



The Banach manifold of measures and the Lagrange multipliers of statistical mechanics

S. Dostoglou¹ · A. M. Hughes¹

Received: 26 March 2021 / Revised: 6 September 2021 / Accepted: 18 October 2021 /
Published online: 3 November 2021
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. 2021

Abstract

Using a Banach manifold structure on the space of finite positive measures it is shown that all critical points of the Gibbs/information entropy are grand canonical equilibria when the constraints are scalar, and local equilibria when the constraints are integrable functions. This provides a rigorous derivation of equilibrium and local equilibrium Gibbs measures via Lagrange multipliers.

Keywords Gibbs ensembles · Lagrange multipliers · Local equilibrium · Banach manifold

1 Introduction

In classical statistical mechanics one seeks to derive the Gibbs equilibrium distributions as critical points of the information entropy over a space of measures satisfying certain constraints. Gibbs himself showed via an elementary argument that such distributions minimize information entropy for given energy, see [7, Theorem 2, p. 130]. See also also [13, p. 746] and (27) below.

Arguments using Lagrange multipliers to show that *all* critical points of entropy under constraints are Gibbs distributions are widespread. For examples see Huang's classic text [11, p. 82] or the recent text [9, p. 24] and articles such as [12] or [24]. For the local equilibrium case see [27, p. 66]. These formal applications of Lagrange multiplier theory on unspecified function spaces do not address the differentiability of

¹ For Gibbs, the main feature of these distributions is that “when the system consists of parts with separate energies, the [distribution] of the separate parts are of the same nature,” [7, p. 33].

This work started while both authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California during the Fall 2018 semester, supported by NSF Grant No. DMS-1440140 and NSA Grant no. H98230-18-1-0269 respectively.

✉ S. Dostoglou
dostoglous@missouri.edu

¹ Department of Mathematics, University of Missouri, Columbia, MO 65211, USA

the entropy or of the constraints. They face the problem that strictly positive distribution functions do not tend to form open sets in function spaces and therefore it is not always clear how to differentiate.

The answer, of course, is to endow the relevant space of measures with a differentiable manifold structure. In order to apply Lagrange multiplier theory on Banach manifolds, as in [1] for example, we set the following minimum requirements for the manifold structure we seek:

- The information entropy must be at least C^1 ,
- all constraints must also be at least C^1 , and
- the manifold charts must map the measures we eventually work with into open sets.

We show here that a Banach manifold structure closely related to the one introduced in [22] and further developed in [3] satisfies these requirements. We can then apply standard theorems for constrained optimization on Banach manifolds as in [1] to conclude that all critical points¹ are indeed Gibbs distributions. For other constructions of statistical manifolds see [16, 18, 19, 25, 26].

Here we shall insist on a manifold structure on spaces of *finite* measures as in [3] and will impose the condition that we have probability measures as one of the constraints. Starting with *probability* measures would introduce an extra complication in the model space of functions (the functions would have to be “centered” as in [22, p. 1553]) making the derivative formulas more difficult to work with. It is also the custom in Physics to impose probability as a constraint.

Our main motivation is to characterize *local equilibrium Gibbs ensembles*, i.e. critical points of entropy under constraints that are *functions on space domains* rather than constants. We have found this to be a necessary step in the further development of ideas from Statistical Mechanics via Information Geometry, along the lines of [4] for example. Local equilibrium measures also play a prominent role both in Morrey’s seminal work on the derivation of hydrodynamic equations from microscopic dynamics [17], as well as in one of the most important advances on the same problem to date, [20]. For the role of local equilibrium ensembles in non-equilibrium statistical mechanics in general see [5, §3].

It turns out that the same manifold structure works for both equilibrium measures and local equilibrium measures: when the constraints are scalar the result is equilibrium Gibbs and scalar Lagrange multipliers (inverse temperature, chemical potential, etc.); when the constraints are functions the result is local equilibrium Gibbs and the multipliers are elements of the space dual to the function constraints (bounded functions). An argument completely analogous to the equilibrium case shows that the local equilibrium critical points are again minimizers. We find this to be one of the main features of the method here: changing the target of the constraints gives the correct multipliers while the manifold structure remains the same.

With these understood, the main results of this article are:

- the realization that the space of Definition 1 satisfies our requirements and

¹ We always use the term “critical point” in the sense of calculus of variations and not in the sense of phase transitions.

- Theorem 2 that characterizes local Gibbs equilibria in terms of entropy.

In Sect. 2 we show in a general setting how to define a manifold structure on finite measures on some arbitrary space so that both entropy and scalar constraints are continuously differentiable and we obtain the equilibrium Gibbs formulas for the critical points.

Section 3 contains the application of the general result to the equilibrium grand canonical Gibbs on the phase space of a physical system.

The case of the local equilibrium appears as Sect. 4.

