag{172}$$

This allows us to show that these variational problems can be restricted to closed balls.

Lemma 3.20 (Consequences of the minimal linear growth of $g_\pm^*$)
Assume Condition B1. Let $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$ be non-zero lower semicontinuous and convex functions with $\mathrm{dom}(g_+^*) = \mathcal{X}_+^*$.
$(\flat)$ *If (172) holds for $g = g_-$ and $\lambda = \|\tau_-\|_\infty$, then there is $R \in \mathbb{R}^+$ such that, for all $y_+ \in \mathcal{X}_+^*$,*

$$P^\flat(y_+) = \inf_{y_- \in B_-(0,R)} P_{\mathrm{NL}}(y_+, y_-) < \inf_{y_- \in \mathcal{X}_-^* \backslash B_-(0,R)} P_{\mathrm{NL}}(y_+, y_-) \,.$$

(♯) *If (172) holds for $g = g_+$ and $\lambda = \|\tau_+\|_\infty$, then there is $R \in \mathbb{R}^+$ such that, for all $y_- \in \mathrm{dom}(g_-^*)$,*

$$P^\sharp(y_-) = \sup_{y_+ \in B_+(0,R)} P_{\mathrm{NL}}(y_+, y_-) > \sup_{y_+ \in \mathcal{X}_+^* \setminus B_+(0,R)} P_{\mathrm{NL}}(y_+, y_-) \ .$$

(♯♭) *If (172) holds for $g = g_\pm$ and $\lambda = \|\tau_\pm\|_\infty$, then there is $R \in \mathbb{R}^+$ such that*

$$\sup_{y_+ \in B_+(0,R)} P^\flat(y_+) > \sup_{y_+ \in \mathcal{X}_+^* \setminus B_+(0,R)} P^\flat(y_+) \ .$$

(♭♯) *If (172) holds for $g = g_\pm$ and $\lambda = \|\tau_\pm\|_\infty$, then there is $R \in \mathbb{R}^+$ such that*

$$\inf_{y_- \in B_-(0,R)} P^\sharp(y_-) < \inf_{y_- \in \mathcal{X}_-^* \setminus B_-(0,R)} P^\sharp(y_-) \ .$$

Proof. Assertions (♭) and (♯) are direct consequences of Lemma 3.17. Note that the strict inequality of (♯) cannot be satisfied when $y_- \notin \mathrm{dom}(g_-^*)$, because, in this case, $P_{\mathrm{NL}}(y_+, y_-) = \infty$ for all $y_+ \in \mathcal{X}_+^*$. To prove (♯♭) take any $x_- \in \mathrm{dom}(g_-^*) \neq \emptyset$ (see (123)) and note from Lemma 3.17 and Equation (172) for $g = g_+$ and $\lambda = \|\tau_+\|_\infty$ that

$$\begin{aligned} \lim_{R\to\infty} \sup_{y_+ \in \mathcal{X}_+^* \setminus B_+(0,R)} P^\flat(y_+) &= \lim_{R\to\infty} \sup_{y_+ \in \mathcal{X}_+^* \setminus B_+(0,R)} \inf_{y_- \in \mathcal{X}_-^*} P_{\mathrm{NL}}(y_+, y_-) \\ &\leq \lim_{R\to\infty} \sup_{y_+ \in \mathcal{X}_+^* \setminus B_+(0,R)} \left\{ P_{\mathrm{L}}(y_+, x_-) + g_-^*(x_-) - g_+^*(y_+) \right\} = -\infty \ . \end{aligned}$$

We meanwhile infer from (♭) the existence of some strictly positive radius $\tilde{R} \in \mathbb{R}^+$ such that

$$P^\flat(y_+) = \inf_{y_- \in B_-(0,\tilde{R})} P_{\mathrm{NL}}(y_+, y_-) \ , \qquad y_+ \in \mathcal{X}_+^* \ .$$

Thus, by the weak* compactness of the closed ball $B_-(0, \tilde{R})$ (cf. the Banach-Alaoglu theorem [57, Theorem 3.15]) and Lemma 3.19 (−),

$$P^\flat(y_+) = \min_{y_- \in B_-(0,\tilde{R})} P_{\mathrm{NL}}(y_+, y_-) = \min_{y_- \in B_-(0,\tilde{R}) \cap \mathrm{dom}(g_-^*)} P_{\mathrm{NL}}(y_+, y_-) \in \mathbb{R} \tag{173}$$

for all $y_+ \in \mathcal{X}_+^*$, keeping in mind that $\mathrm{dom}(g_-^*) \neq \emptyset$ (see (123)). In particular, $P^\flat(y_+) > -\infty$ for all $y_+ \in \mathcal{X}_+^*$ and, as a consequence, for some $R < \infty$,

$$\sup_{y_+ \in B_+(0,R)} P^\flat(y_+) > \sup_{y_+ \in \mathcal{X}_+^* \setminus B_+(0,R)} P^\flat(y_+) \ .$$

