Uniqueness of Gibbs Measures

Stein Andreas Bethuelsen

Department of Mathematics

Utrecht University

Supervisor: Roberto Fernández, Utrecht University
Co-reader: Karma Dajani, Utrecht University

A thesis submitted for the degree of

Master of Science

July 4, 2012

Preface

This master's thesis was written during the last semester of my master of science degree at Utrecht University. It marks the end of the two year master program ”Mathematical Sciences”.

The work has been interesting and rewarding and has provided challenges both in theoretical and practical work. By being part of the probability group of Prof. Roberto Fernandez at Utrecht University I have experienced on close range the frontier of knowledge, which has been highly stimulating and instructive. Moreover, by conducting independent research I have learned more than I could ever imagine, and which have inspired me to proceed further for an academic carrier. It is therefore with much enthusiasm I am looking forward to continue conducting research in mathematical statistical mechanics and related fields.

There are many people who have helped me in the process of writing this thesis, and I would like to extend my deep gratitude to you all. In particular, I would like to thank my advisor Roberto Fernandez at Utrecht University for introducing me to this fascinating subject and for all his support while working on this project. Special thanks also goes to Siamaak Taati for fruitful discussions and for showing such interest in my work.

Stein Andreas Bethuelsen
Utrecht, June 2012

Abstract

We study various conditions for uniqueness of Gibbs measures on a general Polish space embedded on a countable lattice. In particular, we study the Dobrushin uniqueness condition in its original version via two dual techniques: by a coupling argument and by averaging over observables. These are shown to be equivalent. Furthermore, we state the Dobrushin uniqueness condition in a more general form essentially requiring that the interaction matrix is contractive in any norm. This form also implies a dual condition to the original Dobrushin condition, and gives a way to compare different Gibbs measures.

Following these general ideas, we improve the Dobrushin condition to yield for any cover of the lattice. By a simple trick the new approach is shown to extend the known Dobrushin-Shlosman uniqueness condition beyond translation invariance and finite range. This again gives a positive answer to the question whether or not a condition obtained by Lieb and Aizenman by averaging methods is equivalent to the Dobrushin-Shlosman condition. Moreover, the proof is given by a dusting interpretation as invented by Aizenman and is seen to be relatively simple.

Lastly, our new condition is compared to other existing Dobrushin-alike uniqueness conditions. They are seen to be more or less equivalent, though our new condition extends the previous known ones beyond finite range. Moreover, the conditions are proven by dual methods, and thus our new results are seen to yield an important contribution to the question as to whether coupling techniques are superior averaging techniques.

Contents

Chapter 1

Introduction

"A Gibbs measure is a mathematical idealization of an equilibrium state of a physical system which consist of a very large number of interacting components. In the language of Probability Theory, a Gibbs measure is simply the distribution of a stochastic process which, instead of being induced by time, is parametriced by the sites of a spatial lattice, and has the special feature of admitting prescribed versions of the conditional distributions with respect to the configuration outside finite regions."[1]

A central issue in statistical mechanics is the co-existence of several phases and the phenomena of phase transitions. In mathematical terms a co-existence of several phases is described by the existence of several Gibbs measures. Hence, uniqueness of Gibbs measures in the setup of statistical mechanics implies that no such "strange" phenomenon is happening, and the focus of this thesis will be conditions implying such lack of drama. We would like to emphasize that in the following presentation little or no knowledge of statistical mechanics is needed, however much of the jargon has for historical reasons its origin from physics. The setup is more general and is applicable to various fields, spanning from Combinatorial Optimization and Computer Science,[2] to Differential Geometry [3].

1.1 Project description

The basis for the thesis is a uniqueness condition first presented by R. L. Dobrushin in [6] and further generalized in [7] by the same author. By estimating how a flip of the spin at a site y affects the spin value at x, as quantified by the number $\alpha_{x,y}$, one can show uniqueness. The condition is satisfied if the total effect of flipping the spins in the neighborhood of x is small, essentially saying that a site only depends weakly on its neighboring sites. This is quantified by demanding $\sup_x \sum_y \alpha_{x,y} < 1$.

