

**ENTROPY MINIMIZATION, CONVERGENCE, AND
GIBBS ENSEMBLES (LOCAL AND GLOBAL)**

A Dissertation
presented to
the Faculty of the Graduate School
University of Missouri

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
ALEXANDER M. HUGHES
Dr. Stamatis Dostoglou, Dissertation Advisor

May 2021

©Copyright by Alexander Hughes 2021
All Rights Reserved

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled

ENTROPY MINIMIZATION, CONVERGENCE, AND
GIBBS ENSEMBLES (LOCAL AND GLOBAL)

presented by Alexander M. Hughes,

a candidate for the degree of Doctor of Philosophy and hereby certify that in their opinion it is worthy of acceptance.

Professor Stamatis Dostoglou

Professor Adam Helfer

Professor Jan Segert

Professor Giovanni Vignale

Professor Samuel Walsh

ACKNOWLEDGMENTS

I would like to express my gratitude to my dissertation advisor, Stamatis Dostoglou, for his support throughout the dissertation process. Without him, this dissertation would not have been possible.

All material in Part I is joint work completed with Stamatis Dostoglou and Jianfei Xue and has been published in the Journal of Statistical Physics (see [DHX]). All material in Part II is joint work completed with Stamatis Dostoglou and is currently under review with the journal Information Geometry.

Contents

Acknowledgments	ii
Abstract	vi
1 Introduction	1
I Convex Analysis Approach to Statistical Mechanics	3
2 Introduction	4
3 Minimum Entropy on Bounded Domains	9
3.1 Measures and Notation	9
3.2 Information Entropy	12
3.3 Constrained Entropy Minimization on Fixed Volumes	13
3.4 Homeomorphism between \mathcal{J} and the set of solvability \mathfrak{S}_Λ	15
3.5 Characterization of the set of solvability \mathfrak{S}_Λ	16
3.6 Essential Range of (N, \vec{P}, H)	16
3.7 Results on the set of solvability \mathfrak{S}_Λ	19
3.8 Information Entropy and log-Partition Function	23
4 Thermodynamic Limits	25

4.1	Microcanonical Entropy	26
4.2	Pressure	27
4.3	Strict Convexity of Pressure	29
5	Convergence	32
5.1	Convergence of Epigraphs	32
5.2	Comparison with Maximum Likelihood Estimators	36
5.3	Comparison with Lanford	38
6	Local Homeomorphism at the Thermodynamic Limit	39
 II Differential Manifold Approach to Statistical Mechanics		
	ics	41
7	Introduction	42
7.1	Orlicz Spaces	44
7.2	An equivalence relation	46
7.3	e-convergence	49
8	Finite Measures	54
8.1	Preliminaries	54
8.2	Manifold Structure	62
9	Lagrange Multipliers, Equilibrium Case	75
9.1	Preliminaries	75
9.2	Differentiability of H	77

9.3	\mathbf{G} is C^1 Mapping	85
9.4	\mathbf{G} is a Submersion	88
10	Application: The Grand Canonical	94
10.1	The Grand Canonical Gibbs distribution	94
11	Constraint Functionals as an Open Mapping	99
11.1	Constraint Functionals and the Local Surjectivity Theorem	99
12	Lagrange Multipliers, Local Equilibrium Case	102
12.1	The Local Grand Canonical Gibbs Distribution	102
III	Appendix	113
13	Appendix	114
13.1	Orlicz Space Propositions	114
13.2	ϵ -convergence	116
13.3	Differentiation on a Banach Manifold	118
13.4	Local Surjectivity Theorem	119
13.5	Proof of Lagrange Multiplier Theorem	124
13.6	Miscellaneous	125
	Bibliography	128
	Vita	133

Entropy minimization, convergence, and
Gibbs ensembles (local and global)

Alexander M. Hughes

Dr. Stamatis Dostoglou, Dissertation Advisor

ABSTRACT

We approach the subject of Statistical Mechanics from two different perspectives. In Part I we adopt the approach of Lanford and Martin-Lof. We examine the minimization of information entropy for measures on the phase space of bounded domains, subject to constraints that are averages of grand canonical distributions. We describe the set of all such constraints and show that it equals the set of averages of all probability measures absolutely continuous with respect to the standard measure on the phase space. We also investigate how the set of constraints relates to the domain of the microcanonical thermodynamic limit entropy. We then show that, for fixed constraints, the parameters of the corresponding grand canonical distribution converge, as volume increases, to the corresponding parameters (derivatives, when they exist) of the thermodynamic limit entropy.

In Part II, we use the Banach manifold structure on the space of finite positive measures to show that the critical points of the Gibbs entropy are grand canonical equilibria when the constraints are scalar, and local equilibria when the constraints are integrable functions. This provides a rigorous justification of the derivation of the Gibbs measures that appears often in literature.

Chapter 1

Introduction

The concept of information entropy and its optimization subject to constraints are central subjects in Statistical Mechanics. It is well known that in thermodynamic equilibrium, the minimizing probability distribution is the Gibbs distribution. Despite its prominence, no mathematically rigorous derivation of the Gibbs distribution from the perspective of Lagrange multipliers has been given. Instead, the proof of Gibbs as a minimizer utilizes convex analysis, see [K] for example. Further work in the fields of statistical mechanics and thermodynamics also tend to rely on convex analysis, as in [L] and [M-L].

In Part I we approach Statistical Mechanics from the perspective of convex analysis. In Chapter 3 we give the definitions of many of the basic quantities and give a characterization of \mathfrak{S}_Λ , the set of physical quantities that can be the expectation of a Gibbs distribution (the set of solvability). Furthermore, we can show the existence of a homeomorphism between \mathfrak{S}_Λ and set of possible thermodynamic parameters in the Gibbs distribution (the Lagrange Multipliers). Thus, each possible macroscopic quantity corresponds to a unique thermodynamic parameter and unique Gibbs distribution. We also show the image of the expectation values with respect to the Gibbs distributions maps to an open set. In Chapter 4, we introduce the concepts of

microcanonical entropy and thermodynamic pressure as well as prove several useful facts about these quantities. In Chapter 5 we use ideas of epi-convergence and the Legendre transform from convex analysis to prove the convergence of microcanonical entropy to thermodynamic entropy. We also show that when the microcanonical and thermodynamic entropies converge while corresponding to the same macroscopic quantities, we have convergence of the thermodynamic parameters as well. We conclude Part I by using the thermodynamic pressure to show the existence of a local homeomorphism between the thermodynamic parameters and macroscopic physical quantities in the thermodynamic limit.

In Part II, we change our approach and use the tools of differential geometry on Banach manifolds to provide a rigorous derivation of the Gibbs distribution as a Lagrange multiplier problem. After providing an overview of Orlicz spaces and ϵ -convergence in Chapter 7, we construct a manifold structure on the appropriate set so that both information entropy and the constraint functionals are C^1 mappings (Chapter 8). Chapter 9 consists of verifying the necessary conditions to solve a Lagrange multiplier problem on a Banach manifold. We then apply the manifold structure to derive the Grand Canonical Gibbs distribution in Chapter 10. Chapter 11 uses the manifold structure to provide an alternate proof of the open mapping result from Part I. Chapter 12 concludes Part II by applying the Banach manifold structure to the problem of deriving the local Gibbs distribution. In this setting the constraint functionals no longer map into \mathbb{R} but into a function space. The same manifold structure developed in Chapter 8 gives the necessary smoothness for the constraint functionals.

Part I

Convex Analysis Approach to Statistical Mechanics

Chapter 2

Introduction

For systems with a large number of components, information entropy optimization subject to a few given constraints is widely used in several fields, see [BKM] for example. In statistical mechanics, for bounded domains in space, one minimizes the information entropy of measures, absolutely continuous with respect to the standard measure on the phase space, subject to appropriate macroscopic constraints for density, velocity, and energy, see [J], [Z]. The minimizing measure, a grand-canonical Gibbs distribution, is characterized by parameters conjugate to the constraints, in the sense that differentiating the logarithm of the partition function of the measure with respect to these parameters returns the constraints. On the other hand, it is standard how, as the volume increases, one defines microcanonical entropy in terms of thermodynamic limits as a function of the given macroscopic quantities. Convex conjugation (or derivatives, when they exist) now define the thermodynamic parameters at the limit.

Motivated by:

- the need to know in advance in numerical simulations the range of constraints that can be used when minimizing information entropy on bounded domains, as well as

- the comparison of hydrodynamic equations from statistical mechanics on bounded domains, which heavily relies on minimizing information entropy (see [BAR] and [Z]), to the standard hydrodynamic limit approach (see [OVY]),

we ask the following:

- what are the appropriate constraints for a given domain and how does the set of these constraints change with domain?
- do the thermodynamic parameters for finite volume converge to the thermodynamic limit parameters while the macroscopic constraints remain fixed?

In this article we address these questions. In doing so we also obtain some other results which, to our knowledge, are new to the literature.

To start, in Chapter 3, we examine rigorously the constrained information entropy minimization over a bounded domain. It is standard that if we assume the constraints are averages of some grand canonical distribution this grand canonical distribution is the unique minimizer, cf. [Gibbs]. We give an explicit description of the set of solvability, \mathfrak{S}_Λ , of grand-canonical distributions on bounded domains, cf. Theorem 3.7.1. Furthermore, Theorem 3.7.3 asserts that this set in fact coincides with the averages of all measures that are absolutely continuous with respect to the standard measure on phase space (with the exception of a trivial case: the measure concentrated on the empty configuration). This shows that the usual assumption on the constraints is not restrictive. It is worth mentioning that the usual formal Lagrange multipliers argument (cf. [Z]) is not employed here. In fact, we find the application of Lagrange multipliers in this setting questionable, cf. Section 3.3.

We then proceed to thermodynamic limits in Chapter 4. After recalling definitions and standard properties of the thermodynamic limit microcanonical entropy and pressure we discuss the strict convexity property of the thermodynamic limit pressure. The arguments presented here rely on the Gibbs variational principle. Notice here that, by the convex conjugacy between pressure and microcanonical entropy, the strict convexity of pressure gives differentiability of entropy which plays a role in Chapters 5 and 6.

In Chapter 5 we answer the question of convergence of thermodynamic parameters. We start by showing in Theorem 5.1.1 that the domain of the thermodynamic limit entropy (where it is not $-\infty$) is nonempty and also give it an explicit characterization which improves the results in [L] and [M-L]. Then Theorem 5.1.2 shows that, whenever we fix macroscopic constraints in the thermodynamic limit entropy, the finite volume parameters of the corresponding grand-canonical distributions converge, as the domain tends to infinity, to the thermodynamic limit parameters corresponding to the same macroscopic quantities.

The main point here is that when the domain is bounded information entropy and the log-partition function are related via convex conjugation (Legendre-Fenchel transform). But at the thermodynamic limit the convex conjugate of pressure is microcanonical entropy. As pressure is the limit of log-partition functions, to relate information entropy (for macroscopic constraints that do not change as the domain increases) to microcanonical entropy (at the same macroscopic values), we need to know convergence of the conjugates to the conjugate of the limit. We show that the standard results for this type of convergence from convex analysis, involving

convergence of epigraphs of the functions in question, apply. In the presence of derivatives at the limit, already established in work by Georgii [Gi] for example, the results follow.

Related work on determining parameters of Gibbs states via maximum likelihood methods on bounded domains appears in [CG], [Gi], [DL]. We compare this to our results in section 5.2. Note that convergence of *microcanonical entropy* on bounded domains to the thermodynamic limit microcanonical entropy is known, see [RP]. In addition, an argument that appears in [L] and [M-L] shows that the information entropy of the *canonical* Gibbs distributions for a *fixed* inverse temperature converges to the microcanonical thermodynamic entropy evaluated at the energy corresponding to that fixed temperature. We compare this convergence to our results in section 5.3.

Finally, Chapter 6 uses strict convexity results from Chapter 4 to show that there is indeed a local homeomorphism between macroscopic and thermodynamic limit parameters. The existence of such a homeomorphism is often used, and is in fact a cornerstone of some seminal work [OVY, p. 530, p. 556]. For the 1-dimensional case see [LR, Proposition 5.3], [V, Theorem 10.2].

Sign Conventions: We do not use a negative sign for the coefficient of the energy in the exponent of the grand canonical distribution and we define information entropy to be the average of the logarithm of the distribution function. In this way the natural domain for the grand canonical distribution is on negative inverse temperatures and the information entropy on finite volumes here is that of [K], but the opposite of [Z]. These conventions avoid the awkward minus sign in front of the energy expectation

and render the conjugate of information entropy as the log-partition function. But, since we work with convex functions on finite volumes, at the limit we get the opposite of the concave microcanonical entropy.

We present our results for space dimension 3, but it will be clear that they hold in any dimension.

Chapter 3

Minimum Entropy on Bounded Domains

It is standard that a finite grand-canonical Gibbs distribution is the unique minimizer of information (Gibbs) entropy among all distributions with the same average as the grand canonical, see [Gibbs, p. 130]. The main point of this section is to describe the set of all such averages in Theorem 3.7.1 on bounded domains and how this set depends on the domain. Theorem 3.7.3 shows that this set includes the averages of all measures absolutely continuous with respect to the standard measure on phase space (with the exception of the measure concentrated on the empty configuration). The rest of the section collects standard results on information entropy, its minimizations, and the log-partition function, with proofs where these are not available in the references.

3.1 Measures and Notation

For $\Lambda \subset \mathbb{R}^3$ bounded and measurable set with volume $|\Lambda| > 0$, we work on the phase space

$$\mathfrak{X}_\Lambda = \bigcup_{n \geq 0} (\Lambda^n \times \mathbb{R}^{3n}) = \{\emptyset\} \sqcup (\Lambda \times \mathbb{R}^3) \sqcup (\Lambda^2 \times \mathbb{R}^6) \sqcup \dots \quad (3.1)$$

where $\{\emptyset\}$ indicates the empty configuration consisting of no particles. Unless otherwise specified, we assume the underlying σ -algebra to be $\bigcup_{n \geq 0} \mathcal{B}(\Lambda^n \times \mathbb{R}^{3n})$ where $\mathcal{B}(\Lambda^n \times \mathbb{R}^{3n})$ is the Borel σ -algebra on $\Lambda^n \times \mathbb{R}^{3n}$. In general q and p stand for the position and velocity of a particle, respectively. We take the reference measure on \mathfrak{X}_Λ to be

$$\omega = \sum_{n \geq 0} \omega_n(dq_1, \dots, dq_n, dp_1, \dots, dp_n) \quad (3.2)$$

where

$$\begin{aligned} \omega_0(\{\emptyset\}) &= 1, \\ \omega_n(dq_1, \dots, dq_n, dp_1, \dots, dp_n) &= \frac{1}{n!} dq_1 \dots dq_n dp_1 \dots dp_n, \quad n \geq 1. \end{aligned} \quad (3.3)$$

Then

$$\int_{\mathfrak{X}_\Lambda} f d\omega = \sum_{n \geq 0} \int_{\Lambda^n \times \mathbb{R}^{3n}} f(q_1, \dots, q_n, p_1, \dots, p_n) d\omega_n. \quad (3.4)$$

Throughout, \mathcal{P}_Λ denotes the space of probability measures on \mathfrak{X}_Λ .

For any configuration $(\tilde{q}, \tilde{p}) \in \mathfrak{X}_\Lambda$, the total energy $H(\tilde{q}, \tilde{p})$ consists of both kinetic energy and potential energy:

$$H : \mathfrak{X}_\Lambda \rightarrow \mathbb{R}, \quad (q_1, \dots, q_n, p_1, \dots, p_n) \mapsto \frac{1}{2} \sum_{i=1}^n |p_i|^2 + U(q_1, \dots, q_n). \quad (3.5)$$

Throughout this article we will assume the following for the interaction potential U :

- U is **stable** in the sense that there is $L > 0$ with

$$U(q_1, \dots, q_n) \geq -nL, \quad n \in \mathbb{N}. \quad (3.6)$$

- The potential energy of the empty or single particle configurations vanish:

$$U(\emptyset) = U(q_1) = 0.$$

- U is shift invariant: $U(q_1, \dots, q_n) = U(q_1 + h, \dots, q_n + h)$, for all $h \in \mathbb{R}^3$, and all $n \geq 1$.
- U is symmetric: $U(q_1, \dots, q_n) = U(q_{\sigma(1)}, \dots, q_{\sigma(n)})$ for any permutation σ and all $n \geq 1$.

An abundance of examples of interaction potentials satisfying these assumptions can be found in [R, pp. 34-39]. These include pairwise interaction potentials of the form

$$U(q_1, \dots, q_n) = \sum_{i \neq j} \Phi(|q_i - q_j|), \quad (3.7)$$

for appropriate $\Phi : [0, \infty) \rightarrow \mathbb{R}$. Chapter 4 here will require further assumptions on U .

Define also the particle number and total momentum by

$$\begin{aligned} N : \mathfrak{X}_\Lambda &\rightarrow \mathbb{N}, & (q_1, \dots, q_n, p_1, \dots, p_n) &\mapsto n, \\ \vec{P} : \mathfrak{X}_\Lambda &\rightarrow \mathbb{R}^3, & (q_1, \dots, q_n, p_1, \dots, p_n) &\mapsto \sum_{i=1}^n p_i. \end{aligned} \quad (3.8)$$

In the above definitions, for the empty configuration, i.e. when $n = 0$, we take $(N, \vec{P}, H) = (0, \vec{0}, 0)$. For $(\mu, \vec{\lambda}, \beta)$ in $\mathfrak{J} := \mathbb{R} \times \mathbb{R}^3 \times \mathbb{R}_{<0}$, define

$$g_{(\mu, \vec{\lambda}, \beta), \Lambda}(\tilde{q}, \tilde{p}) = \frac{1}{Z_{(\mu, \vec{\lambda}, \beta), \Lambda}} \exp\left((\mu, \vec{\lambda}, \beta) \cdot (N, \vec{P}, H)(\tilde{q}, \tilde{p})\right), \quad (3.9)$$

where $Z_{(\mu, \vec{\lambda}, \beta), \Lambda}$ is the normalization constant (partition function):

$$Z_{(\mu, \vec{\lambda}, \beta), \Lambda} = \int_{\mathfrak{X}_\Lambda} \exp\left((\mu, \vec{\lambda}, \beta) \cdot (N, \vec{P}, H)(\tilde{q}, \tilde{p})\right) \omega(d\tilde{q}, d\tilde{p}). \quad (3.10)$$

The above choice of $(\mu, \vec{\lambda}, \beta) \in \mathfrak{J}$ is justified by the following simple lemma.

Lemma 3.1.1. *Let $\mathfrak{J}_\Lambda \subset \mathbb{R} \times \mathbb{R}^3 \times \mathbb{R}$ be the set of $(\mu, \vec{\lambda}, \beta)$ such that $Z_{(\mu, \vec{\lambda}, \beta), \Lambda} < \infty$.*

Then $\mathfrak{J}_\Lambda = \mathfrak{J}$.

Proof. Complete the square for the p 's in the exponent and use the stability condition (3.6) to see that $\beta < 0$ is sufficient. Conversely, for $\beta \geq 0$, and regardless of $\vec{\lambda}$, the term of the integral for $N = 1$ gives

$$\int_{\Lambda \times \mathbb{R}^3} \exp\left(\mu + \vec{\lambda} \cdot \vec{p} + \beta \frac{|\vec{p}|^2}{2}\right) dqdp = +\infty. \quad (3.11)$$

□

For each $(\mu, \vec{\lambda}, \beta) \in \mathfrak{J}$, the probability measure $g_{(\mu, \vec{\lambda}, \beta), \Lambda} \omega$ is the Gibbs grand-canonical distribution over Λ with parameters $(\mu, \vec{\lambda}, \beta)$. We will use $\mathbb{E}_{(\mu, \vec{\lambda}, \beta), \Lambda}[F]$ for the expectation of a function F on \mathfrak{X}_Λ with respect to $g_{(\mu, \vec{\lambda}, \beta), \Lambda} \omega$, i.e.,

$$\mathbb{E}_{(\mu, \vec{\lambda}, \beta), \Lambda}[F] = \int_{\mathfrak{X}_\Lambda} F(\tilde{q}, \tilde{p}) \frac{\exp\left(\left(\mu, \vec{\lambda}, \beta\right) \cdot (N, \vec{P}, H)(\tilde{q}, \tilde{p})\right)}{Z_{(\mu, \vec{\lambda}, \beta), \Lambda}} \omega(d\tilde{q}, d\tilde{p}). \quad (3.12)$$

3.2 Information Entropy

For any $\nu \in \mathcal{P}_\Lambda$, define the information (Gibbs) entropy by

$$s_\Lambda(\nu) = \begin{cases} \frac{1}{|\Lambda|} \int_{\mathfrak{X}_\Lambda} (\log f) f d\omega, & \nu \ll \omega, f := \frac{d\nu}{d\omega} \\ +\infty & \text{o/w.} \end{cases} \quad (3.13)$$

In particular, if f_n is the density of ν with respect to ω_n , i.e. $\nu = \sum_{n \geq 0} f_n \omega_n$, then

$$s_\Lambda(\nu) = \frac{1}{|\Lambda|} \sum_{n \geq 0} \int_{\Lambda^n \times \mathbb{R}^{3n}} \log f_n f_n d\omega_n. \quad (3.14)$$

It is standard that s_Λ is bounded below by 0 (using the convexity of the function $t \mapsto t \log t$ and Jensen's inequality), therefore it is a proper¹ convex function on \mathcal{P}_Λ .

¹Recall that a convex function is proper if it never takes the value $-\infty$ and it is not identically $+\infty$.

3.3 Constrained Entropy Minimization on Fixed Volumes

In statistical mechanics it is standard to minimize the information entropy s_Λ over \mathcal{P}_Λ subject to suitable constraints

$$\int_{\mathfrak{x}_\Lambda} \frac{N}{|\Lambda|} d\nu = \rho, \quad \int_{\mathfrak{x}_\Lambda} \frac{\vec{P}}{|\Lambda|} d\nu = \vec{u}, \quad \int_{\mathfrak{x}_\Lambda} \frac{H}{|\Lambda|} d\nu = E, \quad (3.15)$$

and derive the grand canonical distributions as the minimizer. For this, it is common to argue via Lagrange multipliers, see for example [Z, p. 66]. This is a formal application of the Lagrange multiplier theory on spaces of functions. It is not at all clear to us how the Lagrange multipliers in infinite dimensions as for example in [Zei, §43.8], [BCM], can be used to make such arguments rigorous. (For example, one would need to choose a Banach space of functions in which the set of probability densities forms an open set.) On the other hand, such arguments can find solid ground if one employs some differential structure in the space of measures, as for example in [PS]. We shall come back to this point in Part II. For the moment we recall with minor modifications the argument from [K]: to minimize s_Λ , it clearly suffices to look only at measures absolutely continuous with respect to ω , i.e. $\nu = f\omega$. Assume there exists $(\mu, \vec{\lambda}, \beta) \in \mathfrak{J}$ such that

$$(\rho, \vec{u}, E) = \mathbb{E} \left[\frac{1}{|\Lambda|} (N, \vec{P}, H) \right], \quad (3.16)$$

an assumption that is not at all restrictive, see comments after equation (3.18) below.

Then for $g_{(\mu, \vec{\lambda}, \beta), \Lambda}$ the grand canonical density as in (3.9), by (3.16) we have that

$\int_{\mathfrak{X}_\Lambda} f \log g_{(\mu, \vec{\lambda}, \beta), \Lambda} d\omega$ is finite, therefore

$$\begin{aligned}
|\Lambda|s_\Lambda(f\omega) &= \int_{\mathfrak{X}_\Lambda} f \log \left(\frac{f}{g_{(\mu, \vec{\lambda}, \beta), \Lambda}} \right) d\omega + \int_{\mathfrak{X}_\Lambda} f \log(g_{(\mu, \vec{\lambda}, \beta), \Lambda}) d\omega \\
&= \int_{\mathfrak{X}_\Lambda} f \log \left(\frac{f}{g_{(\mu, \vec{\lambda}, \beta), \Lambda}} \right) d\omega + (\mu, \vec{\lambda}, \beta) \cdot (\rho, \vec{u}, E) - \log Z_{(\mu, \vec{\lambda}, \beta), \Lambda} \\
&\geq (\mu, \vec{\lambda}, \beta) \cdot (\rho, \vec{u}, E) - \log Z_{(\mu, \vec{\lambda}, \beta), \Lambda} = |\Lambda|s_\Lambda(g_{(\mu, \vec{\lambda}, \beta), \Lambda}\omega).
\end{aligned} \tag{3.17}$$

In the last inequality we have used that the relative entropy $\int_{\mathfrak{X}_\Lambda} f \log \left(\frac{f}{g_{(\mu, \vec{\lambda}, \beta), \Lambda}} \right) d\omega$ is always greater than or equal to zero (by Jensen's inequality). As the relative entropy is zero if and only if $f = g_{(\mu, \vec{\lambda}, \beta), \Lambda}$ a.e., we conclude the standard

Entropy Minimization: For (ρ, \vec{u}, E) such that there exists $(\mu, \vec{\lambda}, \beta)$ with (3.16) satisfied the grand-canonical distribution $g_{(\mu, \vec{\lambda}, \beta), \Lambda}\omega$ is the unique minimizer to s_Λ over all ν 's in \mathcal{P}_Λ satisfying (3.15).

Notice in the above process of entropy minimization the assumptions on the constraints (ρ, \vec{u}, E) : we require that $(\rho, \vec{u}, E) \in \mathfrak{S}_\Lambda$ (the set of solvability) where

$$\mathfrak{S}_\Lambda = \left\{ (\rho, \vec{u}, E) : \text{there is } (\mu, \vec{\lambda}, \beta) \in \mathfrak{I} \text{ such that (3.16) holds} \right\}. \tag{3.18}$$

In the next two sections, we provide a detailed description of the set of solvability \mathfrak{S}_Λ and will see that the assumption $(\rho, \vec{u}, E) \in \mathfrak{S}_\Lambda$ is in fact reasonable: if $(\rho, \vec{u}, E) = \int_{\mathfrak{X}_\Lambda} \frac{1}{|\Lambda|} (N, \vec{P}, H) d\nu$ for some $\nu \in \mathcal{P}_\Lambda$ absolutely continuous with respect to ω then $(\rho, \vec{u}, E) \in \mathfrak{S}_\Lambda$, unless ν is the Dirac measure at the empty configuration, see Theorem 3.7.3 and Theorem 3.7.1.

3.4 Homeomorphism between \mathfrak{J} and the set of solvability \mathfrak{S}_Λ

We now introduce the log-partition function:

$$\Phi_\Lambda(\mu, \vec{\lambda}, \beta) = \begin{cases} \frac{1}{|\Lambda|} \log Z_{(\mu, \vec{\lambda}, \beta), \Lambda}, & (\mu, \vec{\lambda}, \beta) \in \mathfrak{J} \\ +\infty, & \text{o/w.} \end{cases} \quad (3.19)$$

A straightforward application of Fatou's lemma shows that Φ_Λ is closed, in the sense that it has closed epigraph². A standard calculation shows that Φ_Λ is strictly convex on \mathfrak{J} . Therefore $\nabla\Phi_\Lambda$ defines a bijection from \mathfrak{J} to \mathfrak{S}_Λ , see [RV, Theorem B, p. 99]. That this bijection is in fact a homeomorphism follows from

Lemma 3.4.1. *Let f be a proper and closed convex function with domain an open set. If f is differentiable and strictly convex on its domain, then ∇f defines a homeomorphism on the image of ∇f .*

Proof. The domain of the subdifferential of f is between $\text{intdom}(f)$ and $\text{dom}(f)$, see [Rock, p. 253]. Since the domain is open, this means that the subdifferential is identical to ∇f on $\text{dom}(f)$ and empty everywhere else. Then f is essentially differentiable by [Rock, Theorem 26.1]. As f is also strictly convex, by assumption, the statement follows from [Rock, Theorem 26.5]. \square

Notice now that for $(\mu, \vec{\lambda}, \beta)$ in \mathfrak{J}

$$\mathbb{E} \left[\frac{1}{|\Lambda|} \left(N, \vec{P}, H \right) \right] = \frac{1}{|\Lambda|} \nabla_{\mu, \vec{\lambda}, \beta} \log Z_{(\mu, \vec{\lambda}, \beta), \Lambda} = \nabla_{\mu, \vec{\lambda}, \beta} \Phi_\Lambda. \quad (3.20)$$

We therefore have:

²Recall that the epigraph of f is $\{(x, a) : f(x) \leq a\}$.

Proposition 3.4.2. $\mathfrak{S}_\Lambda = \text{Image}(\nabla\Phi_\Lambda)$ and the map

$$(\mu, \vec{\lambda}, \beta) \mapsto \mathbb{E} \left[\frac{1}{|\Lambda|} (N, \vec{P}, H) \right] \quad (3.21)$$

defines a homeomorphism between \mathfrak{I} and \mathfrak{S}_Λ .

3.5 Characterization of the set of solvability \mathfrak{S}_Λ

In this section, we describe the set of solvability, \mathfrak{S}_Λ (see (3.18)). It is easy to find some rough bounds:

Lemma 3.5.1. $\mathfrak{S}_\Lambda \subset \{(\rho, \vec{u}, E) \in \mathbb{R} \times \mathbb{R}^3 \times \mathbb{R} : -L\rho \leq E, \rho > 0\}$, for L the stability constant as in (3.6).

Proof. For $\rho = \mathbb{E} \left[\frac{N}{|\Lambda|} \right]$, using the stability condition and working as in the proof of Lemma 3.1.1, it follows that $0 < \rho < +\infty$. Furthermore

$$\mathbb{E} \left[\frac{1}{|\Lambda|} H \right] \geq \mathbb{E} \left[\frac{1}{|\Lambda|} U \right] \geq \mathbb{E} \left[\frac{1}{|\Lambda|} (-LN) \right] = -L\rho, \quad (3.22)$$

after using (3.6) again. □

3.6 Essential Range of (N, \vec{P}, H)

Theorem 3.7.1 below will describe the set of solvability, \mathfrak{S}_Λ , in more detail after some preparation. Let f be a measurable map between some measure space X to \mathbb{R}^n . Recall first that the essential range of f is all points $x \in \mathbb{R}^n$ such that $\omega(f^{-1}(B)) > 0$, for any neighborhood B of x . (Alternatively, the essential range is the support of the distribution of f .) When f is a function on \mathfrak{X}_Λ or $\Lambda^n \times \mathbb{R}^{3n}$, unless otherwise specified, we assume the reference measure to be ω or ω_n , respectively.

For Λ fixed and for each $n \geq 0$, use the stability of U to define $L_\Lambda^{(n)}$ as the smallest constant satisfying

$$-nL_\Lambda^{(n)} \leq U(q_1, \dots, q_n), \text{ for almost all } (q_1, \dots, q_n) \in \Lambda^n \quad (3.23)$$

w.r.t. Lebesgue measure on Λ^n .

Notation 1. From now on \mathcal{E}_Λ denotes the essential range of $(N, \vec{P}, H)/|\Lambda|$ over \mathfrak{X}_Λ , \mathcal{C}_Λ the convex hull of \mathcal{E}_Λ , and $\mathring{\mathcal{C}}_\Lambda$ the interior of \mathcal{C}_Λ .

Similarly, for each $n \geq 0$, $\mathcal{E}_{\Lambda,n}$ denotes the essential range of $(N, \vec{P}, H)/|\Lambda|$ over $\Lambda^n \times \mathbb{R}^{3n}$ and $\mathcal{C}_{\Lambda,n}$ the convex hull of $\mathcal{E}_{\Lambda,n}$. Notice that \mathcal{C}_Λ is the convex hull of $\bigcup_{n \geq 0} \mathcal{C}_{\Lambda,n}$. Clearly, when $n = 0$, $\mathcal{C}_{\Lambda,0} = \mathcal{E}_{\Lambda,0} = \{(0, \vec{0}, 0)\}$.

Lemma 3.6.1. For each $n \geq 1$,

$$\mathcal{C}_{\Lambda,n} = \left\{ (\rho, \vec{u}, E)/|\Lambda| : \rho = n, E \geq -nL_\Lambda^{(n)}, |\vec{u}|^2 \leq 2n(E + nL_\Lambda^{(n)}) \right\}. \quad (3.24)$$

Proof. By definition of $L_\Lambda^{(n)}$, for ω_n -almost all $(\tilde{q}, \tilde{p}) \in \Lambda^n \times \mathbb{R}^{3n}$ we have

$$H(\tilde{q}, \tilde{p}) = \frac{1}{2} \sum_{i=1}^n |p_i|^2 + U(q_1, \dots, q_n) \geq \frac{1}{2} \sum_{i=1}^n |p_i|^2 - nL_\Lambda^{(n)}. \quad (3.25)$$

Using $|\sum_{k=1}^n p_i|^2 \leq n \sum_{i=1}^n |p_i|^2$ we have that

$$\mathcal{E}_{\Lambda,n} \subset \left\{ (\rho, \vec{u}, E)/|\Lambda| : \rho = n, E \geq -nL_\Lambda^{(n)}, |\vec{u}|^2 \leq 2n(E + nL_\Lambda^{(n)}) \right\}. \quad (3.26)$$

Let

$$\mathcal{I}_{\Lambda,n} = \left\{ (\rho, \vec{u}, E) : \rho = n, E \geq -nL_\Lambda^{(n)}, |\vec{u}|^2 = 2n(E + nL_\Lambda^{(n)}) \right\}. \quad (3.27)$$

It remains to prove that $\mathcal{I}_{\Lambda,n}$ belongs to the essential range of (N, \vec{P}, H) .

Take any $(n, \vec{u}_0, E_0) \in \mathcal{I}_{\Lambda,n}$. Fix any $\epsilon > 0$. We will show that the inverse image of the set

$$S_\epsilon = \{(n, \vec{u}, E) : |\vec{u} - \vec{u}_0| \leq \epsilon, |E - E_0| \leq \epsilon\} \quad (3.28)$$

has nonzero ω_n -measure.