As explained in [22, p. 1548], one is naturally led to model the space of finite positive measures on Orlicz spaces. Here, we shall always set

$$\Phi(x) = e^{|x|} - 1 \tag{1}$$

and we shall use the Orlicz space (for Young function Φ)

$$L^\Phi(\mu) := \left\{ f : \Omega \rightarrow \mathcal{R} \text{ measurable} : \int_\Omega \Phi(\alpha f(x))\mu(dx) < \infty \text{ for some } \alpha > 0 \right\} \tag{2}$$

The norm

$$\|f\|_{L^\Phi(\mu)} = \inf \left\{ k > 0 : \int_\Omega \Phi\left(\frac{f(x)}{k}\right)\mu(dx) \leq 1 \right\}, \tag{3}$$

renders $L^\Phi(\mu)$ a Banach space [23, Chapter 3, Theorem 10]. We shall rely on the fact that $L^\Phi(\mu)$, for this Φ , continuously embeds into $L^p(\mu)$ for all $p \geq 1$:

$$L^\Phi(\mu) \hookrightarrow \bigcap_{p \geq 1} L^p(\mu), \tag{4}$$

see [3, Equation 3.104], [22, Proposition 2.3].

To ensure that entropy is also continuously differentiable, we shall work on those components of the manifold consisting of measures that have the logarithm of their density (with respect to the fixed reference measure) in the corresponding Orlicz space. This should be compared to the discussion in [21, p. 4055] regarding the differentiability of entropy on the subspace of *probability* measures.

2 The general case for real-valued constraints

2.1 Manifold structure

For a fixed space Ω , with μ_0 fixed positive (but not necessarily finite) measure, and for

$$\mathbf{C} : \Omega \rightarrow \mathcal{R}^n; \quad \mathbf{C}(x) = (C_1(x), \dots, C_n(x)) \tag{5}$$

a measurable vector-valued function on Ω (out of which the constraints will be defined), we are interested in a manifold structure (eventually rendering information entropy and constraints differentiable) on the set of finite measures with strictly positive densities

$$\mathcal{F} := \left\{ \mu : \mu \ll \mu_0, \frac{d\mu}{d\mu_0} > 0, \int_{\Omega} |C_i(x)| \mu(dx) < +\infty, i = 0, 1, \dots, n \right\}. \quad (6)$$

Note that here, and for the rest of this article, we set $C_0(x) \equiv 1$ to include the finiteness of the measures. To construct the manifold structure we use (and adapt) the approach from [3, §3.3.3] for all of the finite measures $|C_i|\mu$ simultaneously:

Definition 1 Define

$$X_{\mu} = \bigcap_{i=0}^n L^{\Phi}(|C_i|\mu), \quad (7)$$

with norm

$$\|f\|_{X_{\mu}} = \max_{0 \leq i \leq n} (\|f\|_{L^{\Phi}(|C_i|\mu)}), \quad (8)$$

using the notation of (3).

A few remarks regarding the intersection of (7) are in order. The intersection is taken in the space of measurable functions. Given that all of the C_i 's do not vanish off a set of μ_0 measure zero, the convergence in each $L^{\Phi}(|C_i|\mu)$ implies pointwise convergence up to subsequence, therefore a sequence that is Cauchy with respect to each of the norms has the same limit in each space. (This argument was suggested by a diligent referee.)

Alternatively, use the standard metric for convergence in measure (see for example [8, p. 63, Exercise 32]) and apply [2, p. 57] to show that X_{μ} is Banach.

In any case, from now on we will be working under the following

Assumption For all i and μ_0 -almost all x :

$$C_i(x) \neq 0. \quad (9)$$

It will be easy to check that this assumption is satisfied in the applications that follow.

Definition 2 Let $\mu, \mu' \in \mathcal{F}$. Define $\mu' \leq \mu$ if $\mu' = \psi\mu$, $\psi \in L^p(|C_i|\mu)$ for some $p > 1$ and all $i = 0, 1, \dots, n$. Write $\mu' \sim \mu$ when both $\mu \leq \mu'$ and $\mu' \leq \mu$ hold.

Then \sim is an equivalence relation: this follows from the argument preceding Proposition 3.11 in [3, p. 178], repeated for each of the finite measures $|C_i|\mu$. Note that once we have a ψ_i in $L^{p_i}(|C_i|\mu)$ from this argument we can choose the smallest p_i to satisfy Definition 2.

X_μ will be the model space for each equivalence class of this equivalence relation and the equivalence classes will be connected components with the same tangent space with respect to the manifold structure we are after. More precisely:

Proposition 1 *If $\mu' \leq \mu$ then the identity map continuously embeds X_μ into $X_{\mu'}$. In particular, $\mu' \sim \mu$ implies $X_\mu = X_{\mu'}$ as Banach spaces, i.e. they consist of the same elements and the norms are equivalent.*

Proof Proposition 3.11 in [3, p. 178] shows the equality of each of the spaces in the intersection of Definition 2. □

Proposition 2 *For μ in \mathcal{F} and K the equivalence class of μ , the map*

$$s_\mu : \mu' = \psi \mu \mapsto \log \psi \tag{10}$$

has open and convex image $s_\mu(K)$ in X_μ .

Proof That the image is in X_μ follows directly from Definition 2 of the equivalence relation once the comments at the beginning of section 3.3.3 at [3, p. 176] are repeated for each of the spaces in the intersection (7).