The proof of (♭♯) is similar: Note from Lemma 3.17 and Equation (172) for $g = g_-$ and $\lambda = \|\tau_-\|_\infty$ that, for any $x_+ \in \mathcal{X}_+^*$,

$$\begin{aligned} \lim_{R\to\infty} \inf_{y_- \in \mathcal{X}_-^* \setminus B_-(0,R)} P^\sharp(y_-) &= \lim_{R\to\infty} \inf_{y_- \in \mathcal{X}_-^* \setminus B_-(0,R)} \sup_{y_+ \in \mathcal{X}_+^*} P_{\mathrm{NL}}(y_+, y_-) \\ &\geq \lim_{R\to\infty} \inf_{y_- \in \mathcal{X}_-^* \setminus B_-(0,R)} \left\{ P_{\mathrm{L}}(y_+, x_-) + g_-^*(y_-) - g_+^*(x_+) \right\} = \infty \ . \end{aligned}$$

By Assertion (♯), there is $\tilde{R} \in \mathbb{R}^+$ such that

$$P^\sharp(y_-) = \sup_{y_+ \in B_+(0,\tilde{R})} P_{\mathrm{NL}}(y_+, y_-) \ , \qquad y_- \in \mathrm{dom}(g_-^*) \ ,$$

which, combined with the weak* compactness of the closed ball $B_+(0,\tilde{R})$ and Lemma 3.19 $(+)$, yields

$$P^\sharp(y_-) = \max_{y_+\in B_+(0,\tilde{R})} P_{\mathrm{NL}}(y_+,y_-) \in \mathbb{R}\ , \qquad y_- \in \mathrm{dom}(g_-^*)\ . \tag{174}$$

Since $\mathrm{dom}(g_-^*) \neq \emptyset$ (see (123)) and $P^\sharp(y_-) = \infty$ when $y_- \notin \mathrm{dom}(g_-^*)$, we conclude that, for some $R < \infty$,

$$\inf_{y_-\in B_-(0,R)} P^\sharp(y_-) = \inf_{y_-\in B_-(0,R)\cap\mathrm{dom}(g_-^*)} P^\sharp(y_-) < \inf_{y_-\in \mathcal{X}_-^*\backslash B_-(0,R)} P^\sharp(y_-)\ .$$

■

If the Legendre-Fenchel transforms $g_\pm^*$ have minimal linear growth $\|\tau_\pm\|_\infty$, one infers from Equations (173)–(174) that $P^\flat$ and $P^\sharp$ define real-valued functions on $\mathcal{X}_+^*$ and $\mathrm{dom}(g_-^*) \subseteq \mathcal{X}_-^*$, respectively. Additionally, these two functions have the following properties:

Lemma 3.21 (Properties of the functions $P^\sharp$ and $P^\flat$)
Assume Condition B1. Let $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$ be non-zero lower semicontinuous and convex functions with $\mathrm{dom}(g_+^*) = \mathcal{X}_+^*$.
($\flat$) *If (172) holds for $g = g_-$ and $\lambda = \|\tau_-\|_\infty$, then $P^\flat : \mathcal{X}_+^* \to \mathbb{R}$ is weak*-upper semicontinuous on $B_+(0,R)$ for any $R \in \mathbb{R}^+$. If $\mathcal{X}_+$ is a Banach space then $P^\flat$ is weak*-upper semicontinuous on $\mathcal{X}_+^*$.*
($\sharp$) *If (172) holds for $g = g_+$ and $\lambda = \|\tau_+\|_\infty$, then $P^\sharp : \mathcal{X}_-^* \to \mathbb{R}\cup\{\infty\}$ is convex and weak*-lower semicontinuous on $B_-(0,R)$ for any $R \in \mathbb{R}^+$. Its restriction to (the nonempty convex set) $\mathrm{dom}(g_-^*)$ is convex and real-valued. If $\mathcal{X}_-$ is a Banach space then $P^\sharp$ is weak*-lower semicontinuous on $\mathcal{X}_-^*$.*

Proof. Fix $R \in \mathbb{R}^+$. If (172) holds for $g = g_-$ and $\lambda = \|\tau_-\|_\infty$, then we deduce from Equation (173) that there is $\tilde{R} \in \mathbb{R}^+$ such that the function $P^\flat : B_+(0,R) \to \mathbb{R}$ is the infimum of the family

$$\{y_+ \mapsto P_{\mathrm{NL}}(y_+,y_-)\}_{y_-\in B_-(0,\tilde{R})\cap\mathrm{dom}(g_-^*)}$$

of weak*-upper semicontinuous real-valued functions from $B_+(0,R)$ to $\mathbb{R}$, thanks to Lemma 3.19 $(+)$. The real-valued function $P^\flat : B_+(0,R) \to \mathbb{R}$ is therefore weak*-upper semicontinuous. The Banach case is proven in the same way. Mutatis mutandis for ($\sharp$), thanks to Lemma 3.19 $(-)$ and Equation (174). To prove the convexity of $P^\sharp$, observe that this mapping is the supremum over $y_+ \in \mathcal{X}_+^*$ of the convex mappings