As $-nL_\Lambda^{(n)}$ is the essential infimum of $U(q_1, \dots, q_n)$, we know that

$$A_\epsilon = \left\{ (q_1, \dots, q_n) \in \Lambda^n : 0 \leq U(q_1, \dots, q_n) + nL_\Lambda^{(n)} \leq \frac{\epsilon}{2} \right\} \quad (3.29)$$

has nonzero Lebesgue measure. Let

$$B_\epsilon = \left\{ (p_1, \dots, p_n) \in \mathbb{R}^n : \left| \sum_k p_k - \vec{u}_0 \right| \leq \epsilon, \left| \frac{1}{2} \sum_k |p_k|^2 - \frac{|\vec{u}_0|^2}{2n} \right| \leq \frac{\epsilon}{2} \right\}. \quad (3.30)$$

Notice that the image of $A_\epsilon \times B_\epsilon$ under (N, \vec{P}, H) is a subset of S_ϵ . It remains to show that B_ϵ has nonzero Lebesgue measure. In fact, take $\tilde{p}_0 \in \mathbb{R}^{3n}$ such that

$$\tilde{p}_0 = \frac{1}{n}(\vec{u}_0, \dots, \vec{u}_0). \quad (3.31)$$

Clearly, $\tilde{p}_0 \in B_\epsilon$. As $\sum_{k=1}^n p_k$ and $\sum_{k=1}^n |p_k|^2$ are continuous functions on \mathbb{R}^{3n} , then there exists $\delta > 0$ such that

$$B_{\epsilon, \delta} := \{ \tilde{p} \in \mathbb{R}^{3n} : |\tilde{p} - \tilde{p}_0| \leq \delta \} \subset B_\epsilon. \quad (3.32)$$

As $B_{\epsilon, \delta}$ has nonzero Lebesgue measure, the proof is now complete. \square

Remark 3.6.2. *In fact, for $n \geq 2$, the essential range $\mathcal{E}_{\Lambda, n}$ equals the convex hull $\mathcal{C}_{\Lambda, n}$. The proof is similar to that of Lemma 3.6.1. As this is not relevant to the rest of the article, we omit the details.*

Let $\overline{\mathcal{C}_\Lambda}$ be the closure of \mathcal{C}_Λ . As $\mathcal{C}_\Lambda = \text{conv} \left(\bigcup_{n \geq 0} \mathcal{C}_{\Lambda, n} \right)$, from the structure of $\mathcal{C}_{\Lambda, n}$ we see that if $(\rho, \vec{u}, E) \in \overline{\mathcal{C}_\Lambda}$, then $(\rho, \vec{u}, E + a) \in \overline{\mathcal{C}_\Lambda}$ for any $a > 0$. A useful consequence of this is: for any $(\rho_0, \vec{u}_0, E_0) \notin \overline{\mathcal{C}_\Lambda}$, we can find $(\alpha_0, \vec{\alpha}, \alpha_4)$ with $\alpha_4 > 0$ and $\varepsilon > 0$ so that for all $(\rho, \vec{u}, E) \in \overline{\mathcal{C}_\Lambda}$

$$[(\rho, \vec{u}, E) - (\rho_0, \vec{u}_0, E_0)] \cdot (\alpha_0, \vec{\alpha}, \alpha_4) > \varepsilon. \quad (3.33)$$

3.7 Results on the set of solvability \mathfrak{S}_Λ

Now we are ready to state

Theorem 3.7.1. *The set of solvability \mathfrak{S}_Λ is the interior of the convex hull of the essential range \mathcal{E}_Λ , i.e. $\mathfrak{S}_\Lambda = \mathring{\mathcal{C}}_\Lambda$.*

The rest of this section will give a proof for this theorem. To this end, we first prove a result for the essential range of (N, \vec{P}, H) under more general probability measures on \mathfrak{X}_Λ .

Lemma 3.7.2. *For $n \geq 1$ and ν any probability measure on $\Lambda^n \times \mathbb{R}^{3n}$ which is absolutely continuous with respect to ω_n , we have that the average of $(N, \vec{P}, H)/|\Lambda|$ under ν is in the relative interior of $\mathcal{C}_{\Lambda, n}$ (i.e. $\mathcal{C}_{\Lambda, n}$ is treated as a subset in \mathbb{R}^4) and therefore in the interior of \mathcal{C}_Λ (as a subset in \mathbb{R}^5).*

Proof. Let $Q_0 = (n, \vec{P}_0, H_0) = \int_{\Lambda^n \times \mathbb{R}^{3n}} (N, \vec{P}, H) d\nu$. Assume that $Q_0/|\Lambda|$ is on the boundary of $\mathcal{C}_{\Lambda, n}$ (as a subset of \mathbb{R}^4), then $Q_0 \in \mathcal{I}_{\Lambda, n}$. By the strict convexity of $\mathcal{C}_{\Lambda, n}$, there exists $\vec{\alpha} \in \mathbb{R}^5$ such that for all $Q \in \mathcal{C}_{\Lambda, n}$ such that $Q \neq Q_0$

$$\vec{\alpha} \cdot (Q - Q_0) > 0. \quad (3.34)$$

Notice that the inverse image of Q_0 with respect to the map (N, \vec{P}, H) is a subset of the set

$$\left\{ (\tilde{q}, \tilde{p}) \in \Lambda^n \times \mathbb{R}^{3n} : \sum_{k=1}^n p_k = \vec{P}_0 \right\} \quad (3.35)$$

which has zero ω_n -measure, therefore also zero ν -measure. We obtain now

$$\int_{\Lambda^n \times \mathbb{R}^{3n}} \vec{\alpha} \cdot ((N, \vec{P}, H) - Q_0) d\nu > 0, \quad (3.36)$$

clearly contradicting the fact $Q_0 = (n, \vec{P}_0, H_0) = \int_{\Lambda^n \times \mathbb{R}^{3n}} (N, \vec{P}, H) d\nu$.

On the other hand, if Q_0 is in the exterior of $\mathcal{C}_{\Lambda, n}$ (as a subset of \mathbb{R}^4), by convexity of $\mathcal{C}_{\Lambda, n}$, we can find $\vec{\alpha}$ such that (3.34) holds for all $Q \in \mathcal{C}_{\Lambda, n}$. Therefore we have (3.36) which leads again to a contradiction. \square

Theorem 3.7.3. *Let ν be any probability measure on \mathfrak{X}_Λ absolutely continuous with respect to ω . Assume that $\nu(\{\emptyset\}) < 1$, i.e. ν is not the Dirac measure on the empty configuration. Then the average of $(N, \vec{P}, H)/|\Lambda|$ with respect to ν is in the interior of \mathcal{C}_Λ .*

Proof. Let $Q = (\rho, \vec{u}, E)$ be the average of $(N, \vec{P}, H)/|\Lambda|$ with respect to ν . As ν is not the Dirac at the empty configuration, we know that Q is a convex combination of some family of $\{Q_{n_k}\}_k$, $n_k \geq 1$, for Q_n is the average of $(N, \vec{P}, H)/|\Lambda|$ conditioned on configurations of n particles.

From Lemma 3.7.2, we know that all Q_{n_k} are in the interior of \mathcal{C}_Λ , and then so is their convex combination Q . \square

For any $(\mu, \vec{\lambda}, \beta) \in \mathfrak{I}$, apply Theorem 3.7.3 to the grand-canonical distribution $g_{(\mu, \vec{\lambda}, \beta), \Lambda} \omega$ to conclude that $\mathfrak{S}_\Lambda \subset \mathring{\mathcal{C}}_\Lambda$. To prove Theorem 3.7.1, it remains to show that for any $(\rho, \vec{u}, E) \in \mathring{\mathcal{C}}_\Lambda$, there exists $(\mu, \vec{\lambda}, \beta) \in \mathfrak{I}$ such that (3.16) holds. The contents of the following Lemma appear in [L, Lemma A4.6] without proof:

Lemma 3.7.4. *For any $(N_0, \vec{P}_0, H_0) \in |\Lambda| \mathring{\mathcal{C}}_\Lambda$, i.e. the interior of the convex hull of the essential range of (N, \vec{P}, H) over \mathfrak{X}_Λ , there is ε such that for all unit vectors \vec{e} in \mathbb{R}^5 the following holds:*

$$\omega \left(\left\{ (\tilde{q}, \tilde{p}) \in \mathfrak{X}_\Lambda : \left[(N, \vec{P}, H)(\tilde{q}, \tilde{p}) - (N_0, \vec{P}_0, H_0) \right] \cdot \vec{e} > \varepsilon \right\} \right) > \varepsilon. \quad (3.37)$$

Proof. Whenever $(N_0, \vec{P}_0, H_0) \in |\Lambda| \mathring{\mathcal{C}}_\Lambda$, for any unit vector \vec{e} there is some (N_*, \vec{P}_*, H_*) in the essential range of (N, \vec{P}, H) that belongs to the half-space through (N_0, \vec{P}_0, H_0) with (inward) normal \vec{e} ; otherwise the whole convex hull of the essential range would be on one side of (N_0, \vec{P}_0, H_0) , i.e. (N_0, \vec{P}_0, H_0) would not be in its interior.

We then have

$$v := \left[(N_*, \vec{P}_*, H_*) - (N_0, \vec{P}_0, H_0) \right] \cdot \vec{e} > 0. \quad (3.38)$$

Let \mathcal{N} be a neighborhood of (N_*, \vec{P}_*, H_*) such that for all (N', \vec{P}', H') in \mathcal{N} we have

$$\left[(N', \vec{P}', H') - (N_0, \vec{P}_0, H_0) \right] \cdot \vec{e} > \frac{v}{2}. \quad (3.39)$$

Then the ω measure of the inverse image of \mathcal{N} is not zero, say $m > 0$. Take $\varepsilon = \min\{v/2, m\}$ and notice that this ε satisfies (3.37) for \vec{e} .

By continuity of projection, for the same \mathcal{N} , there is a neighborhood \mathcal{U} of \vec{e} on the unit sphere such that for any unit vector \vec{e}' in \mathcal{U} and any (N', \vec{P}', H') in \mathcal{N} we have

$$\left[(N', \vec{P}', H') - (N_0, \vec{P}_0, H_0) \right] \cdot \vec{e}' > \frac{v}{4}. \quad (3.40)$$

Redefining $\varepsilon = \min\{v/4, m\}$, we now have (3.37) satisfied on \mathcal{U} .

By compactness of the unit sphere, we only need to repeat this finitely many times and take the smallest ε to make sure that (3.37) is satisfied for all unit vectors. \square

We are ready to show that for any $(\rho, \vec{u}, E) \in \mathring{\mathcal{C}}_\Lambda$, there exists $(\mu, \vec{\lambda}, \beta) \in \mathfrak{I}$ such that (3.16) holds. This step builds on [L, pp. 44-47], see also [Kh, §16]. For $(\rho, \vec{u}, E) \in \mathring{\mathcal{C}}_\Lambda$ fixed, let $K : \mathfrak{I} \rightarrow \mathbb{R}_{>0}$ be given by

$$(\mu, \vec{\lambda}, \beta) \mapsto \int_{\mathfrak{X}_\Lambda} \exp \left((\mu, \vec{\lambda}, \beta) \cdot (N - \rho|\Lambda|, \vec{P} - |\Lambda|\vec{u}, H - |\Lambda|E) \right) d\omega. \quad (3.41)$$

Clearly, $\log K = |\Lambda| \left[\Phi_\Lambda - (\mu, \vec{\lambda}, \beta) \cdot (\rho, \vec{u}, E) \right]$ and therefore $\log K$ is strictly convex on \mathfrak{J} . Note that

$$\nabla_{\mu, \vec{\lambda}, \beta} \log K = 0 \Leftrightarrow (\rho, \vec{u}, E) = \nabla_{\mu, \vec{\lambda}, \beta} \Phi_\Lambda(\mu, \vec{\lambda}, \beta). \quad (3.42)$$

To conclude there is unique critical point of $\log K$ in \mathfrak{J} , it suffices to show that K , and therefore $\log K$, goes to $+\infty$ as $(\mu, \vec{\lambda}, \beta)$ approaches the boundary of \mathfrak{J} . Precisely, we show that $K(\mu, \vec{\lambda}, \beta) \rightarrow \infty$ if $|(\mu, \vec{\lambda}, \beta)| \rightarrow \infty$ or $(\mu, \vec{\lambda}, \beta) \rightarrow (\mu_0, \vec{\lambda}_0, 0)$ for any finite μ_0 and $\vec{\lambda}_0$.

To this end, we first apply Lemma 3.7.4 for

$$(N_0, \vec{P}_0, H_0) = |\Lambda|(\rho, \vec{u}, E) \text{ and } \vec{e} = (\mu, \vec{\lambda}, \beta) / |(\mu, \vec{\lambda}, \beta)| \quad (3.43)$$

to find $\varepsilon > 0$ such that for any $(\mu, \vec{\lambda}, \beta)$, there exists a set of \mathfrak{X}_Λ of ω measure at least ε where

$$(\mu, \vec{\lambda}, \beta) \cdot \left((N, \vec{P}, H) - |\Lambda|(\rho, \vec{u}, E) \right) > \varepsilon |(\mu, \vec{\lambda}, \beta)|. \quad (3.44)$$

It follows that

$$K(\mu, \vec{\lambda}, \beta) \geq \varepsilon \exp \left(|(\mu, \vec{\lambda}, \beta)| \varepsilon \right). \quad (3.45)$$

Therefore $K(\mu, \vec{\lambda}, \beta) \rightarrow +\infty$, as $|(\mu, \vec{\lambda}, \beta)| \rightarrow \infty$ while staying in \mathfrak{J} , and so does $\log K(\mu, \vec{\lambda}, \beta)$.

Next, if $(\mu_m, \vec{\lambda}_m, \beta_m) \rightarrow (\mu, \vec{\lambda}, 0)$, as $m \rightarrow \infty$, Fatou's lemma gives

$$\begin{aligned} & \liminf_{m \rightarrow \infty} \int_{\mathfrak{X}_\Lambda} \exp \left((\mu_m, \vec{\lambda}_m, \beta_m) \cdot (N - \rho|\Lambda|, \vec{P} - |\Lambda|\vec{u}, H - |\Lambda|E) \right) d\omega \\ & \geq \int_{\mathfrak{X}_\Lambda} \exp \left((\mu, \vec{\lambda}) \cdot (N - \rho|\Lambda|, \vec{P} - |\Lambda|\vec{u}) \right) d\omega = +\infty, \end{aligned} \quad (3.46)$$

showing that $\log K$ explodes when $(\mu, \vec{\lambda}, \beta)$ approaches the hyperplane $\beta = 0$. The proof of Theorem 3.7.1 is now complete.

3.8 Information Entropy and log-Partition Function

We shall compare eventually the information entropy s_Λ with the entropy that appears at the thermodynamic limit using convex conjugates. For this define now the convex conjugate of the log partition function, Φ_Λ from (3.19),

$$\Phi_\Lambda^*(\rho, \vec{u}, E) = \sup_{(\mu, \vec{\lambda}, \beta) \in \mathbb{R}^5} \left[(\rho, \vec{u}, E) \cdot (\mu, \vec{\lambda}, \beta) - \Phi_\Lambda(\mu, \vec{\lambda}, \beta) \right]. \quad (3.47)$$

(For standard convex analysis consult [Rock].) The following describes the relation of Φ_Λ to s_Λ .

Proposition 3.8.1. Φ_Λ^* takes values as follows:

1. For (ρ, \vec{u}, E) in \mathfrak{S}_Λ (see (3.18)),

$$\Phi_\Lambda^*(\rho, \vec{u}, E) = s_\Lambda \left(g_{(\mu, \vec{\lambda}, \beta), \Lambda} \omega \right) = (\rho, \vec{u}, E) \cdot (\mu, \vec{\lambda}, \beta) - \Phi_\Lambda(\mu, \vec{\lambda}, \beta), \quad (3.48)$$

where $(\mu, \vec{\lambda}, \beta) \in \mathfrak{I}$ is the unique solution of

$$\mathbb{E} \left[\frac{(N, \vec{P}, H)}{|\Lambda|} \right] = (\rho, \vec{u}, E). \quad (3.49)$$

2. For $(\rho, \vec{u}, E) \notin \overline{\mathfrak{S}_\Lambda}$, $\Phi_\Lambda^*(\rho, \vec{u}, E) = +\infty$.

3. And for $(\rho, \vec{u}, E) \in \partial(\mathfrak{S}_\Lambda)$, $\Phi_\Lambda^*(\rho, \vec{u}, E)$ is the limit along any interior segment as we approach (ρ, \vec{u}, E) :

$$\Phi_\Lambda^*(\rho, \vec{u}, E) = \lim_{t \rightarrow 1} \Phi_\Lambda^*((1-t)(\rho', \vec{u}', E') + t(\rho, \vec{u}, E)), \quad (3.50)$$

for any $(\rho', \vec{u}', E') \in \mathfrak{S}_\Lambda$.

Proof. For (1) notice that the critical point equation of

$$(\mu, \vec{\lambda}, \beta) \mapsto (\rho, \vec{u}, E) \cdot (\mu, \vec{\lambda}, \beta) - \Phi_\Lambda(\mu, \vec{\lambda}, \beta) \quad (3.51)$$

is

$$(\rho, \vec{u}, E) = \nabla_{(\mu, \vec{\lambda}, \beta)} \Phi_\Lambda(\mu, \vec{\lambda}, \beta) = \mathbb{E} \left[\frac{1}{|\Lambda|} (N, \vec{P}, H) \right], \quad (3.52)$$

which, by Theorem 3.7.1 has solution if and only if (ρ, \vec{u}, E) is in the set of solvability, \mathfrak{S}_Λ .

For (2), when (ρ, \vec{u}, E) is not in the $\overline{\mathfrak{S}_\Lambda}$, by (3.33), we find $(\alpha_0, \vec{\alpha}, \alpha_4)$ with $\alpha_4 > 0$ and $\varepsilon > 0$ so that almost always

$$\left[(N, \vec{P}, H) - |\Lambda|(\rho, \vec{u}, E) \right] \cdot (\alpha_0, \vec{\alpha}, \alpha_4) > \varepsilon. \quad (3.53)$$

Notice that for any $(\mu, \vec{\lambda}, \beta) \in \mathfrak{J}$,

$$\begin{aligned} & (\rho, \vec{u}, E) \cdot (\mu, \vec{\lambda}, \beta) - \Phi_\Lambda(\mu, \vec{\lambda}, \beta) \\ &= -\frac{1}{|\Lambda|} \log \int_{\mathfrak{x}_\Lambda} \exp \left\{ \left[(N, \vec{P}, H) - |\Lambda|(\rho, \vec{u}, E) \right] \cdot (\mu, \vec{\lambda}, \beta) \right\} d\omega. \end{aligned} \quad (3.54)$$

Then for $\vartheta > 0$, with choosing $(\mu, \vec{\lambda}, \beta) = -\vartheta(\alpha_0, \vec{\alpha}, \alpha_4) - (0, \vec{0}, 1)$, we have

$$\begin{aligned} \Phi_\Lambda^*(\rho, \vec{u}, E) &\geq (\rho, \vec{u}, E) \cdot (\mu, \vec{\lambda}, \beta) - \Phi_\Lambda(\mu, \vec{\lambda}, \beta) \\ &\geq -\frac{1}{|\Lambda|} \log \int_{\mathfrak{x}_\Lambda} \exp \{ -\vartheta\varepsilon - (H - |\Lambda|E) \} d\omega \\ &= \frac{1}{|\Lambda|} \vartheta\varepsilon - \frac{1}{|\Lambda|} \log \int_{\mathfrak{x}_\Lambda} \exp \{ -(H - |\Lambda|E) \} d\omega. \end{aligned} \quad (3.55)$$

Letting $\vartheta \rightarrow \infty$, we conclude that $\Phi_\Lambda^* = +\infty$ whenever (ρ, \vec{u}, E) is not in $\overline{\mathfrak{S}_\Lambda}$.

For (3), when (ρ, \vec{u}, E) is on the boundary of the convex hull, $\Phi_\Lambda^*(\rho, \vec{u}, E)$ is the limit of the values along any segment that has (ρ, \vec{u}, E) as end point and lies in the interior of the convex hull otherwise, as is the case for any convex closed function, see [Rock, Theorem 7.5]. (For similar results in a different setting and with different proofs, see [WJ, Theorem 3.4].) □

Chapter 4

Thermodynamic Limits

We now recall the definition of microcanonical thermodynamic limit entropy. In the next section we will see how it relates to information entropy on bounded domains. This section uses Martin-Löf's formalism in [M-L] for thermodynamic limits as the volume of the domain becomes infinite, see also [L]¹.

As already mentioned, we shall compare entropies via their convex conjugates. We examined convex conjugates of information entropy in the previous section. For the microcanonical thermodynamic limit entropy it is well known that its convex conjugate is related to the (grand canonical) thermodynamic limit pressure, the limit of the log-partition functions Φ_Λ 's as Λ increases. We are especially interested in the strict convexity of pressure: we use it in this section for the differentiability of the limit entropy (which we shall use in the main result of the next section) and we also use it to show a one-to-one correspondence between thermodynamic parameters and macroscopic quantities at the thermodynamic limit in Chapter 6.

¹Note that the formulas in [L] and [M-L] are related as follows: the microcanonical thermodynamic limit entropy at (ρ, E) in [M-L] equals the microcanonical thermodynamic limit entropy in [L] at $(\rho, E/\rho)$ multiplied by ρ .

4.1 Microcanonical Entropy

We first introduce the temperedness condition for interaction potential: U is called **tempered** if for some constants δ greater than the space dimension, $K > 0$, and $R > 0$ it holds that, for \tilde{q}_1 and \tilde{q}_2 configurations consisting of N_1 and N_2 position points respectively,

$$|U(\tilde{q}_1, \tilde{q}_2) - U(\tilde{q}_1) - U(\tilde{q}_2)| \leq K \frac{N_1 N_2}{(d(\tilde{q}_1, \tilde{q}_2))^\delta}, \quad (4.1)$$

whenever \tilde{q}_1 and \tilde{q}_2 have distance $d(\tilde{q}_1, \tilde{q}_2) > R$. (Lennard-Jones type potentials and finite range interactions are tempered, Coulomb potentials are not.)

To define the microcanonical entropy, we will need to specify a special sequence of Λ 's that increases to infinity.

Notation 2. For any $l \in \mathbb{N}$, Λ_l will denote a box in \mathbb{R}^3 with sides of length 2^l . And $\Lambda'_l \subset \Lambda_l$ will denote a smaller box with side $2^l - 2R_l$, where $R_l = R_0 2^{\rho l}$, for some fixed R_0 and $\rho \in [0, 1)$. (The choice of R_0 and ρ depends on the interaction potential U and $\rho = 0$ is when U is of finite range, cf. [M-L, p. 105, p. 88].)

For any $(\rho, \vec{u}, E) \in \mathbb{R}^5$, we fix a sequence of open convex sets $\{A_k\}_{k \in \mathbb{N}}$ such that A_k shrinks to (ρ, \vec{u}, E) . For interactions potentials which are both stable and tempered, the microcanonical entropy at (ρ, \vec{u}, E) is defined as

$$\begin{aligned} s(\rho, \vec{u}, E) &= \lim_{k \rightarrow \infty} \lim_{l \rightarrow \infty} \frac{1}{|\Lambda_l|} \log \omega \left(\left\{ (\tilde{q}, \tilde{p}) \in \mathfrak{X}_{\Lambda'_l} : \frac{1}{|\Lambda_l|} (N, \vec{P}, H)(\tilde{q}, \tilde{p}) \in A_k \right\} \right), \end{aligned} \quad (4.2)$$

cf. [M-L, §3.4.2].

We also know that $s(\rho, \vec{u}, E)$ is an upper-semi continuous concave function on \mathbb{R}^5 and that $s(\rho, \vec{u}, E) < \infty$, cf. [M-L, p. 45, p. 96]. Therefore the set $\text{dom}(s) = \{(\rho, \vec{u}, E) :$

$s(\rho, \vec{u}, E) > -\infty\}$, i.e. the domain of s , is convex. We will give a description of $\text{int dom}(s)$, i.e. the interior of domain of s , in Theorem 5.1.1. To this end, we will need the following notation:

Notation 3. \mathcal{E}'_{Λ_l} will denote the essential range of $\frac{1}{|\Lambda_l|}(N, \vec{P}, H)$ as a map on \mathfrak{X}_{Λ_l} . (Compared with \mathcal{E}_{Λ_l} as in Notation 1, \mathcal{E}'_{Λ_l} takes into account only configurations in Λ'_l .) Also $\text{conv}(\mathcal{E}'_{\Lambda_l})$ denotes the convex hull of \mathcal{E}'_{Λ_l} and $\text{int conv}(\mathcal{E}'_{\Lambda_l})$ is the interior of $\text{conv}(\mathcal{E}'_{\Lambda_l})$. As $\mathcal{E}'_{\Lambda_l} = \frac{\Lambda'_l}{\Lambda_l} \mathcal{E}_{\Lambda'_l}$, from the proof of Lemma 3.6.1, it is easy to see that

$$\text{conv}(\mathcal{E}'_{\Lambda_l}) = \frac{|\Lambda'_l|}{|\Lambda_l|} \text{conv}(\mathcal{E}_{\Lambda'_l}) = \frac{|\Lambda'_l|}{|\Lambda_l|} \mathcal{C}_{\Lambda'_l}. \quad (4.3)$$

In this notation, the following will be crucial in the proof of Theorem 5.1.1:

Lemma 4.1.1. For stable, tempered interaction potentials the domain of s is related to \mathcal{E}'_{Λ_l} as follows:

$$\overline{\text{dom}(s)} = \overline{\bigcup_l \mathcal{E}'_{\Lambda_l}}. \quad (4.4)$$

Proof. See [M-L, p. 93, p. 95] for finite range, stable interaction potentials and [M-L, pp. 105–111] for infinite range, stable, tempered interaction potentials. Note that the claim there is for \mathcal{E}_{Λ} instead of \mathcal{E}'_{Λ_l} (in our notation). \square

4.2 Pressure

For any $(\mu, \vec{\lambda}, \beta) \in \mathbb{R}^5$, define now the thermodynamic limit pressure by

$$\Xi(\mu, \vec{\lambda}, \beta) = \lim_{\Lambda \rightarrow \infty} \Phi_{\Lambda}(\mu, \vec{\lambda}, \beta) \quad (4.5)$$

where the sequence of Λ 's approaches infinity in the sense of “strong van Hove” [M-L, p. 107] and “approximable by cubes” [M-L, p. 91]. The sequence $\{\Lambda_l\}$ as in Notation

2 is an example of such Λ 's. See [M-L, p. 111] for the existence and properties of this limit². Trivially, by (4.5), $\Xi(\mu, \vec{\lambda}, \beta) = \infty$ for $(\mu, \vec{\lambda}, \beta) \notin \mathfrak{I}$.

Lemma 4.2.1. *For $(\mu, \vec{\lambda}, \beta) \in \mathfrak{I}$, we have $0 \leq \Xi(\mu, \vec{\lambda}, \beta) < \infty$. (Therefore Ξ is a proper convex function with domain $\text{dom}(\Xi) = \mathfrak{I}$.)*

Proof. Fix any $(\mu, \vec{\lambda}, \beta)$ in \mathfrak{I} . By the stability (3.6) and completing squares for the velocity terms, we easily see that $\Xi(\mu, \vec{\lambda}, \beta) < \infty$. On the other hand, as $Z_{(\mu, \vec{\lambda}, \beta), \Lambda} \geq 1$ for all Λ , we have $\Phi_\Lambda(\mu, \vec{\lambda}, \beta) \geq 0$, therefore $\Xi(\mu, \vec{\lambda}, \beta) \geq 0$. \square

The thermodynamic limit microcanonical entropy s and the pressure Ξ are related via convex conjugation:

$$\Xi(\mu, \vec{\lambda}, \beta) = \sup_{(\rho, \vec{u}, E) \in \mathbb{R}^5} \{s(\rho, \vec{u}, E) + (\rho, \vec{u}, E) \cdot (\mu, \vec{\lambda}, \beta)\}, \quad (4.6)$$

(see [M-L, p. 45, Lemma 5]³) i.e.

$$\Xi(\mu, \vec{\lambda}, \beta) = (-s)^*(\mu, \vec{\lambda}, \beta), \quad (4.7)$$

with $*$ still denoting the convex conjugate. As s is concave and upper-semi continuous, we can conjugate once again to get

$$\Xi^*(\rho, \vec{u}, E) = -s(\rho, \vec{u}, E). \quad (4.8)$$

²In [M-L]'s notation, our $\Xi(\mu, \vec{\lambda}, \beta)$ is $s(\mathbb{R} \times \mathbb{R}^3 \times \mathbb{R}_{>0}; \mu, \vec{\lambda}, \beta)$. [M-L] also comments on differentiability of Ξ for finite range interactions but leaves open the differentiability for stable potentials. We therefore resort to our own differentiability arguments here.

³To go from Martin-Löf's Lemma 5 to (4.6) notice that the limit on the left of (4.6) is the supremum of what Martin-Löf calls $s(u, a)$ in Lemma 5, p. 45, and that the *pointwise* formula (in Martin-Löf's notation) $s(u, a) = s(u) - a \cdot u$ holds by [M-L, Lemma 5a].

As mentioned in the introduction to this section, it will be important to know when Ξ is strictly convex. We assume strict convexity for the moment and provide a discussion in section 4.3.

Lemma 4.2.2. *Assume that $\Xi(\mu, \vec{\lambda}, \beta)$ is strictly convex on \mathfrak{J} . Then s is essentially differentiable, i.e. differentiable in the interior of its domain and $|\nabla s| \rightarrow \infty$ on the boundary of its domain.*

Proof. As $\text{dom}(\Xi) = \mathfrak{J}$, we have that Ξ , as an extended-reals-valued function, is essentially strictly convex (for the general definition see [Rock, p. 253]). This observation, (4.8), and [Rock, Theorem 26.3] immediately imply the lemma (which should be compared to [Gi, Remark 3.7, Remark 6.5] where differentiability of s on its domain was shown). □

4.3 Strict Convexity of Pressure

Strict convexity of pressure for lattice systems was shown by Griffiths and Ruelle in [GR]. For continuous systems, this is the case in general when the Gibbs variational principle (see [D], [Gi] for a definition) holds. As here is the only place we use the concept of Gibbs states, we only sketch the argument. For details see [P, Proposition 8.5] and [Gi, Remark 3 and §5].

Let \mathfrak{X} be the the space of locally finite configurations on \mathbb{R}^3 marked with velocities. For a state ν (a probability measure on \mathfrak{X}), one can define specific entropy $\mathfrak{s}(\nu)$, density $\mathfrak{r}(\nu)$ and average energy $\mathfrak{u}(\nu)$ (when a suitable interaction potential is given). In general, for any $\mu \in \mathbb{R}$ and $\beta < 0$, it holds that

$$-\Xi(\mu, \vec{0}, \beta) \leq \mathfrak{s}(\nu) - \beta \mathfrak{u}(\nu) - \mu \mathfrak{r}(\nu), \tag{4.9}$$

see [G, §3] for precise definitions and proof. The Gibbs variational principle states that the equality in (4.9) holds if and only if ν is a Gibbs state indexed by μ and β . If we assume that $\Xi(\mu, \vec{0}, \beta)$ is not strictly convex on $\mathbb{R} \times \mathbb{R}_{<0}$ then there exists distinct (μ_1, β_1) and (μ_2, β_2) such that

$$\Xi(\mu_t, \vec{0}, \beta_t) = t \Xi(\mu_1, \vec{0}, \beta_1) + (1 - t) \Xi(\mu_2, \vec{0}, \beta_2), \quad (4.10)$$

where $(\mu_t, \beta_t) = t(\mu_1, \beta_1) + (1 - t)(\mu_2, \beta_2)$ and $0 < t < 1$. Let ν be a Gibbs state for (μ_t, β_t) . Then applying Gibbs variational principle to ν ,

$$\begin{aligned} t [\mathfrak{s}(\nu) - \beta_1 \mathbf{u}(\nu) - \mu_1 \mathbf{r}(\nu)] + (1 - t) [\mathfrak{s}(\nu) - \beta_2 \mathbf{u}(\nu) - \mu_2 \mathbf{r}(\nu)] \\ = -t \Xi(\mu_1, \vec{0}, \beta_1) - (1 - t) \Xi(\mu_2, \vec{0}, \beta_2). \end{aligned} \quad (4.11)$$

Again by the Gibbs variational principle, we have now that ν is also a Gibbs state of both (μ_1, β_1) and (μ_2, β_2) . However, it is standard that the Gibbs states are identifiable, in the sense that the sets of Gibbs states for different (μ, β) are disjoint, cf. [G, Remark 3.7]. Therefore we conclude strict convexity of $\Xi(\mu, \vec{0}, \beta)$ under the assumption of Gibbs variational principle.

The Gibbs principle holds at least for the following cases:

- For superstable, finite range pairwise potentials as in [D, Theorem 1].
- For regular, non-integrably divergent pairwise potentials as in [G, p. 1344].

Note that Lennard-Jones type interactions satisfy the assumptions from [G] and are tempered.

The above argument was written for $\Xi(\mu, \vec{0}, \beta)$ which is the case in [G]. It is easy to see that the dependence on $\vec{\lambda}$ does not spoil strict convexity:

Proposition 4.3.1. *If $(\mu, \beta) \mapsto \Xi(\mu, \vec{0}, \beta)$ is strictly convex on $\mathbb{R} \times \mathbb{R}_{<0}$ (as in the above cases) then $\Xi(\mu, \vec{\lambda}, \beta)$ is strictly convex on \mathfrak{I} .*

Proof. For any $(\mu, \vec{\lambda}, \beta) \in \mathfrak{I}$, by completing squares, we obtain $\Phi_\Lambda(\mu, \vec{\lambda}, \beta) = \Phi_\Lambda(\mu + \tau, \vec{0}, \beta)$ where $\tau = \tau(\vec{\lambda}, \beta) = -\frac{|\vec{\lambda}|^2}{2\beta}$. Therefore $\Xi(\mu, \vec{\lambda}, \beta) = \Xi(\mu + \tau, \vec{0}, \beta)$. Elementary calculus shows that τ is convex on $\mathbb{R}^3 \times \mathbb{R}_{<0}$. Then the strict convexity of $\Xi(\mu, \vec{\lambda}, \beta)$ follows from the fact that $\Xi(\mu, \vec{0}, \beta)$ is increasing in μ and the strict convexity of $(\mu, \beta) \mapsto \Xi(\mu, \vec{0}, \beta)$. □

Chapter 5

Convergence

We are now ready for the main results. The main point here is that, thanks to the properties of the functions at hand, as the log-partition functions for finite volumes converge to the thermodynamic pressure their convex conjugates (finite volume information entropies) converge to the convex conjugate of the limit (microcanonical entropy), and so do their derivatives (the thermodynamic parameters). The chapter also includes comparisons with related results.

5.1 Convergence of Epigraphs

Recall again that the epigraph of a function f is the set $\{(x, a) : f(x) \leq a\}$, that a convex function is called proper if it is not identically $+\infty$ and it never has the value $-\infty$ (with obvious modifications for concave functions), and that proper and closed is synonymous to proper and lower semi-continuous in our setting.

For functions in general, convergence of epigraphs (for convergence of sets as in [RW, §4B]) is used to preserve critical points at the limit. For convex functions, pointwise convergence is not too far from convergence of epigraphs, in a sense that is made precise as we recall the following facts :

Fact 1. *For f_n, f convex functions on \mathbb{R}^d with $f_n \rightarrow f$, assume that f is closed and*

the interior of the domain of f is not empty. Then the epigraphs of f_n converge to the epigraph of f as sets.