Given this, the proof of [3, Theorem 3.4] applies verbatim to show that the image is open and convex. □

Remark 1 Both s_μ and its inverse on $s_\mu(K)$, given by $s_\mu^{-1}(u) = e^u \mu$, are continuous with respect to the following modification of the standard e-convergence, cf. [3, Definition 3.13 and Proposition 3.13]:

Definition 3 For $\{g_k \mu_0, k \in \mathcal{N}\}$ and $g \mu_0$ in \mathcal{F} we say that $g_k \mu_0$ is e-convergent to $g \mu_0$ if for all $i = 0, 1, \dots, n$ and all $p \geq 1$ we have

$$\lim_{k \rightarrow \infty} \int_\Omega \left| \left(\frac{g_k}{g} \right) - 1 \right|^p |C_i| g \mu_0 = 0, \quad \text{and} \quad \lim_{k \rightarrow \infty} \int_\Omega \left| \left(\frac{g}{g_k} \right) - 1 \right|^p |C_i| g \mu_0 = 0. \tag{11}$$

Changing from μ to $\mu' = \psi \mu$ in the same equivalence class replaces s_μ with a homeomorphism $s_{\mu'}$ from the same K to $X_{\mu'}$. The transition function

$$s_\mu^{-1} \circ s_{\mu'} : s_\mu(X_\mu) \rightarrow s_{\mu'}(X_{\mu'}) \tag{12}$$

between open sets in equivalent Banach spaces is

$$u \mapsto u + \log \psi, \tag{13}$$

clearly a C^∞ diffeomorphism.

This provides the Banach manifold structure we shall use: \mathcal{F} is a disjoint union of open, connected components, the equivalence classes, and every component is modeled by any of the equivalent Banach spaces X_μ , for μ in the corresponding equivalence

class, cf. [3, Remark 3.13]. Via the chart s_μ the tangent space $T_\mu\mathcal{F}$ of \mathcal{F} at μ (defined as equivalence classes of curves) can be identified with $X_\mu = \bigcap_{i=0}^n L^\Phi(|C_i|\mu)$, with the norm (8) or any of the equivalent norms $\|f\|_{X_{\mu'}}$, $\mu' \sim \mu$, via an isomorphism (the restriction on fibres of a bundle isomorphism $\bigcup_{\mu \in K} T_\mu\mathcal{F} \simeq s_\mu(K) \times X_\mu$ corresponding to s_μ , see [1, p. 155]).

2.2 Entropy and constraints

Consider now the information entropy²

$$\begin{aligned}
 H : \mathcal{F} &\rightarrow \mathcal{R} \cup \{+\infty\} \\
 H(\phi\mu_0) &= \int_\Omega \phi \log \phi \mu_0.
 \end{aligned}
 \tag{14}$$

It turns out that if $\mu' \sim \mu$ with $\mu = \phi\mu_0$ and $\mu' = \tau\mu_0$ then $\log \phi \in L^\Phi(\mu)$ iff $\log \tau \in L^\Phi(\mu')$ and $H(\phi\mu_0)$ is finite iff $H(\tau\mu_0)$ is, see [21, p. 4055, §4.3]. Recalling standard facts from [1, Corollary 2.4.10, Definition 3.2.5], to show that H is C^1 it is enough to show that $H \circ s_\mu^{-1}$ is C^1 -Gateaux on X_μ .

Proposition 3 *Let $\mu = \phi\mu_0$ be such that $\log \phi \in L^\Phi(\mu)$. Then H is a real valued C^1 function on K and*

$$d_0(H \circ s_\mu^{-1})[v] = \int (\log \phi(x) + 1)v(x)\phi(x)\mu_0(dx).
 \tag{15}$$

Proof Following [1, Corollary 2.4.20 and Definition 3.2.5] it is enough to check that $H \circ s_\mu^{-1}$ is C^1 -Gateaux on $s_\mu(K_\mu)$. For this, calculate

$$\begin{aligned}
 d_u(H \circ s_\mu^{-1})[v] &= \left. \frac{d}{dt} \right|_{t=0} (H \circ s_\mu^{-1})(u + tv) \\
 &= \int v e^u [1 + u + \log \phi] \mu(dx)
 \end{aligned}
 \tag{16}$$

for $u, v \in X_\mu$. Applying Hölder on the last two terms and using that $\log \phi \in L^\Phi(\mu)$ and (4) we get

$$\left| d_u(H \circ s_\mu^{-1})[v] \right| \leq \|u\|_{X_{\mu'}} \|v\|_{X_{\mu'}} + C \|v\|_{X_{\mu'}},
 \tag{17}$$

for $\mu' = e^u\mu$. Then use Proposition 1 to see that $d_u H$ is bounded. The continuous dependence on u follows from (16) with the use of (11). □

Definition 4 Define \mathcal{F}_e to be the union of all the components K of \mathcal{F} such that for $\mu = \phi\mu_0$ in K the condition $\log \phi$ in $L^\Phi(\mu)$ holds.

² For the relation between this information entropy and thermodynamic entropy, see formula (14) of [14, p. 63].

Define now, for C_i as in (5),

$$\begin{aligned}
 G_i : \mathcal{F} &\rightarrow \mathcal{R}, & G_i(\mu) &= \int C_i(x)\mu(dx) \\
 \mathbf{G} : \mathcal{F} &\rightarrow \mathcal{R}^{n+1}, & \mathbf{G}(\mu) &= (G_0(\mu), \dots, G_n(\mu)).
 \end{aligned}
 \tag{18}$$

Proposition 4 \mathbf{G} is C^1 on \mathcal{F} and

$$d_0(\mathbf{G} \circ s_\mu^{-1})[v] = \int v(x)\mathbf{C}(x)\mu(dx).
 \tag{19}$$

Proof First calculate

$$d_0(G_i \circ s_\mu^{-1})[v] = \left. \frac{d}{dt} \right|_{t=0} \int C_i e^{tv} \mu = \int C_i v \mu,
 \tag{20}$$

and repeat the argument of Proposition 3 for the measure $|C_i|\mu$ for each i . □

Remark 2 The Propositions above provide C^1 differentiability which is sufficient for applying Proposition 5 below. Higher differentiability is likely, but we shall not insist on it here.