$$y_- \mapsto P_{\mathrm{NL}}(y_+,y_-) \doteq P_{\mathrm{L}}(y_+,y_-) + g_-^*(y_-) - g_+^*(y_+)$$

all defined on the same (nonempty) convex domain $\mathrm{dom}(g_-^*)$. ■

For all $y_\pm \in \mathcal{X}_\pm^*$, we study now the subsets (154) and (155) of solutions to the variational problems $P^\flat(y_+)$ and $P^\sharp(y_-)$, i.e.,

$$\begin{aligned} M^\flat(y_+) &\doteq \left\{x_- \in \mathcal{X}_-^* : P^\flat(y_+) = P_{\mathrm{NL}}(y_+,x_-)\right\} \subseteq \mathcal{X}_-^*\ , \\ M^\sharp(y_-) &\doteq \left\{x_+ \in \mathcal{X}_+^* : P^\sharp(y_-) = P_{\mathrm{NL}}(x_+,y_-)\right\} \subseteq \mathcal{X}_+^*\ . \end{aligned}$$

Lemma 3.22 (Solutions to the variational problems $P^\flat(y_+)$ and $P^\sharp(y_-)$)
Assume Condition B1. Let $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$ be non-zero lower semicontinuous and convex functions with $\mathrm{dom}(g_+^*) = \mathcal{X}_+^*$.
($\flat$) *If (172) holds for $g = g_-$ and $\lambda = \|\tau_-\|_\infty$, then, for all $y_+ \in \mathcal{X}_+^*$, $M^\flat(y_+)$ is nonempty, convex and weak*-compact. There is $R \in \mathbb{R}^+$ such that $M^\flat(y_+) \subseteq B_-(0,R) \cap \mathrm{dom}(g_-^*)$ for all $y_+ \in \mathcal{X}_+^*$.*
($\sharp$) *If (172) holds for $g = g_+$ and $\lambda = \|\tau_+\|_\infty$, then, for all $y_- \in \mathrm{dom}(g_-^*)$, $M^\sharp(y_-)$ is nonempty and weak*-compact. There is $R \in \mathbb{R}^+$ such that $M^\sharp(y_-) \subseteq B_+(0,R)$ for all $y_- \in \mathrm{dom}(g_-^*)$.*

Proof. Recall that $\mathrm{dom}(g_-^*) \neq \emptyset$, thanks to (123). For all $y_+ \in \mathcal{X}_+^*$, $M^\flat(y_+) \subseteq \mathrm{dom}(g_-^*)$ because $P_{\mathrm{NL}}(y_+, y_-) = \infty$ when $y_- \notin \mathrm{dom}(g_-^*)$. If (172) holds for $g = g_-$ and $\lambda = \|\tau_-\|_\infty$ then, by Lemma 3.20 ($\flat$), $M^\flat(y_+) \subseteq B_-(0, R) \cap \mathrm{dom}(g_-^*)$ for all $y_+ \in \mathcal{X}_+^*$. Additionally, in this case, the mapping $y_- \mapsto P_{\mathrm{NL}}(y_+, y_-)$ is weak*-lower semicontinuous, thanks to Lemma 3.19 ($-$). Bearing in mind again that closed balls of the dual space of a normed space are weak*-compact (cf. the Banach-Alaoglu theorem [57, Theorem 3.15]), we deduce that the sets $M^\flat(y_+) \subseteq B_-(0, R)$, $y_+ \in \mathcal{X}_+^*$, are nonempty and weak*-compact, as they are always weak*-closed. Their convexity is a direct consequence of the convexity of the function $P_{\mathrm{NL}}(y_+, \cdot)$ (Lemma 3.19 ($-$)). The case ($\sharp$) is proven in a similar way, from the weak*-upper semicontinuity of $P_{\mathrm{NL}}(\cdot, y_-)$. See Lemma 3.19 ($+$). ■

Now we consider the conservative values

$$\mathrm{P}^\flat \doteq \sup_{y_+ \in \mathcal{X}_+^*} P^\flat(y_+) \doteq \sup_{y_+ \in \mathcal{X}_+^*} \inf_{y_- \in \mathcal{X}_-^*} P_{\mathrm{NL}}(y_+, y_-) , \tag{175}$$

$$\mathrm{P}^\sharp \doteq \inf_{y_- \in \mathcal{X}_-^*} P^\sharp(y_-) \doteq \inf_{y_- \in \mathcal{X}_-^*} \sup_{y_+ \in \mathcal{X}_+^*} P_{\mathrm{NL}}(y_+, y_-) , \tag{176}$$

of the thermodynamic game. See Section 3.4, in particular Equation (151). We also study the corresponding sets (156)–(157) of optimizers, i.e.,

$$\mathrm{M}^\flat \doteq \left\{x_+ \in \mathcal{X}_+^* : \mathrm{P}^\flat = P^\flat(x_+)\right\} \subseteq \mathcal{X}_+^* ,$$
$$\mathrm{M}^\sharp \doteq \left\{x_- \in \mathcal{X}_-^* : \mathrm{P}^\sharp \doteq P^\sharp(x_-)\right\} \subseteq \mathcal{X}_-^* ,$$

which are nothing but the sets of conservative strategies of the thermodynamic game.