The importance of the convergence of epigraphs also lies in that it is inherited by conjugates:

Fact 2. *If f_n, f are proper, closed, convex functions on \mathbb{R}^d , then the epigraphs of f_n converge to the epigraph of f if and only if the epigraphs of f_n^* converge to the epigraph of f^* .*

Finally, convergence of epigraphs implies pointwise convergence as in the following:

Fact 3. *Let f_n and f be convex functions on \mathbb{R}^d with the epigraphs of f_n converging to the epigraph of f . Assume that f is closed and the interior of the domain of f is not empty. Then $f_n \rightarrow f$ uniformly on any compact set in the interior of the domain of f .*

Facts 1 and 3 are included in [RW, Theorem 7.17], while Fact 2 is Theorem 11.34 in the same reference. Recalling the definition of \mathcal{E}'_{Λ_l} from Notation 3 and the definition of the set of solvability \mathfrak{S}_Λ from (3.18), we can now state:

Theorem 5.1.1. *For interaction U stable and tempered:*

1. *The interior of the domain of s is not empty.*
2. *Fix any sequence of Λ 's that approaches infinity in the strong van Hove sense and is approximable by cubes. Whenever (ρ, \vec{u}, E) is in the interior of the domain of s then (ρ, \vec{u}, E) is in \mathfrak{S}_Λ , for all Λ large enough. Therefore $\text{int dom}(s) \subset \liminf_{\Lambda} \mathfrak{S}_\Lambda$.*

$$3. \text{ int dom}(s) = \bigcup_l \text{int conv}(\mathcal{E}'_{\Lambda_l}).$$

Proof. For 1, by (4.4) and the fact that $\text{dom}(s)$ is convex we get

$$\overline{\text{dom}(s)} = \overline{\bigcup_l \mathcal{E}'_{\Lambda_l}} = \overline{\bigcup_l \text{conv}(\mathcal{E}'_{\Lambda_l})}. \quad (5.1)$$

Again by convexity of $\text{dom}(s)$, we have $\text{int dom}(s) = \text{int } \overline{\text{dom}(s)}$ (cf. [Rock, Theorem 6.3]). It therefore follows that

$$\text{int dom}(s) = \text{int} \left(\overline{\bigcup_l \text{conv}(\mathcal{E}'_{\Lambda_l})} \right) \supset \bigcup_l \text{int conv}(\mathcal{E}'_{\Lambda_l}). \quad (5.2)$$

As $\text{int conv}(\mathcal{E}'_{\Lambda_l})$ is not empty (cf. Notation 3), the proof of 1 is complete.

For 2: The interior of the domain of Ξ is not empty, by Lemma 4.2.1 and Ξ is closed since it is a conjugate as in (4.7)(cf. [Rock, Thm. 12.2]). Therefore, according to Fact 1, the pointwise converge (4.5) implies that the epigraphs of Φ_Λ converge to the epigraph of Ξ .

Therefore, by Fact 2, the epigraphs of Φ_Λ^* converge as sets to the epigraph of Ξ^* . Equation (4.8) gives $\Xi^*(\rho, \vec{u}, E) = -s(\rho, \vec{u}, E)$. In other words, the epigraphs of Φ_Λ^* converge to the epigraph of $-s$. Then apply Fact 3 to obtain that that for (ρ, \vec{u}, E) in the interior of the domain of s we have that Φ_Λ^* converges uniformly to $-s$ on any ball in the interior of the domain that contains (ρ, \vec{u}, E) . In particular, (ρ, \vec{u}, E) is in the interior of the domain of Φ_Λ^* for all Λ large enough.

For 3, observe that we have already shown in part 1 that

$$\text{int dom}(s) \supset \bigcup_l \text{int conv}(\mathcal{E}'_{\Lambda_l}). \quad (5.3)$$

Conversely, let $x \in \text{int dom}(s)$. For any sequence of Λ 's that approaches infinity in the strong van Hove sense and approximable by cubes, part 2 shows that $x \in \mathfrak{S}_\Lambda =$

int conv(\mathcal{E}_Λ) for all Λ large enough. Now we fix the sequence of Λ 's as $\{\Lambda'_l\}$. We therefore have $x \in \text{int conv}(\mathcal{E}_{\Lambda'_l})$ for l large—note that here the prime is on Λ and not on \mathcal{E} . Indeed, using the convexity of the sets conv($\mathcal{E}_{\Lambda'_l}$), a whole neighborhood of x , say B_x , is in conv($\mathcal{E}_{\Lambda'_l}$) for l large enough. Using (4.3) and $\lim_{l \rightarrow \infty} \frac{|\Lambda'_l|}{|\Lambda_l|} = 1$, we have that B_x is in conv(\mathcal{E}'_{Λ_l}) for l large, therefore $x \in \bigcup_l \text{int conv}(\mathcal{E}'_{\Lambda_l})$.

□

To prepare for the statement of the next theorem, we introduce:

Notation 4. For $(\rho, \vec{u}, E) \in \mathbb{R}^5$, set $\tilde{s}_\Lambda(\rho, \vec{u}, E) = -\Phi_\Lambda^*(\rho, \vec{u}, E)$.

This notation is partly justified by the relation between \tilde{s}_Λ and the information entropy s_Λ . In fact, when (ρ, \vec{u}, E) is in \mathfrak{S}_Λ , by part 1 of Proposition 3.8.1, we have $\tilde{s}_\Lambda(\rho, \vec{u}, E) = -s_\Lambda(g_{((\mu_\Lambda, \vec{\lambda}_\Lambda, \beta_\Lambda), \Lambda)})$ for $(\mu_\Lambda, \vec{\lambda}_\Lambda, \beta_\Lambda)$ the unique thermodynamic parameter corresponding to (ρ, \vec{u}, E) such that (3.16) holds. The following theorem shows that, as Λ tends to infinity, $\tilde{s}_\Lambda(\rho, \vec{u}, E)$ converges to thermodynamic limit microcanonical entropy $s(\rho, \vec{u}, E)$ and $(\mu_\Lambda, \vec{\lambda}_\Lambda, \beta_\Lambda)$ converges to $(\mu, \vec{\lambda}, \beta)$, the thermodynamic parameter corresponding to (ρ, \vec{u}, E) in the limit.

Theorem 5.1.2. Fix any sequence of Λ 's that approaches infinity in the strong van Hove sense and approximable by cubes. For (ρ, \vec{u}, E) in the interior of the domain of s , we have

1. $\tilde{s}_\Lambda(\rho, \vec{u}, E) \rightarrow s(\rho, \vec{u}, E)$, as Λ tends to infinity.
2. If s is differentiable on some open convex set C containing (ρ, \vec{u}, E) (e.g. as in section 4.2), for $(\mu, \vec{\lambda}, \beta) = -\nabla s(\rho, \vec{u}, E)$,

$$(\mu_\Lambda, \vec{\lambda}_\Lambda, \beta_\Lambda) \rightarrow (\mu, \vec{\lambda}, \beta) \text{ as } \Lambda \text{ tends to infinity.} \quad (5.4)$$

$((\mu_\Lambda, \vec{\lambda}_\Lambda, \beta_\Lambda)$ is well defined for all Λ large.)

Proof. As $\tilde{s}_\Lambda = -\Phi_\Lambda^*$, then $\tilde{s}_\Lambda(\rho, \vec{u}, E) \rightarrow -s(\rho, \vec{u}, E)$ follows from proof of Part 2 in Theorem 5.1.1.

By Part 2 of Theorem 5.1.1, (ρ, \vec{u}, E) is in \mathfrak{S}_Λ for all Λ large. Therefore for such Λ 's, $(\mu_\Lambda, \vec{\lambda}_\Lambda, \beta_\Lambda)$ is well defined. As Φ_Λ is strictly convex on its domain \mathfrak{J} , \tilde{s}_Λ is differentiable on \mathfrak{S}_Λ , cf. [Rock, Theorem 26.3]. By convexity of \mathfrak{S}_Λ and Part 2 of Theorem 5.1.1, we may find open convex set C' containing (ρ, \vec{u}, E) such that $C' \subset C$ and $C' \subset \mathfrak{S}_\Lambda$ for all Λ large. Now we may use [Rock, Theorem 25.7] (roughly: for convex functions pointwise convergence implies convergence of derivatives) to conclude that $\nabla \tilde{s}_\Lambda(\rho, \vec{u}, E) \rightarrow \nabla s(\rho, \vec{u}, E)$, i.e. $(\mu_\Lambda, \vec{\lambda}_\Lambda, \beta_\Lambda) \rightarrow (\mu, \vec{\lambda}, \beta)$. \square

Remark 5.1.3. *Of less practical use is the following: In general, for any $(\rho, \vec{u}, E) \in \text{int dom}(s)$, we know that the subdifferential set ∂s of s at (ρ, \vec{u}, E) is not empty, cf. [Rock, Theorem 23.4]. Then for any $(\mu, \vec{\lambda}, \beta)$ such that $-(\mu, \vec{\lambda}, \beta) \in \partial s(\rho, \vec{u}, E)$ there is sequence $(\rho_\Lambda, \vec{u}_\Lambda, E_\Lambda) \rightarrow (\rho, \vec{u}, E)$ such that $\nabla \tilde{s}_\Lambda(\rho_\Lambda, \vec{u}_\Lambda, E_\Lambda) \rightarrow -(\mu, \vec{\lambda}, \beta)$. This follows from Attouche's Theorem: For f_n, f proper, closed, convex functions, if f_n converges to f epigraphically then ∂f_n converges to ∂f graphically. See [RW, §5E and Theorem 12.35] for definitions and proof.*

5.2 Comparison with Maximum Likelihood Estimators

Several articles address the consistency of maximum likelihood estimators for Gibbs point processes, for example [CG], [Gi], [DL]. The main point there is: given a Gibbs state with parameter β , let ω , a locally finite configuration on \mathbb{R}^3 , be a realization of

it. For ω_Λ the restriction of ω on Λ , maximize the likelihood

$$\beta_\Lambda = \operatorname{argmax}_\beta \frac{\exp(\beta U(\omega_\Lambda))}{Z_\Lambda(\beta)}, \quad (5.5)$$

for U the interaction and $\beta < 0$ to match our conventions here. Then [CG], [Gi], [DL] show that almost always, $\beta_\Lambda \rightarrow \beta$, as $\Lambda \rightarrow \infty$.

For exponential families of measures, maximizing likelihood and minimizing entropy are closely related in general. Let $x_i, i = 1, \dots, n$ be independent realizations of a member of an exponential family of probability measures $\mu_\vartheta = \exp(\vartheta t(x) - \log Z(\vartheta)) \mu_0$. Then maximizing the likelihood $\prod_i \exp(\vartheta t(x_i) - \log Z(\vartheta))$ is the same as solving

$$\frac{d}{d\vartheta} \log \prod_i \exp(\vartheta t(x_i) - \log Z(\vartheta)) = 0 \Leftrightarrow \frac{1}{n} \sum_i t(x_i) = \frac{d}{d\vartheta} \log Z(\vartheta). \quad (5.6)$$

Note that the right hand side is $\mathbb{E}_\vartheta[t]$. In other words, for exponential families of the form $\exp(\vartheta t(x) - \log Z(\vartheta))$, the maximum likelihood estimator ϑ for $\{x_i\}_{i=1}^n$ is the same as the ϑ of minimizing entropy with constraint $\sum t(x_i)/n$. See [Kull, p. 82, p. 94] and the references there for more.

It is clear then how our convergence of thermodynamics parameters differs from the consistency problem of maximum likelihood estimators: in our case, the constrains (i.e. ρ, \vec{u}, E) are given and fixed for all volumes Λ whereas in the consistency problems the constrains come from the restriction of the realized configuration on Λ . Therefore, in the consistency problems the constraints not only change with Λ , but they also depend on ω , i.e. the estimated thermodynamic parameters are a sequence of random variables.

5.3 Comparison with Lanford

Lanford in [L, p. 63] shows (and [M-L, p. 57] explains a crucial step in the proof) that for β the derivative of the microcanonical thermodynamic limit entropy s at energy E (when this exists even in the absence of kinetic energy) for fixed density ρ , the information entropy of the *canonical* distribution with parameter β converges to $-s(\rho, E)$. Note carefully that for Lanford β stays the same over all volumes, whereas in our work we show convergence while the parameters change with volume. Furthermore, the measure in Lanford's argument is canonical, rather than grand canonical. This is because ρ is now fixed throughout and is not on the same footing as E . In particular, Lanford's β is a function of ρ .

Chapter 6

Local Homeomorphism at the Thermodynamic Limit

We show here that at the thermodynamic limit the bijection between thermodynamic parameters and macroscopic quantities as in Theorem 3.7.1 holds at least locally in some region. The existence of such a bijection for continuous systems is folklore in the theory of hydrodynamic limits, see for example [OVY, p. 530, p. 556]. For the case of particle configurations on the line see [LR, Proposition 5.3], [V, Theorem 10.2].

Throughout this chapter, in addition to stability and temperedness, we will assume pair interactions: $U(q_1, \dots, q_n) = \sum_{i \neq j} \Phi(q_i - q_j)$. Notice that the stability (3.6) and temperedness as in (4.1) imply that

$$C(\beta) := \int_{\mathbb{R}^3} |\exp(\beta\Phi(x)) - 1| dx < +\infty, \quad \text{for all } \beta < 0, \quad (6.1)$$

cf. [R, p. 32, p. 72]. To establish the local homeomorphism, we will need the thermodynamic limit pressure Ξ to be both strictly convex (as in section 4.3) and differentiable. It is standard that, when not taking velocities into account, i.e. $\mathfrak{X}_\Lambda = \bigcup_{n \geq 0} \Lambda^n$, $H = U$ and $\vec{\lambda} = 0$, Ξ is analytic in μ and β in the low density region

$$\{(\mu, \beta) : \beta < 0, \mu < 2\beta L - 1 - \log C(\beta)\}. \quad (6.2)$$

Analyticity with respect to μ is shown in [R, Theorem 4.3.1]. Combining this with

results from [LP], analyticity with respect to β also follows. This is presented in detail in [X, Appendix D].

Including the kinetic energy and $\vec{\lambda}$, it follows easily that we have differentiability of $\Xi(\mu, \vec{\lambda}, \beta)$ for

$$\mathcal{R} = \left\{ (\mu, \vec{\lambda}, \beta) : \beta < 0, \mu < 2\beta L - 1 - \log C(\beta) - \log \int_{\mathbb{R}^3} \exp \left(\vec{\lambda} \cdot \vec{p} + \beta \frac{|\vec{p}|^2}{2} \right) d\vec{p} \right\}. \quad (6.3)$$

Proposition 6.0.1. *Let U be a stable, tempered, pair interaction potential with Ξ strictly convex (e.g. as in section 4.3) and let K be convex, open subset of \mathcal{R} . Then $\nabla\Xi : K \rightarrow \mathbb{R}^5$ is a homeomorphism onto $\nabla\Xi(K)$ with $(\nabla\Xi)^{-1} = -\nabla s$.*

Proof. On such a K the pressure Ξ is both differentiable and strictly convex which implies that $\nabla\Xi$ is one-to-one from K to $\nabla\Xi(K)$, using [RV, Theorem B, p. 99] again. $\nabla\Xi$ is also continuous, see [Rock, Theorem 25.5].

$\nabla\Xi(\mu, \vec{\lambda}, \beta) = (\rho, \vec{u}, E)$ if and only if $(\mu, \vec{\lambda}, \beta)$ is in the subdifferential of $-s$ at (ρ, \vec{u}, E) ([Rock, Theorem 23.5]). Therefore, as $-s$ is essentially differentiable by Lemma 4.2.2, the range $\nabla\Xi(K)$ can only be a subset of the interior of the domain of $-s$, where $-s$ is differentiable, and $\nabla\Xi = (\nabla(-s))^{-1}$. Since ∇s is also continuous, $\nabla\Xi$ is a homeomorphism. □

Part II

Differential Manifold Approach to Statistical Mechanics

Chapter 7

Introduction

The physical derivation of the canonical Gibbs distribution at thermal equilibrium for a Hamiltonian E is as follows. For f be a probability density function and $d\mathbf{q}d\mathbf{p} = dq_1 \dots dq_N dp_1 \dots dp_N$ the Lebesgue measure on phase space we minimize $\int f \ln f d\mathbf{q}d\mathbf{p}$ subject to $\int E f d\mathbf{q}d\mathbf{p} = \text{constant}$. In particular, for δf a small perturbation such that $\int \delta f d\mathbf{q}d\mathbf{p} = 0$ we solve

$$\left. \frac{d}{d\epsilon} \right|_{\epsilon=0} \int (f + \epsilon \delta f) \ln (f + \epsilon \delta f) d\mathbf{q}d\mathbf{p} = \left. \frac{d}{d\epsilon} \right|_{\epsilon=0} \lambda \int E (f + \epsilon \delta f) d\mathbf{q}d\mathbf{p} \quad (7.1)$$

for any δf , giving $f = e^{\lambda E - 1}$. After relabeling and applying the condition $\int f d\mathbf{q}d\mathbf{p} = 1$, we have the Gibbs distribution:

$$f = \frac{e^{\lambda E}}{\int e^{\lambda E} d\mathbf{q}d\mathbf{p}} \quad (7.2)$$

However, there is an issue with the above derivation. The space of probability measures is not a linear space. Therefore, while the above steps are intuitively correct, mathematically they are not justified. Our goal is to construct a framework in which the process of deriving the Gibbs density makes sense.

To this end, we shall construct a differentiable manifold structure on the set of finite measures on some Ω with respect to which a given vector valued function

$$\mathbf{C} : \Omega \rightarrow \mathbb{R}^n; \quad \mathbf{C}(\Omega) = (C_1(\Omega), \dots, C_n(\Omega)). \quad (7.3)$$

is integrable. More precisely, let μ_0 be some fixed positive measure on a topological space Ω and let

$$\mathcal{F} := \{ \text{positive measure } \mu : \mu \ll \mu_0, \frac{d\mu}{d\mu_0} > 0, \left| \int_{\Omega} C_i \mu \right| < +\infty \ i = 1, \dots, n \}, \quad (7.4)$$

where $C_0 \equiv 1$ corresponds to the requirement that each $\mu \in \mathcal{F}$ is a finite measure. Note that the components of \mathbf{C} need not be positive, hence the $C_i \mu$'s are signed measures.

Our goal is to solve the Lagrange multiplier problem where the information entropy functional $H : \mathcal{F} \rightarrow \mathbb{R}$ is

$$H(p\mu_0) = \int_{\Omega} p \log p \mu_0 \quad (7.5)$$

is minimized subject to given constraint functionals $G_i : \mathcal{F} \rightarrow \mathbb{R}$ defined by

$$G_i(p\mu_0) = \int_{\Omega} C_i p \mu_0. \quad (7.6)$$

This requires that both (7.5) and (7.6) define differentiable functionals on \mathcal{F} . For this we shall construct a manifold structure on \mathcal{F} , mapping open subsets of \mathcal{F} to a linear model space. The question is then what should our model space be. Part of the issue with \mathcal{F} is that probability density functions are positive everywhere. A natural mapping to a space where functions are both positive and negative is the natural log function. The inverse mapping from the model space back to \mathcal{F} will then involve the exponential function in some way. Thus, we require the functions in our model space to be well behaved when exponentiated. Additionally, we wish to take derivatives along paths through some $p\mu_0 \in \mathcal{F}$. These paths in \mathcal{F} will be written as

$$\gamma(t) = e^{tu} p\mu_0 \quad (7.7)$$

where u is in the model space and $p\mu_0$ corresponds to $t = 0$. In order to take a derivative at p , the above must be well defined in a neighborhood around $t = 0$. In other words, both e^{-tu} and e^{tu} must be integrable for some $t > 0$. This is equivalent to requiring that our model space contains u 's such that $e^{t|u|}$ is integrable for some $t > 0$. This naturally leads us to consider Orlicz spaces with Young function $e^{|t|} - 1$.

However, we also require that the constraint functionals are differentiable. If one assumes that the model space consists of u 's in the above Orlicz space and in $L^1(C_i p\mu_0)$, problems appear with proving openness of the image of the chart. Furthermore, the differentiability of $\int C_i p\mu_0$ at p requires $\int_{\Omega} C_i e^{tu} p\mu_0$ is well defined in a neighborhood around $t = 0$. These point to a model space consisting of functions in Orlicz spaces against the finite measures $p\mu_0$ and $C_i p\mu_0$.

After providing some preliminary definitions (Chapter 7), we construct a manifold on \mathcal{F} (Chapter 8). We then apply the manifold structure to the above Lagrange multiplier problem, showing that the result is a Gibbs distribution (Chapter 9). Additionally, we apply the manifold structure to the problem of proving the constraint functionals are open mappings (Chapter 11). Last, we apply the manifold structure of Chapter 8 to the problem of local Gibbs.

7.1 Orlicz Spaces

We first review some standard theory regarding Orlicz spaces. First, consider the following (see [AJVLS, p. 172]):

Definition 7.1.1. *A function $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is called a Young function if $\phi(0) = 0$, ϕ is even, convex, strictly increasing on $[0, \infty)$ and $\lim_{t \rightarrow \infty} \frac{\phi(t)}{t} = \infty$. Given a Young*

function ϕ and a finite measure space (Ω, μ) , we define the Orlicz space

$$L^\phi(\mu) := \left\{ f : \Omega \rightarrow \mathbb{R} : \int_{\Omega} \phi(\epsilon f(x)) \mu(dx) < \infty \text{ for some } \epsilon > 0 \right\}. \quad (7.8)$$

As required of a model space, Orlicz spaces are linear, (see [RR, Chapter 3 Theorem 2(ii)] or Proposition 13.1.1).

Note that for any $f \in L^\phi(\mu)$ there is ϵ such that [PS, p. 173]

$$\int_{\Omega} \phi(\epsilon f) \mu \leq 1. \quad (7.9)$$

In particular the following definition is justified:

Definition 7.1.2. *The norm on $L^\phi(\mu)$ is given by*

$$\|f\|_\phi = \inf \left\{ a > 0 : \int_{\Omega} \phi\left(\frac{f(x)}{a}\right) \mu(dx) \leq 1 \right\}. \quad (7.10)$$

$L^\phi(\mu)$ with this norm is a Banach space ([RR, Chapter 3, Theorem 10]). Note that the infimum in the above definitions are attained (see [AJVLS, p. 173]). The Orlicz norm is the smallest number that f can be divided by to make the integral less than 1. Note that dividing by larger numbers will also make the resulting integral smaller than 1 (see Proposition 13.1.3). Convergence in Orlicz space can be defined in the following equivalent ways (see [AJVLS, Proposition 3.8])

Proposition 7.1.3. *Let f_n be a sequence in $L^\phi(\mu)$. Then the following are equivalent:*

1. $\lim_{n \rightarrow \infty} f_n = f$ in Orlicz norm.
2. For all $c > 0$, $\lim_{n \rightarrow \infty} \int_{\Omega} \phi(c(f_n - f)) \mu = 0$

Additionally, it will be helpful to relate Orlicz spaces corresponding to different Young functions (see [AJVLS, Proposition 3.9]).

Proposition 7.1.4. *Let $\phi_1(t)$ and $\phi_2(t)$ be two Young functions and (Ω, μ) a finite measure space. If*

$$\lim_{t \rightarrow \infty} \frac{\phi_1(t)}{\phi_2(t)} = c \in (0, \infty), \quad (7.11)$$

then $L^{\phi_1}(\mu) = L^{\phi_2}(\mu)$ and the Orlicz norms $\|f\|_{L^{\phi_1}(\mu)}$ and $\|f\|_{L^{\phi_2}(\mu)}$ are equivalent.

Additionally, Orlicz spaces are continuously embedded in L^p spaces (see [AJVLS, Equation 3.104], or [PS, Proposition 2.3]):

Proposition 7.1.5. *The Orlicz space $L^\phi(\mu), \phi(t) = e^{|t|} - 1$, is continuously embedded in $L^p(\mu)$ for all $p \geq 1$,*

$$L^\phi(\mu) \hookrightarrow \bigcap_{p \geq 1} L^p(\mu). \quad (7.12)$$

7.2 An equivalence relation

Let C_i be the components of \mathbf{C} for $i = 1, \dots, n$. Recall $C_0 \equiv 1$. Note that since $\mathbb{E}_\mu[C_i] < \infty$ for $\mu \in \mathcal{F}$,

$$\int_{\Omega} C_i \mu = \int_{\Omega} C_i^+ \mu - \int_{\Omega} C_i^- \mu \quad (7.13)$$

are both finite integrals. Thus,

$$\int_{\Omega} |C_i| \mu = \int_{\Omega} C_i^+ \mu + \int_{\Omega} C_i^- \mu < \infty \quad (7.14)$$

and so we may regard each $|C_i| \mu$ as a positive finite measure. This will be used in the construction of the model space in Chapter 8, as well as in the following definitions that are adapted from [AJVLS, Equation 3.105].

Definition 7.2.1. *Let $\mu, \mu' \in \mathcal{F}$. Define $\mu' \lesssim \mu$ if $\mu' = \phi \mu$, $\phi \in L^{p_i}(|C_i| \mu)$ for some $p_i > 1$, $i = 0, 1, \dots, n$.*

When both $\mu \lesssim \mu'$ and $\mu' \lesssim \mu$ we denote this by $\mu' \sim \mu$.

Note that since $|C_i|\mu$ are finite measures, $\phi \in L^{p_i}(|C_i|\mu)$ implies $\phi \in L^r(|C_i|\mu)$ for all $1 \leq r \leq p_i$. Without loss of generality, we can take $p = \min_i p_i$. Thus, $\mu' \lesssim \mu$ implies

$$\mu' = \phi\mu \text{ s.t. } \phi \in \bigcap_{i=0}^n L^p(|C_i|\mu) \text{ for some } p > 1. \quad (7.15)$$

As these are in fact equivalent, we shall use (7.15) from now on.

The following is adapted from [AJVLS, p. 178].

Proposition 7.2.2. $\mu \sim \mu'$ is an equivalence relation.

Proof. We verify the three required properties:

1. Reflexivity: $\mu = (1)\mu$, with $1 \in L^p(|C_i|\mu)$, $p > 1$ by the fact $|C_i|\mu$ is a finite measure.
2. Symmetry: obvious.
3. Transitivity: Suppose $\mu \sim \mu'$ and $\mu' \sim \mu''$. Then $\mu' = \phi\mu$ and $\mu = \phi'\mu'$ with

$$\begin{aligned} \phi &\in \bigcap_{i=0}^n L^p(|C_i|\mu) \text{ for some } p > 1 \\ \phi' &\in \bigcap_{i=0}^n L^q(|C_i|\mu') \text{ for some } q > 1 \end{aligned} \quad (7.16)$$

and $\mu'' = \psi'\mu'$, $\mu' = \psi''\mu''$ with

$$\begin{aligned} \psi' &\in \bigcap_{i=0}^n L^{p'}(|C_i|\mu') \text{ for some } p' > 1 \\ \psi'' &\in \bigcap_{i=0}^n L^{q'}(|C_i|\mu'') \text{ for some } q' > 1. \end{aligned} \quad (7.17)$$

Observe $\mu'' = \psi'\mu' = \psi'\phi\mu$. Let $p'' := \frac{pp'}{p+p'-1}$. Note that $p'' > 1$ since $p, p' > 1$. Let $\lambda := \frac{p'-1}{p+p'-1} \in (0, 1)$. Then $p'' = p'(1-\lambda)$ and $p'' = 1+\lambda(p-1)$.

Fixing i , by Holder's inequality (with conjugates $1/\lambda$ and $1/(1-\lambda)$)

$$\begin{aligned}
\int (\psi' \phi)^{p''} |C_i| \mu &= \int (\psi')^{p''} \phi^{p''} |C_i| \mu \\
&= \int (\psi')^{p'(1-\lambda)} \phi^{1+\lambda(p-1)} |C_i| \mu \\
&= \int (\psi'^{p'} \phi)^{1-\lambda} (\phi^p)^\lambda |C_i| \mu \\
&\leq \left(\int \psi'^{p'} \phi |C_i| \mu \right)^{1-\lambda} \left(\int \phi^p |C_i| \mu \right)^\lambda \\
&= \left(\underbrace{\int \psi'^{p'} |C_i| \mu'}_{\psi' \in L^{p'}(|C_i| \mu')} \right)^{1-\lambda} \left(\underbrace{\int \phi^p |C_i| \mu}_{\phi \in L^p(|C_i| \mu)} \right)^\lambda < \infty.
\end{aligned} \tag{7.18}$$

Similarly, $\mu = \phi' \mu' = \phi' \psi'' \mu''$. Let $q'' := \frac{qq'}{q+q'-1}$ and $\lambda := \frac{q-1}{q+q'-1}$. It can be shown $q'' = q(1-\lambda)$ and $q'' = 1 + \lambda(q'-1)$. Fixing i and applying Holder's inequality

$$\begin{aligned}
\int (\phi' \psi'')^{q''} |C_i| \mu'' &= \int (\psi'')^{q''} (\phi')^{q''} |C_i| \mu'' \\
&= \int (\psi'')^{1+\lambda(q'-1)} (\phi')^{q(1-\lambda)} |C_i| \mu'' \\
&= \int (\psi'' \phi'^q)^{1-\lambda} ((\psi'')^{q'})^\lambda |C_i| \mu'' \\
&\leq \left(\int (\phi')^q \psi'' |C_i| \mu'' \right)^{1-\lambda} \left(\int (\psi'')^{q'} |C_i| \mu'' \right)^\lambda \\
&= \left(\underbrace{\int (\phi')^q |C_i| \mu'}_{\phi' \in L^q(|C_i| \mu')} \right)^{1-\lambda} \left(\underbrace{\int (\psi'')^{q'} |C_i| \mu''}_{\psi'' \in L^{q'}(|C_i| \mu'')} \right)^\lambda < \infty.
\end{aligned} \tag{7.19}$$

Thus, $\phi' \psi'' \in \bigcap_{i=0}^n L^{q''}(|C_i| \mu'')$, $q'' > 1$. We conclude $\mu'' \sim \mu$. \square

Remark 7.2.3. Note that $\mu' = \phi \mu$ where $\phi \in L^p(|C_0| \mu)$ implies $\mu' \ll \mu$ with ϕ the Radon-Nikodym derivative. Similarly, $\mu = \phi' \mu'$ implies that $\mu \ll \mu'$ with ϕ' the Radon-Nikodym derivative of μ with respect to μ' . Then $(\phi)(\phi') = 1$ almost everywhere with respect to either μ or μ' (see [F, Corollary 3.10]). Thus, $\phi' = \frac{1}{\phi}$.

Similarly, $\mu := f\mu_0$ and $\mu' := g\mu_0$ are equivalent elements of \mathcal{F} . Then $\mu = f\mu_0$, $f = \frac{d\mu}{d\mu_0}$ and $\mu' = g\mu_0$, $g = \frac{d\mu'}{d\mu_0}$ and $\mu' \ll \mu_0$. Additionally, $\mu \ll \mu'$ with $\mu = \phi'\mu'$, $\phi' = \frac{d\mu}{d\mu'}$. Then $\mu \ll \mu_0$ and $\frac{d\mu}{d\mu_0} = \frac{d\mu}{d\mu'} \frac{d\mu'}{d\mu_0}$ implies $f = \phi'g$. (see [F, Proposition 3.9]). Equivalently $\phi' = \frac{f}{g}$. A similar argument applying $\mu' \ll \mu$, $\phi = \frac{d\mu'}{d\mu}$ implies $\phi = \frac{g}{f}$. This details the relationship between ϕ and the probability density functions f, g , which will be used below. Additionally, this shows the connection between the chart defined below in Chapter 8 (and [AMR, Theorem 3.4]) and [PS].

We will apply Remark 7.2.3 to simplify the application of the equivalence relation.

Note that the equivalence class of μ is non-empty. Let ϕ be a function bounded away from the origin: $0 < \alpha \leq \phi \leq \beta$. Let $p > 1$. Then

$$\alpha^p |C_i| \mu(\Omega) \leq \int_{\Omega} \phi^p |C_i| \mu \leq \beta^p |C_i| \mu(\Omega) \quad (7.20)$$

implies $\phi \in \bigcap_{i=0}^n L^p(|C_i| \mu)$ by the fact that each $|C_i| \mu$ is a finite measure. Similarly, since ϕ is bounded away from 0, $0 < \frac{1}{\beta} \leq \frac{1}{\phi} \leq \frac{1}{\alpha}$. This implies

$$\frac{1}{\beta^p} |C_i| \mu(\Omega) \leq \int_{\Omega} \left(\frac{1}{\phi}\right)^p |C_i| \mu \leq \frac{1}{\alpha^p} |C_i| \mu(\Omega). \quad (7.21)$$

and $\frac{1}{\phi} \in \bigcap_{i=0}^n L^p(|C_i| \mu)$. Thus, the equivalence class is non-empty.

7.3 e-convergence

We consider a modification to the notion of e -convergence found in [AJVLS, Definition 3.13].

Definition 7.3.1. Let g_n, g be measurable functions such that g_n and g are positive μ_0 a.e. We say that a sequence g_n is e -convergent to g in \mathcal{F} if

$$\lim_{n \rightarrow \infty} \int_{\Omega} \left| \left(\frac{g_n}{g} \right) - 1 \right|^p |C_i| g \mu_0 = 0 \quad \forall p \geq 1, i = 0, 1, \dots, n \quad (7.22)$$

and

$$\lim_{n \rightarrow \infty} \int_{\Omega} \left| \left(\frac{g}{g_n} \right) - 1 \right|^p |C_i| g \mu_0 = 0 \quad \forall p \geq 1, \quad i = 0, 1, \dots, n \quad (7.23)$$

We denote e -convergence by $g_n \xrightarrow{e} g$.

Following [AJVLS, Proposition 3.7], e -convergence can be restated in several different ways (see Section 13.2 for proof). Note that we will only use (1) and (2) in the below proposition. The proof given in the appendix reflects this.

Proposition 7.3.2. *The following are equivalent:*

1. *The sequence g_n is e -convergent to g .*
2. *For all $p \geq 1$, we have*

$$\lim_{n \rightarrow \infty} \int \left| \left(\frac{g_n}{g} \right)^p - 1 \right| |C_i| g \mu_0 = \lim_{n \rightarrow \infty} \int \left| \left(\frac{g}{g_n} \right)^p - 1 \right| |C_i| g \mu_0 = 0 \quad (7.24)$$

for $i = 0, 1, \dots, n$.

3. *The following conditions hold:*

(a) *g_n converges to g in $L^1(|C_i| \mu_0)$, $i = 0, 1, \dots, n$.*

(b) *for all $p \geq 1$, we have*

$$\begin{aligned} \limsup_{n \rightarrow \infty} \int \left(\frac{g_n}{g} \right)^p |C_i| g \mu_0 &< \infty \\ \limsup_{n \rightarrow \infty} \int \left(\frac{g}{g_n} \right)^p |C_i| g \mu_0 &< \infty \end{aligned} \quad (7.25)$$

$i = 0, 1, \dots, n$

Note the following relationship between the equivalence relation \sim and e -convergence.