The original condition was proven via coupling methods and a so-called Surgery argument. However, later, different approaches appeared. In particular a dual approach via averaging over observables (functions), as for instance seen in [22] by H. Föllmer. One of the central issues of this thesis is to clarify and compare these two approaches, as we do in chapter 4. The theory illuminates in the Theorem 4.4.1 where we show that the absence of several phases is implied. This holds as long as we are working on a Polish spin space S and the interaction matrix consisting of the $\alpha_{x,y}$ estimates is less then 1 in any metric norm.

[1] From [23], p. 5
[2] For instance the papers [42] and [43]
[3] See for instance [28]

Another central issue of study is the relation between the so-called Dobrushin-Shlosman uniqueness condition and a dual condition by Lieb and Aizenman, which were claimed to be equivalent. The condition by Dobrushin and Shlosman was presented in [10] as a generalization of the Dobrushin condition to larger volumes, however restricted to translation-invariant models of finite range. The condition was only shown via coupling methods, while the Lieb-Aizenman condition was written in a dual description. Seemingly no published version of this condition appeared other than a version for the stochastic Ising model, as seen in [1] by M. Aizenman and R. Holley. On the other hand, we were in possession of some lecture notes by R. Fernandez, [19], discussing this condition in more generality. An in-depth comparison of the two conditions has, however, to our knowledge never been put forward, and one of our main tasks is to study their relation.

This is mainly done in chapter 5 where we show that they are essentially equivalent. While working on these issues we discovered that the condition could be pushed even further. By simple methods the existing conditions are extended to hold even beyond translation-invariance and finite range. Moreover, all of our techniques yield for general quasilocal systems, assuming S to be a Polish space only. The most general statement is given in Theorem 5.4.2, which as for the Dobrushin condition states that for an improved version of the interaction matrix being less then 1 in any metric norm implies uniqueness.

In the paper [42] by D. Weitz some new conditions of similar character were given, seemingly improving the previous known ones. In chapter 6 we introduce his conditions and compare them with the conditions shown in chapter 5. As summarized in the tables 6.1 and 6.2 the conditions by Weitz are shown to yield better estimates then already known condition, however only for Markovian systems. On the other hand, comparing Weitz's conditions with our new condition it is seen that they are more or less equivalent. However, our condition extends the conditions beyond finite range and is written via a simple statement, unifying the two conditions given by Weitz. Being proven via dual methods, our new condition is also seen to yield a important contribution to whether coupling methods are superior to the dual methods by averaging over observables, a question raised by A. Sokal in [35].

1.2 Preliminaries

The following paper demands a good understanding of measure theoretic probability theory, for instance as presented in the Mastermath course ”MTP”[4]. Some familiarity with topics in topology and functional analysis is also recommended, but not essential. We refer to the book by [14] for more on the notions of such character mention in this paper. The theory on Gibbs measures needed for our purpose is put forward in chapter 2 and chapter 3, where the duality between the coupling method and the averaging method also is introduced.

[4] See the notes [36]

Chapter 2

Gibbs Measures

In this chapter we introduce the concept of a Gibbs Measure as it was formalized by Dobrushin, Ruelle and Lanlord via the DLR equations in the late 60's (see for instance [6] for one of the very first papers on this approach). We go through the basics of what we will consider as a mathematical model of statistical mechanics and much of the notation that we use later is for the first time introduced in this chapter. We do not, however, attempt to give a full overview of the theory of Statistical Mechanics, but more to serve a foundation of what follows. For a nice introduction with some emphasis on the physical theory we recommend the first chapter of the book [4] by A. Bovier. This chapter relies extensively on the paper [18] by A.C.D. van Enter, R. Fernandez and A. D. Sokal. Other references we have used are mainly the standard books [23] by H.O. Georgii and [34] by B. Simon and the notes [21] by R. Fernandez.

2.1 The General Framework

The mathematical theory of classical statistical mechanics in equilibrium as seen as a branch of probability theory is, as commented in [18], in general based on the following four ingredients:

- A configuration space Ω.
- A σ-algebra $\mathscr{F}$ of subsets of Ω.
- Observables, that is, real-valued $\mathscr{F}$-measurable functions on Ω.
- A probability measure on μ on $(\Omega, \mathscr{F})$

In this paper, however, we will restrict ourselves to models defined on a countable infinite lattice.