Recall now that a closed subspace F of a Banach space E is “split” if there is a closed subspace H of E such that $E = F \oplus H$, where $F \oplus H$ is the direct sum of F and H (i.e. the product of F and H with the product topology and any of the equivalent norms that produce this topology).

Also recall that a map f between Banach manifolds M and N is a submersion if for each m in M , $d_m f$ is surjective with split kernel.

Lemma 1 $d_\mu \mathbf{G} : T_\mu \mathcal{F} \rightarrow \mathcal{R}^{n+1}$ is surjective for all μ in \mathcal{F} .

Proof If not, there is i such that for all v in X_μ

$$\int C_i(x) v(x)\mu(dx) = 0.
 \tag{21}$$

This implies $C_i \equiv 0$ as the characteristic function of the set $M = \{x : C_i(x) > 0\}$ is in X_μ , contradicting our standing Assumption (9). □

Lemma 2 $T_\mu \mathcal{F} = \ker d_\mu \mathbf{G} \oplus F$, with $F, \ker d_\mu \mathbf{G}$ closed subspaces.

Proof Since $d_\mu \mathbf{G}$ is continuous, it is standard that $\ker d_\mu \mathbf{G}$ is closed, cf. [8, Chapter 5, Exercise 15]). Take F to be the span of $y_0, y_1, \dots, y_n \in T_\mu \mathcal{F}$ such that $d_\mu \mathbf{G}[y_i] = \mathbf{e}_i$ for all i , and \mathbf{e}_i the standard basis vectors in \mathcal{R}^{n+1} . The statement follows from the fact that

$$P(x) = \sum_{i=0}^n d_\mu G_i[x]y_i
 \tag{22}$$

is a projection with F as fixed points and the characterization of direct sums in terms of projections as in [1, Proposition 2.2.17]. □

We are now ready to apply the following, see [1, Proposition 3.5.24]:

Proposition 5 *Let M and P be Banach manifolds, let $f : M \rightarrow \mathcal{R}$ be C^1 , and let $g : M \rightarrow P$ be a C^1 submersion. For $N = g^{-1}(p_0)$ for some p_0 in P a point n in N is a critical point of $f|_N$ if and only if there exists λ in $(T_{p_0}P)^*$, a Lagrange multiplier, such that $d_n f = \lambda \circ d_n g$.*

Recalling Definition 4, we then have

Theorem 1 *Let $\mathbf{G} : \mathcal{F} \rightarrow \mathcal{R}^{n+1}$ as in (18). Then $\mu = \phi\mu_0$ is a critical point of the Gibbs entropy H as in (14) in $\mathbf{G}^{-1}(\gamma) \cap \mathcal{F}_e$, for some γ in the image of \mathbf{G} and \mathcal{F}_e as in Definition 4, if and only if there exist λ_i 's in \mathcal{R} such that*

$$\phi = \frac{\exp\left(\sum_{i=1}^n \lambda_i C_i(x)\right)}{\int_{\Omega} \exp\left(\sum_{i=1}^n \lambda_i C_i(x)\right) \mu_0(dx)}. \tag{23}$$

Proof Apply Proposition 5 to \mathbf{G} and H at $\phi\mu_0$ using the formulas for the derivative from (19) and (15) respectively to get $\lambda \in \mathcal{R}^{n+1}$ such that

$$\int_{\Omega} [\log \phi(x) + 1] v(x) \mu(dx) = \lambda \cdot \int_{\Omega} \mathbf{C}(x) v(x) \mu(dx) \tag{24}$$

for all $v \in X_{\mu}$. Equivalently,

$$\int_{\Omega} \left(\log \phi(x) + 1 - \sum_{i=0}^n \lambda_i C_i(x) \right) v(x) \mu(dx) = 0 \tag{25}$$

for all $v \in X_{\mu}$. As in the proof of Lemma 1, this implies

$$\log \phi(x) = -1 + \lambda_0 + \sum_{i=1}^n \lambda_i C_i(x). \tag{26}$$

Since the first constraint is 1, this implies the statement. □

Recall here the standard argument from [13] to see that the the Gibbs measure (23) is an entropy minimizer: for any other $\mu = \tau \mu_0$ satisfying the same constraints

$$\begin{aligned} H(\tau \mu_0) &= \int \left(\tau \log \frac{\tau}{\phi} + \tau \log \phi \right) \mu_0 \\ &\geq \int \tau \log \phi \mu_0 \\ &= \int \tau \sum \lambda_i C_i \mu_0 - \log Z = H(\phi \mu_0), \end{aligned} \tag{27}$$

for Z the denominator of (23).

3 The grand canonical ensemble in statistical mechanics

3.1 The grand canonical ensembles as entropy minimizers

We now apply the previous section to derive the grand canonical Gibbs measure in Statistical Mechanics as a critical point of Gibbs entropy under function constraints, as explained in the Introduction.