Proposition 7.3.3. *Let $f\mu_0 \sim g\mu_0$ and g_n a sequence e -converging to g . Then, for all $p \geq 1$,*

$$\lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p |C_i|g\mu_0 = 0 \iff \lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p |C_i|f\mu_0 = 0 \quad (7.26)$$

Proof. \Rightarrow : Suppose

$$\lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p |C_i|g\mu_0 = 0 \quad (7.27)$$

for all $p \geq 1$. Since $f\mu_0 \sim g\mu_0$, there exists $\frac{1}{\phi} \in L^{q'}(|C_i|g\mu_0)$, $q' > 1$ such that $\frac{1}{\phi}g\mu_0 = f\mu_0$. Then, for all $p \geq 1$,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p |C_i|f\mu_0 \\ &= \lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p \frac{1}{\phi} |C_i|g\mu_0 \\ &\leq \lim_{n \rightarrow \infty} \left(\int \left(\left| \frac{g_n}{g} - 1 \right|^p \right)^{p'} |C_i|g\mu_0 \right)^{1/p'} \left(\int \left(\frac{1}{\phi} \right)^{q'} |C_i|g\mu_0 \right)^{1/q'} \end{aligned} \quad (7.28)$$

where the last line follows from Holder's inequality. Note $\int \left(\frac{1}{\phi} \right)^{q'} |C_i|g\mu_0 < \infty$ since $\frac{1}{\phi} \in L^{q'}(|C_i|g\mu_0)$. By hypothesis

$$\lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p |C_i|g\mu_0 = 0 \quad (7.29)$$

for all $p \geq 1$, therefore,

$$\lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p |C_i|f\mu_0 = 0 \quad (7.30)$$

\Leftarrow : Now suppose

$$\lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p |C_i|f\mu_0 = 0 \quad (7.31)$$

for all $p \geq 1$. Since $f\mu_0 \sim g\mu_0$, there exists $\phi \in L^{q'}(|C_i|f\mu_0)$, $q' > 1$ such that

$\phi f \mu_0 = g \mu_0$. Then, for all $p \geq 1$,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p |C_i| g \mu_0 \\ &= \lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p \phi |C_i| f \mu_0 \\ &\leq \lim_{n \rightarrow \infty} \left(\int \left(\left| \frac{g_n}{g} - 1 \right|^p \right)^{p'} |C_i| f \mu_0 \right)^{1/p'} \left(\int (\phi)^{q'} |C_i| f \mu_0 \right)^{1/q'} \end{aligned} \quad (7.32)$$

where the last line follows from Holder's inequality. Note $\int (\phi)^{q'} |C_i| f \mu_0 < \infty$ since $\phi \in L^{q'}(|C_i| f \mu_0)$. By hypothesis

$$\lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p |C_i| f \mu_0 = 0 \quad (7.33)$$

for all $p \geq 1$, therefore,

$$\lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p |C_i| g \mu_0 = 0 \quad (7.34)$$

□

Thus, so long as two measures are equivalent, e -convergence against one implies convergence against the other. Last, note that the above holds for any two measures in the equivalence class, not just the representative of the equivalence class ($f \mu_0$) and the measure the sequence converges to ($g \mu_0$). In other words:

Corollary 7.3.4. *Let $[f]$ be the equivalence class of the measure $f \mu_0$. Let g_n be a sequence converging to some $g \in [f]$. Let $f_1, f_2 \in [f]$. Then g_n e -converges to g with respect to f_1 if and only if g_n e -converges to g with respect to f_2 .*

Note that there is a small difference between the e -convergence defined in [AJVLS] and [PS]: in [PS], the definition of e -convergence is given as g_n converging to g in μ_0 measure and $\limsup \mathbb{E}_g \left[\left(\frac{g_n}{g} \right)^p \right] < \infty$ and $\limsup \mathbb{E}_g \left[\left(\frac{g}{g_n} \right)^p \right] < \infty$, for all $p \geq 1$. Note that this is similar to (3) in Proposition 7.3.2.

It is of course not true that convergence in L^1 is equivalent to convergence in measure. However, if g_n and g are restricted to probability density functions, it is true that convergence in measure implies convergence in L^1 . This can be shown one of two ways.

The first approach is to apply Scheffe's Lemma (see [S]). The second approach is to recall that if a sequence of functions f_n converges pointwise μ -a.e. to f , and if $\int |f_n| \mu \rightarrow \int |f| \mu$, then f_n converges to f in L^1 , see [F, Chapter 6 Exercise 10]. Note that the second condition is always satisfied if f_n and f are probability density functions. Given convergence in measure, the first condition will be true up to subsequence. It can then be argued by contradiction that the entire sequence converges in L^1 to f . Note that in both approaches it is necessary to know that the limiting function f is a probability density function beforehand.

Since [AJVLS] is concerned with finite, not probability, densities it cannot be said that convergence in measure implies convergence in L^1 . Therefore, while the definition of e -convergence in [AJVLS] implies the e -convergence of [PS], the reverse is not true. Since we are interested in finite densities our definition of e -convergence follows [AJVLS].

Chapter 8

Finite Measures

In this chapter we build the manifold structure of \mathcal{F} . We first construct the model space (8.3). After, proving a number of useful properties, we then show this model space endows \mathcal{F} with a C^∞ manifold structure (Theorem 8.2.7). Note much of what follows is adapted from [AJVLS, Section 3.3].

8.1 Preliminaries

Recall

$$\mathcal{F} = \left\{ \mu : \mu \ll \mu_0, \quad \frac{d\mu}{d\mu_0} > 0 \quad \left| \int_{\Omega} C_i \mu \right| < \infty \quad i = 0, \dots, n \right\}. \quad (8.1)$$

Note that the condition on C_0 ensures that the density function itself is finite. We first construct the set, X_μ that will be the model space for \mathcal{F} .

Consider the following definitions (c.f [AJVLS, p. 176]).

Definition 8.1.1. For $\mu \in \mathcal{F}$, let

$$\begin{aligned}
\widehat{B}_\mu &= \{f : \Omega \rightarrow \overline{\mathbb{R}} : e^f \in L^1(|C_i|\mu), i = 0, 1, \dots, n\}, \\
B_\mu &= \widehat{B}_\mu \cap (-\widehat{B}_\mu) \\
&= \{f : \Omega \rightarrow \overline{\mathbb{R}} : e^{\pm f} \in L^1(|C_i|\mu), i = 0, 1, \dots, n\} \\
&\stackrel{\text{Lem 8.1.6}}{=} \{f : \Omega \rightarrow \overline{\mathbb{R}} : e^{|f|} \in L^1(|C_i|\mu), i = 0, 1, \dots, n\} \\
B_\mu^0 &= \{f \in B_\mu : (1+s)f \in B_\mu \text{ for some } s > 0\}.
\end{aligned} \tag{8.2}$$

Finally, we define the proposed model space:

$$X_\mu = \{f : \Omega \rightarrow \overline{\mathbb{R}} : tf \in B_\mu \text{ for some } t \neq 0\}. \tag{8.3}$$

Remark 8.1.2. The description of the model space X_μ can be greatly simplified as a result of Lemma 8.1.7. See (8.25).

Note the above spaces are constructed with respect to non-negative measures. As noted in Equations 7.13 and 7.14, using $|C_i|\mu$ as our finite measure is equivalent to taking $C_i\mu$ because of the integrability of C_i .

Lemma 8.1.3. \widehat{B}_μ is a convex set.

Proof. Let $f, g \in \widehat{B}_\mu$ so that $\int e^f |C_i|\mu$ and $\int e^g |C_i|\mu$ are finite for all i . By the convexity of the exponential function we have

$$\int e^{\lambda f + (1-\lambda)g} |C_i|\mu \leq \lambda \int e^f |C_i|\mu + (1-\lambda) \int e^g |C_i|\mu < +\infty \tag{8.4}$$

for all i . We conclude \widehat{B}_μ is a convex set. \square

Example 8.1.4. Let μ_0 be the Lebesgue measure on \mathbb{R} and $\Omega = [0, 1]$.

First, consider $f(x) = \ln(x)$. Note that

$$\int_{[0,1]} e^{\ln(x)} dx = \int_{[0,1]} x dx < +\infty \tag{8.5}$$

but that

$$\int_{[0,1]} e^{-\ln(x)} dx = \int_{[0,1]} \frac{1}{x} dx = +\infty. \quad (8.6)$$

Thus, $f \in \widehat{B}_\mu$ but $f \notin B_\mu$.

Similarly, consider $f(x) = \ln(x^{2/3})$. Then $(1+1/2)f = \ln(x)$, so that $(1+1/2)f \notin B_\mu$. However, $(1+1/4)f = \ln x^{5/6}$ so that

$$\begin{aligned} \int_{[0,1]} e^{(1+1/4)f} dx &= \int_{[0,1]} x^{5/6} dx < +\infty \\ \int_{[0,1]} e^{(1+1/4)f} dx &= \int_{[0,1]} \frac{1}{x^{5/6}} dx < +\infty. \end{aligned} \quad (8.7)$$

Then there exists $s > 0$ such that $(1+s)f \in B_\mu$. Thus, $f \in B_\mu^0$.

Lemma 8.1.5. X_μ is a linear space.

Proof. Let $\alpha, \beta \in \mathbb{R}$. Let f and g be such that

$$\begin{aligned} \int_{\Omega} e^{|t_1 f|} |C_i| \mu &< +\infty \\ \int_{\Omega} e^{|t_2 g|} |C_i| \mu &< +\infty \end{aligned} \quad (8.8)$$

for some $t_1, t_2 \neq 0$ for $i = 0, 1, \dots, n$. By the triangle inequality and Holder's

$$\begin{aligned} \int_{\Omega} e^{|t(\alpha f + \beta g)|} |C_i| \mu &\leq \int_{\Omega} e^{|t\alpha f|} e^{|t\beta g|} |C_i| \mu \\ &\leq \left(\int_{\Omega} e^{|2t\alpha f|} |C_i| \mu \right)^{1/2} \left(\int_{\Omega} e^{|2t\beta g|} |C_i| \mu \right)^{1/2}. \end{aligned} \quad (8.9)$$

Taking $t = \min\left(\frac{|t_1|}{2|\alpha|}, \frac{|t_2|}{2|\beta|}\right)$, each of the above integrals are finite. This follows from $e^{2\alpha t|f|} \leq e^{|t_1 f|}$ and $e^{2\beta t|g|} \leq e^{|t_2 g|}$, in addition to the fact $e^{|tx|}$ is strictly increasing in t .

Thus, X_μ is a linear space. \square

Lemma 8.1.6.

$$\{f : \Omega \rightarrow \overline{\mathbb{R}} : e^{\pm f} \in L^1(|C_i| \mu)\} = \{f : \Omega \rightarrow \overline{\mathbb{R}} : e^{|f|} \in L^1(|C_i| \mu)\} \quad (8.10)$$

for $i = 0, 1, \dots, n$.

Proof. First, assume $e^{\pm f} \in L^1(|C_i|\mu)$ for all $i = 0, 1, \dots, n$, then

$$\int_{f \geq 0} e^f |C_i| \mu + \int_{f < 0} e^f |C_i| \mu < \infty \quad (8.11)$$

and

$$\int_{f \geq 0} e^{-f} |C_i| \mu + \int_{f < 0} e^{-f} |C_i| \mu < \infty. \quad (8.12)$$

Therefore,

$$\int_{\Omega} e^{|f|} |C_i| \mu = \int_{f \geq 0} e^f |C_i| \mu + \int_{f < 0} e^{-f} |C_i| \mu < \infty. \quad (8.13)$$

Next, assume $\int e^{|f|} |C_i| \mu < +\infty$ for all i . Then

$$\int_{\Omega} e^{|f|} |C_i| \mu = \int_{f \geq 0} e^f |C_i| \mu + \int_{f < 0} e^{-f} |C_i| \mu < \infty. \quad (8.14)$$

Observe that

$$\int_{f < 0} e^f |C_i| \mu \leq \int_{f < 0} |C_i| \mu \leq \int_{\Omega} |C_i| \mu < \infty, \quad (8.15)$$

and that

$$\int_{f \geq 0} e^{-f} |C_i| \mu \leq \int_{f \geq 0} |C_i| \mu \leq \int_{\Omega} |C_i| \mu < \infty. \quad (8.16)$$

Therefore, both

$$\int_{\Omega} e^f |C_i| \mu = \int_{f \geq 0} e^f |C_i| \mu + \int_{f < 0} e^f |C_i| \mu < \infty \quad (8.17)$$

and

$$\int_{\Omega} e^{-f} |C_i| \mu = \int_{f \geq 0} e^{-f} |C_i| \mu + \int_{f < 0} e^{-f} |C_i| \mu < \infty. \quad (8.18)$$

□

Note that in the definition of B_{μ}^0 , we could instead require that there exist s_1, \dots, s_n all strictly positive, such that $e^{\pm(1+s_i)f} \in L^1(|C_i|\mu)$ for all i . In particular, we could have taken

$$B_{\mu}^0 = \{f : \Omega \rightarrow \overline{\mathbb{R}} : e^{(1+s_i)|f|} \in L^1(|C_i|\mu) \text{ for some } s_i > 0, i = 0, \dots, n\} \quad (8.19)$$

However, each $|C_i|\mu$ is a finite measure, so that $f \in L^p(|C_i|\mu)$ implies $f \in L^q(|C_i|\mu)$ for all $q \leq p$. Therefore, $e^f \in L^{1+s_i}(|C_i|\mu)$ implies $e^f \in L^{1+s}(|C_i|\mu)$ when $s \leq s_i$ and we simplify the definition by taking s to be the minimum of s_i .

The following proposition gives the relationship between Orlicz spaces and the model space, X_μ .

Lemma 8.1.7. *Let μ be a finite measure on Ω . Then $f \in X_\mu$ if and only if*

$$\int_{\Omega} (e^{|tf|} - 1) |C_i|\mu < \infty \text{ for } i = 0, 1, \dots, n. \quad (8.20)$$

for some $t > 0$.

Proof. The forward direction follows from

$$\begin{aligned} f \in X_\mu &\implies tf \in B_\mu \text{ for some } t \neq 0 \\ &\implies \int_{\Omega} e^{|tf|} |C_i|\mu < \infty \end{aligned} \quad (8.21)$$

for all $i = 0, 1, \dots, n$. It follows

$$\int_{\Omega} (e^{|tf|} - 1) |C_i|\mu < \infty \quad (8.22)$$

for all $i = 0, 1, \dots, n$. Thus, there exists $t' = |t| > 0$ such that

$$\int_{\Omega} (e^{|t'f|} - 1) |C_i|\mu < +\infty \quad (8.23)$$

for all i .

For the reverse direction, suppose $\int_{\Omega} (e^{|tf|} - 1) |C_i|\mu < \infty$ for some $t > 0$ for all $i = 0, 1, \dots, n$. It follows immediately that $\int_{\Omega} e^{|tf|} |C_i|\mu < \infty$ for $t > 0$ for all $i = 0, 1, \dots, n$. Thus, $tf \in B_\mu$ and $f \in X_\mu$. \square

Lemma 8.1.7 shows that X_μ is the intersection of the Orlicz spaces, $L^{e^{t| \cdot |}-1}(|C_i| \mu)$ $i = 0, 1, \dots, n$, (see Definition 7.1.1) where

$$\begin{aligned} & L^{e^{t| \cdot |}-1}(|C_i| \mu) \\ &= \left\{ f : \Omega \rightarrow \overline{\mathbb{R}} : \int_{\Omega} (e^{t|f|} - 1) |C_i| \mu < +\infty \text{ for some } t > 0 \right\}, \end{aligned} \quad (8.24)$$

i.e.

$$X_\mu = \bigcap_{i=0}^n L^{e^{t| \cdot |}-1}(|C_i| \mu). \quad (8.25)$$

We will consider X_μ as an intersection of Banach spaces with the norm

$$\|f\|_{L(\mu)} = \max_{0 \leq i \leq n} \left(\|f\|_{L^{e^{t| \cdot |}-1}(|C_i| \mu)} \right) \quad (8.26)$$

(see Proposition 13.6.1) where (as in Definition 7.1.1)

$$\|f\|_{L^{e^{t| \cdot |}-1}(|C_i| \mu)} = \inf \left\{ \alpha > 0 : \int_{\Omega} (e^{|f/\alpha|} - 1) |C_i| \mu \leq 1 \right\}. \quad (8.27)$$

are the standard Orlicz norms.

Remark 8.1.8. *We know X_μ is non-empty since it will contain functions of compact support.*

We now prove a number of useful properties regarding our proposed model space. First, we show that for elements of the same equivalence class within \mathcal{F} (see Definition 7.2.1), the model spaces are equivalent (i.e. equal as sets and with equivalent norms). The following is modeled after [AJVLS, Proposition 3.11].

Proposition 8.1.9. *Let $\mu' \lesssim \mu$ (see Definition 7.2.1). Then $X_\mu \subset X_{\mu'}$ is continuously embedded with respect to the L norm (Equation 8.26) on these tangent spaces. In particular, if $\mu' \sim \mu$, then $X_\mu = X_{\mu'}$ as Banach spaces (equal as sets and with equivalent norms).*

Proof. Let $\mu' \lesssim \mu$ and $u \in X_\mu$. Then $\int_\Omega (e^{t|u|} - 1) |C_i| \mu < \infty$ for some $t > 0$ for all $i = 0, 1, \dots, n$. Additionally, $\mu' = \phi \mu$ with $\phi \in L^p(|C_i| \mu)$ for some $p > 1$ for all i . Let $q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. By Holder's inequality

$$\begin{aligned} \int (e^{t|u|} - 1) |C_i| \mu' &= \int (e^{t|u|} - 1) |C_i| \phi \mu \\ &\leq \|\phi\|_{L^p(|C_i| \mu)} \left(\int (e^{t|u|} - 1)^q |C_i| \mu \right)^{1/q}. \end{aligned} \quad (8.28)$$

Fix i . Define $\psi(t) = \|\phi\|_{L^p(|C_i| \mu)}^q (e^{t|u|} - 1)^q$, noting that ψ is still a Young function (see Definition 7.1.1) and that $u \in L^\psi(|C_i| \mu)$. This follows from

$$\lim_{t \rightarrow \infty} \frac{\psi(t)}{e^{q|t|}} = \|\phi\|_{L^p(|C_i| \mu)}^q \in (0, \infty) \quad (8.29)$$

implying by Proposition 7.1.4 that $L^\psi(|C_i| \mu) = L^{e^{q|t|}-1}(|C_i| \mu)$. However, [AJVLS, Lemma 3.10] states that Young functions $\phi(t)$ and $\phi(\lambda t)$ for $\lambda > 0$ have equivalent Orlicz spaces. In particular, $L^{e^{q|t|}-1}(|C_i| \mu) = L^{e^{t|u|}-1}(|C_i| \mu)$ with equivalent norm. Thus, $u \in L^{e^{t|u|}-1}(|C_i| \mu)$ implies $u \in L^\psi(|C_i| \mu)$. Then (8.28) implies that $u \in L^{e^{t|u|}-1}(|C_i| \mu')$. This shows $L^{e^{t|u|}-1}(|C_i| \mu) \subset L^{e^{t|u|}-1}(|C_i| \mu')$. Since this holds for any i , $\bigcap_{i=0}^n L^{e^{t|u|}-1}(|C_i| \mu) \subset \bigcap_{i=0}^n L^{e^{t|u|}-1}(|C_i| \mu')$, or $X_\mu \subset X_{\mu'}$.

We now show this inclusion is a continuous embedding. Set $a := \|u\|_{L^\psi(|C_i| \mu)}$ and let $t = \frac{1}{a}$, then

$$\|\phi\|_{L^p(|C_i| \mu)} \left(\int (e^{t|u|} - 1)^q |C_i| \mu \right)^{1/q} \leq 1 \quad (8.30)$$

by the definition of the Orlicz norm for ψ (see Definition 7.1.1). Thus, by (8.28)

$$\int (e^{t|u|} - 1) |C_i| \mu' \leq 1. \quad (8.31)$$

implying $\|u\|_{L^\psi(|C_i| \mu)} = a \geq \|u\|_{L^{e^{t|u|}-1}(|C_i| \mu')}$. Therefore, $L^\psi(|C_i| \mu) \subset L^{e^{t|u|}-1}(|C_i| \mu')$

is a continuous inclusion, i.e.

$$L^\psi(|C_i| \mu) \hookrightarrow L^{e^{t|u|}-1}(|C_i| \mu'). \quad (8.32)$$

Since $L^\psi(|C_i|\mu) = L^{e|qt|-1}(|C_i|\mu) = L^{e|t|-1}(|C_i|\mu)$, we have

$$L^{e|t|-1}(|C_i|\mu) \hookrightarrow L^{e|t|-1}(|C_i|\mu'). \quad (8.33)$$

with $\|u\|_{L^{e|t|-1}(|C_i|\mu')} \leq K_i \|u\|_{L^{e|t|-1}(|C_i|\mu)}$, for some $K_i > 0$. Therefore

$$\begin{aligned} \|u\|_{L^{e|t|-1}(|C_i|\mu')} &\leq K_i \|u\|_{L^{e|t|-1}(|C_i|\mu)} \\ &\leq K_i \max_i \left(\|u\|_{L^{e|t|-1}(|C_i|\mu)} \right) \\ &= K_i \|u\|_{L(\mu)} \end{aligned} \quad (8.34)$$

Note that i was arbitrary. Thus the above holds for all i , implying

$$\|u\|_{L(\mu')} \leq K \|u\|_{L(\mu)}, \quad (8.35)$$

with $K = \max_i K_i$. We conclude $X_\mu \hookrightarrow X_{\mu'}$. \square

Remark 8.1.10. *The above argument also shows $L^{e|t|-1}(|C_i|\mu) \hookrightarrow L^{e|t|-1}(|C_i|\mu')$ for individual i .*

The following lemma (see [AJVLS, Page 177]) will be useful in Theorem 8.2.3.

Lemma 8.1.11. *The set*

$$B_\mu^0 \subset X_\mu \quad (8.36)$$

contains the open unit ball in X_μ .

Proof. Let B denote the open unit ball in X_μ :

$$B = \left\{ f \in X_\mu : \|f\|_{L(\mu)} < 1 \right\}. \quad (8.37)$$

Let $f \in B$. Then

$$\max_{0 \leq i \leq n} (\|f\|_{L^{e|t|-1}(|C_i|\mu)}) < 1. \quad (8.38)$$

Set

$$c_i := \inf \left\{ \beta > 0 : \int_{\Omega} (e^{|f/\beta|} - 1) |C_i| \mu < 1 \right\}. \quad (8.39)$$

In particular, $c_i < 1$ imply $\frac{1}{c_i} = 1 + s_i$ for some $s_i > 0$. Letting $s = \min_i s_i$, and recalling $f \in L^{1+s_i}(|C_i|\mu)$ implies $f \in L^{1+s}(|C_i|\mu)$ since $|C_i|\mu$ are finite measures, we conclude $(1+s)f \in B_{\mu}$. Therefore, $f \in B_{\mu}^0$. We conclude $B \subset B_{\mu}^0$. \square

8.2 Manifold Structure

We now build the manifold structure on \mathcal{F} . Let $\mu \in \mathcal{F}$. Our model space is

$$X_{\mu} = \bigcap_{i=0}^n L^{e^{|t|}-1}(|C_i|\mu). \quad (8.40)$$

with norm

$$\|x\|_{L(\mu)} = \max_{0 \leq i \leq n} \|x\|_{L^{e^{|t|}-1}(|C_i|\mu)}. \quad (8.41)$$

The chart between \mathcal{F} and its model space is defined now.

Definition 8.2.1. *Define the map from \mathcal{F} to X_{μ} which on the equivalence class of μ (see Definition 7.2.1) is given by*

$$\log_{\mu}(\mu') = \log_{\mu}(\phi\mu) = \log(\phi). \quad (8.42)$$

where $\phi \in \bigcap_{i=0}^n L^p(|C_i|\mu)$ for some $p > 1$.

Our chart between \mathcal{F} and its model space X_{μ} will be:

Definition 8.2.2. *Let $\mu \in \mathcal{F}$ and let $K_{\mu} = [\mu]$ be the equivalence class of μ . Then the chart at μ is a map $s_{\mu} : K_{\mu} \rightarrow \text{Im}(s_{\mu})$ defined by*

$$s_{\mu}(\mu') = \log_{\mu}(\mu') = \log \phi. \quad (8.43)$$

Since the above mapping is defined via the natural log function it is one to one on each equivalence class. This mapping is then a bijection. We first show that the image of $\log_\mu(K)$ is in fact in X_μ and is an open set. We will then show that s_μ and its inverse are continuous and therefore s_μ is a homeomorphism.

The following is based on [AJVLS, Theorem 3.4]:

Theorem 8.2.3. *Let $K \subset \mathcal{F}$ be an equivalence class with respect to \sim . For a fixed $\mu \in K$, consider X_μ . Then:*

1. $\log_\mu(K) \subset X_\mu$
2. $\log_\mu(K)$ is convex.
3. $\log_\mu(K)$ is open.

In particular, $\log_\mu(K)$ is convex implies, $\log_\mu(K)$ is connected.

Proof. Let $f \in \log_\mu(K)$, then f is the image of some $\mu' \in K$, with $\mu' \sim \mu$. By definition (see Definition 7.2.1) $\mu' \sim \mu$ implies

$$\begin{aligned} \mu' &= \phi_1 \mu, & \phi_1 &\in \bigcap_{i=0}^n L^{p_1}(|C_i|\mu) \\ \mu &= \frac{1}{\phi_1} \mu', & \frac{1}{\phi_1} &\in \bigcap_{i=0}^n L^{p_1}(|C_i|\mu') \end{aligned} \tag{8.44}$$

for some $p_1 > 1$ (where we have set p_1 to be the minimum of the two values that appear in Definition 7.2.1) and f as the image of some μ' implies

$$f = \log_\mu(\mu') = \log_\mu(\phi_1 \mu) = \log(\phi_1) \implies e^f = \phi_1. \tag{8.45}$$

Then, for $s_1 > 0$ such that $p_1 = 1 + s_1$,

$$\begin{aligned}
\int e^{-s_1 f} |C_i| \mu &= \int \phi_1^{-s_1} |C_i| \mu \\
&= \int \left(\frac{1}{\phi_1} \right)^{1+s_1} \phi_1 |C_i| \mu \\
&= \int \left(\frac{1}{\phi_1} \right)^{p_1} |C_i| \mu' < \infty, \quad i = 0, 1, \dots, n \\
&\implies -s_1 f \in \widehat{B}_\mu
\end{aligned} \tag{8.46}$$

and

$$\begin{aligned}
\int e^{(1+s_1)f} |C_i| \mu &= \int \phi_1^{1+s_1} |C_i| \mu \\
&= \int \phi_1^{p_1} |C_i| \mu < \infty \quad i = 0, 1, \dots, n \\
&\implies (1 + s_1)f \in \widehat{B}_\mu.
\end{aligned} \tag{8.47}$$

where $s_1 > 0$. Thus, $s_1 f \in \widehat{B}_\mu$ by the convexity of \widehat{B}_μ (Lemma 8.1.3). Together with $-s_1 f \in \widehat{B}_\mu$ shown above, we have $s_1 f \in B_\mu$. Therefore, $f \in L^{e^{|t|}-1}(|C_i| \mu)$ for all i .

We conclude $f \in X_\mu$.

Next, we argue that $\log_\mu(K)$ is a convex set. Let $f \in \log_\mu(K)$. We first show $\lambda f \in \log_\mu(K)$ for all $\lambda \in [0, 1]$. Noting that

$$\lambda f = \log(\phi_1^\lambda) \tag{8.48}$$

we must show $\mu_\lambda = \phi_1^\lambda \mu$ is equivalent to μ , i.e. $\mu_\lambda \sim \mu$. First, fix $\lambda \in (0, 1)$ and note by Holder's inequality (with conjugates $1/\lambda$ and $1/(1-\lambda)$) that

$$\int (\phi_1^\lambda)^{p_1} |C_i| \mu \leq \left[\int \phi_1^{p_1} |C_i| \mu \right]^\lambda \left[\int 1^{1/(1-\lambda)} |C_i| \mu \right]^{1-\lambda}. \tag{8.49}$$

That the above is finite follows from the fact $\phi_1 \in L^{p_1}(|C_i| \mu)$ for all i and each $|C_i| \mu$ is a finite measure. Thus, $\phi_1^\lambda \in \bigcap_{i=0}^n L^{p_1}(|C_i| \mu)$.

Similarly,

$$\begin{aligned}
\int (\phi_1^{-\lambda})^{p_1} |C_i| \phi_1^\lambda \mu &= \int \phi_1^{-\lambda(p_1-1)} |C_i| \mu \\
&\leq \left[\int \phi_1^{1-p_1} |C_i| \mu \right]^\lambda \left[\int 1^{1/(1-\lambda)} |C_i| \mu \right]^{1-\lambda} \\
&= \left[\int e^{-s_1 f} |C_i| \mu \right]^\lambda \left[\int |C_i| \mu \right]^{1-\lambda}
\end{aligned} \tag{8.50}$$

is finite since $-s_1 f \in \widehat{B}_\mu$ as shown above and the $|C_i| \mu$ are finite measures. We conclude $\frac{1}{\phi_1^\lambda} \in \bigcap_{i=0}^n L^{p_1}(|C_i| \mu_\lambda)$.

For $\lambda \in (0, 1)$ we conclude $\mu_\lambda \sim \mu$. If $\lambda = 0$, then $\mu_\lambda = \mu$ and $\mu \sim \mu$ by the reflexivity of the equivalence relation. If $\lambda = 1$ then $\mu_\lambda = \phi \mu = \mu'$ which is equivalent to μ by hypothesis. We conclude that if $f \in \log_\mu(K)$ then $\lambda f \in \log_\mu(K)$ for $\lambda \in [0, 1]$.

Now we argue $tf + (1-t)g \in \log_\mu(K)$, where

$$\begin{aligned}
\mu'' &= \phi_2 \mu, \quad \phi_2 \in \bigcap_{i=0}^n L^{p_2}(|C_i| \mu) \\
\mu &= \frac{1}{\phi_2} \mu', \quad \frac{1}{\phi_2} \in \bigcap_{i=0}^n L^{p_2}(|C_i| \mu'')
\end{aligned} \tag{8.51}$$

for some $p_2 > 1$, such that

$$g = \log_\mu(\mu'') = \log(\phi_2). \tag{8.52}$$

Note also that $g \in \log_\mu(K)$ implies $-s_2 g \in \widehat{B}_\mu$ for some $s_2 > 0$ such that $p_2 = 1 + s_2$.

Similar to above, observing

$$tf + (1-t)g = \log \phi_1^t \phi_2^{1-t}, \tag{8.53}$$

we must show $\mu_t := \phi_1^t \phi_2^{1-t} \mu$ is equivalent to μ , i.e. $\mu_t \sim \mu$. Let $t \in (0, 1)$. First, without loss of generality let $p = \min(p_1, p_2) = p_1$, and apply Holder's inequality to get

$$\int (\phi_1^t \phi_2^{1-t})^p |C_i| \mu \leq \left[\int \phi_1^{p_1} |C_i| \mu \right]^t \left[\int \phi_2^{p_1} |C_i| \mu \right]^{1-t}. \tag{8.54}$$

The right hand side is finite since $\phi_1 \in L^{p_1}(|C_i|\mu)$ for all i and $\phi_2 \in L^{p_2}(|C_i|\mu)$.

Since $|C_i|\mu$ is a finite measure and $p_1 \leq p_2$, $\phi_2 \in L^{p_1}(|C_i|\mu)$. We conclude $\phi_1^t \phi_2^{1-t} \in \bigcap_{i=1}^n L^p(|C_i|\mu)$. For the other direction of the equivalence relation

$$\begin{aligned} \int \left(\phi_1^{-t} \phi_2^{-(1-t)} \right)^p |C_i| \phi_1^t \phi_2^{1-t} \mu &= \int \phi_1^{-t(p-1)} \phi_2^{-(1-t)(p-1)} |C_i| \mu \\ &\leq \left[\int \phi_1^{-(p-1)} |C_i| \mu \right]^t \left[\int \phi_2^{-(p-1)} |C_i| \mu \right]^{1-t}. \end{aligned} \quad (8.55)$$

Since $p = \min(p_1, p_2) = p_1$, $-(p-1) = -(p_1-1) = -s_1$, we have

$$\int \left(\phi_1^{-t} \phi_2^{-(1-t)} \right)^p |C_i| \phi_1^t \phi_2^{1-t} \mu \leq \left[\int e^{-s_1 f} |C_i| \mu \right]^t \left[\int e^{-s_1 g} |C_i| \mu \right]^{1-t}. \quad (8.56)$$

The first term on the right hand side is finite since $-s_1 f \in \widehat{B}_\mu$. For the second term, we know $-s_2 g \in \widehat{B}_\mu$. It remains to argue this implies $-s_1 g \in \widehat{B}_\mu$. Note that since $p_1 \leq p_2$, $s_1 \leq s_2$. Observe

$$\int e^{-s_1 g} |C_i| \mu = \int_{g \leq 0} e^{-s_1 g} |C_i| \mu + \int_{g > 0} e^{-s_1 g} |C_i| \mu \quad (8.57)$$

and note $g \leq 0$ implies $e^{-s_1 g} \geq 1$ while $g > 0$ implies $e^{-s_1 g} < 1$. Therefore

$$\int e^{-s_1 g} |C_i| \mu \leq \int_{g \leq 0} e^{-s_1 g} |C_i| \mu + \int_{g > 0} |C_i| \mu. \quad (8.58)$$

Finally, note that for $g \leq 0$, $0 < s_1 \leq s_2$ implies $e^{-s_1 g} \leq e^{-s_2 g}$ so that

$$\int e^{-s_1 g} |C_i| \mu \leq \int_{g \leq 0} e^{-s_2 g} |C_i| \mu + \int_{g > 0} |C_i| \mu. \quad (8.59)$$

The first integral is finite since $-s_2 g \in \widehat{B}_\mu$ and the second since $|C_i|\mu$ is a finite measure. Therefore, $-s_1 g \in \widehat{B}_\mu$ and

$$\int \left(\phi_1^{-t} \phi_2^{-(1-t)} \right)^p |C_i| \phi_1^t \phi_2^{1-t} \mu < \infty. \quad (8.60)$$

We conclude $\phi_1^{-t}\phi_2^{-(1-t)} \in L^p(|C_i|\mu_t)$. Thus, for $t \in (0, 1)$ $\mu_t \sim \mu$ and $tf + (1-t)g$ is the image of an element of $\log_\mu(K)$. Note for the case $t = 0$ or $t = 1$, μ_t equals μ'' and μ' , respectively. We conclude $\log_\mu(K)$ is convex.

Last, we show $\log_\mu(K)$ is open. We first argue that for $\mu \sim \mu'$, $\log_{\mu'}(K)$ is a shift of $\log_\mu(K)$. Let $g \in \log_\mu(K)$ so that

$$g = \log_\mu(\mu') = \log_\mu(\phi\mu) = \log \phi \quad (8.61)$$

with $\mu' \sim \mu$. This implies $e^g = \phi$, or $\mu' = e^g\mu$. Thus, $g \in \log_\mu(K)$ implies $\mu' = e^g\mu \in K$ and $\mu' = e^g\mu \in K$ implies $\log_\mu(e^g\mu) = g$. In particular, $g \in \log_\mu(K)$ if and only if $\mu' = e^g\mu \in K$. Similarly, $f \in \log_{\mu'}(K)$ if and only if $\mu = e^f\mu' \in K$. Combining these gives the following chain of equivalences:

$$\begin{aligned} f \in \log_{\mu'}(K) &\iff e^f\mu' \in K \\ &\iff e^{f+g}\mu \in K \\ &\iff f + g \in \log_\mu(K). \end{aligned} \quad (8.62)$$

Therefore, $f \in \log_{\mu'}(K)$ if and only if $f + g \in \log_\mu(K)$. However, $f \in \log_{\mu'}(K)$ implies $f + g \in \log_\mu(K)$ results in $\log_{\mu'}(K) \subset g + \log_\mu(K)$. Alternatively, $f + g \in \log_\mu(K)$ implies $f \in \log_{\mu'}(K)$ results in $g + \log_\mu(K) \subset \log_{\mu'}(K)$. Thus, for some fixed $g \in \log_\mu(K)$, $\log_{\mu'}(K) = g + \log_\mu(K)$.