Lemma 3.23 (Sets $\mathrm{M}^\sharp$ and $\mathrm{M}^\flat$ of optimizers)
Assume Condition B1. Let $g_\pm : \mathcal{X}_\pm \to \mathbb{R}$ be non-zero lower semicontinuous and convex functions for which $\mathrm{dom}(g_+^*) = \mathcal{X}_+^*$ and (172) holds for $g = g_\pm$ and $\lambda = \|\tau_\pm\|_\infty$.*
($\flat$) $\mathrm{P}^\flat \in \mathbb{R}$ *and the set* $\mathrm{M}^\flat \subseteq \mathcal{X}_+^*$ *is nonempty, norm-bounded and weak*-compact.*
($\sharp$) $\mathrm{P}^\sharp \in \mathbb{R}$ *and the set* $\mathrm{M}^\sharp \subseteq \mathcal{X}_-^*$ *is a nonempty, convex, norm-bounded and weak*-compact subset of* $\mathrm{dom}(g_-^*)$.

Proof. By Lemma 3.20 ($\sharp\flat$), $\mathrm{M}^\flat \subseteq B_+(0, R) \subseteq \mathcal{X}_+^*$ is norm-bounded. From Lemma 3.21 ($\flat$) and the weak* compactness of $B_+(0, R)$ one concludes that $\mathrm{P}^\flat \in \mathbb{R}$ and that $\mathrm{M}^\flat$ is nonempty and weak*-compact. Mutatis mutandis for that case ($\sharp$), thanks to Lemmata 3.20 ($\flat\sharp$) and 3.21 ($\sharp$). $\mathrm{M}^\sharp$ is convex, because $P^\sharp$ is convex, thanks to Lemma 3.21 ($\sharp$). Note that $\mathrm{M}^\sharp \subseteq \mathrm{dom}(g_-^*)$, because $\mathrm{P}^\sharp < \infty$ and $P^\sharp(y_-) = \infty$ when $y_- \notin \mathrm{dom}(g_-^*) \neq \emptyset$. ■

3.7 Appendix

This appendix contains some useful standard results. These could help non-specialists to understand the present paper more easily. We present the von Neumann minimax theorem (Section 3.7.1) and the Choquet theorem (Section 3.7.2).

3.7.1 The von Neumann minimax theorem

This theorem was first proven by John von Neumann in 1928 [63]. He proved it in relation to two-person zero-sum games. This laid the foundations of game theory, even if the theorem has applications that go far beyond this field of mathematics. It gives a general criterion for the existence of *saddle points*.

Recall first that saddle points are defined as follows: Let M and N be two sets. The element $(x_0, y_0) \in M \times N$ is, by definition, a saddle point of the real-valued function $f : M \times N \to \mathbb{R}$ if

$$\sup_{y\in N} f(x_0, y) = \inf_{x\in M}\sup_{y\in N} f(x, y) = \sup_{y\in N}\inf_{x\in M} f(x, y) = \inf_{x\in M} f(x, y_0) \ .$$

A saddle point $(x_0, y_0) \in M \times N$ in particular satisfies the equalities

$$f(x_0, y_0) = \inf_{x\in M}\sup_{y\in N} f(x, y) = \sup_{y\in N}\inf_{x\in M} f(x, y) \ .$$

The von Neumann minimax theorem, which is used here in an essential way, refers to the following statement about the existence of saddle points of functions on topological vector spaces:

Theorem 3.24 (von Neumann)
Let M and N be two (nonempty) compact convex space. Assume that $f : M \times N \to \mathbb{R}$ is a real-valued function such that, for all $y \in N$, the mapping $x \mapsto f(x, y)$ is convex and lower semicontinuous, whereas, for all $x \in M$, the mapping $y \mapsto f(x, y)$ is concave and upper semicontinuous. Then there exists a saddle point $(x_0, y_0) \in M \times N$ of f.

There are many different proofs of this assertion available in the literature and we recommend, for example, [4, Chapter 8] for a concise review on two-person zero-sum games, including a proof of the von Neumann minimax theorem [4, Theorem 8.2].