Recall first that a classical system of N particles in an open bounded domain Λ of \mathcal{R}^3 is described at any time by the positions q_1, \dots, q_N of the particles and the corresponding momenta p_1, \dots, p_N , with the q_i 's in Λ and the p_i 's in \mathcal{R}^3 . In classical statistical mechanics, see [7], one examines *ensembles* of systems, with each system in the ensemble occurring according to some probability law. This means that one works with $\Lambda^N \times (\mathcal{R}^3)^N$. And if one does not fix the number of particles then the relevant space is $\bigsqcup_{N=1}^{\infty} (\Lambda \times \mathcal{R}^3)^N$. At equilibrium the probability does not change. Prominent among equilibrium distributions are the canonical (for fixed N) and the grand canonical Gibbs distributions (when N is not fixed). We examine here the grand canonical - the argument for the canonical follows by restricting on a fixed N component.

Then let Λ be a finite volume, open set in \mathcal{R}^3 and \mathcal{X}_Λ the corresponding phase space

$$\mathcal{X}_\Lambda = \bigsqcup_{N \geq 1} (\Lambda^N \times \mathcal{R}^{3N}) = (\Lambda \times \mathcal{R}^3) \sqcup (\Lambda^2 \times \mathcal{R}^6) \sqcup \dots \tag{28}$$

Each of the components of this disjoint union has the topology from \mathcal{R}^{6N} . The underlying σ -algebra will be $\bigcup_{N \geq 1} \mathcal{B}(\Lambda^N \times \mathcal{R}^{3N})$ where $\mathcal{B}(\Lambda^N \times \mathcal{R}^{3N})$ is the Borel σ -algebra on $\Lambda^N \times \mathcal{R}^{3N}$.

The reference measure μ_0 on \mathcal{X}_Λ will be

$$\mu_0 = \sum_{N=1}^{\infty} \frac{dq_1 \dots dq_N}{N!} dp_1 \dots dp_N. \tag{29}$$

Then for f measurable on \mathcal{X}_Λ

$$\int_{\mathcal{X}_\Lambda} f \mu_0 = \sum_{N \geq 1} \int_{\Lambda^N \times \mathcal{R}^{3N}} f(q_1, \dots, q_N, p_1, \dots, p_N) \frac{dq_1 \dots dq_N}{N!} dp_1 \dots dp_N. \tag{30}$$

For more on \mathcal{X}_Λ see [15, p. 18].

To describe the constraints, start with U a *stable interaction potential* (measuring the potential energy of a particle configuration), i.e.

$$\begin{aligned} U : \mathcal{X}_A &\rightarrow \mathcal{R}, \text{ measurable and symmetric,} \\ (q_1, \dots, q_N) &\rightarrow U(q_1, \dots, q_N), \text{ for each } N \\ U(q_1, \dots, q_N) &\geq -NL, \end{aligned} \quad (31)$$

for some $L > 0$ and for all $N \in \mathcal{N}$, and define the total energy of a configuration $(\mathbf{q}, \mathbf{p}) = (q_1, \dots, q_N, p_1, \dots, p_N)$ as the sum of the potential and the kinetic energy

$$\begin{aligned} E : \mathcal{X}_A &\rightarrow \mathcal{R}, \\ E(q_1, \dots, q_N, p_1, \dots, p_N) &= \frac{1}{2} \sum_{i=1}^N |p_i|^2 + U(q_1, \dots, q_N). \end{aligned} \quad (32)$$

(We are implicitly setting all masses equal to 1.)

The total momentum and particle number are defined as

$$\mathbf{P} : \mathcal{X}_A \rightarrow \mathcal{R}^3, \mathbf{P}(q_1, \dots, q_N, p_1, \dots, p_N) = \sum_{i=1}^N p_i \quad (33)$$

and

$$\mathbf{N} : \mathcal{X}_A \rightarrow \mathcal{N}, \mathbf{N}(q_1, \dots, q_N, p_1, \dots, p_N) = N, \quad (34)$$

respectively. The constraints now are

$$\begin{aligned} \mathcal{E}[\mu] &:= \int_{\mathcal{X}_A} E(\mathbf{q}, \mathbf{p}) \mu(d\mathbf{q}, d\mathbf{p}) = e_0, \\ \mathcal{P}[\mu] &:= \int_{\mathcal{X}_A} \mathbf{P}(\mathbf{q}, \mathbf{p}) \mu(d\mathbf{q}, d\mathbf{p}) = \mathbf{p}_0, \\ \tilde{\mathbf{N}}[\mu] &:= \int_{\mathcal{X}_A} \mathbf{N}(\mathbf{q}, \mathbf{p}) \mu(d\mathbf{q}, d\mathbf{p}) = n_0, \end{aligned} \quad (35)$$

with derivatives

$$\begin{aligned} d_\mu \mathcal{E}[v] &= \int_{\mathcal{X}_A} E(\mathbf{q}, \mathbf{p}) v(\mathbf{q}, \mathbf{p}) \mu(d\mathbf{q}, d\mathbf{p}), \\ d_\mu \mathcal{P}[v] &= \int_{\mathcal{X}_A} \mathbf{P}(\mathbf{q}, \mathbf{p}) v(\mathbf{q}, \mathbf{p}) \mu(d\mathbf{q}, d\mathbf{p}), \\ d_\mu \tilde{\mathbf{N}}[v] &= \int_{\mathcal{X}_A} \mathbf{N}(\mathbf{q}, \mathbf{p}) v(\mathbf{q}, \mathbf{p}) \mu(d\mathbf{q}, d\mathbf{p}), \end{aligned} \quad (36)$$

respectively. The Gibbs grand canonical distribution for paramaters β, λ, ν is defined as $\phi_{\beta, \lambda, \nu} \mu_0$ with

$$\phi_{\beta, \lambda, \nu}(\mathbf{q}, \mathbf{p}) = \frac{\exp(\beta E(\mathbf{q}, \mathbf{p}) + \lambda \cdot \mathbf{P}(\mathbf{q}, \mathbf{p}) + \nu N(\mathbf{q}, \mathbf{p}))}{Z_{\beta, \lambda, \nu}}, \tag{37}$$

$$Z_{\beta, \lambda, \nu} = \int_{\mathcal{X}_A} e^{\beta E + \lambda \cdot \mathbf{P} + \nu E} \mu_0.$$

We now show that the requirement of Theorem 1 does not eliminate the Gibbs grand canonical distribution.