Since $\log_{\mu'}(K)$ is a shift of $\log_\mu(K)$ by some fixed $g \in \log_\mu(K)$, it suffices to show that 0 is an interior point of $\log_{\mu'}(K)$ for any $\mu' \in K$. Then for any point $f \in \log_\mu(K)$ we shift f to 0 and a neighborhood of 0 will be contained in $\log_{\mu'}(K)$. Shifting back will give a neighborhood of f contained in $\log_\mu(K)$, i.e. any point of $\log_\mu(K)$ is contained in an open set in $\log_\mu(K)$. Note by Lemma 8.1.11 that $B_{\mu'}^0$ contains the unit ball in the $X_{\mu'}$ norm. Thus, 0 and a neighborhood of 0 are contained in $B_{\mu'}^0$.

Thus, if $B_{\mu'}^0 \subset \log_{\mu'}(K)$, $\log_{\mu'}(K)$ contains a neighborhood of 0, i.e. 0 is an interior point of $\log_{\mu'}(K)$. We argue $B_{\mu'}^0 \subset \log_{\mu'}(K)$.

Let $f \in B_{\mu'}^0$ so that

$$(1 + s')f \in B_{\mu'} \text{ for some } s' > 0 \quad (8.63)$$

by definition, implying

$$e^{\pm(1+s')f} \in L^1(|C_i|\mu') \quad i = 0, 1, \dots, n. \quad (8.64)$$

In particular

$$e^f \in L^{1+s'}(|C_i|\mu'), \quad i = 0, 1, \dots, n. \quad (8.65)$$

Similarly,

$$\begin{aligned} & \int e^{-(1+s')f} |C_i|\mu' < \infty \quad i = 0, 1, \dots, n \\ \implies & \int e^{-(2+s')f} e^f |C_i|\mu' < \infty \quad i = 0, 1, \dots, n \\ \implies & \int (e^{-f})^{2+s'} e^f |C_i|\mu' < \infty \quad i = 0, 1, \dots, n \end{aligned} \quad (8.66)$$

giving

$$e^{-f} \in L^{2+s'}(|C_i|e^f\mu') \quad i = 0, 1, \dots, n. \quad (8.67)$$

By 8.65 and 8.67, $e^f\mu' \sim \mu'$. Therefore $\mu' \sim \mu$ implies $e^f\mu' \in K$ and so $f \in \log_{\mu'}(K)$. \square

Next, we define the inverse of s_μ :

Definition 8.2.4. Let $\mu \in \mathcal{F}$. Then the mapping $s_\mu^{-1} : \text{Im}(s_\mu) \rightarrow K_\mu$ defined by

$$s_\mu^{-1}(u) = e^u \mu \quad (8.68)$$

is the inverse of Definition 8.2.2 (see Lemma 8.2.5).

Lemma 8.2.5. *Let $\mu \in \mathcal{F}$. Then the inverse of s_μ in Definition 8.2.2 is $s_\mu^{-1} : \text{Im}(s_\mu) \rightarrow K_\mu$ defined by*

$$s_\mu^{-1}(u) = e^u \mu. \quad (8.69)$$

Proof. First, we note $e^u \mu$ is well defined since

$$u = \log \phi \quad (8.70)$$

for some $\phi \in \bigcap_{i=0}^n L^p(|C_i| \mu)$ with $\frac{1}{\phi} \in \bigcap_{i=0}^n L^p(|C_i| \mu')$, $p > 1$, for $\mu' = \phi \mu$. Then

$$\phi = e^u \quad (8.71)$$

and $\phi \in L^p(\mu)$ implies $e^u \in L^p(\mu)$. Therefore, e^u is finite almost everywhere.

Next, $\mu, \mu' \in K_\mu$ so that $\mu' = \phi \mu$, $\phi \in \bigcap_{i=0}^n L^p(|C_i| \mu)$, $p > 1$. Then

$$s_\mu^{-1}(s_\mu(\mu')) = e^{\log_\mu(\phi \mu)} = e^{\log \phi} = \phi. \quad (8.72)$$

Last,

$$s_\mu(s_\mu^{-1}(u)) = \log_\mu(e^u \mu) = \log(e^u) = u. \quad (8.73)$$

We conclude s_μ^{-1} is the inverse of s_μ . \square

Last, we prove the chart in Definition 8.2.1 and its inverse in Definition 8.2.4 are continuous (c.f [AJVLS, Proposition 3.13]).

Proposition 8.2.6. *The sequence g_n e -converges to g_0 ($g_n \xrightarrow{e} g_0$) if and only if $g_n \mu_0 \sim g_0 \mu_0$ for n large enough and $u_n \rightarrow u_0$ in the $L(g_0 \mu_0)$ norm as $n \rightarrow \infty$ (see 8.41).*

Proof. \Rightarrow : Let $g_n \xrightarrow{e} g_0$. Then

$$\int \left| \left(\frac{g_n}{g_0} \right) - 1 \right|^p |C_i| g_0 \mu_0 \rightarrow 0 \quad \forall p \geq 1, \quad i = 0, 1, \dots, n \quad (8.74)$$

and

$$\int \left| \left(\frac{g_0}{g_n} \right) - 1 \right|^p |C_i| g_0 \mu_0 \rightarrow 0 \quad \forall p \geq 1, i = 0, 1, \dots, n. \quad (8.75)$$

Then by reverse triangle inequality, for large n , $\left(\frac{g_n}{g_0} \right), \left(\frac{g_0}{g_n} \right) \in L^p(|C_i| g_0 \mu_0)$ for all $p \geq 1$ and for all i . In particular, this implies $\phi = \frac{g_n}{g_0} \in L^p(|C_i| g_0 \mu_0)$ for all i .

Therefore, $g_n \mu_0 \lesssim g_0 \mu_0$.

The other direction of the equivalence relation follows from (8.75):

$$\begin{aligned} & \int \left(\frac{g_0}{g_n} \right)^p |C_i| g_0 \mu_0 < \infty \quad i = 0, \dots, n \\ \implies & \int \left(\frac{g_0}{g_n} \right)^{p+1} |C_i| g_n \mu_0 < \infty \quad i = 0, \dots, n \end{aligned} \quad (8.76)$$

implying $\phi' = \frac{g_0}{g_n} \in L^{p+1}(|C_i| g_n \mu_0)$ for all i . Thus, $g_0 \mu_0 \lesssim g_n \mu_0$ and we conclude $g_0 \mu_0 \sim g_n \mu_0$ for n large enough.

Let $u_n = \log g_n \in X_{g_n \mu_0}$. Then $g_0 \mu_0 \lesssim g_n \mu_0$ implies $X_{g_n \mu_0} \subset X_{g_0 \mu_0}$ for large enough n (see Proposition 8.1.9). Thus, $u_n \in X_{g_0 \mu_0}$ for n large enough. It remains to show $u_n \rightarrow u_0$.

Since, $g_n \xrightarrow{e} g_0$ implies e -convergence against all $|C_i| g_0$, by Proposition 7.3.2 we have

$$\lim_{n \rightarrow \infty} \int \left| \left(\frac{g_n}{g_0} \right)^p - 1 \right| |C_i| g_0 \mu_0 = 0 \quad (8.77)$$

and

$$\lim_{n \rightarrow \infty} \int \left| \left(\frac{g_0}{g_n} \right)^p - 1 \right| |C_i| g_0 \mu_0 = 0 \quad (8.78)$$

for all $p \geq 1$ and $i = 0, 1, \dots, n$. Therefore,

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \int \left[\left| \left(\frac{g_n}{g_0} \right)^p - 1 \right| + \left| \left(\frac{g_0}{g_n} \right)^p - 1 \right| \right] |C_i| g_0 \mu_0 \\ &= \lim_{n \rightarrow \infty} \int \left[|e^{p(u_n - u_0)} - 1| + |e^{p(u_0 - u_n)} - 1| \right] |C_i| g_0 \mu_0 \\ &= \lim_{n \rightarrow \infty} \int 2 \sinh(p|u_n - u_0|) |C_i| g_0 \mu_0 \quad \forall p \geq 1 \text{ and } i = 1, \dots, n. \end{aligned} \quad (8.79)$$

This also holds for $p \in [0, 1)$ since $\sinh(|t|)$ is an increasing function for $t \geq 0$.

Proposition 7.1.3 implies $u_n \rightarrow u_0$ in $L^{2\sinh|t|}(|C_i|g_0\mu_0)$ for all i , and

$$\lim_{t \rightarrow \infty} \frac{2\sinh|t|}{e^{|t|} - 1} = 1 \in (0, \infty) \quad (8.80)$$

implies $L^{2\sinh|t|}(|C_i|g_0\mu_0) = L^{e^{|t|}-1}(|C_i|g_0\mu_0)$, by Proposition 7.1.4. Thus we have $u_n \rightarrow u_0$ in $L^{e^{|t|}-1}(|C_i|g_0\mu_0)$ for all i .

Therefore, $u_n \rightarrow u_0$ in $X_{g_0\mu_0}$.

\Leftarrow : Now suppose that $g_n\mu_0 \sim g_0\mu_0$ for n large enough and $u_n \rightarrow u_0$ in $X_{g_0\mu_0}$. We want to show $g_n \xrightarrow{e} g_0$. Since $u_n \rightarrow u_0$ in $X_{g_0\mu_0}$ we have $u_n \rightarrow u_0$ in the Orlicz space $L^{e^{|t|}-1}(g_0|C_i|\mu_0)$ which is equivalent to $L^{2\sinh|t|}(g_0|C_i|\mu_0)$, implying

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \int 2\sinh(p|u_n - u_0|) |C_i|g_0\mu_0 \quad \forall p \geq 1, \quad i = 0, 1, \dots, n \\ &= \lim_{n \rightarrow \infty} \int [|e^{p(u_n - u_0)} - 1| + |e^{p(u_0 - u_n)} - 1|] |C_i|g_0\mu_0 \\ &= \lim_{n \rightarrow \infty} \int \left[\left| \left(\frac{g_n}{g_0} \right)^p - 1 \right| + \left| \left(\frac{g_0}{g_n} \right)^p - 1 \right| \right] |C_i|g_0\mu_0, \end{aligned} \quad (8.81)$$

giving $g_n \xrightarrow{e} g_0$ by Proposition 7.3.2. \square

We conclude s_μ is a homeomorphism. Moreover, by Theorem 8.2.3 we know the image of K_μ is connected. Since K_μ and its image are homeomorphic, K_μ is also connected. Thus the equivalence relation on \mathcal{F} partitions \mathcal{F} into connected components, each of which is homeomorphic to an intersection of Orlicz spaces. It remains to show that the collection (s_μ, K_μ) forms an atlas on \mathcal{F} .

Theorem 8.2.7. *The collection (K_μ, s_μ) is an atlas on \mathcal{F} (see [AMR], Definition 3.1.1, Page 123).*

Proof. Note that since the transition map is a composition of homeomorphisms, it remains only to show that it is C^∞ . Let $K_\mu, K_{\mu'}$ be subsets of \mathcal{F} such that $K_\mu \cap K_{\mu'} \neq$

\emptyset . Since K_μ and $K_{\mu'}$ are equivalence relations, this implies $\tilde{K} := K_\mu \cap K_{\mu'} = K_{\mu'} = K_\mu$ and $\mu' \sim \mu$ where $\mu' = \phi\mu$, with $\phi \in L^p(\mu)$ and $\frac{1}{\phi} \in L^p(\mu')$ for some $p > 1$. Let $u \in s_{\mu'}(\tilde{K}) = s_{\mu'}(K_\mu)$. Then the transition map

$$s_\mu \circ s_{\mu'}^{-1} : s_{\mu'}(\tilde{K}) \rightarrow s_\mu(\tilde{K}) \quad (8.82)$$

is given by

$$\begin{aligned} s_\mu(s_{\mu'}^{-1}(u)) &= \log_\mu(e^u \mu') \\ &= \log_\mu(e^u \phi\mu) \\ &= \log(e^u \phi) \\ &= u + \log \phi \end{aligned} \quad (8.83)$$

We argue the transition mapping is a C^∞ diffeomorphism. For ease of notation, define

$$I[u] = u + \log \phi. \quad (8.84)$$

Let $v \in s_{\mu'}(\tilde{K}) \subset X_{\mu'}$ then

$$\begin{aligned} d_u I[v] &= \left. \frac{d}{dt} \right|_{t=0} ((u + tv) + \log \phi) \\ &= v \end{aligned} \quad (8.85)$$

is the Gateaux derivative in the direction of v . Note that

$$\begin{aligned} \lim_{\|v\|_{L(\mu')} \rightarrow 0} \frac{\|I[u+v] - (I[u] + d_u I[v])\|_{L(\mu)}}{\|v\|_{L(\mu')}} &= \lim_{\|v\|_{L(\mu')} \rightarrow 0} \frac{0}{\|v\|_{L(\mu')}} \\ &= 0. \end{aligned} \quad (8.86)$$

Thus the mapping $u \rightarrow d_u I = v$ is the derivative of I by uniqueness. Note that derivative is continuous: let $u_n \rightarrow u$ in $s_{\mu'}(\tilde{K})$. Then

$$\begin{aligned} \left| \sup_{\|x\|_{L(\mu')}=1} d_{u_n} I[x] - d_u I[x] \right| &= \left| \sup_{\|x\|_{L(\mu')}=1} v - v \right| \\ &= 0. \end{aligned} \quad (8.87)$$

Therefore, I is C^1 . Next, we take the second order directional (Gateaux) derivative in the direction of v :

$$\begin{aligned} d_u^2 I(v, v) &= \left. \frac{d^2}{dt^2} \right|_{t=0} ((u + tv) + \log \phi) \\ &= \left. \frac{d}{dt} \right|_{t=0} v \\ &= 0. \end{aligned} \tag{8.88}$$

Note that

$$\begin{aligned} &\lim_{\|w\|_{L(\mu')} \rightarrow 0} \frac{\|d_{u+w} I[v] - (d_u I[v] - d_u^2 I(v, v))\|_{L(\mu)}}{\|w\|_{L(\mu')}} \\ &= \lim_{\|w\|_{L(\mu')} \rightarrow 0} \frac{\|d_{u+w} I[v] - d_u I[v]\|_{L(\mu)}}{\|w\|_{L(\mu')}} \\ &= 0 \end{aligned} \tag{8.89}$$

Thus, I is twice differentiable. Clearly, I will also be C^2 . Since the second derivative is 0, all higher order derivatives will also be 0 and continuous. We conclude I is C^∞ and that the collection (K_μ, s_μ) is an atlas on \mathcal{F} . \square

We have now shown \mathcal{F} is endowed with a Banach manifold structure with X_μ serving as the model space. Before applying this manifold structure to the problem of deriving the Gibbs distribution (Chapter 9), we detail the connection between the model space X_μ and the tangent space of \mathcal{F} at μ : $T_\mu \mathcal{F}$. Recall the tangent space to a point μ on a manifold is the set of equivalence classes $[\gamma]_\mu$ of curves at μ , where two curves are equivalent if they have the same tangent vector, with the same tangent vector (see [AMR, Definition 3.4.3]):

$$T_\mu(\mathcal{F}) = \{[\gamma]_\mu : \gamma \text{ is a curve at } \mu\}. \tag{8.90}$$

Alternatively, we can represent each equivalence class $[\gamma]$ by the tangent vector $u \in X_\mu$ common to all curves in that class.

For the model space, let c be a curve through some $u \in s_\mu(K)$. It can be shown (see [AMR, Lemma 3.4.4]) that for each equivalence class of curves through u , there is a unique $\mathbf{e} \in X_\mu$ such that the curve $c_{u,\mathbf{e}}(t) = u + t\mathbf{e}$ is tangent to c : $[c]_u = [c_{u,\mathbf{e}}]_u$. Then ([AMR, Lemma 3.4.4]) $i : s_\mu(K) \times X_\mu \rightarrow T(s_\mu(K))$ defined by $i(u, \mathbf{e}) = [c_{u,\mathbf{e}}]_u$ is an isomorphism. Thus, the tangent space $T_\mu\mathcal{F}$ can be identified with the fibres of the bundle isomorphism corresponding to s_μ : $\{u\} \times X_\mu \simeq T_\mu\mathcal{F}$.

Further, in Chapter 9 we need to show functions defined on \mathcal{F} are C^1 . Showing a mapping between manifolds is C^1 requires we show the local representation of that map is C^1 (see [AMR, Definition 3.2.5]). In our case for a map $f : \mathcal{F} \rightarrow \mathbb{R}$, we need to show $f \circ s_\mu^{-1}$ is C^1 as a mapping from X_μ into \mathbb{R} . Our strategy below will be to calculate the directional derivative at some point $u_0 \in X_\mu$ in the direction of u : $\left. \frac{d}{dt} \right|_{t=0} (f \circ s_\mu^{-1})(u_0 + tu)$ and then show this directional derivative is in fact the standard (Frechet) derivative, i.e. the bounded linear functional that best approximates the mapping ([AMR, Definition 2.3.3]).

Chapter 9

Lagrange Multipliers, Equilibrium Case

In the following chapter we apply the manifold structure built in Chapter 8 to solve the Lagrange multiplier problem outlined in Chapter 7. To do so, we apply Theorem 9.1.4 from [AMR]. In Sections 9.2, 9.3 and 9.4 we prove the necessary properties to apply Theorem 9.1.4. This allows us to conclude the Gibbs distribution is a minimum of information entropy in Theorem 9.4.5 and Corollary 9.4.6.

9.1 Preliminaries

The definitions and theorems in this section can be found in [AMR], see also Section 13.3. Let $\mathcal{L}(E, F)$ denote the set of bounded linear functions from E to F . Consider the Boltzmann-Gibbs entropy functional

$$H : \mathcal{F} \rightarrow \mathbb{R} \cup \{+\infty\} \tag{9.1}$$

defined by

$$H(\mu) = H(p\mu_0) = \int_{\Omega} p \log p \mu_0. \tag{9.2}$$

Let $\mathbf{G} : \mathcal{F} \rightarrow \mathbb{R}^n$, where $\mathbf{G}(\mu) = (G_0(\mu), \dots, G_n(\mu))$, with

$$G_i(\mu) = \int_{\Omega} C_i \mu. \tag{9.3}$$

C_i is a measurable function mapping from Ω into $\overline{\mathbb{R}}$. Our goal is to find $\mu \in \mathcal{F}$ that minimizes $E(\mu)$ subject to the constraints $G_i(\mu) = K_i$, where K_i is a fixed constant. This requires the following definitions (see [AMR] 2.1.14 and 3.5.3).

Definition 9.1.1. *The closed subspace F of the Banach space E is said to be split (or complemented) if there is a closed subspace $H \subset E$ such that $E = F \oplus H$, where $F \oplus H$ denotes the direct sum of F and H .*

Definition 9.1.2. *Suppose M and N are manifolds with $f : M \rightarrow N$ of class C^r , $r \geq 1$. A point $n \in N$ is called a regular value of f if $T_m f$ is surjective with split kernel for each $m \in f^{-1}(\{n\})$.*

Definition 9.1.3. *If, for each m in a set S , $T_m f$ is surjective with split kernel, we say f is a submersion on S .*

Our goal is to apply the following theorem (see [AMR] 3.5.24 for a similar proposition). The proof is given in the appendix (Section 13.5).

Theorem 9.1.4. *Let $g : M \rightarrow P$ be a C^1 submersion, $N = g^{-1}(p_0)$ and let $f : M \rightarrow \mathbb{R}$ be C^1 . A point $n \in N$ is a critical point of $f|_N$ if and only if there exists $\lambda \in (T_{p_0} P)^*$, called a Lagrange Multiplier, such that $T_n f = \lambda \circ T_n g$.*

Therefore, the above Lagrange multiplier problem requires the following:

1. H must be a differentiable mapping.
2. \mathbf{G} must be C^1
3. \mathbf{G} must be a submersion.

We prove the above conditions are satisfied on partitions of \mathcal{F} . Thus, for all p and q below, we have $p\mu_0 \sim q\mu_0$. Throughout this section, let $\mu := p\mu_0$.

9.2 Differentiability of H

We first look at $H(\mu)$. Recall

$$H(p\mu_0) = \int_{\Omega} \log pp\mu_0. \quad (9.4)$$

Our strategy is to show $H(\mu)$ is C^1 Gateaux (see Definition 13.3.4) and then apply the following (see [AMR, Proposition 2.4.10]):

Proposition 9.2.1. *If $f : U \subset E \rightarrow F$ is C^1 -Gateaux, then it is C^1 and the two derivatives coincide.*

Before showing H is C^1 , note that it is not necessarily true that H is finite for all measures in \mathcal{F} . Recall that \mathcal{F} is partitioned into connected components by the equivalence relation (see Definition 7.2.1). It can be shown that, under certain conditions, if H is finite for one element of an equivalence class, then it is finite for all elements of the equivalence class.(c.f. [P, p. 4055]).

Lemma 9.2.2. *Let $\mu, \mu' \in \mathcal{F}$ be equivalent (see Definition 7.2.1). Let $\mu := p\mu_0$ and $\mu' := q\mu_0$. Then $\log p \in L^{e^{|\iota|-1}}(\mu)$ if and only if $\log q \in L^{e^{|\iota|-1}}(\mu)$.*

Proof. Starting from $\mu' = \phi\mu$, with $\phi = \frac{q}{p}$ (see Remark 7.2.3). We know $\log \phi \in X_{\mu}$ by Theorem 8.2.3. In particular,

$$\left\| \log \frac{q}{p} \right\|_{L^{e^{|\iota|-1}}(\mu)} < +\infty. \quad (9.5)$$

Thus, if $\|\log p\|_{L^{e^{|\iota|-1}}(\mu)} < +\infty$

$$\left\| \log \frac{q}{p} \right\|_{L^{e^{|\iota|-1}}(\mu)} + \|\log p\|_{L^{e^{|\iota|-1}}(\mu)} < +\infty. \quad (9.6)$$

This implies

$$\left\| \log \frac{q}{p} + \log p \right\|_{L^{e|t|-1}(\mu)} < +\infty \quad (9.7)$$

by the triangle inequality. Therefore,

$$\|\log q\|_{L^{e|t|-1}(\mu)} < +\infty. \quad (9.8)$$

For the other direction, $\mu = \frac{1}{\phi}\mu'$ with $\frac{1}{\phi} = \frac{p}{q}$ (see Remark 7.2.3). We know $\log \frac{1}{\phi} \in X_{\mu'}$ and

$$\left\| \log \frac{p}{q} \right\|_{L^{e|t|-1}(\mu')} < +\infty. \quad (9.9)$$

Additionally, $\mu \sim \mu'$ implies $L^{e|t|-1}(\mu) = L^{e|t|-1}(\mu')$ with equivalent norms. Therefore

$$\left\| \log \frac{p}{q} \right\|_{L^{e|t|-1}(\mu)} < +\infty. \quad (9.10)$$

Thus, if $\|\log q\|_{L^{e|t|-1}(\mu)} < +\infty$

$$\left\| \log \frac{p}{q} \right\|_{L^{e|t|-1}(\mu)} + \|\log q\|_{L^{e|t|-1}(\mu)} < +\infty. \quad (9.11)$$

This implies

$$\left\| \log \frac{p}{q} + \log q \right\|_{L^{e|t|-1}(\mu)} < +\infty. \quad (9.12)$$

Therefore,

$$\|\log p\|_{L^{e|t|-1}(\mu)} < +\infty. \quad (9.13)$$

□

Lemma 9.2.3. *Let $\mu := p\mu_0 \in \mathcal{F}$ such that $\log p \in L^{e|t|-1}(\mu)$. Then $H(p\mu_0)$ is finite and $H(q\mu_0)$ is finite for all $q\mu_0 \sim p\mu_0$.*

Proof. If $\log p \in L^{e|t|-1}(\mu)$, then $\log p \in L^1(\mu)$ by the continuous embedding of $L^{e|t|-1}(\mu)$ into all L^p spaces (see Proposition 7.1.5). Thus $H(p\mu_0) = \int p \log p \mu_0 < \infty$.

Similarly, $\log p \in L^{e^{|\iota|-1}}(\mu)$ and Lemma 9.2.2 imply $\log q \in L^{e^{|\iota|-1}}(\mu)$. Additionally, $q\mu_0 \sim p\mu_0$ implies $L^{e^{|\iota|-1}}(p\mu_0) = L^{e^{|\iota|-1}}(q\mu_0)$ (see Remark 8.1.10) so that $\log q \in L^{e^{|\iota|-1}}(q\mu_0)$. The above embedding argument then allows use to conclude $H(q\mu_0) < \infty$. \square

We will solve the Lagrange multiplier problem only on those connected components where H is finite.

Notation: Below we use non-standard notation to denote the derivatives of the local representation of H (and later \mathbf{G}). When writing the derivative of $H \circ s_\mu^{-1}$ at u_0 , it would be standard to express this quantity as $d_{u_0}H$. However, we express this quantity as $d_\mu H$, where μ is the element of \mathcal{F} mapped to $u_0 \in X_\mu$ by an arbitrary chart $s_{\mu'}: s_{\mu'}(\mu) = u_0$.

Remark 9.2.4. *When calculating the derivative of the local representations at some $u_0 \in X_\mu$ in Propositions 9.2.5, 9.2.6, 9.3.1 and 9.3.2 (as well as the corresponding propositions in Chapter 12), we show they are C^1 with respect to an arbitrary chart $s_{\mu'}$. Then $s_{\mu'}^{-1}(u_0) = \mu$. Since the manifold is C^∞ it would also have been sufficient (and easier) to show the local representations are C^1 with respect to the "natural" chart, i.e. the chart that maps to u_0 to $0: s_\mu^{-1}(0) = \mu$.*

Proposition 9.2.5. *Let $\mu = p\mu_0 \in \mathcal{F}$. Then H is Gateaux differentiable on the components of \mathcal{F} satisfying $\log p \in L^{e^{|\iota|-1}}(p\mu_0)$. Furthermore, $d_\mu H$ is a bounded linear functional.*

Proof. First, we find the Gateaux derivative of H . We must show H is a bounded linear functional from X_μ to \mathbb{R} with respect to any chart.

Let $s_{\mu'}, \mu' \sim \mu$, be a chart such that $s_{\mu'}^{-1}(u_0) = \mu$ for some $u_0 \in X_{\mu'}$. Then

$$\begin{aligned}
d_{\mu}H(u) &= \left. \frac{d}{dt} \right|_{t=0} H \circ s_{\mu'}^{-1}(u_0 + tu) \\
&= \left. \frac{d}{dt} \right|_{t=0} \int e^{u_0+tu} q \log e^{u_0+tu} q \mu_0 \\
&= \left. \frac{d}{dt} \right|_{t=0} \int e^{u_0+tu} q [u_0 + tu + \log q] \mu_0 \\
&= \int u e^{u_0} q [u_0 + \log q] + e^{u_0} q [u] \mu_0.
\end{aligned} \tag{9.14}$$

Using $p\mu_0 = e^{u_0} q \mu_0$ and $\mu = p\mu_0$, the above becomes

$$\begin{aligned}
d_{\mu}H(u) &= \int (u [u_0 + \log q] p + up) \mu_0 \\
&= \int (u_0 + \log q + 1) u \mu
\end{aligned} \tag{9.15}$$

Equivalently,

$$d_{\mu}H(u) = \int u_0 u \mu + \int \log q u \mu + \int u \mu. \tag{9.16}$$

We now show $d_{\mu}H$ is a bounded linear functional. First,

$$|d_{\mu}H(u)| \leq \int |u_0 u| \mu + \int |\log q u| \mu + \int |u| \mu. \tag{9.17}$$

We look at each of these terms individually. For the first term, Holder's inequality $\left(\frac{1}{r} + \frac{1}{s} = 1\right)$ implies

$$\int |u_0 u| \mu \leq \|u_0\|_{L^r(\mu)} \|u\|_{L^s(\mu)}. \tag{9.18}$$

Since $u_0 \in X_{\mu'}$, Proposition 8.1.9 implies $u_0 \in X_{\mu}$. Proposition 7.1.5 then implies that

$\|u_0\|_{L^r(\mu)} = K < \infty$ and that there exists $K' > 0$ such that $\|u\|_{L^s(\mu)} < K' \|u\|_{L^{e^{|\ell|}-1}(\mu)}$.

Last, recall

$$\|u\|_{L^{e^{|\ell|}-1}(\mu)} \leq \max_{0 \leq i \leq n} \|u\|_{L^{e^{|\ell|}-1}(C_i|\mu)} = \|u\|_{L(\mu)}. \tag{9.19}$$

Therefore,

$$\int |u_0 u| q \mu_0 \leq K K' \|u\|_{L(\mu)} = K_1 \|u\|_{L(\mu)}. \tag{9.20}$$

The second term of (9.17) can be bounded using the same arguments:

$$\int |\log q u| \mu \leq \|\log q\|_{L^r(\mu)} \|u\|_{L^s(\mu)} \quad (9.21)$$

Keep in mind we have restricted ourselves to components of \mathcal{F} where $\log p \in L^{e|t|-1}(p\mu_0)$.

Then Proposition 7.1.5 implies $\log p \in L^r(\mu)$ and there exists $K' > 0$ such that

$\|u\|_{L^s(\mu)} \leq K' \|u\|_{L^{e|t|-1}(\mu)}$. Therefore

$$\begin{aligned} \int |\log q u| \mu &\leq \|\log q\|_{L^r(\mu)} K' \|u\|_{L^{e|t|-1}(\mu)} \\ &= K_2 \|u\|_{L^{e|t|-1}(\mu)} \\ &\leq K_2 \|u\|_{L(\mu)}. \end{aligned} \quad (9.22)$$

For the third term:

$$\begin{aligned} \int |u| \mu &\leq K_3 \|u\|_{L^{e|t|-1}(\mu)} \\ &= K_3 \|u\|_{L(\mu)} \end{aligned} \quad (9.23)$$

where we have again used Proposition 7.1.5.

Thus, $d_\mu H$ is a bounded linear functional at each μ in the direction of u . We conclude that H is Gateaux differentiable on those partitions where $\log p \in L^{e|t|-1}(p\mu_0)$.

□

Next, we prove that the Gateaux derivatives are continuous.

Proposition 9.2.6. *Let $\mu = p\mu_0 \in \mathcal{F}$. The mapping $\mu \rightarrow d_\mu H$ is continuous on the components of \mathcal{F} where $\log p \in L^{e|t|-1}(p\mu_0)$. In particular, H is C^1 with Frechet derivative*

$$d_\mu H[u] = \int (u_0 + \log q + 1) u \mu \quad (9.24)$$

where $\mu' = q\mu_0$ is equivalent to μ and $u_0 \in X_{\mu'}$ such that $s_{\mu'}^{-1}(u_0) = \mu$.

Proof. We want to show

$$\left| \sup \left\{ d_{\mu_n} H[x] - d_{\mu} H[x] : \|x\|_{L(\mu)} = 1 \right\} \right| \rightarrow 0 \quad (9.25)$$

as $\mu_n \rightarrow \mu$. Let $\mu' \in K_{\mu}$ and $s_{\mu'}$ be an arbitrary chart as in the proof of Proposition 9.2.5. Since $\mu_n \rightarrow \mu$, we know by Proposition 8.2.6 that $\mu_n := p_n \mu_0 \in K_{\mu}$ (i.e. $p_n \mu_0 \sim p \mu_0$) for n large enough. Then there exists $u_n \in X_{\mu'}$ such that $s_{\mu'}^{-1}(u_n) = \mu_n$ and Proposition 8.2.6 implies that $u_n \rightarrow u_0$ in $X_{\mu'}$. Then we must show

$$\left| \sup_{\|x\|_{L(\mu)}=1} \left\{ \int (u_n + \log q + 1) x \mu_n - \int (u_0 + \log q + 1) x \mu \right\} \right| \rightarrow 0 \quad (9.26)$$

as $n \rightarrow \infty$. Equivalently, we look at the difference

$$\underbrace{\int u_n x \mu_n - \int u_0 x \mu}_{(1)} + \underbrace{\int \log q x \mu_n - \int \log q x \mu}_{(2)} + \underbrace{\int x \mu_n - \int x \mu}_{(3)} \quad (9.27)$$

and show that each of the three groupings go to 0 as n goes to infinity.

Consider grouping (1):

$$\int u_n x \mu_n - \int u_0 x \mu. \quad (9.28)$$

Adding and subtracting $\int u_n x \mu$ gives

$$\begin{aligned} \left| \int u_n x \mu_n - \int u_0 x \mu \right| &= \left| \int u_n x (\mu_n - \mu) + \int (u_n - u_0) x \mu \right| \\ &\leq \int |u_n| \cdot |x| \left| \frac{p_n}{p} - 1 \right| p \mu_0 + \int |u_n - u_0| \cdot |x| \mu. \end{aligned} \quad (9.29)$$

Applying Holder's inequality (twice for the first integral a, b and c, d are conjugates)

$$\begin{aligned} &\left| \int u_n x \mu_n - \int u_0 x \mu \right| \\ &\leq \underbrace{\left(\int |u_n|^{ac} p \mu_0 \right)^{1/ac}}_{1.)} \underbrace{\left(\int |x|^{ad} p \mu_0 \right)^{1/ad}}_{2.)} \underbrace{\left(\int \left| \frac{p_n}{p} - 1 \right|^b p \mu_0 \right)^{1/b}}_{3.)} \\ &\quad + \underbrace{\left(\int |x|^a \mu \right)^{1/a}}_{4.)} \underbrace{\left(\int |u_n - u_0|^b \mu \right)^{1/b}}_{5.}). \end{aligned} \quad (9.30)$$

Now

1. $u_n \in L^{e^{|\iota|-1}}(q\mu_0)$, $q\mu_0 \sim p\mu_0$ implies $u_n \in L^{e^{|\iota|-1}}(p\mu_0)$ (Proposition 8.1.9).

Hence $u_n \in L^{ac}(q\mu_0)$ (by Proposition 7.1.5) and 1.) is finite.

2. $x \in L^{e^{|\iota|-1}}(p\mu_0)$, hence $x \in L^{ad}(p\mu_0)$ and 2.) is finite.

3. p_n e -converges to p . Thus, 3.) goes to 0.

4. $x \in L^{e^{|\iota|-1}}(p\mu_0)$. Hence $x \in L^a(\mu)$ (by Proposition 7.1.5) and 4.) is finite.

5. p_n e -converges to p implies $u_n \rightarrow u$ in $L^{e^{|\iota|-1}}(p\mu_0)$ (see Proposition 8.2.6). Thus,

$u_n \rightarrow u$ in $L^b(q\mu_0)$ (by Proposition 7.1.5). Thus, 5.) goes to 0.

Thus, the first grouping of terms in the difference $d_{\mu_n}H[x] - d_{\mu}H[x]$ goes to 0.