3.7.2 The Choquet theorem

A classical result of Minkowski's states that, in finite dimensions, any element $x \in K$ in a (nonempty) compact convex subset $K \subseteq \mathcal{X}$ can be decomposed into a convex combination of a finite number of extreme points $\hat{x}_1, \ldots, \hat{x}_k \in \mathcal{E}(K)$, that is,

$$x = \sum_{j=1}^{k} \lambda_j \hat{x}_j \ , \tag{177}$$

where $\lambda_1, \ldots, \lambda_k \geq 0$ are positive numbers satisfying $\Sigma_{j=1}^{k} \lambda_j = 1$. Recall that the extreme points of a convex set are the elements that cannot be written as (non-trivial) convex combinations of other elements in K. Their existence is ensured for all compact sets in a locally convex spaces $\mathcal{X}$. In fact, it is well-known that in this case, any compact convex set $K \subseteq \mathcal{X}$ is the closure of the convex hull of the (nonempty) set $\mathcal{E}(K)$ of its extreme points, according to the Krein-Milman theorem [57, Theorems 3.4 (b) and 3.21].

To the decomposition (177) we can naturally associate a probability measure, i.e., a normalized positive Borel regular measure, ξ on K: Take the probability measure ξ_x on K defined by

$$\xi_x \doteq \sum_{j=1}^{k} \lambda_j \delta_{\hat{x}_j}$$

with δ_y being the Dirac (or point) measure[27] at y and rewrite (177) as the following (weak) integral

$$x = \int_K \hat{x}\ \xi_x(\mathrm{d}\hat{x}) \ . \tag{178}$$

That is, the point x is the so-called *barycenter* of the probability measure ξ_x. This notion is defined in the general case as follows (cf. [55, p. 1]):

[27] δ_y is the Borel measure such that for any Borel subset $B \in \mathfrak{B}$ of K, $\delta_y(B) = 1$ if $y \in B$ and $\delta_y(B) = 0$ if $y \notin B$.

Definition 3.25 (Barycenters of a measure)
Let $\mathcal{X}$ be any real topological vector space, $K \subseteq \mathcal{X}$ a nonempty compact subset and ξ a (Borel) probability measure on K. An element $x \in \mathcal{X}$ is a barycenter of ξ if, for any continuous linear functional $x^* \in \mathcal{X}^*$,

$$x^*(x) = \int_K x^*(\hat{x})\, \xi(\mathrm{d}\hat{x}) \ .$$

Note that if $\mathcal{X}$ if locally convex then ξ has at most one barycenter, for $\mathcal{X}^*$ separates the points of $\mathcal{X}$, and in this case we refer to *the* barycenter of ξ. Its existence is ensured by the following result [55, Propositions 1.1–1.2]:

Proposition 3.26 (Existence of barycenters)
Let $\mathcal{X}$ be any locally convex vector space and $K \subseteq \mathcal{X}$ a nonempty compact subset. If the closed convex hull $\overline{\mathrm{co}}K \subseteq \mathcal{X}$ is also compact, then every normalized Borel measure on K has a (unique) barycenter. Additionally, for all $x \in \mathcal{X}$, $x \in \overline{\mathrm{co}}K$ iff x is the barycenter of some (not necessarily unique) probability measure on K.

The Krein-Milman theorem says that any nonempty compact *convex* $K \subseteq X$ of a locally convex vector space has extreme points forming a set $\mathcal{E}(K)$ which satisfies $K = \overline{\mathrm{co}}\, \mathcal{E}(K)$, see [57, Theorems 3.4 (b) and 3.21]. In particular, by Proposition 3.26, any element $x \in K$ is the barycenter of some normalized Borel measure on the closure $\overline{\mathcal{E}}(K)$ of the set $\mathcal{E}(K)$ of all extreme points of K. In infinite dimension one generically [22] has that $\overline{\mathcal{E}}(K) = K$ and this property is thus useless in such a situation. If the relative topology of the given compact subset $K \subseteq \mathcal{X}$ is metrizable, the Choquet theorem strengthens the Krein-Milman theorem by stating that the measure representing an arbitrary element $x \in K$ as its barycenter can always be chosen such that it is supported in $\mathcal{E}(K)$ (and not just in the closure $\overline{\mathcal{E}}(K)$). Indeed, in this case, $\mathcal{E}(K)$ is a Borel set [55, Proposition 1.3]:

Lemma 3.27 (Extreme points form a G_δ-set)
Let $\mathcal{X}$ be any topological vector space and $K \subseteq \mathcal{X}$ a compact convex subset. If the relative topology of K is metrizable, then $\mathcal{E}(K)$ is a G_δ-set with respect to the relative topology of K. In particular, it is a Borel set.

We are now able to state the Choquet theorem, first proven in 1956 by Gustave Choquet [26]:

Theorem 3.28 (Choquet)
Let $\mathcal{X}$ be any locally convex vector space and $K \subseteq \mathcal{X}$ a nonempty, compact and convex subset, whose relative topology is metrizable. For any $x \in K$, there is a (not necessarily unique) probability measure ξ_x on K, which is supported on $\mathcal{E}(K) \subseteq K$ (i.e., $\xi_x(K \backslash \mathcal{E}(K)) = 0$) and whose barycenter is x.