Proposition 6 For U stable potential as in (31) and $\beta < 0$, $\log \phi_{\beta, \lambda, \nu}$ is in $L^\Phi(\phi_{\beta, \lambda, \nu} \mu_0)$.

Proof As in [3, Proposition 3.9 of §3.3.2] it suffices to show $\log \phi_{\beta, \lambda, \nu}$ is in $L^{\cosh(t)-1}(\phi_{\beta, \lambda, \nu} \mu_0)$. We show there exists $\varepsilon > 0$ such that

$$\int_{\mathcal{X}_A} (\cosh(\varepsilon \log \phi_{\beta, \lambda, \nu}) - 1) \phi_{\beta, \lambda, \nu} \mu_0 < +\infty. \tag{38}$$

For this, after disposing of constants, it is enough to show that there exists $\varepsilon > 0$ such that

$$\int_{\mathcal{X}_A} \left(e^{(1+\varepsilon)[\beta E + \lambda \cdot \mathbf{P} + \nu N]} + e^{(1-\varepsilon)[\beta E + \lambda \cdot \mathbf{P} + \nu N]} \right) \mu_0 < +\infty. \tag{39}$$

For the second term in the integral complete the square to get

$$\int_{\mathcal{X}_A} e^{(1-\varepsilon) \left[\beta U + \sum_{i=1}^N \frac{\beta}{2} \left(p_i + \frac{\lambda}{2\beta} \right)^2 - \frac{\lambda^2}{2\beta} + \nu N \right]} \mu_0 \tag{40}$$

and integrate with respect to $dp_1 \dots dp_N$ to get, up to constant, when $\varepsilon < 1$,

$$\sum_N \left(\frac{2\pi}{(1-\varepsilon)\beta} \right)^{N/2} e^{(1-\varepsilon)\lambda^2/2\beta} e^{(1-\varepsilon)\nu N} \int_{\Lambda^N} e^{(1-\varepsilon)\beta U} \frac{dq_1 \dots dq_N}{N!}. \tag{41}$$

The stability condition gives

$$\begin{aligned} \int_{\Lambda^N} e^{(1-\varepsilon)\beta U} \frac{dq_1 \dots dq_N}{N!} &\leq \int_{\Lambda^N} e^{(1-\varepsilon)\beta LN} \frac{dq_1 \dots dq_N}{N!} \\ &= \frac{e^{(1-\varepsilon)\beta LN} |\Lambda|^N}{N!} < +\infty, \end{aligned} \tag{42}$$

where $|\Lambda|$ denotes the Lebesgue measure of Λ . Therefore (41) is finite for any $\varepsilon < 1$. The argument repeats for the first term in (39), this time without restrictions on ε . \square

Apply Theorem 1 to get

Corollary 1 *A probability measure μ is a critical point for the entropy H amongst all probability measures in \mathcal{F}_e with fixed \mathcal{E} , \mathcal{P} , and \mathbf{N} if and only if μ is a grand canonical Gibbs $\phi_{\beta,\lambda,\nu}\mu_0$ for some β, λ, ν .*

3.2 The range of the constraints

For applications, it is important to know the possible values that we can give the constraints (35). An answer to this was given in [6], without the use of any manifold structure, where an intricate step was to show that the set of all such values is open. Given what we have already developed here, we can show directly the openness of this set, in the general setting of the previous section.

Proposition 7 *Let*

$$\mathbf{G} : \mathcal{F} \rightarrow \mathcal{R}^{n+1}, \quad \mathbf{G}[\mu] = \int \mathbf{C}(x)\mu(dx), \tag{43}$$

for \mathbf{C} as in (5). Then \mathbf{G} has an open image.

Proof We have already established the surjectivity of $d_\mu \mathbf{G}$ under the assumptions of the Proposition in 1. The openness then follows from the local surjectivity theorem [1, Proposition 2.5.9]. □

The uniqueness of (β, λ, ν) given (e_0, \mathbf{p}_0, n_0) was also shown in [6, Proposition 1] using the Legendre transform of H (the log-partition function). In future work we shall address this point in the framework of information geometry.

4 The local equilibrium

The main application of what we have developed so far is that we can rigorously address the case of local equilibrium Gibbs measures, i.e. we can find critical points of the same entropy function subject to different constraints to get measures on the X_A of the form

$$\frac{1}{Z} \exp \left\{ \sum_{i=1}^N \beta(q_i)E(q_1, \dots, q_N, p_1, \dots, p_N) + \sum_{i=1}^N \nu(q_i) + \sum_{i=1}^N \lambda(q_i)p_i \right\} \mu_0, \tag{44}$$

for Z the partition function (normalizing constant).