Next, we consider grouping (2):

$$\int \log q \, x \mu_n - \int \log q \, x \mu. \quad (9.31)$$

We have

$$\begin{aligned} \left| \int \log q \, x \mu_n - \int \log q \, x \mu \right| &= \left| \int \log q \, x (p_n \mu_0 - p \mu_0) \right| \\ &\leq \int |\log q| \, |x| \left| \frac{p_n}{p} - 1 \right| p \mu_0. \end{aligned} \quad (9.32)$$

Applying Holder's inequality twice (a, b and c, d are conjugate) the above is bounded

by

$$\left(\int |\log q|^{ac} p \mu_0 \right)^{1/ac} \left(\int |x|^{ad} p \mu_0 \right)^{1/ad} \left(\int \left| \frac{p_n}{p} - 1 \right|^b p \mu_0 \right)^{1/b} \quad (9.33)$$

Recall that we are restricted to components of \mathcal{F} where $\log p \in L^{e^{|\iota|-1}}(p\mu_0)$, then by

Proposition 7.1.5 $\log p \in L^{ac}(p\mu_0)$. Similarly, $x \in L^{e^{|\iota|-1}}(p\mu_0)$ implies $\int |x|^{ad} p \mu_0 <$

∞ . Last $\int \left| \frac{p_n}{p} - 1 \right|^b p \mu_0 \rightarrow 0$ by the e -convergence of μ_n to μ .

Last, for grouping (3):

$$\int x\mu_n - \int x\mu \quad (9.34)$$

we have

$$\left| \int x\mu_n - \int x\mu \right| = \left| \int x \left(\frac{p_n}{p} - 1 \right) p\mu_0 \right| \leq \int |x| \left| \frac{p_n}{p} - 1 \right| p\mu_0. \quad (9.35)$$

Apply Holder's inequality to get

$$\left| \int x\mu_n - \int x\mu \right| \leq \left(\int |x|^a p\mu_0 \right)^{1/a} \left(\int \left| \frac{p_n}{p} - 1 \right|^b p\mu_0 \right)^{1/b}. \quad (9.36)$$

Since $x \in L^{e^{|t|}-1}(p\mu_0)$, $\int |x|^a p\mu_0 < \infty$ by Proposition 7.1.5. Additionally,

$\int \left| \frac{p_n}{p} - 1 \right|^b p\mu_0 \rightarrow 0$ by the e -convergence of μ_n to μ .

We conclude

$$\left| \sup \left\{ d_{\mu_n} H[x] - d_{\mu} H[x] : \|x\|_{L(\mu)} = 1 \right\} \right| \rightarrow 0 \quad (9.37)$$

and that $\mu \rightarrow d_{\mu} H$ is a continuous mapping. \square

We have shown H is C^1 with respect to any arbitrary chart. When solving the Lagrange multiplier problem it will be helpful to use the chart $s_{\mu}(u) = e^u \mu$ that maps μ to 0 in the model space. Since we have shown H is C^1 we can calculate its derivative with respect to any chart. This leads to the following corollary.

Corollary 9.2.7. *Let $\mu = p\mu_0 \in \mathcal{F}$. The tangent $d_{\mu} H$ of H , on those components on \mathcal{F} where $\log p \in L^{e^{|t|}-1}(\mu)$, is given by*

$$d_{\mu} H[u] = \int (\log p + 1) u \mu \quad (9.38)$$

Proof. Since we have shown H is C^1 we can calculate its derivative with respect to any chart. Consider the chart $s_\mu(u) = e^u \mu$ so that μ corresponds to $u_0 = 0$. Then

$$\begin{aligned}
d_\mu H(u) &= \left. \frac{d}{dt} \right|_{t=0} H \circ s_\mu(0 + tu) \\
&= \left. \frac{d}{dt} \right|_{t=0} \int e^{tu} p \log e^{tu} p \mu_0 \\
&= \left. \frac{d}{dt} \right|_{t=0} \int e^{tu} p [tu + \log p] \mu_0 \\
&= \int u p [\log p] + p [u] \mu_0
\end{aligned} \tag{9.39}$$

Recalling $\mu = p\mu_0$, the above becomes

$$d_\mu H(u) = \int [\log p + 1] u \mu \tag{9.40}$$

□

9.3 \mathbf{G} is C^1 Mapping

Next, we show that \mathbf{G} is C^1 . We start by proving each component $G_i : \mathcal{F} \rightarrow \mathbb{R}$ is C^1 .

Recall $\mu := p\mu_0$. Again we argue G_i is C^1 Gateaux and apply Proposition 9.2.1.

Proposition 9.3.1. *Let $\mu = p\mu_0 \in \mathcal{F}$. Then G_i is Gateaux differentiable and the Gateaux derivative is a bounded linear functional.*

Proof. As in the proof of Proposition 9.2.5, let $s_{\mu'}, \mu' \sim \mu$, be a chart such that $s_{\mu'}^{-1}(u_0) = \mu$ for some $u_0 \in X_{\mu'}$, with $\mu' = q\mu_0$.

Then

$$\begin{aligned}
d_\mu G_i(u) &= \left. \frac{d}{d\epsilon} \right|_{\epsilon=0} \int C_i e^{u_0 + \epsilon u} q \mu_0 \\
&= \int C_i u e^{u_0} q \mu_0 \\
&= \int C_i u \mu
\end{aligned} \tag{9.41}$$

It remains to show $d_\mu G_i[u]$ is a bounded linear functional. Note

$$|d_\mu G_i[u]| \leq \int |u| |C_i| \mu. \quad (9.42)$$

In particular $u \in L^{e^{|\iota|-1}}(|C_i| \mu)$. Then by Proposition 7.1.5, $u \in L^1(|C_i| \mu)$ and there exists $K > 0$ such that $\|u\|_{L^1(|C_i| \mu)} \leq K \|u\|_{L^{e^{|\iota|-1}}(|C_i| \mu)}$. Therefore,

$$|d_\mu G_i[u]| \leq K \|u\|_{L^{e^{|\iota|-1}}(|C_i| \mu)} \leq K \|u\|_{L(\mu)} \quad (9.43)$$

We conclude the Gateaux derivative is a bounded linear functional. \square

As in Section 9.2, we next prove the mapping $\mu \rightarrow d_\mu G_i$ is continuous.

Proposition 9.3.2. *Let $\mu = p\mu_0 \in \mathcal{F}$. Then the mapping $\mu \rightarrow d_\mu G_i$ is continuous.*

In particular, G_i is C^1 with Frechet derivative

$$d_\mu G_i[u] = \int C_i u \mu \quad (9.44)$$

Proof. We want to show

$$\left| \sup \left\{ d_{\mu_n} G_i[x] - d_\mu G_i[x] : \|x\|_{L(\mu)} = 1 \right\} \right| \rightarrow 0 \quad (9.45)$$

as $\mu_n \rightarrow \mu$. Let $\mu' \in K_\mu$ and $s_{\mu'}$ be an arbitrary chart. Since $\mu_n \rightarrow \mu$, we know by Proposition 8.2.6 that $\mu_n := p_n \mu_0 \in K_\mu$ (i.e. $p_n \mu_0 \sim p \mu_0$) for n large enough and there exists $u_n \in X_{\mu'}$ such that $s_{\mu'}^{-1}(u_n) = \mu_n$. Then we must show

$$\left| \sup \left\{ \int C_i x \mu_n - \int C_i x \mu : \|x\|_{L(\mu)} = 1 \right\} \right| \rightarrow 0 \quad (9.46)$$

as $\mu_n \rightarrow \mu$. We look at

$$\begin{aligned} \left| \int C_i x \mu_n - \int C_i x \mu \right| &= \left| \int x C_i (\mu_n - \mu) \right| \\ &\leq \int |x| \left| \frac{p_n}{p} - 1 \right| |C_i| p \mu_0. \end{aligned} \quad (9.47)$$

Apply Holder's inequality, giving

$$\begin{aligned} & \left| \int C_i x \mu_n - \int C_i x \mu \right| \\ & \leq \left(\int |x|^a |C_i| p \mu_0 \right)^{1/a} \left(\int \left| \frac{p_n}{p} - 1 \right|^b |C_i| p \mu_0 \right)^{1/b}. \end{aligned} \quad (9.48)$$

Note that $x \in X_\mu$ implies $x \in L^{e^{t|} - 1}(|C_i| p \mu_0)$. Hence $x \in L^a(|C_i| p \mu_0)$ by Proposition 7.1.5. By the e -convergence of $\mu_n \rightarrow \mu$, $\int \left| \frac{p_n}{p} - 1 \right|^b |C_i| p \mu_0 \rightarrow 0$. We conclude

$$\left| \sup \left\{ d_{\mu_n} G_i[x] - d_\mu G_i[x] : \|x\|_{L(\mu)} = 1 \right\} \right| \rightarrow 0 \quad (9.49)$$

and that $\mu \rightarrow d_\mu G_i$ is a continuous mapping. \square

Last, note that $\mathbf{G} : \mathcal{F} \rightarrow \mathbb{R}^{k+1}$ given by $\mathbf{G}(\mu) = (G_0(\mu), \dots, G_k(\mu))$ is C^1 by the following Proposition (see [AMR], Proposition 2.3.5):

Proposition 9.3.3. *Let $f_i : U \subset E \rightarrow F$, $1 \leq i \leq n$ be r times differentiable mappings. Then $f = f_1 \times \dots \times f_n : U \subset E \rightarrow F_1 \times \dots \times F_n$ defined by $f(\mathbf{u}) = (f_1(\mathbf{u}), \dots, f_n(\mathbf{u}))$ is r times differentiable and $D^r f = D^r f_1 \times \dots \times D^r f_n$.*

By the above, we know that $d_\mu \mathbf{G} = (d_\mu G_0, \dots, d_\mu G_k)$ is C^1 :

$$\begin{aligned} & \sup_{\|x\|=1} \|d_{\mu_n} \mathbf{G}[x] - d_\mu \mathbf{G}[x]\|_{\mathbb{R}^{n+1}} \\ & = \sup_{\|x\|=1} \sqrt{|d_{\mu_n} G_0[x] - d_\mu G_0[x]|^2 + \dots + |d_{\mu_n} G_k[x] - d_\mu G_k[x]|^2} \\ & \leq \sqrt{\sup_{\|x\|=1} |d_{\mu_n} G_0[x] - d_\mu G_0[x]|^2 + \dots + |d_{\mu_n} G_k[x] - d_\mu G_k[x]|^2} \\ & \rightarrow 0 \end{aligned} \quad (9.50)$$

as $n \rightarrow \infty$, since for each $1 \leq i \leq k+1$, $|d_{\mu_n} G_i[x] - d_\mu G_i[x]|$ goes to 0, as shown above. Thus, $u \rightarrow d_\mu \mathbf{G}$ is continuous and we conclude the vector valued constraint functional \mathbf{G} is a C^1 mapping.

As in the section above, we can now identify the tangent in its simplest form by using the chart that maps μ to 0:

Corollary 9.3.4. *Let $\mu = p\mu_0 \in \mathcal{F}$. The derivative $d_\mu \mathbf{G} : X_\mu \rightarrow \mathbb{R}^{n+1}$ of \mathbf{G} is given by*

$$d_\mu \mathbf{G}[u] = \int \mathbf{C} u \mu \quad (9.51)$$

Proof. Since we have shown \mathbf{G} is C^1 we can calculate its derivative with respect to any chart. Consider the chart $s_\mu(u) = e^u \mu$ so that μ corresponds to 0. Then

$$\begin{aligned} d_\mu \mathbf{G}([\gamma]) &= \left. \frac{d}{dt} \right|_{t=0} \mathbf{G} \circ s_\mu^{-1}(0 + tu) \\ &= \left. \frac{d}{dt} \right|_{t=0} \int \mathbf{C} e^{tu} \mu \\ &= \int \mathbf{C} u \mu. \end{aligned} \quad (9.52)$$

□

9.4 \mathbf{G} is a Submersion

To be able to apply Theorem 9.1.4, it remains to show \mathbf{G} is a submersion. Let $\bar{\mathbf{C}} \in \mathbb{R}^{n+1}$ be the constraints, i.e. $\mathbf{G}(\mu) = \bar{\mathbf{C}}$. For $\mu \in \mathbf{G}^{-1}\{\bar{\mathbf{C}}\}$, $\mu = p\mu_0$, we must show

1. $d_\mu \mathbf{G} : T_\mu \mathcal{F} \rightarrow \mathbb{R}$ is surjective.
2. $T_\mu \mathcal{F} = \ker d_\mu \mathbf{G} \oplus F$, where F and $\ker d_\mu \mathbf{G}$ are closed subspaces.

where p is such that $\log p \in L^{e^{|\mu|}-1}(\mu)$.

First we argue that $d_\mu \mathbf{G} = (d_\mu G_1, \dots, d_\mu G_n)$ is surjective.

Lemma 9.4.1. *Let $\mu \in \mathbf{G}^{-1}\{\bar{\mathbf{C}}\}$ and let $C_i : \Omega \rightarrow \mathbb{R}$ be not identically equal to zero, $C_i \not\equiv 0$ for all $0 \leq i \leq n$. Then $d_\mu \mathbf{G} : T_\mu \mathcal{F} \rightarrow \mathbb{R}^{n+1}$ is surjective.*

Proof. By way of contradiction, suppose $d_\mu \mathbf{G}$ is not surjective. This implies there exists $Y \subset \mathbb{R}^{n+1}$ such that $\mathbb{R}^{n+1} = \text{Im}(d_\mu \mathbf{G}) \oplus Y$. Then for some $0 < k \leq n + 1$,

$\dim Y = k$. Then for k of the standard basis vectors $\mathbf{e}_i \in \mathbb{R}^{n+1}$, there does not exist x_i mapping to these e_i . If there were, then $\text{Im}(d_\mu \mathbf{G})$ would be of higher dimension for each additional \mathbf{e}_i included. In particular,

$$d_\mu \mathbf{G}[u] \neq \mathbf{e}_i \quad \forall u \in X_\mu \quad (9.53)$$

implying

$$\int_{\Omega} C_i u \mu \neq 1 \quad \forall u \in X_\mu. \quad (9.54)$$

Therefore

$$\int_{\Omega} C_i u \mu = 0 \quad \forall u \in X_\mu, \quad (9.55)$$

since one u mapping to a non-zero number $c_0 \in \mathbb{R}$ implies $\frac{c}{c_0}u$ maps to any $c \in \mathbb{R}$.

We claim this implies $C_i \equiv 0$. Suppose $C_i \not\equiv 0$. Define the sets

$$\begin{aligned} M &= \{\omega \in \Omega : C_i(\omega) > 0\} \\ N &= \{\omega \in \Omega : C_i(\omega) < 0\}. \end{aligned} \quad (9.56)$$

Suppose M has non-zero μ measure. Define $u_1 = \chi_M \in X_\mu$. Then

$$\int_{\Omega} C_i u_1 \mu = \int_M C_i \mu > 0. \quad (9.57)$$

This contradicts $\int_{\Omega} C_i u \mu = 0$ for all $u \in X_\mu$. We conclude $\mu(M) = 0$.

A similar argument holds for N , allowing us to conclude $\mu(N) = 0$ and implying $C_i \equiv 0$, contradicting our hypothesis. We conclude $d_\mu \mathbf{G}$ is surjective. \square

Remark 9.4.2. *Lemma 9.4.1 implies that any G_i will be regular (have non identically zero derivative) so long as $C_i \not\equiv 0$.*

It remains only to show the derivative has split kernel, to do so requires the following Proposition (see [AMR, Corollary 2.2.18]):

Proposition 9.4.3. *Let E be a Banach space and F a closed subspace of E . Then F is split if and only if there exists $P \in \mathcal{L}(E, F)$ such that $P \circ P = P$ and $F = \{e \in E : Pe = e\}$.*

Proposition 9.4.4. *$T_\mu \mathcal{F} = \ker d_\mu \mathbf{G} \oplus F$, where F and $\ker d_\mu \mathbf{G}$ are closed subspaces.*

Proof. A continuous linear mapping has a closed kernel (see [F, Chapter 5, Exercise 15]). Therefore the subspace $\ker d_\mu \mathbf{G}$ is closed since $d_\mu \mathbf{G}$ is continuous.

Next, note that by the surjectivity of $d_\mu \mathbf{G}$, there exists $y_1, \dots, y_n \in T_\mu \mathcal{F}$ such that $d_\mu \mathbf{G}[y_i] = \mathbf{e}_i$ for all i , where \mathbf{e}_i is the standard basis vector in \mathbb{R}^{n+1} . Define $F = \langle y_1, \dots, y_n \rangle$. Note that F is closed (a finite dimensional subspace of a normed space is closed). Fix $x \in T_\mu \mathcal{F}$ and let $d_\mu \mathbf{G}[x] = \mathbf{k}$ (if $\mathbf{k} = \mathbf{0}$ then $x \in \ker d_\mu \mathbf{G}$ and we are done). Then

$$d_\mu \mathbf{G}[x - k_1 y_1 - \dots - k_n y_n] = \mathbf{k} - k_1 \mathbf{e}_1 - \dots - k_n \mathbf{e}_n = \mathbf{0}. \quad (9.58)$$

Hence $x - k_1 y_1 - \dots - k_n y_n \in \ker d_\mu \mathbf{G}$. Additionally, $k_1 y_1 + \dots + k_n y_n \in F$. We conclude

$$x = \underbrace{(x - k_1 y_1 - \dots - k_n y_n)}_{\in \ker d_\mu \mathbf{G}} + \underbrace{(k_1 y_1 + \dots + k_n y_n)}_{\in F}, \quad (9.59)$$

implying $T_\mu \mathcal{F} = \ker d_\mu \mathbf{G} + F$. To show $T_\mu \mathcal{F}$ is topologically equivalent to $\ker d_\mu \mathbf{G} \times F$, we apply Proposition 9.4.3.

Define $P : T_\mu \mathcal{F} \rightarrow F$ by $P(x) = d_\mu \mathbf{G}[x] \cdot \mathbf{y}$. We argue P is a projection on F . Let $\mathbf{k} \in \mathbb{R}^{n+1}$ and note that

$$d_\mu \mathbf{G}[\mathbf{k} \cdot \mathbf{y}] = d_\mu \mathbf{G}[k_1 y_1 + \dots + k_n y_n] = \mathbf{k}. \quad (9.60)$$

Then

$$\begin{aligned} P(x) &= d_\mu \mathbf{G}[x] \cdot \mathbf{y} \\ &= \mathbf{k} \cdot \mathbf{y} \end{aligned} \quad (9.61)$$

for some $\mathbf{k} \in \mathbb{R}^{n+1}$. Then

$$\begin{aligned} P^2(x) &= P(\mathbf{k} \cdot \mathbf{y}) \\ &= d_\mu \mathbf{G}[\mathbf{k} \cdot \mathbf{y}] \cdot \mathbf{y} \\ &= \mathbf{k} \cdot \mathbf{y} \end{aligned} \tag{9.62}$$

so that $P^2 = P$. Next, let $x \in F$ so that $x = \mathbf{k} \cdot \mathbf{y}$ for some $\mathbf{k} \in \mathbb{R}^{n+1}$. Then

$$P(x) = d_\mu \mathbf{G}[\mathbf{k} \cdot \mathbf{y}] \cdot \mathbf{y} = \mathbf{k} \cdot \mathbf{y} = x. \tag{9.63}$$

Thus, $P(x) = x$ for $x \in F$. We conclude P is a projection of F and by Proposition 9.4.3 $T_\mu \mathcal{F} = \ker d_\mu \mathbf{G} \oplus F$. □

Having satisfied the conditions of Theorem 9.1.4, we now apply it to H and \mathbf{G} (see (9.1) and (9.3)).

Theorem 9.4.5. *Let $C_i : \Omega \rightarrow \mathbb{R}$ be non-zero functions. Then the critical point of Gibbs entropy, on components of \mathcal{F} where $\log p \in L^{e^{|\mathbf{t}|-1}}(p\mu_0)$, is the Gibbs distribution:*

$$p = \frac{\exp\left(\sum_{i=1}^n \lambda_i C_i\right)}{\int_{\Omega} \exp\left(\sum_{i=1}^n \lambda_i C_i\right) \mu} \tag{9.64}$$

with $\lambda_i \in \mathbb{R}$.

Proof. Let $\mu = p\mu_0$ be a critical point. From Corollary 9.2.7 and Corollary 9.3.4 we know

$$d_\mu H(u) = \int_{\Omega} [\log p + 1] u \mu \tag{9.65}$$

and

$$d_\mu \mathbf{G}(u) = \int_{\Omega} \mathbf{C} u \mu. \tag{9.66}$$

By Lemma 9.4.1 and Proposition 9.4.4 we know \mathbf{G} is a submersion. We therefore can apply Theorem 9.1.4 to conclude there exists $\boldsymbol{\lambda} \in \mathbb{R}^{n+1}$ such that

$$\int_{\Omega} [\log p + 1] u \mu = \boldsymbol{\lambda} \cdot \int_{\Omega} \mathbf{C} u \mu \quad (9.67)$$

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

$$\int_{\Omega} \left(\log p + 1 - \sum_{i=0}^n \lambda_i C_i \right) u \mu = 0 \quad (9.68)$$

for all $u \in X_{\mu}$.

It remains to argue that the above implies $\log p + 1 - \sum_{i=1}^k \lambda_i C_i = 0$. We show that if $\log p - \sum_{i=1}^n \lambda_i C_i \neq 0$ then it is possible to construct $u \in X_{\mu}$ such that Equation (9.68) is non-zero.

We proceed in a similar manner to Lemma 9.4.1. Define

$$\begin{aligned} M &= \left\{ \omega \in \Omega : \log p + 1 - \sum_{i=0}^n \lambda_i C_i > 0 \right\} \\ N &= \left\{ \omega \in \Omega : \log p + 1 - \sum_{i=0}^n \lambda_i C_i < 0 \right\}. \end{aligned} \quad (9.69)$$

Suppose one has non-zero μ -measure. WLOG, suppose M is the set of non-zero μ -measure. Define $u_1 = \chi_M \in X_{\mu}$. Then

$$\int_{\Omega} \left(\log p + 1 - \sum_{i=0}^n \lambda_i C_i \right) u_1 \mu = \int_M \left(\log p + 1 - \sum_{i=0}^n \lambda_i C_i \right) \mu > 0. \quad (9.70)$$

However, this contradicts $\int_{\Omega} \left(\log p + 1 - \sum_{i=0}^n \lambda_i C_i \right) u \mu = 0$ for all $u \in X_{\mu}$. We conclude $\mu(M) = 0$. If N also has non-zero μ -measure a similar argument would also show $\mu(N) = 0$. Therefore

$$\log p + 1 - \sum_{i=0}^n \lambda_i C_i = 0. \quad (9.71)$$

Solving gives the Gibbs distribution

$$p = \frac{\exp\left(\sum_{i=1}^n \lambda_i C_i\right)}{\int_{\Omega} \exp\left(\sum_{i=1}^n \lambda_i C_i\right) \mu} \quad (9.72)$$

where we have used the condition $\int \mu = 1$ to get the normalizing constant. \square

We conclude this chapter by noting that the argument given in (3.17) can be generalized to show the generalized Gibbs distribution given in (9.64) minimizes information entropy. This is summarized in the following corollary:

Corollary 9.4.6. *The distribution given by (9.64):*

$$p = \frac{\exp\left(\sum_{i=1}^n \lambda_i C_i\right)}{\int_{\Omega} \exp\left(\sum_{i=1}^n \lambda_i C_i\right) \mu} \quad (9.73)$$

minimizes information entropy (7.5):

$$H(p\mu_0) = \int_{\Omega} p \log p \mu_0. \quad (9.74)$$

Chapter 10

Application: The Grand Canonical

In this chapter we apply the results of Chapter 9. After recalling the definitions and setting first given in Part I, deriving the Gibbs distribution follows almost immediately from the proceeding chapter. We conclude by verifying that reasonable physical assumptions imply that $\log p \in L^{e^{|\Lambda|}-1}(p\mu_0)$.

10.1 The Grand Canonical Gibbs distribution

We now apply Theorem 9.4.5 to derive the Grand Canonical. Let Λ be a bounded set in \mathbb{R}^3 . Let Ω be the phase space \mathcal{X}_Λ , where

$$\mathcal{X}_\Lambda = \bigsqcup_{N=0}^{\infty} (\Lambda \times \mathbb{R}^3)^N. \quad (10.1)$$

Take the reference measure μ_0 on \mathcal{X}_Λ to be

$$\mu_0 = \sum_{N=0}^{\infty} \frac{dq_1 \dots dq_N}{N!} dp_1 \dots dp_N. \quad (10.2)$$

Let $(\mathbf{q}, \mathbf{p}) = (q_1, \dots, q_N, p_1, \dots, p_N)$. Next we outline the constraints. Let U be the interaction potential

$$U : \mathcal{X}_\Lambda \rightarrow \mathbb{R}, (q_1, \dots, q_N) \rightarrow U(q_1, \dots, q_N). \quad (10.3)$$

It will be shown below that we require the assumption that U is stable: there exists $L > 0$ such that

$$U(q_1, \dots, q_N) \geq -NL \quad (10.4)$$

for all $N \in \mathbb{N}$.

The energy of a configuration is given by the Hamiltonian $E : \mathcal{X}_\Lambda \rightarrow \mathbb{R}$ defined by

$$E(q_1, \dots, q_N, p_1, \dots, p_N) = \frac{1}{2} \sum_{i=1}^N |p_i|^2 + U(q_1, \dots, q_N). \quad (10.5)$$

We also define the total momentum and particle number as

$$\vec{P} : \mathcal{X}_\Lambda \rightarrow \mathbb{R}^3, \quad \vec{P}(q_1, \dots, q_N, p_1, \dots, p_N) = \sum_{i=1}^N p_i \quad (10.6)$$

and

$$\mathcal{N} : \mathcal{X}_\Lambda \rightarrow \mathbb{N}, \quad \mathcal{N}(q_1, \dots, q_N, p_1, \dots, p_N) = N. \quad (10.7)$$

Recall from Chapter 9 that the derivative of

$$G_i(\mu) = \int_{\Omega} C_i \mu \quad (10.8)$$

is

$$d_\mu G_i[u] = \int_{\Omega} C_i u \mu_0. \quad (10.9)$$

For E, \vec{P} and \mathcal{N} the corresponding constraints are

$$\bar{E}(\mu) = \int_{\mathcal{X}_\Lambda} E \mu, \quad \vec{\bar{P}}(\mu) = \int_{\mathcal{X}_\Lambda} \vec{P} \mu, \quad \bar{\mathcal{N}}(\mu) = \int_{\mathcal{X}_\Lambda} \mathcal{N} \mu \quad (10.10)$$

with derivatives

$$d_\mu \bar{E}[u] = \int_{\mathcal{X}_\Lambda} E u \mu, \quad d_\mu \vec{\bar{P}}[u] = \int_{\mathcal{X}_\Lambda} \vec{P} u \mu, \quad d_\mu \bar{\mathcal{N}}[u] = \int_{\mathcal{X}_\Lambda} \mathcal{N} u \mu \quad (10.11)$$

respectively.

Additionally, Theorem 9.4.5 holds on regions of the underlying manifold where $\log p$ is in the corresponding Orlicz space. We verify that this requirement does not eliminate the Gibbs Grand Canonical distribution under reasonable physical assumptions.

Proposition 10.1.1. *Let U be a stable potential (see Equation (10.4)) and*

$$p(\mathbf{q}, \mathbf{p}) = \frac{e^{\beta E(\mathbf{q}, \mathbf{p}) + \vec{\lambda} \cdot \vec{P}(\mathbf{q}, \mathbf{p}) + \nu N(\mathbf{q}, \mathbf{p})}}{Z_{\beta, \vec{\lambda}, \nu}} \quad (10.12)$$

the Grand Canonical density function with $\beta < 0$, where

$$Z_{\beta, \vec{\lambda}, \nu} = \int_{\mathcal{X}_\Lambda} e^{\beta E + \vec{\lambda} \cdot \vec{P} + \nu N} \mu_0. \quad (10.13)$$

Then $\log p \in L^{e^{|\epsilon|}-1}(p\mu_0)$.

Proof. By Proposition 7.1.4 it is sufficient to show $\log p \in L^{\cosh(t)-1}(p\mu_0)$. We show there exists $\epsilon > 0$ such that

$$\int_{\mathcal{X}_\Lambda} (\cosh(\epsilon \log p) - 1) p \mu_0 < +\infty. \quad (10.14)$$

In particular, we show there exists $\epsilon > 0$ such that

$$\int_{\mathcal{X}_\Lambda} \left(\frac{1}{2} \left(e^{\epsilon \beta E + \epsilon \vec{\lambda} \cdot \vec{P} + \epsilon \nu N - \epsilon \log Z_{\beta, \vec{\lambda}, \nu}} + e^{-\epsilon \beta E - \epsilon \vec{\lambda} \cdot \vec{P} - \epsilon \nu N + \epsilon \log Z_{\beta, \vec{\lambda}, \nu}} \right) - 1 \right) \frac{e^{\beta E + \vec{\lambda} \cdot \vec{P} + \nu N}}{Z_{\beta, \vec{\lambda}, \nu}} \mu_0 < +\infty. \quad (10.15)$$

Since $\log Z_{\beta, \vec{\lambda}, \nu}$ is a constant it will not cause effect whether the above is finite. Similarly, we can use the fact that $p = \frac{e^{\beta E + \vec{\lambda} \cdot \vec{P} + \nu N}}{Z_{\beta, \vec{\lambda}, \nu}}$ is a probability measure to simplify the above to showing

$$\int_{\mathcal{X}_\Lambda} \left(e^{\epsilon \beta E + \epsilon \vec{\lambda} \cdot \vec{P} + \epsilon \nu N} + e^{-\epsilon \beta E - \epsilon \vec{\lambda} \cdot \vec{P} - \epsilon \nu N} \right) \frac{e^{\beta E + \vec{\lambda} \cdot \vec{P} + \nu N}}{Z_{\beta, \vec{\lambda}, \nu}} \mu_0 < +\infty \quad (10.16)$$

for some $\epsilon > 0$. Multiplying through and recognizing $\frac{1}{Z_{\beta, \vec{\lambda}, \nu}}$ will not effect the convergence, we must show

$$\int_{\mathcal{X}_\Lambda} \left(e^{(1+\epsilon)[\beta E + \vec{\lambda} \cdot \vec{P} + \nu N]} + e^{(1-\epsilon)[\beta E + \vec{\lambda} \cdot \vec{P} + \nu N]} \right) \mu_0 < +\infty. \quad (10.17)$$

Consider each integral separately, and recall $E = U + \frac{1}{2} \sum_{i=1}^N p_i^2$, implying the integral of the first term can be written as

$$\int_{\mathcal{X}_\Lambda} e^{(1+\epsilon)[\beta U + \beta/2 \sum_{i=1}^N p_i^2 + \epsilon \vec{\lambda} \cdot \sum_{i=1}^N p_i + \nu N]} \mu_0. \quad (10.18)$$

Completing the square with respect to p implies

$$\int_{\mathcal{X}_\Lambda} e^{(1+\epsilon)[\beta U + \sum_{i=1}^N \frac{\beta}{2} \left(p_i + \frac{\vec{\lambda}}{2\beta} \right)^2 - \frac{\vec{\lambda}^2}{2\beta} + \nu N]} \mu_0. \quad (10.19)$$

If we integrate with respect to $dp_1 \dots dp_N$ the above becomes

$$\sum_N \left(\frac{2\pi}{(1+\epsilon)|\beta|} \right)^{N/2} e^{-(1+\epsilon)\vec{\lambda}^2/2\beta} e^{(1+\epsilon)\nu N} \int_{\Lambda^N} e^{(1+\epsilon)\beta U} \frac{dq_1 \dots dq_N}{N!}, \quad (10.20)$$

where $\left(\frac{2\pi}{(1+\epsilon)|\beta|} \right)^{N/2}$ is the constant resulting from integrating the N dimensional Gaussian. Note that the above integral will be a decaying exponential (and hence finite) when $U \geq 0$. When $U < 0$ we can use the stability condition (keeping in mind $\beta < 0$) to bound the integral

$$\begin{aligned} \int_{\Lambda^N} e^{(1+\epsilon)\beta U} \frac{dq_1 \dots dq_N}{N!} &\leq \int_{\Lambda^N} e^{(1+\epsilon)\beta L N} \frac{dq_1 \dots dq_N}{N!} \\ &= \frac{e^{(1+\epsilon)\beta L N} |\Lambda|^N}{N!} < +\infty \end{aligned} \quad (10.21)$$

where $|\Lambda|$ denotes the Lebesgue measure of Λ . This shows Equation (10.20) is finite for any $\epsilon > 0$. A similar argument will show

$$\int_{\mathcal{X}_\Lambda} e^{(1-\epsilon)[\beta E + \vec{\lambda} \cdot \vec{P} + \nu N]} \mu_0 < +\infty. \quad (10.22)$$

so long as $\epsilon < 1$ (otherwise the integral with respect to \mathbf{p} will no longer be Gaussian).

We conclude $\log p \in L^{\cosh(t)-1}(p\omega)$. □

Remark 10.1.2. *The Lagrange multiplier β corresponds physically to inverse temperature, $1/T$. Note that in the standard derivation of the Gibbs distribution entropy is $-\int p \log p \mu_0$ and the resulting β is positive. Our omission of the negative sign explains this seeming physical discrepancy.*

With all the conditions of Theorem 9.4.5 satisfied, we conclude the Gibbs Grand Canonical minimizes Gibbs entropy.

Chapter 11

Constraint Functionals as an Open Mapping

We now use the manifold structure to provide an alternate proof to the open mapping proved in Part I.

11.1 Constraint Functionals and the Local Surjectivity Theorem

This alternate proof is based on the following theorem (see [AMR, Proposition 2.5.9], [Luen, Section 9.2, Theorem 1] or Section 13.4).

Theorem 11.1.1. (*Local Surjectivity Theorem*) *Let D be an open subset of X and let $T : D \rightarrow Y$ be continuously differentiable. Let $x_0 \in D$ be a regular point, i.e. $T'(x_0) : X \rightarrow Y$ is a surjective map. Then there is an open neighborhood $N'(y_0)$ of the point $y_0 = T(x_0)$ such that the equation $T(x) = y$ has a solution for every $y \in N'(y_0)$. Furthermore, there exists an open set in D containing x_0 such that T maps this open set onto $N'(y_0)$. In particular, if $T'(x)$ is onto for $x \in D$, then f is an open mapping.*

We now apply this theorem to the manifold constructed in Chapter 8 to prove the constraint functionals are open mappings.

Proposition 11.1.2. Define $\mu := p\mu_0$. The mapping $G : \mathcal{F} \rightarrow \mathbb{R}$, defined by

$$G_i(\mu) = \int C_i \mu, \quad (11.1)$$

where $C_i \not\equiv 0$, has an open image.