We recommend [55] for a concise review on the Choquet theorem and its generalization to non-metrizable cases. The Choquet theorem stated above is proven on page 14 of these lecture notes [55].

Such a measure ξ_x, as given by Theorem 3.28, is called here a *Choquet measure* associated with the element $x \in K$. If any point $x \in K$ has exactly one associated Choquet measure then the convex subset $K \subseteq X$ is a so-called *Choquet simplex*.

Acknowledgments: This work is supported by FAPESP (grant 2025/12824-5), CNPq (grant 303682/2025-6), the Basque Government through the BERC 2022-2025 program, as well as the following grant: Grant PID2024-156184NB-I00 funded by MICIU/AEI/10.13039/501100011033 and cofunded by the European Union.

References

[1] D. Aguiar, L. Cioletti and R. Ruviaro, A Variational Principle for the Specific Entropy for Symbolic Systems with Uncountable Alphabets, *Math. Nachr.* **291**(17-18) (2018) 2506-2515.

[2] E. M. Alfsen, *Compact convex sets and boundary integrals*. Ergebnisse der Mathematik und ihrer Grenzgebiete – Band 57. Springer-Verlag, 1971.

[3] Ch. D. Aliprantis, K. C. Border, *Infinite Dimensional Analysis. A Hitchhiker's Guide*. *3rd Edition*. Springer-Verlag, 2006.

[4] J.-P. Aubin, *Optima and Equilibria: An Introduction to Nonlinear Analysis,* Berlin-Heidelberg, Springer-Verlag, 1998.

[5] J.-P. Aubin, *Mathematical Methods of Game and Economic Theory.* Dover Publications Inc., 2009.

[6] V. Baladi, *Positive Transfer Operators and decay of correlations*, World Scientific, 2000.

[7] L. Barreira and C. Holanda, Higher-dimensional nonlinear thermodynamical formalism, *J. Stat. Phys*. **187** (2022) 18 (article number).

[8] A. Biś, M. Carvalho, M. Mendes and P. Varandas, A Convex Analysis Approach to Entropy Functions, Variational Principles and Equilibrium States, *Commun. Math. Phys.* **394**(1) (2022) 215-256. Correction: *Commun. Math. Phys.* **401** (2023) 3335–3342.

[9] N.N. Bogoliubov, On the theory of superfluidity, *J. Phys. (USSR)* **11** (1947) 23-32.

[10] N.N. Bogoliubov Jr., On model dynamical systems in statistical mechanics, *Physica* **32**(5) (1966) 933-944.

[11] N.N. Bogoliubov Jr., J.G. Brankov, V.A. Zagrebnov, A.M. Kurbatov and N.S. Tonchev, *Metod approksimiruyushchego gamil'toniana v statisticheskoi fizike*[28]. Sofia: Izdat. Bulgar. Akad. Nauk[29], 1981.

[12] N.N. Bogoliubov Jr., J.G. Brankov, V.A. Zagrebnov, A.M. Kurbatov and N.S. Tonchev, Some classes of exactly soluble models of problems in Quantum Statistical Mechanics: the method of the approximating Hamiltonian, *Russ. Math. Surv.* **39** (1984) 1-50.

[13] R. Bowen, Equilibrium states and the ergodic theory of Anosov diffeomorphisms, *Lect. Notes Math.* **470** (1975).

[14] J.G. Brankov, N.S. Tonchev and V.A. Zagrebnov, A nonpolynomial generalization of exactly soluble models in statistical mechanics, *Ann. Phys. (N. Y.)* **107**(1-2) (1977) 82-94.

[15] J.G. Brankov, N.S. Tonchev and V.A. Zagrebnov, On a class of exactly soluble statistical mechanical models with nonpolynomial interactions, *J. Stat. Phys.* **20**(3) (1979) 317-330.

[16] J.G. Brankov, D.M. Danchev and N.S. Tonchev, *Theory of Critical Phenomena in Finite–size Systems: Scaling and Quantum Effects,* Singapore-New Jersey-London-Hong Kong, World Scientific, 2000.

[28]The Approximating Hamiltonian Method in Statistical Physics.
[29]Publ. House Bulg. Acad. Sci.

[17] J.-B. Bru, W. de Siqueira Pedra, Remarks on the Γ–regularization of Non-convex and Non-Semi-Continuous Functionals on Topological Vector Spaces, *J. Convex Anal.* **19**(3) (2012) 467-483.

[18] J.-B. Bru, W. de Siqueira Pedra, Non-cooperative Equilibria of Fermi Systems With Long Range Interactions, *Memoirs of the AMS* **224**(1052) (2013).

[19] J.-B. Bru, W. de Siqueira Pedra, Classical Dynamics from Self-Consistency Equations in Quantum Mechanics, *J. Math. Phys.* **63** (2022) 052101 (33 pages). See also arXiv:2009.04969 [math-ph] which is an extended version (72 pages).