To achieve this, we start from the same space of measures, with the difference that now the constraints are *functions* rather than constants, and the (unaltered) definition of \mathcal{F} renders these functions *integrable*. The function constraints are defined as follows, where we always use the max norm on the product of Banach spaces: let

$$C(q_1, \dots, q_N, p_1, \dots, p_N) = (E, \mathbf{P}, \mathbf{N})(q_1, \dots, q_N, p_1, \dots, p_N), \tag{45}$$

for $E, \mathbf{P}, \mathbf{N}$ as in (32), (33), (34), and define now

$$\mathbf{G} : \mathcal{F} \rightarrow \mathcal{R}^+ \times L^1(\Lambda, dx) \times \dots \times L^1(\Lambda, dx),$$

$$\mathbf{G}(\mu)(x) = \left(\int \mu(d\mathbf{q}, d\mathbf{p}), \int \sum_{i=1}^N C(q_1, \dots, q_N, p_1, \dots, p_N) \delta_x(q_i) \mu(d\mathbf{q}, d\mathbf{p}) \right), \tag{46}$$

where we use the Dirac $\delta_x(q_i)$ merely as a shorthand for replacing x for q_i and integrating with respect to all q_j 's except q_i . The result is a function of x . Integrating with respect to the x variable restores the missing q_i integration and the definition of \mathcal{F} shows that the last five components of \mathbf{G} are indeed in $L^1(\Lambda)$ with respect to dx .

Proposition 8 \mathbf{G} is C^1 on \mathcal{F} with

$$d_0(\mathbf{G} \circ s_\mu^{-1}) : X_\mu \rightarrow \mathcal{R}^+ \times L^1(\Lambda, dx) \times \dots \times L^1(\Lambda, dx)$$

$$d_0(\mathbf{G} \circ s_\mu^{-1})[v] = \left(\int_{X_\Lambda} v \mu(d\mathbf{q}, d\mathbf{p}), \int_{X_\Lambda} \sum_{i=1}^N C v \delta_x(q_i) \mu(d\mathbf{q}, d\mathbf{p}) \right). \tag{47}$$

Proof As in the proof of Proposition 4,

$$d_0(G_j \circ s_\mu^{-1})[v] = \frac{d}{dt} \Big|_{t=0} \int_{X_\Lambda} \sum_{i=1}^N C_j e^{tv} \delta(q_i - x) \mu$$

$$= \int_{X_\Lambda} \sum_{i=1}^N C_j v \delta(q_i - x) \mu. \tag{48}$$

Then

$$\|d_0(G_j \circ s_\mu^{-1})[v]\|_{L^1(dx)} \leq \int_\Lambda \int_{X_\Lambda} \sum_{i=1}^N |C_j| |v| \delta(q_i - x) \mu dx$$

$$\leq \sum_{N \geq 1} N \int |v| |C_j| \mu. \tag{49}$$

Use now (4) and apply Proposition 1 for $\mu \sim N\mu$ to conclude as in Proposition 3. \square

Define now \mathcal{R} to be the set of regular values of \mathbf{G} , i.e. values of \mathbf{G} with pre-image consisting of measures where the derivative of \mathbf{G} is surjective with split kernel.

Theorem 2 A measure μ is a critical point for the entropy H as in (14) amongst all measures $\mu' = \phi\mu_0$ in \mathcal{F}_e which satisfy $\mathbf{G}(\mu') = (1, f_1, \dots, f_5) \in \mathcal{R}$ if and only if μ is a local equilibrium Gibbs probability measure as in (44) for some $L^\infty(\Lambda, dx)$ functions β, λ, v .

Proof Let $\mu = \phi\mu_0$ be a critical point. By Theorem 5, there exist

$$(\lambda_0, \beta(x), \lambda(x), v(x)) \in (\mathcal{R} \times L^1(\Lambda, dx) \times \dots \times L^1(\Lambda, dx))^*$$

$$= \mathcal{R} \times L^\infty(\Lambda, dx) \times \dots \times L^\infty(\Lambda, dx) \tag{50}$$

(for the sum norm on the last product) such that

$$\begin{aligned} \int_{X_A} (\log \phi + 1) v \phi \mu_0 &= \lambda_0 \int_{X_A} v \mu(d\mathbf{q}, d\mathbf{p}) \\ &+ \int_A \beta(x) \int_{X_A} \sum_i E v \delta_x(q_i) \mu(d\mathbf{q}, d\mathbf{p}) dx \\ &+ \int_A \lambda(x) \int_{X_A} \sum_i \mathbf{P} v \delta_x(q_i) \mu(d\mathbf{q}, d\mathbf{p}) dx \\ &+ \int_A v(x) \int_{X_A} \sum_i \mathbf{N} v \delta_x(q_i) \mu(d\mathbf{q}, d\mathbf{p}) dx, \end{aligned} \quad (51)$$

for all v . Replace q_i for x to get

$$\int_{X_A} \left(\log \phi + 1 - \lambda_0 - \sum_{i=1}^N \beta(q_i) E - \sum_{i=1}^N \lambda(q_i) \mathbf{P} - \sum_{i=1}^N v(q_i) \mathbf{N} \right) v \mu = 0, \quad (52)$$

for all v . As in the proof of Theorem 1, for $\int \mu = 1$ this implies

$$\phi = \frac{1}{Z} \exp \left(\sum_{i=1}^N \beta(q_i) E + \sum_{i=1}^N \lambda(q_i) \mathbf{P} + \sum_{i=1}^N v(q_i) \mathbf{N} \right). \quad (53)$$

□

An argument completely analogous to the equilibrium case shows that the critical points are again minimizers.