Proof. Recall the derivative of $G_i(\mu)$ is (see Corollary 9.3.4)

$$d_\mu G_i[u] = \int_\Omega C_i u \mu. \quad (11.2)$$

We have already shown the derivative at μ is surjective on \mathbb{R} so long as $C_i \not\equiv 0$ (see Lemma 9.4.1). By the Local Surjectivity Theorem, for every $y_0 \in \mathbb{R}$ there exists an open neighborhood $N'(y_0)$ such that there exists an open neighborhood of μ , \mathcal{O}_μ with $G_i(\mathcal{O}_\mu)$ mapped onto $N'(y_0)$: $G_i(\mathcal{O}_\mu) = N'(y_0)$. Note this equality implies $G_i(\mathcal{O}_\mu)$ is open. Additionally, $\mu \in \mathcal{F}$ was arbitrary.

Thus,

$$G_i(\mathcal{F}) = \bigcup_{\mu \in \mathcal{F}} G_i(\mathcal{O}_\mu) \quad (11.3)$$

is the union of open sets. We conclude $G_i(\mathcal{F})$ is an open set. Finally, since $\mathbf{G}(\mathcal{F}) = (G_1(\mathcal{F}), \dots, G_n(\mathcal{F}))$ and the product of a finite number of opens sets is open, we conclude $\mathbf{G}(\mathcal{F})$ is open. \square

We now apply Proposition 11.1.2 to the grand canonical ensemble (see Chapter 10). Recall our base measure μ_0 in this setting is

$$\mu_0 = \sum_{N \geq 0} \frac{dq_1 \dots dq_N}{N!} dp_1 \dots dp_N. \quad (11.4)$$

Note that by convention, $\mu_0(\emptyset) = 1$. Define \mathcal{F}_{μ_0} be as in (7.4) with respect to the particular μ_0 above. Observe that the Dirac measure on the null set

$$\delta_\emptyset = \begin{cases} 1 & \text{on } \emptyset \\ 0 & \text{o.w.} \end{cases} \quad (11.5)$$

is absolutely continuous with respect to μ_0 . However, δ_\emptyset fails to meet the conditions of the local surjectivity theorem. Let $u \in X_{\delta_\emptyset}$. Then

$$d_{\delta_\emptyset} G_i[u] = \int_{\Omega} C_i u \delta_\emptyset = 0 \quad (11.6)$$

for all u and for all i . Therefore, the derivative of G_i at δ_\emptyset is not onto. Additionally, note that δ_\emptyset is not a connected component of the manifold \mathcal{F}_{μ_0} . There is no $\mu \in \mathcal{F}_{\mu_0}$ such that $\delta_\emptyset \sim \mu$. If this was the case, then there would exist some $\phi \in L^p(\mu)$ for some $p > 1$ such that

$$\delta_\emptyset = \phi \mu. \quad (11.7)$$

Thus, \mathcal{F}_{μ_0} is the union of a singleton and all other measures absolutely continuous with respect to μ_0 .

Define $\overline{\mathcal{F}}_{\mu_0} = \mathcal{F}_{\mu_0} / \{\delta_\emptyset\}$. Applying the above proposition on the on the constraint functionals of the Grand Canonical, restricted to $\overline{\mathcal{F}}_{\mu_0}$ gives $\mathbf{G}(\overline{\mathcal{F}}_{\mu_0})$ is an open set. Last, we we restrict $\overline{\mathcal{F}}_{\mu_0}$ further to only Gibbs measures, the resulting image is still a union of open sets, giving an alternative proof of the result from Part I.

Chapter 12

Lagrange Multipliers, Local Equilibrium Case

We would now like to extend the result of Chapter 9 to local equilibrium. In particular, our constraint functionals will now map into some function space, rather than into \mathbb{R} . We will use the application to phase space in Chapter 10 as a starting point.

12.1 The Local Grand Canonical Gibbs Distribution

Define E , \vec{P} and \mathcal{N} the same as (10.5), (10.6), and (10.7). The reference measure μ_0 also stays the same as (10.2). Recall $C_0 \equiv 1$, let $(\mathbf{q}, \mathbf{p}) = (q_1, \dots, q_N, p_1, \dots, p_N)$, and defined

$$\begin{aligned} G_0 : \mathcal{F} &\rightarrow \mathbb{R}^+ \\ G_0(\mu) &= \int |C_0| \mu(d\mathbf{q}, d\mathbf{p}) \end{aligned} \tag{12.1}$$

The constraint functionals are now defined as

$$\begin{aligned} G_j : \mathcal{F} &\rightarrow L^1(dx) \\ G_j(\mu) &= \sum_{N \geq 0} \int \sum_{i=1}^N (E, \vec{P}, \mathcal{N})(\mathbf{q}, \mathbf{p}) \delta(q_i - x) \mu(d\mathbf{q}, d\mathbf{p}). \end{aligned} \tag{12.2}$$

Let $\mathbf{G}(\mu) := (G_0(\mu), G_i(\mu))$. Similar to Chapter 10, the goal is to solve the corresponding Lagrange multiplier problem and derive the local equilibrium Gibbs mea-

sure:

$$\frac{1}{Z} \exp \left(\sum_{i=1}^N \beta(q_i) E(\mathbf{q}, \mathbf{p}) + \sum_{i=1}^N \vec{\lambda}(q_i) \vec{P}(\mathbf{q}, \mathbf{p}) + \sum_{i=1}^N \nu(q_i) \mathcal{N}(\mathbf{q}, \mathbf{p}) \right) \quad (12.3)$$

where Z is the corresponding partition function.

The main difficulty with mapping into a function space (as opposed to \mathbb{R} , see Lemma 9.4.1) is ensuring the functions mapped to are regular values. Below we will assume the constraint functions in $(L^1(dx))^5$ that $G_j(\mu)$ maps to are regular points. Under this assumption we show (12.3) is a critical point of Gibbs entropy (9.1). A characterization of the regular values of G_j will be the focus of later work. As in Chapter 9 we show \mathbf{G} is C^1 by calculating the Gateaux derivative, and then showing this derivative is C^1 and hence is the Frechet derivative. First we prove $\mathbf{G}(\mu)$ maps into $L^1(dx)$:

Lemma 12.1.1. *Let $\mu = f\mu_0$ be an element of \mathcal{F} . Then*

$$G_j(\mu) = \int_{X_\Lambda} \sum_{i=1}^N C_j \delta(q_i - x) \mu \quad (12.4)$$

maps into $L^1(dx)$.

Proof. Note

$$\begin{aligned} & \int |G_j(\mu)(x)| dx \\ & \leq \int \left(\sum_{N \geq 0} \int \sum_{i=1}^N |C_j| \delta(q_i - x) f(\mathbf{q}, \mathbf{p}) \mu_0 \right) dx \\ & = \int \left(\sum_{N \geq 0} \int |C_j|(x, q_2, \dots, q_N, \mathbf{p}) f(x, q_2, \dots, q_N, \mathbf{p}) \frac{dq_2 \dots dq_N}{N!} d\mathbf{p} \right. \\ & \quad + \dots + \\ & \quad \left. \int |C_j|(q_1, q_2, \dots, x, \mathbf{p}) f(q_1, q_2, \dots, x, \mathbf{p}) \frac{dq_1 \dots dq_{N-1}}{N!} d\mathbf{p} \right) dx. \end{aligned} \quad (12.5)$$

By Tonelli's theorem, iterating the above integrals and letting $x = q_i$ for each term of the sum gives

$$\begin{aligned} \int |G_j(\mu)(x)| dx &\leq \sum_{N \geq 0} N \int |C_j|(\mathbf{q}, \mathbf{p}) f(\mathbf{q}, \mathbf{p}) \frac{d\mathbf{q}}{N!} d\mathbf{p} \\ &= \sum_{N \geq 0} \int |C_j|(\mathbf{q}, \mathbf{p}) f(\mathbf{q}, \mathbf{p}) \frac{d\mathbf{q}}{(N-1)!} d\mathbf{p}. \end{aligned} \quad (12.6)$$

We argue the final summation in (12.6) is finite. Note that $\mu \in \mathcal{F}$ implies by definition

$\sum_{N \geq 0} \int |C_j| \mu$ is finite. Since

$$\begin{aligned} \sum_{N \geq 0} \int |C_j| \mu &= \sum_{N \geq 0} \frac{1}{N!} \underbrace{\int |C_j|(\mathbf{q}, \mathbf{p}) f(\mathbf{q}, \mathbf{p}) dq_1 \dots dq_N dp_1 \dots dp_N}_{=: I_N} \\ &= \sum_{N \geq 0} \frac{I_N}{N!} \end{aligned} \quad (12.7)$$

we conclude $\sum_{N \geq 0} \frac{I_N}{N!}$ is finite. Noting that (12.6) can be rewritten as

$$\int |G_j(\mu)(x)| dx \leq \sum_{N \geq 0} \frac{I_N}{(N-1)!} \quad (12.8)$$

we apply the Ratio test, keeping in mind $I_N < +\infty$ and $I_N \rightarrow 0$ as $N \rightarrow \infty$:

$$\lim_N \left| \frac{\frac{I_N}{N!}}{\frac{I_{N-1}}{(N-1)!}} \right| = \lim_N \frac{I_N}{I_{N-1}} \frac{1}{N} = 0. \quad (12.9)$$

to conclude $G_j(\mu)(x) \in L^1(dx)$. □

This suggests measures of the form $f \frac{d\mathbf{q}}{(N-1)!} d\mathbf{p}$ or equivalently, $Nf\mu_0$ will be important in proving \mathbf{G} is C^1 , leading to the following lemma:

Lemma 12.1.2. *Let $\mu = f\mu_0$ be an element of \mathcal{F} . Define $\hat{\mu} = Nf\mu_0$. Then $\hat{\mu} \in \mathcal{F}$ and $\hat{\mu}$ and μ are equivalent measures.*

Proof. First note that $\mu \in \mathcal{F}$ implies

$$\begin{aligned} \sum_{N \geq 0} \int f \frac{dq_1 \dots dq_N}{N!} dp_1 \dots dp_N &= \sum_{N \geq 0} \frac{1}{N!} \underbrace{\int f dq_1 \dots dq_N dp_1 \dots dp_N}_{=: I_N} \\ &= \sum_{N \geq 0} \frac{I_N}{N!} < +\infty. \end{aligned} \quad (12.10)$$

Therefore, we write

$$\begin{aligned} \int_{X_\Lambda} \hat{\mu} &= \sum_{N \geq 0} \int N f \frac{dq_1 \dots dq_N}{N!} dp_1 \dots dp_N \\ &= \sum_{N \geq 0} \frac{I_N}{(N-1)!}. \end{aligned} \quad (12.11)$$

The sum in (12.11) can be shown to converge by the Ratio test as in the proof of Lemma 12.1.1. Analogous arguments show $\int_{X_\Lambda} f|H|\hat{\mu}$, $\int_{X_\Lambda} f|\vec{P}|\hat{\mu}$ and $\int_{X_\Lambda} f|\mathcal{N}|\hat{\mu}$ are finite. We conclude $\hat{\mu} \in \mathcal{F}$.

To show $\hat{\mu}$ and μ are equivalent (see Definition 7.2.1), note that their ratios are $\phi = \frac{Nf}{f} = N$ and $\frac{1}{\phi} = \frac{f}{Nf} = \frac{1}{N}$. Then $\hat{\mu} = \phi\mu$ and $\mu = \frac{1}{\phi}\hat{\mu}$, where

$$\int_{X_\Lambda} \phi^p \mu = \int_{X_\Lambda} N^p \mu = \sum_{N \geq 0} \frac{N^p I_N}{N!} < +\infty \quad (12.12)$$

for any $p > 1$, and

$$\int_{X_\Lambda} \frac{1}{\phi^p} \mu = \int_{X_\Lambda} \frac{1}{N^p} \mu = \sum_{N \geq 0} \frac{I_N}{N^p N!} < +\infty \quad (12.13)$$

for any $p > 1$. Implying $\phi \in L^p(\mu)$ and $\frac{1}{\phi} \in L^p(\hat{\mu})$ for some $p > 1$. Similar arguments will show $\phi \in L^p(|C_j|\mu)$ and $\frac{1}{\phi} \in L^p(|C_j|\hat{\mu})$. We conclude $\mu \sim \hat{\mu}$. \square

We now prove the constraint functionals are C^1 . First, we show the derivative is a bounded linear functional.

Proposition 12.1.3. *The derivative of the mapping $G : \mathcal{F} \rightarrow (L^1(dx))^5$ is a bounded linear functional, with the derivative of the j^{th} component given by*

$$d_\mu G_j(u) = \int_{X_\Lambda} \sum_{i=1}^N C_j u \delta(q_i - x) \mu. \quad (12.14)$$

Proof. We show the derivative corresponding to the energy functional is a bounded linear functional. Let $\mu' = g\mu_0 \in K_\mu$ and $s_{\mu'} : K_{\mu'} \rightarrow X_{\mu'}$ such that $s_{\mu'}^{-1}(u_0) = \mu$ for some $u_0 \in X_{\mu'}$. Let $u \in X_{\mu'}$. Then

$$\begin{aligned} d_\mu G_j(u) &= \left. \frac{d}{d\epsilon} \right|_{\epsilon=0} \int_{X_\Lambda} \sum_{i=1}^N C_j e^{u_0 + \epsilon u} \delta(q_i - x) g\mu_0 \\ &= \int_{X_\Lambda} \sum_{i=1}^N C_j u \delta(q_i - x) e^{u_0} g\mu_0 \\ &= \int_{X_\Lambda} \sum_{i=1}^N C_j u \delta(q_i - x) \mu. \end{aligned} \tag{12.15}$$

Note that $u \in X_\mu = X_{\hat{\mu}}$ and an analogous argument to that in the proof of Lemma 12.1.1 shows $d_\mu G_j \in L^1(dx)$ and

$$\begin{aligned} \|d_\mu G_j[u]\|_{L^1(dx)} &= \int_\Lambda \left| \int_{X_\Lambda} \sum_{i=1}^N C_j u \delta(q_i - x) \mu \right| dx \\ &\leq \int_\Lambda \int_{X_\Lambda} \sum_{i=1}^N |C_j| |u| \delta(q_i - x) \mu dx. \end{aligned} \tag{12.16}$$

Similar to the calculation in (12.5) and (12.6), we can rewrite the above as

$$\|d_\mu G_j[u]\|_{L^1(dx)} \leq \sum_{N \geq 0} N \int |u| |C_j| \mu = \int_{X_\Lambda} |u| |C_j| \hat{\mu}. \tag{12.17}$$

Since Orlicz spaces are continuously embedded in all L^p spaces, there exists $K > 0$ such that

$$\|d_\mu G_j[u]\|_{L^1(dx)} \leq K \|u\|_{L^{e^{|t|}-1}(|C_j|\hat{\mu})}. \tag{12.18}$$

Last, since $\hat{\mu} \sim \mu$ (Lemma 12.1.2), we know the norms Orlicz norms with respect to μ and $\hat{\mu}$ are equivalent. Therefore, there exists some $K' > 0$ such that

$$\|d_\mu G_j[u]\|_{L^1(dx)} \leq K' \|u\|_{L^{e^{|t|}-1}(|C_j|\mu)} \leq K' \|u\|_{L(\mu)}. \tag{12.19}$$

where $\|x\|_{L(\mu)} = \max_{0 \leq j \leq n} \|x\|_{L^{e^{|t|}-1}(|C_j|\mu)}$ defined in (8.41). \square

Proposition 12.1.4. *The mapping $\mu \rightarrow d_\mu G_j$ is continuous.*

Proof. Let $\mu := f\mu_0$ and $\mu_n = f_n\mu_0$. From Proposition 12.1.3 we know

$$d_\mu G_j(u) = \int_{X_\Lambda} \sum_{i=1}^N C_j u \delta(q_i - x) \mu. \quad (12.20)$$

We must show

$$\sup_{\|y\|_{L(\mu)}=1} \|d_{\mu_n} G_j[y] - d_\mu G_j[y]\|_{L^1(dx)} \rightarrow 0 \quad (12.21)$$

as μ_n e -converges to μ (see Definition 7.3.1).

We look at

$$\int_\Lambda \left| \int_{X_\Lambda} \sum_{i=1}^N C_j y \delta(q_i - x) \mu_n - \int_{X_\Lambda} \sum_{i=1}^N C_j y \delta(q_i - x) \mu \right| dx. \quad (12.22)$$

This is bounded above by

$$\int_\Lambda \int_{X_\Lambda} \sum_{i=1}^N |C_j| |y| \delta(q_i - x) |f_n - f| \mu_0 dx. \quad (12.23)$$

Equivalently,

$$\int_\Lambda \int_{X_\Lambda} \sum_{i=1}^N |C_j| |y| \delta(q_i - x) \left| \frac{f_n}{f} - 1 \right| \mu dx. \quad (12.24)$$

As in the proof of Proposition 12.1.3, we integrate over dq_i and then change variables (again note any iterated integrals are justified by Tonelli's theorem):

$$\int_\Lambda \int_{\Lambda^{N-1} \times \mathbb{R}^N} \sum_{i=1}^N |y|(\dots, x, \dots) \left| \frac{f_n}{f}(\dots, x, \dots) - 1 \right| |C_j|(\dots, x, \dots) \widehat{\mu}_i dx. \quad (12.25)$$

Letting $x = q_i$ gives

$$\int_{X_\Lambda} N |y| \left| \frac{f_n}{f} - 1 \right| |C_j| \mu. \quad (12.26)$$

Thus far we have

$$\begin{aligned} \|d_{\mu_n} G_j[y] - d_\mu G_j[y]\|_{L^1(dx)} &\leq \int_{X_\Lambda} N |y| \left| \frac{f_n}{f} - 1 \right| |C_j| \mu \\ &= \int_{X_\Lambda} |y| \left| \frac{f_n}{f} - 1 \right| |C_j| \widehat{\mu}. \end{aligned} \quad (12.27)$$

We proceed as in the proof of Proposition 9.3.3. We apply Holder's inequality (a, b conjugates) and get $\|d_{\mu_n}G_j[y] - d_\mu G_j[y]\|_{L^1(dx)}$ is bounded above by

$$\left(\int_{X_\Lambda} |y|^a |C_j|\widehat{\mu}\right)^{1/a} \left(\int_{X_\Lambda} \left|\frac{f_n}{f} - 1\right|^b |C_j|\widehat{\mu}\right)^{1/b}. \quad (12.28)$$

Note that the first term is finite since $y \in X_\mu = X_{\widehat{\mu}}$ implies $y \in L^{e^{|\mathfrak{t}|-1}}(|C_j|\widehat{\mu})$. Proposition 7.1.5 then implies $y \in L^a(|C_j|\widehat{\mu})$. For the second term, recall that $\mu \sim \widehat{\mu}$ and that e -convergence against one measure implies e -convergence against equivalent measures (see Proposition 7.3.3). Thus, the second term converges to 0 by e -convergence of μ_n to μ . Therefore,

$$\|d_{\mu_n}G_j[y] - d_\mu G_j[y]\|_{L^1(dx)} \rightarrow 0 \quad (12.29)$$

and we conclude $\mu \rightarrow d_\mu G_j$ is a continuous mapping. \square

We now apply Theorem 9.1.4 to show (12.3) is a critical point of the Gibbs entropy.

Theorem 12.1.5. *The critical point of Gibbs entropy, on components of \mathcal{F} where $\log f \in L^{e^{|\mathfrak{t}|-1}}(f\mu_0)$, is the Gibbs distribution:*

$$\frac{1}{Z} \exp \left(\sum_{i=1}^N \beta(q_i)E + \sum_{i=1}^N \vec{\lambda}(q_i)\vec{P} + \sum_{i=1}^N \nu(q_i)\mathcal{N} \right) \quad (12.30)$$

where $\beta(x), \vec{\lambda}(x)$ and $\nu(x)$ are $L^\infty(dx)$ functions.

Proof. Let $\mu = f\mu_0$ be a critical point. By Theorem 9.1.4, we have $(\lambda_0, \beta(x), \vec{\lambda}(x), \nu(x)) \in (\mathbb{R} \times (L^1(dx))^5)^* = \mathbb{R} \times (L^\infty(dx))^5$ such that

$$\begin{aligned} & \int_{X_\Lambda} (\log f + 1) u f \mu_0 \\ &= (\lambda_0, \beta(x), \vec{\lambda}(x), \nu(x)) \circ \left(\int_{X_\Lambda} u |C_0| \mu, \int_{X_\Lambda} \sum_{i=1}^N (E, \vec{P}, \mathcal{N}) u \delta(q_i - x) \mu \right) \end{aligned} \quad (12.31)$$

for all u . In other words,

$$\begin{aligned}
\int_{X_\Lambda} (\log f + 1) u \mu &= \int_{X_\Lambda} \lambda_0 u |C_0| \mu \\
&+ \int_{\Lambda} \beta(x) \left(\int_{X_\Lambda} \sum_{i=1}^N E u \delta(q_i - x) \mu \right) dx \\
&+ \int_{\Lambda} \vec{\lambda}(x) \left(\int_{X_\Lambda} \sum_{i=1}^N \vec{P} u \delta(q_i - x) \mu \right) dx \\
&+ \int_{\Lambda} \nu(x) \left(\int_{X_\Lambda} \sum_{i=1}^N \mathcal{N} u \delta(q_i - x) \mu \right) dx.
\end{aligned} \tag{12.32}$$

Note that

$$\begin{aligned}
&\int_{\Lambda} \beta(x) \left(\int_{X_\Lambda} \sum_{i=1}^N E u \delta(q_i - x) \mu \right) dx \\
&= \int_{\Lambda} \beta(x) \left(\int E(x, \dots) u(x, \dots) f(x, \dots) \frac{dq_2 \dots dq_N}{N!} dp_1 \dots dp_N \right. \\
&\quad \left. + \dots + \int E(\dots, x) u(\dots, x) f(\dots, x) \frac{dq_1 \dots dq_{N-1}}{N!} dp_1 \dots dp_N \right) dx.
\end{aligned} \tag{12.33}$$

Making a change of variables ($x = q_i$) at each integral in the sum gives

$$\begin{aligned}
&\int_{\Lambda} \beta(x) \left(\int_{X_\Lambda} \sum_{i=1}^N E u \delta(q_i - x) \mu \right) dx \\
&= \int_{X_\Lambda} \beta(q_1) E u \mu + \dots + \int_{X_\Lambda} \beta(q_N) E u \mu \\
&= \int_{X_\Lambda} \sum_{i=1}^N \beta(q_i) E u \mu.
\end{aligned} \tag{12.34}$$

Similar calculations hold for the other constraint functionals. Therefore we have

$$\begin{aligned}
\int_{X_\Lambda} (\log f + 1) u \mu &= \int_{X_\Lambda} \lambda_0 u |C_0| \mu + \int_{X_\Lambda} \sum_{i=1}^N \beta(q_i) E u \mu \\
&+ \int_{X_\Lambda} \sum_{i=1}^N \vec{\lambda}(q_i) \vec{P} u \mu + \int_{X_\Lambda} \sum_{i=1}^N \nu(q_i) \mathcal{N} u \mu.
\end{aligned} \tag{12.35}$$

implying ($C_0 \equiv 1$, folding 1 and λ_0 into a new relabeled λ_0)

$$\int_{X_\Lambda} \left(\log f + \lambda_0 - \sum_{i=1}^N \beta(q_i) E - \sum_{i=1}^N \vec{\lambda}(q_i) \vec{P} - \sum_{i=1}^N \nu(q_i) \mathcal{N} \right) u \mu = 0 \tag{12.36}$$

for all u . Proceeding as in Theorem 9.4.5, we can define $u = \chi_M$ where M is the set of all points where the terms in the parenthesis are strictly positive. Then $u \in X_\mu$ and the above integral will be some positive number, a contradiction unless

$$\log f + \lambda_0 - \sum_{i=1}^N \beta(q_i)E - \sum_{i=1}^N \vec{\lambda}(q_i)\vec{P} - \sum_{i=1}^N \nu(q_i)\mathcal{N} = 0. \quad (12.37)$$

A similar argument holds on N the set of negative points. Solving for f and using the condition $\int \mu = 1$ implies

$$f = \frac{1}{Z} \exp \left(\sum_{i=1}^N \beta(q_i)E + \sum_{i=1}^N \vec{\lambda}(q_i)\vec{P} + \sum_{i=1}^N \nu(q_i)\mathcal{N} \right). \quad (12.38)$$

We conclude the local Gibbs (12.3) is a critical point of Gibbs entropy. \square

Next, we provide a proof of the fact that (12.3) minimizes entropy.

Lemma 12.1.6. *The minimizing probability density of information entropy subject to constraints is the local Gibbs distribution given by (12.3).*

Proof. Let $e(x), \vec{p}(x), r(x) \in L^1(dx)$ be the constraint functions mapped to by (12.2). Let g denote the local Gibbs distribution given in (12.3) and let q be some other probability measure such $\mathbf{G}(g\mu_0)$ also maps to $e(x), \vec{p}(x)$ and $r(x)$. Note

$$\begin{aligned} & \int_{X_\Lambda} \sum_{i=1}^N \beta(q_i)E(\mathbf{q}, \mathbf{p})f(\mathbf{q}, \mathbf{p})\mu_0(d\mathbf{q}, d\mathbf{p}) \\ &= \int_{X_\Lambda} \beta(q_1)E(\mathbf{q}, \mathbf{p})f(\mathbf{q}, \mathbf{p})dq_2 \dots dq_N \frac{d\mathbf{p}}{N!}dq_1 \\ &+ \dots + \int_{X_\Lambda} \beta(q_N)E(\mathbf{q}, \mathbf{p})f(\mathbf{q}, \mathbf{p})dq_1 \dots dq_{N-1} \frac{d\mathbf{p}}{N!}dq_N. \end{aligned} \quad (12.39)$$

After making the change of variables $x = q_i$ and using the delta function to condense the result, we have

$$\begin{aligned} & \int_{X_\Lambda} \sum_{i=1}^N \beta(q_i)E(\mathbf{q}, \mathbf{p})f(\mathbf{q}, \mathbf{p})\mu_0(d\mathbf{q}, d\mathbf{p}) \\ &= \int_{\Lambda} \beta(x) \left(\sum_{i=1}^N \int_{X_\Lambda} E(\mathbf{q}, \mathbf{p})f(\mathbf{q}, \mathbf{p})\delta(q_i - x)\mu_0(d\mathbf{q}, d\mathbf{p}) \right) dx. \end{aligned} \quad (12.40)$$

Noting that the quantity in the parenthesis is $G_1(f)$ which is constrained to equal $e(x)$, (12.40) simplifies to

$$\int_{X_\Lambda} \sum_{i=1}^N \beta(q_i) E(\mathbf{q}, \mathbf{p}) f(\mathbf{q}, \mathbf{p}) \mu_0(d\mathbf{q}, d\mathbf{p}) = \int_{\Lambda} \beta(x) e(x) dx. \quad (12.41)$$

Similar arguments show

$$\begin{aligned} \int_{X_\Lambda} \sum_{i=1}^N \vec{\lambda}(q_i) \vec{P}(\mathbf{q}, \mathbf{p}) f(\mathbf{q}, \mathbf{p}) \mu_0(d\mathbf{q}, d\mathbf{p}) &= \int_{\Lambda} \vec{\lambda}(x) \cdot \vec{p}(x) dx \\ \int_{X_\Lambda} \sum_{i=1}^N \nu(q_i) \mathcal{N}(\mathbf{q}, \mathbf{p}) f(\mathbf{q}, \mathbf{p}) \mu_0(d\mathbf{q}, d\mathbf{p}) &= \int_{\Lambda} \nu(x) r(x) dx. \end{aligned} \quad (12.42)$$

Recalling H denotes the information entropy, we have

$$\begin{aligned} H(f\mu_0) &= \int_{X_\Lambda} f \log f \mu_0 \\ &= \int_{X_\Lambda} f \log \frac{f}{g} \mu_0 + \int_{X_\mu} f \log g \mu_0 \\ &= \int_{X_\Lambda} \frac{f}{g} \log \frac{f}{g} g \mu_0 + \int_{X_\mu} f \log g \mu_0 \\ &\geq \int_{X_\mu} f \log g \mu_0 \end{aligned} \quad (12.43)$$

where last line follows from $\int_{X_\Lambda} \frac{f}{g} \log \frac{f}{g} g \mu_0 \geq 0$ as a result of Jensen's inequality on the convex function $\phi(t) = t \log t$. Therefore, applying (12.41) and (12.42) gives

$$\begin{aligned} H(f\mu_0) &\geq \int_{X_\mu} f \log g \mu_0 \\ &= \int_{X_\mu} \left(\sum_{i=1}^N \beta(q_i) E + \sum_{i=1}^N \vec{\lambda}(q_i) \vec{P} + \sum_{i=1}^N \nu(q_i) \mathcal{N} - \log Z \right) f \mu_0 \\ &= \int_{\Lambda} \beta(x) e(x) dx + \int_{\Lambda} \vec{\lambda}(x) \cdot \vec{p}(x) dx + \int_{\Lambda} \nu(x) r(x) dx - \log Z. \end{aligned} \quad (12.44)$$

However, the Gibbs distribution g also maps to the same $e(x), \vec{p}(x), r(x)$ and so (12.41)

and (12.42) also hold for g . Applying these in the reverse direction for g gives

$$\begin{aligned} H(f\mu_0) &\geq \int_{X_\mu} \left(\sum_{i=1}^N \beta(q_i) E + \sum_{i=1}^N \vec{\lambda}(q_i) \vec{P} + \sum_{i=1}^N \nu(q_i) \mathcal{N} - \log Z \right) g \mu_0 \\ &= \int_{X_\mu} g \log g \mu_0 \\ &= H(g\mu_0) \end{aligned} \quad (12.45)$$

We conclude the Gibbs distribution minimizes entropy. \square

Proposition 12.1.7. *Let $\mu = f\mu_0$ be the local Gibbs distribution (12.3):*

$$f = \frac{1}{Z} \exp \left(- \sum_{i=1}^N \beta(q_i)E - \sum_{i=1}^N \vec{\lambda}(q_i)\vec{P} - \sum_{i=1}^N \nu(q_i)\mathcal{N} \right), \quad (12.46)$$

where $\beta(x), \vec{\lambda}(x)$ and $\nu(x)$ are $L^\infty(dx)$ functions. Then $\log f \in L^{e^{|\ell|}-1}(\mu)$.

Proof. As in Proposition 10.1.1, we show $\log f \in L^{e^{|\ell|}-1}(\mu)$. This requires we show

$$\sum_{N \geq 0} \int e^{-(1+\epsilon) \sum_{i=1}^N [\beta(q_i)E + \vec{\lambda}(q_i)\vec{P} + \nu(q_i)\mathcal{N}]} \mu_0 < +\infty \quad (12.47)$$

and

$$\sum_{N \geq 0} \int e^{-(1-\epsilon) \sum_{i=1}^N [\beta(q_i)E + \vec{\lambda}(q_i)\vec{P} + \nu(q_i)\mathcal{N}]} \mu_0 < +\infty \quad (12.48)$$

Starting with (12.47), note that $\beta(x), \vec{\lambda}(x)$ and $\nu(x)$ in $L^\infty(dx)$ implies there exists $M > 0$ such that $|(\beta, \vec{\lambda}, \nu)| \leq M$. Therefore, these terms will not effect the convergence of the integral. Similarly, the summation over i , will only add another factor of N in a decaying exponential. The argument then follows similarly to the proof of Proposition 10.1.1. For the second integral, after the above arguments are taken into account, the approach is again the same as the proof of Proposition 10.1.1, down to the restriction $\epsilon < 1$. We conclude $\log f \in L^{e^{|\ell|}-1}(\mu)$. \square

Part III
Appendix

Chapter 13

Appendix

The following sections give additional details omitted from some of the chapters in Part II.

13.1 Orlicz Space Propositions

Proposition 13.1.1. *The Orlicz space $L^\phi(\mu)$, where ϕ is a Young function, is linear.*

Proof. Letting $\alpha, \beta \in \mathbb{R}$ and $f, g \in L^\phi(\mu)$, there exists c_1 and c_2 such that $\int_\Omega \phi(f/c_1) \mu \leq 1$ and $\int_\Omega \phi(g/c_2) \mu \leq 1$. Then, by the convexity of ϕ ,

$$\begin{aligned} \int_\Omega \phi\left(\frac{\alpha f + \beta g}{\alpha c_1 + \beta c_2}\right) \mu &= \int_\Omega \phi\left(\frac{\alpha c_1}{\alpha c_1 + \beta c_2} \frac{f}{c_1} + \frac{\beta c_2}{\alpha c_1 + \beta c_2} \frac{g}{c_2}\right) \mu \\ &\leq \frac{\alpha c_1}{\alpha c_1 + \beta c_2} \int_\Omega \phi\left(\frac{f}{c_1}\right) \mu + \frac{\beta c_2}{\alpha c_1 + \beta c_2} \int_\Omega \phi\left(\frac{g}{c_2}\right) \mu \\ &\leq \frac{\alpha c_1}{\alpha c_1 + \beta c_2} (1) + \frac{\beta c_2}{\alpha c_1 + \beta c_2} (1) \\ &= 1 \end{aligned} \tag{13.1}$$

Thus, $\alpha f + \beta g \in L^\phi(\mu)$. We conclude $L^\phi(\mu)$ is a linear space. □

Proposition 13.1.2. *Let ϕ be a Young function and $L^\phi(\mu)$ the associated Orlicz space. Then the infimum in the definition of the norm*

$$\|f\|_{L^\phi(\mu)} = \inf \left\{ \alpha > 0 : \int_\Omega \phi\left(\frac{f}{\alpha}\right) \mu \leq 1 \right\} \tag{13.2}$$

is attained.

Proof. Let $f \in L^\phi(\mu)$ and a_n be a sequence decreasing to $\|f\|_{L^\phi(\mu)}$. Since a_n is a decreasing sequence, $\frac{f}{a_n}$ is an increasing sequence. Thus, since ϕ is a Young function (and therefore increasing on $[0, \infty)$), $g_n = \phi\left(\frac{f}{a_n}\right)$ is monotonically increasing, regardless of f being positive or negative, by the evenness of ϕ , to $\phi\left(\frac{f}{\|f\|_{L^\phi(\mu)}}\right)$. By the Monotone Convergence Theorem

$$\lim_{n \rightarrow \infty} \int_{\Omega} \phi\left(\frac{f}{a_n}\right) \mu = \int_{\Omega} \phi\left(\frac{f}{\|f\|_{L^\phi(\mu)}}\right) \mu \leq 1. \quad (13.3)$$

□

Proposition 13.1.3. *Let ϕ be a Young function and $L^\phi(\mu)$ be the associated Orlicz space. Then*

$$\int_{\Omega} \phi\left(\frac{f}{\alpha}\right) \mu \leq 1 \quad (13.4)$$

for all $\alpha \geq \|f\|_{L^\phi(\mu)}$.