[20] J.-B. Bru, W. de Siqueira Pedra and A.O. Lopes, Nonlinear Thermodynamic Formalism: Mean-field Phase Transitions, Large Deviations and Bogoliubov's Variational Principle, arXiv:2511.06975 [math.DS] (2025).

[21] J.-B. Bru, W. de Siqueira Pedra and K. Rodrigues Alves, From Short-Range to Mean-Field Models in Quantum Lattices, *Adv. Theor. Math. Phys.* **28**(1) (2024) 69-159.

[22] J.-B. Bru and W. de Siqueira Pedra, Weak* Hypertopologies with Application to Genericity of Convex Sets, *J. Convex Anal.* **29**(1) (2021) 13-60.

[23] J.-B. Bru and de Siqueira Pedra, *C^* -Algebra and Mathematical Foundations of Quantum Statistical Mechanics*, Latin American Mathematics Series - UFSCar subseries, Springer Nature Switzerland AG, 2023.

[24] H. Bruin, *Topological and Ergodic Theory of Symbolic Dynamics*, Graduate Studies in Mathematics, Vol. 228, AMS, 2022.

[25] J. Buzzi, B. Kloeckner and R. Leplaideur, Nonlinear thermodynamical formalism, *Annales Henri Lebesgue* **6** (2023) 1429-1477.

[26] G. Choquet, Existence et unicité des représentations intégrales au moyen des points extrémaux dans les cônes convexes, *Seminaire Bourbaki* **139** (Dec. 1956) 15 pp.

[27] L. Cioletti, A. O. Lopes and M. Stadlbauert, Ruelle Operator for Continuous Potentials and DLR-Gibbs Measures, *Disc and Cont. Dyn. Syst. A* **40**(8) (2020) 4625-4652.

[28] V. Climenhaga, The thermodynamic approach to multifractal analysis, *Ergodic Theory Dyn. Syst.* **34**:5 (2014) 1409-1450.

[29] V. Climenhaga, D. J. Thompson and K. Yamamoto, Large deviations for systems with non-uniform structure, TAMS, 369 (2017), no. 6, 4167-4192.

[30] M.J. Colbrook, Z. Drmač, A. Horningar, An Introductory Guide to Koopman Learning, ArXiv:2510.22002v1 [math.NA] (2025) (33pp).

[31] H.-O. Georgii, *Gibbs Measures and Phase Transitions* , Berlin-New York, De Gruyter, 2011.

[32] J. Ginibre, On the Asymptotic Exactness of the Bogoliubov Approximation for many Bosons Systems, *Commun. Math. Phys.* **8** (1968) 26-51

[33] P. Giulietti, B. Kloeckner, A. O. Lopes and D. Marcon, The calculus of thermodynamical formalism, *J. Eur. Math. Soc.* **20** (10) (2018) 2357-2412.

[34] J. Gonzalez-Conde, D. Lewis, S.S. Bharadwaj, and M. Sanz, Quantum Carleman linearization efficiency in nonlinear fluid dynamics, *Physical Review Research* **7** (2025) 023254 (16pp).

[35] R. Haydon, E. Odell, H. Rosenthal, *On certain classes of Baire-1 functions with applications to Banach space theory*. In: Odwell, E.E., Rosenthal, H.P. (eds) Functional Analysis. Lecture Notes in Mathematics, vol 1470. Springer, Berlin, Heidelberg, 1991.

[36] O. Kallenberg, *Foundations of Modern Probability*, 2nd Edition, New York-Berlin-Heidelberg, Springer-Verlag, 2001.

[37] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Cambridge University Press, Cambridge, 1995.

[38] J. Katz, G. Muraleedharan and A. Alase, Efficient quantum algorithm for solving differential equations with Fourier nonlinearity via Koopman linearization, arXiv:2512.06488 [quant-ph] (2025) (46pp).

[39] A.S. Kechris, *Classical Descriptive Set Theory*, Springer-Verlag New York. Inc., 1995.

[40] G. Keller, *Equilibrium states in Ergodic Theory*, Cambridge Univ. Press, 1998.

[41] H. Komiya, Elementary Proof For Sion's minimax theorem, *Kodai Math. J.* **11**(1) (1988) 5-7.

[42] T. Kucherenko, Nonlinear thermodynamic formalism through the lens of rotation theory, *Disc. and Cont. Dyn. Systems* **44**(12) (2024) 3760-3773.

[43] E.H. Lieb, R. Seiringer and J. Yngvason, Justification of c–Number Substitutions in Bosonic Hamiltonians, *Phys. Rev. Lett.* **94** (2005) 080401-1-4.

[44] J. Lindenstrauss, G.H. Olsen and Y. Sternfeld, The Poulsen simplex, *Ann. Inst. Fourier (Grenoble)* **28** (1978) 91-114.