In the case of the local equilibrium it turns out to be substantial more difficult to check the conditions for a regular value, i.e. surjectivity and split kernel. As in corresponding finite dimensional cases, regular values are dense provided certain conditions are satisfied, see [1, Appendix E]. In future work we address the problem of characterizing the elements of \mathcal{R} .

Acknowledgements This article is based on a part of the Ph.D. Thesis of the second author [10], who would like to thank Prof. G. Pistone for prompt and helpful communications. Both authors thank Jianfei Xue for constructive discussions

Declarations

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

References

1. Abraham, R., Marsden, J., Ratiu, T.: Manifolds, Tensor Analysis, and Applications. Addison-Wesley, Boston (1983)

2. Aronszajn, N., Gagliardo, E.: Interpolation spaces and methods. *Ann. Mat. Pura Appl.* **68**, 51–118 (1965)
3. Ay, N., Jost, J., Vˆan Lˆe, H., Schwachhöfer, L.: *Information Geometry*. Springer, Berlin (2017)
4. Balian, R., Alhassid, Y., Reinhardt, H.: Dissipation in many-body systems: a geometric approach based on information theory. *Phys. Rep.* **131**, 1–46 (1986)
5. De Masi, A., Ianiro, N., Pellegrinotti, A., Presutti, E.: A survey of the hydrodynamic behavior of many-particle systems. In: Lebowitz, J.L., Montroll, E.W. (eds.) *Nonequilibrium Phenomena II*, from Stochastics to Hydrodynamics, pp. 124–294. North-Holland, Amsterdam (1984)
6. Dostoglou, S., Hughes, A., Xue, Jianfei: On entropy minimization and convergence. *J. Stat. Phys.* **177**(3), 485–505 (2019)
7. Gibbs, J.W.: *Elementary Principles In Statistical Mechanics*. Ox Bow Press, Connecticut (1981)
8. Folland, G.: *Real Analysis; Modern Techniques and Their Applications*, 2nd edn. Wiley Inter-Science, New York (1999)
9. Friedli, S., Velenik, Y.: *Statistical Mechanics of Lattice Systems*. Cambridge University Press, Cambridge (2018)
10. Hughes, A.M.: Ph.D. Thesis, University of Missouri (2021)
11. Huang, K.: *Statistical Mechanics*. Wiley, New York (1987)
12. Jaynes, E.T.: Information theory and statistical mechanics. *Phys. Rev.* **106**(4), 620–630 (1957)
13. Kullback, S.: Certain inequalities in information theory and the Cramer-Rao inequality. *Ann. Math. Stat.* **25**(4), 745–751 (1954)
14. Lanford, O.E.: Entropy and equilibrium states in classical statistical mechanics. In: Lenard A., et al. (eds.) *Statistical mechanics and mathematical problems*, pp. 1–113. Springer, Basel (1973)
15. Lanford, O.E.: Time evolution of large classical systems. In: Moser J. (eds.) *Dynamical Systems, Theory and Applications*, pp. 1–111. Springer, Berlin, Heidelberg, New York (1975)
16. Loaiza, G., Quiceno, H.R.: A q -exponential statistical Banach manifold. *J. Math. Anal. Appl.* **398**, 466–476 (2013)
17. Morrey, C.B.: On the derivation of the equations of hydrodynamics from statistical mechanics. *Commun. Pure Appl. Math.* **8**, 279–326 (1955). ((*Proc. Nat. Acad. Sci. U.S.A.* **40**, 317–322 (1954)))
18. Newton, N.J.: An infinite-dimensional statistical manifold modeled on Hilbert space. *J. Funct. Anal.* **263**, 1661–1681 (2012)
19. Newton, N.J.: Infinite-dimensional statistical manifolds based on a balanced chart. *Bernoulli* **22**, 711–731 (2016)
20. Olla, S., Varadhan, S.R.S., Yau, H.-T.: Hydrodynamical limit for a Hamiltonian system with weak noise. *Commun. Math. Phys.* **155**(3), 523–560 (1993)
21. Pistone, G.: Examples of the application of nonparametric information geometry to statistical physics. *Entropy* **15**, 4042–4065 (2013)
22. Pistone, G., Sempì, C.: An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Stat.* **23**(5), 1543–1561 (1995)
23. Rao, M.M., Ren, Z.D.: *Theory of Orlicz Spaces*. Dekker, Prairie City (1991)
24. Sanov, I.N.: On the probability of large deviations of random variables. *Mat. Sbornik* **42**(84), 11–44 (1957). ((**No. 1**))
25. Santacroce, M., Siri, P., Trivellato, B.: New results on mixture and exponential models by Orlicz spaces. *Bernoulli* **22**(3), 1431–1447 (2016)
26. Vigelis, R.F., Cavalcante, C.C.: On ϕ -families of probability distributions. *J. Theor. Probab.* **26**, 870–884 (2013)
27. Zubarev, D., Morozov, V., Röpke, G.: *Statistical Mechanics of Nonequilibrium Processes*. Vol. 1. Basic Concepts, Kinetic Theory. Akademie Verlag, Berlin (1996)