Proof. Let

$$c := \|f\|_{L^\phi(\mu)} = \inf \left\{ \alpha > 0 : \int_{\Omega} \phi\left(\frac{f}{\alpha}\right) \mu \leq 1 \right\}. \quad (13.5)$$

Note

$$\frac{|f|}{\alpha} \leq \frac{|f|}{c} \quad (13.6)$$

for $\alpha > c > 0$. Since ϕ is strictly increasing function,

$$\phi\left(\frac{|f|}{\alpha}\right) \leq \phi\left(\frac{|f|}{c}\right). \quad (13.7)$$

Since ϕ is an even function, we can disregard the absolute value bars. Thus,

$$\phi\left(\frac{f}{\alpha}\right) < \phi\left(\frac{f}{c}\right) \text{ for all } \alpha > c \quad (13.8)$$

implying

$$\int \phi\left(\frac{f}{\alpha}\right) \mu < \int \phi\left(\frac{f}{c}\right) \mu \leq 1. \quad (13.9)$$

Thus,

$$\int \phi\left(\frac{f}{\alpha}\right) \mu \leq 1 \text{ for all } \alpha > c. \quad (13.10)$$

□

13.2 e -convergence

We now give a proof of Proposition 7.3.2. See [AJVLS, Lemma 3.7 and Proposition 3.9] for an alternate approach.

Proposition 13.2.1. *The following are equivalent:*

1. *The sequence g_n is e -convergent to g .*
2. *For all $p \geq 1$, we have*

$$\lim_{n \rightarrow \infty} \int \left| \left(\frac{g_n}{g} \right)^p - 1 \right| |C_i| g \mu_0 = \lim_{n \rightarrow \infty} \int \left| \left(\frac{g}{g_n} \right)^p - 1 \right| |C_i| g \mu_0 = 0, \quad (13.11)$$

for $i = 0, 1, \dots, n$.

Proof. Let $C_0 \equiv 1$. Fix i .

First, we show (1) \implies (2): Note that $\left| \left(\frac{g_n}{g} \right)^p - 1 \right|$ and $\left| \left(\frac{g}{g_n} \right)^p - 1 \right|$ are increasing in p . Without loss of generality, we assume $p \in \mathbb{N}$. Define $f_n := \frac{g_n}{g} - 1$.

Then

$$\lim_{n \rightarrow \infty} \int |f_n|^p |C_i| g \mu_0 = \lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p |C_i| g \mu_0 = 0 \quad (13.12)$$

for all $p \geq 1$ by assumption. Then

$$\int \left| \left(\frac{g_n}{g} \right)^p - 1 \right| |C_i| g \mu_0 = \int |(1 + f_n)^p - 1| |C_i| g \mu_0. \quad (13.13)$$

Applying the binomial theorem, we see

$$\begin{aligned} \int \left| \left(\frac{g_n}{g} \right)^p - 1 \right| |C_i| g \mu_0 &\leq \int \sum_{k=1}^p \binom{p}{k} |f_n|^k |C_i| g \mu_0 \\ &= \sum_{k=1}^p \binom{p}{k} \int |f_n|^k |C_i| g \mu_0 \rightarrow 0. \end{aligned} \quad (13.14)$$

Similarly, we can define $h_n := \frac{g}{g_n} - 1$ and show

$$\int \left| \left(\frac{g}{g_n} \right)^p - 1 \right| |C_i| g \mu_0 \rightarrow 0. \quad (13.15)$$

Next, we show (2) \implies (1). First note the following facts:

1. if $x \geq 1$, then $|x - 1|^p \leq |x^p - 1|$ for $p \geq 1$.
2. if $x < 1$, then $|x - 1|^p \leq |\frac{1}{x} - 1|^p$ for $p \geq 1$.

By hypothesis

$$\lim_{n \rightarrow \infty} \int \left| \left(\frac{g_n}{g} \right)^p - 1 \right| |C_i| g \mu_0 = \lim_{n \rightarrow \infty} \int \left| \left(\frac{g}{g_n} \right)^p - 1 \right| |C_i| g \mu_0 = 0. \quad (13.16)$$

Then, applying the above two facts implies

$$\begin{aligned} &\lim_{n \rightarrow \infty} \int \left| \frac{g_n}{g} - 1 \right|^p |C_i| g \mu_0 \\ &= \lim_{n \rightarrow \infty} \int_{\{\frac{g_n}{g} \geq 1\}} \left| \frac{g_n}{g} - 1 \right|^p |C_i| g \mu_0 + \int_{\{\frac{g}{g_n} > 1\}} \left| \frac{g_n}{g} - 1 \right|^p |C_i| g \mu_0 \\ &\leq \lim_{n \rightarrow \infty} \int_{\{\frac{g_n}{g} \geq 1\}} \left| \left(\frac{g_n}{g} \right)^p - 1 \right| |C_i| g \mu_0 + \int_{\{\frac{g}{g_n} > 1\}} \left| \frac{g_n}{g} - 1 \right|^p |C_i| g \mu_0 \\ &\leq \lim_{n \rightarrow \infty} \int_{\{\frac{g_n}{g} \geq 1\}} \left| \left(\frac{g_n}{g} \right)^p - 1 \right| |C_i| g \mu_0 + \int_{\{\frac{g}{g_n} > 1\}} \left| \frac{g}{g_n} - 1 \right|^p |C_i| g \mu_0 \\ &\leq \lim_{n \rightarrow \infty} \int_{\{\frac{g_n}{g} \geq 1\}} \left| \left(\frac{g_n}{g} \right)^p - 1 \right| |C_i| g \mu_0 + \int_{\{\frac{g}{g_n} > 1\}} \left| \left(\frac{g}{g_n} \right)^p - 1 \right| |C_i| g \mu_0 \\ &\leq \lim_{n \rightarrow \infty} \int \left| \left(\frac{g_n}{g} \right)^p - 1 \right| |C_i| g \mu_0 + \int \left| \left(\frac{g}{g_n} \right)^p - 1 \right| |C_i| g \mu_0 \\ &= 0 \end{aligned} \quad (13.17)$$

Similarly,

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \int \left| \frac{g}{g_n} - 1 \right|^p |C_i| g \mu_0 \\
&= \lim_{n \rightarrow \infty} \int_{\{\frac{g}{g_n} \geq 1\}} \left| \frac{g}{g_n} - 1 \right|^p |C_i| g \mu_0 + \int_{\{\frac{g_n}{g} > 1\}} \left| \frac{g}{g_n} - 1 \right|^p |C_i| g \mu_0 \\
&\leq \lim_{n \rightarrow \infty} \int_{\{\frac{g}{g_n} \geq 1\}} \left| \left(\frac{g}{g_n} \right)^p - 1 \right| |C_i| g \mu_0 + \int_{\{\frac{g_n}{g} > 1\}} \left| \frac{g}{g_n} - 1 \right|^p |C_i| g \mu_0 \\
&\leq \lim_{n \rightarrow \infty} \int_{\{\frac{g}{g_n} \geq 1\}} \left| \left(\frac{g}{g_n} \right)^p - 1 \right| |C_i| g \mu_0 + \int_{\{\frac{g_n}{g} > 1\}} \left| \frac{g_n}{g} - 1 \right|^p |C_i| g \mu_0 \tag{13.18} \\
&\leq \lim_{n \rightarrow \infty} \int_{\{\frac{g}{g_n} \geq 1\}} \left| \left(\frac{g}{g_n} \right)^p - 1 \right| |C_i| g \mu_0 + \int_{\{\frac{g_n}{g} > 1\}} \left| \left(\frac{g_n}{g} \right)^p - 1 \right| |C_i| g \mu_0 \\
&\leq \lim_{n \rightarrow \infty} \int \left| \left(\frac{g}{g_n} \right)^p - 1 \right| |C_i| g \mu_0 + \int \left| \left(\frac{g_n}{g} \right)^p - 1 \right| |C_i| g \mu_0 \\
&= 0
\end{aligned}$$

We conclude g_n e -converges to g . □

13.3 Differentiation on a Banach Manifold

First, consider what it means for a mapping between Banach manifolds to be C^1 and C^r (see [AMR, Definition 2.3.3]):

Definition 13.3.1. *Let E and F be Banach spaces, $f : U \subset E \rightarrow F$, and $\mathcal{L}(E, F)$ denote the space of linear maps from E to F . Let $u, u_0 \in U$. If there exists $L \in \mathcal{L}(E, F)$ such that*

$$\lim_{u \rightarrow u_0} \frac{\|f(u) - (f(u_0) + L[u - u_0])\|_F}{\|u - u_0\|_E} = 0 \tag{13.19}$$

we say f is Frechet differentiable at u_0 , and define the derivative of f at u_0 to be $d_{u_0}f := L$. If f is differentiable at each $u_0 \in U$, the map

$$\begin{aligned}
Df : I &\rightarrow \mathcal{L}(E, F) \\
u &\rightarrow d_u f
\end{aligned} \tag{13.20}$$

is the derivative of f . Moreover, if $d_u f$ is a continuous map (where $\mathcal{L}(E, F)$ has the norm topology, we say f is of class C^1 .

Definition 13.3.2. Proceeding inductively from Definition 13.3.1, we define

$$d_u^r f = d_u(d_u^{r-1} f) : U \subset E \rightarrow \mathcal{L}^r(E, F) \quad (13.21)$$

if it exists, where $\mathcal{L}^r(E, F) := \mathcal{L}(\underbrace{E \times \dots \times E}_{n \text{ times}}, F)$. If $d_u^r f$ exists and is norm continuous we say f is of class C^r .

Definition 13.3.3. Let E, F be Banach spaces. Let $f : U \subset E \rightarrow F$ and let $u \in U$.

We say that f has a derivative in the direction $e \in E$ at u if

$$\left. \frac{d}{dt} \right|_{t=0} f(u + te) \quad (13.22)$$

exists. We call this element of F the Gateaux derivative.

Note that Frechet differentiability implies Gateaux differentiability (see [AMR, Proposition 2.4.6]). This is analogous to differentiability implying the existence of directional derivatives.

Definition 13.3.4. If f is Gateaux differentiable (see Definition 13.3.3) and if the mapping $E \rightarrow \mathcal{L}(E, F)$ defined by $u \rightarrow d_u f$ is continuous, then E is C^1 Gateaux.

13.4 Local Surjectivity Theorem

Theorem 13.4.1. (Local Surjectivity Theorem) Let D be an open subset of X and let $T : D \rightarrow Y$ be continuously differentiable. Let $x_0 \in D$ be a regular point, i.e. $T'(x_0) : X \rightarrow Y$ is a surjective map. Then there is an open neighborhood $N'(y_0)$ of the point $y_0 = T(x_0)$ such that the equation $T(x) = y$ has a solution for every

$y \in N'(y_0)$. Furthermore, there exists an open set in D containing x_0 such that T maps this open set onto $N'(y_0)$. In particular, if $T'(x)$ is onto for $x \in D$, then f is an open mapping.

Proof. Let $L_0 = \text{null } T'(x_0)$. Since L_0 is closed, X/L_0 is a Banach space. Define $A : X/L_0 \rightarrow Y$ by $A[x] = T'(x_0)x$ where $[x]$ denotes the equivalence class of x , modulo L_0 . A is well defined since equivalent elements yield the same y , i.e.

$$\begin{aligned} x_1 = x_2 &\implies x_1 \equiv x_2 \\ &\implies A[x_1] = A[x_2]. \end{aligned} \tag{13.23}$$

This operator is linear:

$$A[x_1 + x_2] = T'(x_0)(x_1 + x_2) = T'(x_0)x_1 + T'(x_0)x_2 = A[x_1] + A[x_2] \tag{13.24}$$

$$A[kx] = T'(x_0)(kx) = kT'(x_0)x = kA[x]$$

continuous by the continuous differentiability of T , injective (since A is defined on X mod the kernel)

and onto:

$$\begin{aligned} y \in Y &\implies \text{by surjectivity of } T'(x_0), \exists x \text{ s.t. } y = T'(x_0)x \\ &\implies A[x] = y. \end{aligned} \tag{13.25}$$

Hence A^{-1} exists and is continuous (Banach Isomorphism Theorem see [AMR] Page 63). Let $\epsilon > 0$ be given. By the continuity of T' , there exists $r > 0$ such that $\|x - x_0\| < r$ implies $\|T'(x) - T'(x_0)\| < \epsilon$. However, since we want our neighborhood of x_0 to be contained in D , we make a further restriction on the neighborhood of x_0 . Define

$$d = \text{dist}(x_0, D^c) = \inf \{\|u - x_0\| : u \in D^c\} \tag{13.26}$$

and take all $x \in X$ such that $\|x - x_0\| < r'$ where $r' = \min(r, d)$. Note that for such x we still have $\|T'(x) - T'(x_0)\| < \epsilon$. Let $y \in N'(y_0)$ so that y is sufficiently close to

y_0 . In particular,

$$N'(y_0) = \left\{ y \in Y : \|y - y_0\| < \frac{r'}{4\|A^{-1}\|} \right\}. \quad (13.27)$$

We now construct a sequence of elements L_n from X/L_0 and $g_n \in L_n$. Let $g_0 = 0 \in L_0$. We prove by induction that $\|g_n\| < r'$ for all n .

Define L_1 recursively by

$$\begin{aligned} L_1 - L_0 &= A^{-1}(y - T(x_0 + g_0)) \\ &= A^{-1}(y - T(x_0)) \\ &= A^{-1}(y - y_0). \end{aligned} \quad (13.28)$$

Note that $A[x] = T'(x_0)x$, implying

$$\begin{aligned} A[L_i] &= T'(x_0)L_i \\ \implies L_i &= A^{-1}(T'(x_0)L_i). \end{aligned} \quad (13.29)$$

Therefore

$$L_0 = A^{-1}(T'(x_0)L_0) = A^{-1}(T'(x_0) \cdot 0) = A^{-1}(0) = [0]. \quad (13.30)$$

Then $L_1 - L_0 = A^{-1}(y - y_0)$ becomes

$$L_1 = A^{-1}(y - y_0). \quad (13.31)$$

Select $g_1 \in L_1$ such that $\|g_1 - g_0\| \leq 2\|L_1 - L_0\|$ which is possible since $\|L_1 - L_0\| = \inf_{g \in L_1} \|g - g_0\|$. Note that $\|g_1 - g_0\| \leq 2\|L_1 - L_0\|$ is the same as

$$\|g_1 - 0\| \leq 2\|L_1 - [0]\| \quad (13.32)$$

or

$$\|g_1\| \leq 2\|L_1\|. \quad (13.33)$$

Since $L_1 = A^{-1}(y - y_0)$, this implies

$$\|g_1\| \leq 2\|L_1\| \leq 2\|A^{-1}\|\|y - y_0\| < 2\|A^{-1}\|\frac{r'}{4\|A^{-1}\|} = \frac{r'}{2}. \quad (13.34)$$

Next, we prove the n^{th} case, $\|g_n\| < r'$. Assume $\|g_{n-1}\|, \|g_{n-2}\| < r'$. Define

$$L_n - L_{n-1} = A^{-1}(y - T(x_0 + g_{n-1})). \quad (13.35)$$

Select $g_n \in L_n$ such that

$$\|g_n - g_{n-1}\| \leq 2\|L_n - L_{n-1}\|. \quad (13.36)$$

Since $A^{-1}(T'(x_0)g_{n-1}) = [g_{n-1}] = L_{n-1}$,

$$\begin{aligned} L_n &= A^{-1}(y - T(x_0 + g_{n-1})) + L_{n-1} \\ &= A^{-1}(y - T(x_0 + g_{n-1})) + A^{-1}(T'(x_0)g_{n-1}) \\ &= A^{-1}(y - T(x_0 + g_{n-1}) + T'(x_0)g_{n-1}). \end{aligned} \quad (13.37)$$

Similarly,

$$L_{n-1} = A^{-1}(y - T(x_0 + g_{n-2}) + T'(x_0)g_{n-2}). \quad (13.38)$$

Therefore,

$$\begin{aligned} L_n - L_{n-1} &= A^{-1}(y - T(x_0 + g_{n-1}) + T'(x_0)g_{n-1}) \\ &\quad - A^{-1}(y - T(x_0 + g_{n-2}) + T'(x_0)g_{n-2}) \\ &= A^{-1}(-T(x_0 + g_{n-1}) + T(x_0 + g_{n-2}) + T'(x_0)(g_{n-1} - g_{n-2})) \\ &= -A^{-1}(T(x_0 + g_{n-1}) - T(x_0 + g_{n-2}) - T'(x_0)(g_{n-1} - g_{n-2})). \end{aligned}$$

Define $g_t = tg_{n-1} + (1-t)g_{n-2}$. Note that $\|g_{n-1}\|, \|g_{n-2}\| < r'$ implies $\|g_t\| < r'$.

Then

$$\begin{aligned} \|L_n - L_{n-1}\| &\leq \|A^{-1}\| \|T(x_0 + g_{n-1}) - T(x_0 + g_{n-2}) - T'(x_0)(g_{n-1} - g_{n-2})\| \\ &\leq \|A^{-1}\| \sup_{0 \leq t \leq 1} \|T'(x_0 + g_t)(g_{n-1} - g_{n-2}) - T'(x_0)(g_{n-1} - g_{n-2})\| \\ &\leq \|A^{-1}\| \|g_{n-1} - g_{n-2}\| \sup_{0 \leq t \leq 1} \|T'(x_0 + g_t) - T'(x_0)\|. \end{aligned}$$

Since T' is continuous at x_0 , $\|(x_0 + g_t) - x_0\| = \|g_t\| < r$, implying

$\|T'(x_0 + g_t) - T'(x_0)\| < \epsilon$. Therefore,

$$\|L_n - L_{n-1}\| \leq \|A^{-1}\| \|g_{n-1} - g_{n-2}\| \epsilon. \quad (13.39)$$

We have already shown $\|g_n - g_{n-1}\| \leq 2 \|L_n - L_{n-1}\|$, implying

$$\|g_n - g_{n-1}\| \leq 2 \|A^{-1}\| \|g_{n-1} - g_{n-2}\| \epsilon. \quad (13.40)$$

Thus, for ϵ sufficiently small $\left(\epsilon = \frac{1}{4 \|A^{-1}\|}\right)$,

$$\|g_n - g_{n-1}\| \leq \frac{1}{2} \|g_{n-1} - g_{n-2}\|. \quad (13.41)$$

Therefore,

$$\begin{aligned} \|g_n\| &= \|g_1 + (g_2 - g_1) + \cdots + (g_n - g_{n-1})\| \\ &\leq \|g_1\| + \|g_2 - g_1\| + \cdots + \|g_n - g_{n-1}\| \\ &= \|g_1\| + \frac{1}{2} \|g_1\| + \cdots + \frac{1}{2^{n-1}} \|g_1\| \\ &= \left(1 + \frac{1}{2} + \cdots + \frac{1}{2^{n-1}}\right) \|g_1\| \\ &= 2 \|g_1\| \\ &< 2 \left(\frac{r'}{2}\right) \\ &= r'. \end{aligned} \quad (13.42)$$

This completes the induction, we conclude $\|g_n\| < r'$ for all n . Additionally,

$$\|g_n - g_{n-1}\| \leq \frac{1}{2} \|g_{n-1} - g_{n-2}\| \quad (13.43)$$

for all n , implying g_n is Cauchy. Thus, $g_n \rightarrow g$. Similarly, we have

$$\|L_n - L_{n-1}\| \leq \|A^{-1}\| \|g_{n-1} - g_{n-2}\| \epsilon. \quad (13.44)$$

Hence, $L_n \rightarrow L$. Therefore,

$$L_n = L_{n-1} + A^{-1}(y - T(x_0 + g_n)) \quad (13.45)$$

converges to (by the continuity of A^{-1} and T)

$$\begin{aligned} L &= L + A^{-1}(y - T(x_0 + g)) \\ \implies 0 &= A^{-1}(y - T(x_0 + g)). \end{aligned} \quad (13.46)$$

Thus, by the injectivity of A^{-1}

$$0 = y - T(x_0 + g). \quad (13.47)$$

We conclude $T(x_0 + g) = y$. Hence, there exists $N'(y_0)$ such that the image of $T(\{x : \|x - x_0\| < r'\})$ contains $N'(y_0)$, where $\{x : \|x - x_0\| < r'\} \subset D$. Note $x_0 \in D$. Thus, there exists an open set in D containing x_0 such that T maps this set onto $N'(y_0)$. \square

13.5 Proof of Lagrange Multiplier Theorem

We provide a proof of Theorem 9.1.4. First note the following (see [AMR, Theorem 3.5.2(ii)]):

Theorem 13.5.1. (*Local Submersion Theorem*) *Let M and N be manifolds and $f : M \rightarrow N$ be of class C^r , $r \geq 1$. Suppose Tf restricted to the fiber $T_m M$ is surjective to $T_{f(m)} N$. Then, if in addition, the kernel $\ker(T_m f)$ is split in $T_m M$ there are charts (U, ϕ) and (V, ψ) with $m \in U$, $f(U) \subset V$, $\phi : U \rightarrow U' \times V'$, $\phi(m) = (\mathbf{0}, \mathbf{0})$, $\psi : V \rightarrow V'$ and $f_{\phi\psi} : U' \times V' \rightarrow V'$ is the projection onto the second factor.*

The Local Submersion Theorem is used to prove the following:

Theorem 13.5.2. *Let $g : M \rightarrow P$ be a C^1 submersion, $N = g^{-1}(p_0)$ and let $f : M \rightarrow \mathbb{R}$ be C^1 . A point $n \in N$ is a critical point of $f|_N$ if and only if there exists $\lambda \in (T_{p_0} P)^*$, called a Lagrange Multiplier, such that $T_n f = \lambda \circ T_n g$.*

Proof. \Rightarrow : Assume $n \in N$ is a critical point of $f|_N$. We first prove a local version of the theorem. Let $E \subset M$ and $F \subset P$ be open. Since g is a submersion and is C^1 , by

Theorem 13.5.1 there exists charts $(U, \phi), (V, \psi)$

$$\phi : U \rightarrow U_1 \times V_1 \subset E \times F \quad (13.48)$$

$$\psi : V \rightarrow U_1 \subset E$$

with U_1 open in E , V_1 open in F , $g(U) \subset V$, $n \in U$, $\phi(U \cap N) = \{0\} \times V_1$, $\phi(n) = (0, 0)$ and $p_0 \in V$ (since $N = g^{-1}(p_0)$ and $g(U) \subset V$), $\psi(p_0) = 0$ such that

$$g_{\phi\psi}(x, y) = x \quad (13.49)$$

for all $(x, y) \in U_1 \times V_1$. Next, define $f_\phi = f \circ \phi^{-1} : U_1 \times V_1 \rightarrow \mathbb{R}$. Note that since f is differentiable, we have

$$D_2 f_\phi(0, 0) \cdot e = \underbrace{D_2 f(n)}_0 \cdot e = 0 \quad (13.50)$$

for all $e \in E$ since $T_n f \Big|_{T_n N} = 0$. Then, define $\mu = D_1 f_\phi(0, 0)$ and let $e \in E$, $d \in F$, we get

$$\begin{aligned} Df_\phi(0, 0) \cdot (e, d) &= D_1 f_\phi(0, 0)e + \underbrace{D_2 f_\phi(0, 0)}_0 d \\ &= \mu(e) \end{aligned} \quad (13.51)$$

$$= (\mu \circ Dg)(0, 0) \cdot (e, d)$$

where the last line follows from $D_1 g_{\phi\psi}(x, y) = 1$ and $D_2 g_{\phi\psi}(x, y) = 0$. Thus,

$$Df_\phi(0, 0) = (\mu \circ Dg)(0, 0) \quad (13.52)$$

To get a global result, we pull back to M and P , let $\lambda = \mu \circ T_{p_0} \psi \in (T_{p_0} P)^*$. Then composing each side of $Df_\phi(0, 0) = (\mu \circ Dg)(0, 0)$ from the right by $T_n \phi$ gives the desired result. \square

13.6 Miscellaneous

Proposition 13.6.1. *Let $Y_i, i = 1, \dots, n$ be a collection of Banach spaces with norm $\|x\|_{Y_i}$. Then the space $X = \bigcap_{i=1}^n Y_i$ with norm $\|x\|_X = \max_{1 \leq i \leq n} \|x\|_{Y_i}$ is a Banach space.*

Proof. First we verify $\|x\|_X$ is norm:

1. We first show that $\|x\|_X = 0$ if and only if $x = 0$. Suppose $\|x\|_X = 0$. Then $\|x\|_{Y_i} = 0$ for all i . Since $\|x\|_{Y_i}$ is already a norm by hypothesis, this implies $x = 0$.

Next, suppose $x = 0$. Then $\|x\|_{Y_i} = 0$ for all i . It follows $\|x\|_X = 0$.

2. Next, letting $\alpha \in \mathbb{R}$, we have

$$\begin{aligned}
 \|\alpha x\|_X &= \max_i \|\alpha x\|_{Y_i} \\
 &= \max_i |\alpha| \|x\|_{Y_i} \\
 &= |\alpha| \max_i \|x\|_{Y_i} \\
 &= |\alpha| \|x\|_X.
 \end{aligned} \tag{13.53}$$

3. Last, we verify the triangle inequality:

$$\begin{aligned}
 \|x + y\|_X &= \max_i \|x + y\|_{Y_i} \\
 &\leq \max_i \|x\|_{Y_i} + \|y\|_{Y_i} \\
 &= \max_i \|x\|_{Y_i} + \max_i \|y\|_{Y_i} \\
 &= \|x\|_X + \|y\|_X.
 \end{aligned} \tag{13.54}$$

It remains to show that X is complete. Let x_n be a sequence in X , converging to x . Our goal is to show that $x \in X$. The sequence x_n is also a sequence in each Y_i . Since each Y_i is a Banach space, this implies x_n converges to $x \in Y_i$ for all i . Thus, $x \in X$, i.e. x_n converges to an $x \in X$. We conclude X is a Banach space under the norm

$$\|x\|_X = \max_{1 \leq i \leq n} \|x\|_{Y_i}. \quad \square$$

Lemma 13.6.2.

$$X \cap \bigcup_{\alpha} Y_{\alpha} = \bigcup_{\alpha} X \cap Y_{\alpha} \tag{13.55}$$

Proof. First, let $x \in X \cap \bigcup_{\alpha} Y_{\alpha}$. Then $x \in X$ and $x \in Y_{\alpha_0}$ for some α_0 . Therefore, $x \in X \cap Y_{\alpha_0}$. We conclude $x \in \bigcup_{\alpha} X \cap Y_{\alpha}$.

Next, let $x \in \bigcup_{\alpha} X \cap Y_{\alpha}$. Then $x \in X \cap Y_{\alpha_0}$ for some α_0 . In particular, $x \in Y_{\alpha_0}$. Thus, $x \in X \cap \bigcup_{\alpha} Y_{\alpha}$. \square

Proposition 13.6.3. *Let X be a normed vector space and X_0 be a closed subspace.*

Let τ_{X_0} denote the norm topology on X_0 and τ_r denote the subspace topology on X_0 .

Then $\tau_{X_0} = \tau_r$.

Proof. Let $B_{\delta}^0(x)$, an open ball of radius $\delta > 0$ centered at x , be an element of τ_{X_0} .

Let τ_X denote the norm topology on X and note $B_{\delta}(x) \in \tau_X$. Then

$$B_{\delta}^0(x) = X_0 \cap B_{\delta}(x). \quad (13.56)$$

Thus, $B_{\delta}^0(x) \in \tau_r$.

Next, let $B \in \tau_r$. Then $B = X_0 \cap A$ where $A \in \tau_X$. Then A can be written as the union of open balls in τ_X , $A = \bigcup_x B_{\delta_x}(x)$, where $B_{\delta_x}(x)$ is a ball of radius δ_x centered at x . Therefore

$$\begin{aligned} B &= X_0 \cap \bigcup_x B_{\delta_x}(x) \\ &= \bigcup_{\alpha} X_0 \cap B_{\delta_x}(x) \end{aligned} \quad (13.57)$$

where the second line follows from Lemma 13.6.2. Since unions of open sets are open, it suffices to show $X_0 \cap B_{\delta_x}(x)$ is open. However,

$$\begin{aligned} X_0 \cap B_{\delta_x}(x) &= \{y \in X : y \in X_0 \text{ and } \|x - y\|_X < \delta\} \\ &= \{y \in X_0 : \|x - y\|_X < \delta\} \end{aligned} \quad (13.58)$$

which is an open set in X_0 . Thus, $X_0 \cap B_{\delta_x}(x)$ is open. \square

Bibliography

- [AMR] Abraham, R., Marsden, J., Ratiu, T.: *Manifolds, Tensor Analysis, and Applications*. Addison-Wesley, (1983)
- [ADM] Ambrosio, L., Da Prato, G., Mennucci, A.: *Introduction to Measure Theory and Integration*. Edizioni Della Normale, (2011)
- [AJVLS] Ay, N., Jost, J., Van Le, H., Schwachhöfer, L.: *Information Geometry*. Springer, (2017)
- [BAR] Balian, R., Alhassid, Y., Reinhardt, H.: *Dissipation in many-body systems: a geometric approach based on information theory*. Phys. Rept. 131, 1–46 (1986)
- [BKM] Barlow, H.B.; Kaushal, T.P., Mitchison, G.J.: *Finding Minimum Entropy Codes*. Neural Computation 1, 412–423 (1989)
- [BCM] Bot, R.I., Csetnek, E.R., Moldovan, A.: *Revisiting Some Duality Theorems via the Quasirelative Interior in Convex Optimization*. J. Optim. Theory Appl. 139, 67–84 (2008)
- [CP] Cena, A., Pistone, G.: *Exponential Statistical Manifold* Springer, (2007)

- [CG] Comets, F., Gidas, B.: *Parameter estimation for Gibbs distributions from partially observed data*. The Annals of Applied Probability 2 (1), 142–70 (1992)
- [D] Dereudre, D.: *Variational Principle for Gibbs point processes with finite range interaction*. Electron. Commun. Probab. 21 (10), 1–11 (2016)
- [DL] Dereudre, D., Lavancier, F.: *Consistency of likelihood estimation for Gibbs point processes*. The Annals of Statistics 45 (2), 744–770 (2017)
- [DHX] Dostoglou, S., Hughes, A., Xue, Jianfei: *On entropy minimization and convergence*. J. Stat. Phys. 177, no. 3, 485-505. (2019)
- [DM] Drabek, P., Milota, J.: *Methods of Nonlinear Analysis: Applications to Differential Equations, Second Edition*. Birkhauser, (2013)
- [ET] Ekeland, I., Temam, R.: *Convex Analysis and Variational Problems*. Studies in Mathematics and its Applications, (1973)
- [F] Folland, G.: *Real Analysis Modern Techniques and Their Applications, Second Edition*. Wiley Inter-Science, (1999)
- [G] Georgii, H.-O.: *The Equivalence of Ensembles for Classical Systems of Particles*. Journal of Statistical Physics 80, (5/6), 1341–1378 (1995)
- [Gibbs] Gibbs, J.W.: *Elementary Principles In Statistical Mechanics*. Ox Bow Press, Connecticut (1981)

- [Gi] Gidas, B.: *Parameter Estimation for Gibbs Distributions from Fully Observed Data*. In: Chellappa, R., Jain, A. (eds.) *Markov Random Fields, Theory and Application*, pp. 471–498. Academic Press (1993)
- [GR] Griffiths, R.B., Ruelle, D.: *Strict convexity (“continuity”) of the pressure in lattice systems*. *Communications in Mathematical Physics* 23 (3), 169–75 (1971)
- [J] Jaynes, E.T.: *Information Theory and Statistical Mechanics*. In: *Statistical Physics. Brandeis Lectures*, vol. 3. Benjamin (1963)
- [Kh] Khinchin, A.I.: *Mathematical Foundations of Statistical Mechanics*. Dover (1949)
- [K] Kullback, S.: *Certain inequalities in information theory and the Cramer-Rao inequality*. *The Annals of Mathematical Statistics* 25 (4), 745–751 (1954)
- [Kull] Kullback, S.: *Information Theory and Statistics*. Dover (1968)
- [L] Lanford, O.E.: *Entropy and equilibrium states in classical statistical mechanics*. In: *Statistical mechanics and mathematical problems*, pp. 1–113. Springer (1973)
- [LR] Lanford, O.E., Ruedin, L.: *Statistical mechanical methods and continued fractions*. *Helvetica Physica Acta* 69(5), 908–948 (1996)

- [LP] Lebowitz, J.L., Penrose, O.: *Analytic and clustering properties of thermodynamic functions and distribution functions for classical lattice and continuum systems*. *Comm. Math. Phys.* 11, 99–124 (1968/1969)
- [Luen] Luenberger, D.G.: *Optimization by Vector Space Methods*. John Wiley, (1969)
- [M-L] Martin-Lof, A.: *Statistical mechanics and the foundations of thermodynamics*. Springer (1979)
- [OVY] Olla, S., Varadhan, S.R., Yau, H.T.: *Hydrodynamical limit for a Hamiltonian system with weak noise*. *Communications in Mathematical Physics*. 155 (3) 523–60 (1993)
- [PS] Pistone, G., Sempi, C.: *An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one*. *The Annals of Statistics* 23 (5), 1543-61 (1995)
- [PIST] Pistone, G.: *Examples of the Application of Nonparametric Information Geometry to Statistical Physics*. *Entropy*, (2013)
- [P] Preston, C.: *Random fields*. *Lecture Notes in Math*. Vol. 534. Springer (1976)
- [RR] Rao, M.M., Ren, Z.D.. *Theory of Orlicz Spaces*. Dekker, (1991)
- [RP] Rechtman, R., Penrose, O.: *Continuity of the Temperature and Derivation of the Gibbs Canonical Distribution in Classical Statistical Mechanics*. *Journal of Statistical Physics*, 19 (4), 359–366 (1978)

- [RV] Roberts, A.W., Varberg, D.E.: *Convex Functions*. Academic Press (1973)
- [Rock] Rockafellar, R.T.: *Convex analysis*. Princeton University press (1970)
- [RW] Rockafellar, R.T., Wets, R.: *Variational Analysis*. Springer (1998)
- [R] Ruelle, D.: *Statistical Mechanics; Rigorous results*. Benjamin (1969)
- [S] Scheffe, H. *A Useful Convergence Theorem for Probability Distributions*.
The Annals of Mathematics Statistics, (1947)
- [V] Varadhan, S.R.S.: *Scaling Limits for Interaction Diffusions*. Comm. Math. Phys. 135, 313–353 (1991)
- [WJ] Wainwright, M.J., Jordan, M.I.: *Graphical Models, Exponential Families, and Variational Inference*. Foundations and Trends in Machine Learning 1 (1–2), 1–305, (2008)
- [X] Xue, Jianfei : *On Hydrodynamic Equations and Their Relation to Kinetic Theory and Statistical Mechanics*. University of Missouri Ph.D. Thesis (2017)
- [Zei] Zeidler, E.: *Nonlinear Functional Analysis and its Applications III*. Springer (1985)
- [Z] Zubarev, D.N., Morozov, V., Ropke, G.: *Statistical Mechanics of Nonequilibrium Processes*. Vol. 1. Akademie Verlag (1996)

VITA

Alexander M. Hughes was born January, 1991 in Ames, Iowa. In 2013 he graduated from Rensselaer Polytechnic Institute with Bachelor of Science degrees in Mathematics and Physics. In 2014 he graduated with a Master's of Applied Mathematics degree from Northwestern University. He then continued his education at University of Missouri, while also participating as a Program Associate at the *Hamiltonian systems, from topology to applications through analysis* program at the Mathematical Sciences Research Institute in Berkeley, California from August to December 2018. This position was supported by an NSA grant. He was also awarded the Michael Taksar Scholarship for Graduate Students in Probability in May 2019 from the University of Missouri, where he plans on graduating with a Ph.D. in Mathematics.