[45] R. Leplaideur and F. Watbled, Generalized Curie-Weiss model and quadratic pressure in ergodic theory, *Bull. Soc. Math. France* **147**(2) (2019) 197-219.

[46] R. Leplaideur and F. Watbled, Curie-Weiss Type Models for General Spin Spaces and Quadratic Pressure in Ergodic Theory, *J. Stat. Phys*. **181** (2020) 263-292.

[47] A.O. Lopes, The Dimension Spectrum of the Maximal Measure, *SIAM Journal on Mathematical Analysis* **20**(5) (1989) 1243-1254.

[48] A. O. Lopes, Entropy and Large Deviation, *NonLinearity* **3**(2) (1990) 527-546.

[49] A.O. Lopes, Thermodynamic Formalism, Maximizing Probabilities and Large Deviations, Text under construction, (2024), see http://mat.ufrgs.br/˜alopes/pub3/notesformteherm.pdf.

[50] A. O. Lopes, J. K. Mengue, J. Mohr and R. R. Souza, Entropy and Variational Principle for one-dimensional Lattice Systems with a general a-priori probability: positive and zero temperature, *Erg. Theory and Dyn Systems* **35**(6) (2015) 1925-1961.

[51] A.O. Lopes, J. K. Mengue and E.R. Oliveira, Idempotent approach to level-2 variational principles in Thermodynamical Formalism, *J. Math. Phys.* **67** (2026) 042701.

[52] A.O. Lopes and E.R. Oliveira, Level-2 IFS Thermodynamic Formalism: Gibbs probabilities in the space of probabilities and the push-forward map, *Dyn. Systems* **39**, Issue 4 (2024) 823–847.

[53] K.R. Parthasarathy, *Probability Measures on Metric Spaces*, New York and London, Academic Press, 1967.

[54] W. Parry and M. Pollicott, Zeta functions and the periodic orbit structure of hyperbolic dynamics, *Astérisque* **187-188** (1990), Chapters 1-4.

[55] R.R. Phelps, *Lectures on Choquet's Theorem*, 2nd Edition, Lecture Notes in Mathematics, Vol. 1757, Berlin-Heidelberg, Springer-Verlag, 2001.

[56] R.R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, 2nd Edition, Lecture Notes in Mathematics, Vol. 1364, Berlin-Heidelberg, Springer-Verlag 1993.

[57] W. Rudin, *Functional Analysis*, McGraw-Hill Science, 1991.

[58] D. Ruelle, *Thermodynamic Formalism: The Mathematical Structures of Classical Equilibrium Statistical Mechanics*, Volumen 5 de Encyclopedia of mathematics and its applications, Addison-Wesley Publishing Company, Advanced Book Program, 1978; 2nd Edition, Cambridge University Press, 2004.

[59] W. Schirotzek, *Nonsmooth Analysis*, Universitext, Berlin-Heidelberg, Springer-Verlag, 2007.

[60] D. Shi and X. Yang, Koopman Spectral Linearization vs. Carleman Linearization: A Computational Comparison Study, Mathematics **12**(14) (2024) 2156 (16pp).

[61] R. R. Souza and V. Vargas, Existence of Gibbs States and Maximizing Measures on a General One-Dimensional Lattice System with Markovian Structure, *Qual. Theory Dyn. Syst.* **21** (2022) article 5.

[62] F. Tennie, S. Laizet, S. Lloyd and Luca Magri, Quantum computing for nonlinear differential equations and turbulence, *Nature Reviews Physics* **7** (2025) 220-230.

[63] J. Von Neumann, Zur Theorie der Gesellschaftsspiele, *Math. Ann.* **100** (1928) 295-320.

[64] M. Viana and K. Oliveira, *Foundations of Ergodic Theory*, Cambridge University Press, 2016.

[65] C. Villani, *Optimal Transport, Old and New*, Berlin-Heidelberg, Springer-Verlag, 2009.

[66] P. Walters, *Introduction to Ergodic Theory*, Graduate Texts in Mathematics, Springer Verlag, 2000.

[67] T. Wang and W. Wu, Thermodynamic Formalism for Non-uniform Systems with Controlled Specification and Entropy Expansiveness, *Commun. Math. Phys.* **32** (2026) 407.

[68] Y. Xu, Z. Kuang, Q. Huang, J. Yang, H. Zahrouni, M. Potier-Ferry, K. Huang, J.-C. Zhang, H. Fan, H. Hu, A robust quantum nonlinear solver based on the asymptotic numerical method, arXiv:2412.03939 [quant-ph] (2024) (35pp).

[69] V.A. Zagrebnov and J.-B. Bru, The Bogoliubov Model of Weakly Imperfect Bose Gas, *Phys. Rep.* **350** (2001) 291-434.

[70] *Thermodynamic Formalism, CIRM Jean-Morlet Chair, Fall 2019*, Lecture Notes in Mathematics, Eds: M. Pollicott, S. Vaienti, Springer Cham, 2021.