Definition 2.1.1. *A mathematical model of statistical mechanics is a system characterized by* $(G,(S,\mathscr{T}_S,\mathscr{F}_S),\Pi)$ *where*

- $G = (\mathbb{L}, E)$ *is a countable locally finite graph with vertex set* $\mathbb{L}$ *and edge set* E*. For historical reasons, elements of* $\mathbb{L}$ *are often referred to as sites. Induced on this graph we consider the graph-distance between sites given by* $d(x,y) = \inf\{|r| \, r \text{ is a path from } x \text{ to } y\}$.
- $(S,\mathscr{T}_S)$ *is a Polish space, i.e. the topology* $\mathscr{T}_S$ *is metrizable by some metric* ρ *for which S is separable and complete. Often, however, we will restrict us to S being in addition compact. The elements in S are called spins, and S the single spin space.* $\mathscr{F}_S$ *is the corresponding Borel-*σ*-algebra, that is, generated by the open sets.*

- Π *is a specification (see definition 2.1.3 below).*

Based on the triple $(G,(S,\mathscr{T}_S),\Pi)$, we construct $\Omega := S^{\mathbb{L}}$, called the configuration space, and equip it with the product topology $\mathscr{T} := \prod_{x\in\mathbb{L}} \mathscr{T}_S$ and the product σ-algebra $\mathscr{F} := \prod_{x\in\mathbb{L}} \mathscr{F}_S$. As S is a Polish space by construction so is Ω. When writing $\omega_\Lambda\sigma^c_\Lambda$ we mean the configuration which equals ω on Λ and σ outside Λ. The phrase "$\sigma = \omega$ off y" means two configurations $\sigma, \omega \in \Omega$ such that $\sigma_x = \omega_x$ for all $x \in \mathbb{L}\setminus\{y\}$. We define $\mathscr{L}$ as the set of finite subset of $\mathbb{L}$, and for each $\Lambda \in \mathscr{L}$ we write $\mathscr{F}_\Lambda \subset \mathscr{F}$ for the sub-σ-algebra corresponding to events measurable within the subset Λ. Elements $A \in \mathscr{F}_\Lambda$, or measurable functions $f \in \mathscr{F}_\Lambda$, will be called local when Λ is finite. To define what we mean by a specification Π we first need to introduce the concept of a probability kernel.

Definition 2.1.2. [1] *A probability kernel from a probability space $(\Omega,\mathscr{F})$ to another probability space $(\Omega',\mathscr{F}')$ is a function $\pi(\cdot|\cdot) : \mathscr{F}'\times\Omega \mapsto [0,1]$ such that*

1. *$\pi(\cdot|\omega)$ is a probability measure on $(\Omega',\mathscr{F}')$ for each $\omega\in\Omega$.*
2. *$\pi(A'|\cdot)$ is $\mathscr{F}$-measurable for each $A'\in\mathscr{F}'$.*

Definition 2.1.3. [2] *A specification is a family of probability kernels $\pi_\Lambda(\cdot|\cdot) : \mathscr{F}\times\Omega\mapsto[0,1]$ from $(\Omega,\mathscr{F})$ to itself defined for each finite $\Lambda\subset\mathscr{L}$, such that*

i) *for each $A\in\mathscr{F}$, the function $\pi_\Lambda(A|\cdot)$ is a $\mathscr{F}_{\Lambda^c}$-measurable function.*

ii) *π_Λ is $\mathscr{F}_{\Lambda^c}$-proper, i.e. for each $B\in\mathscr{F}_{\Lambda^c}$, $\pi_\Lambda(B|\omega) = 1_B(\omega)$.*

iii) *If $\Delta\subset\Lambda$, then $\pi_\Lambda\pi_\Delta = \pi_\Lambda$.*

By the third property a specification is said to be consistent. Written out it says that for any measurable function f,

$$\pi^\sigma_\Lambda f = \pi_\Lambda(f|\sigma) = \int f(\eta_\Lambda\sigma_{\Lambda^c})\pi_\Lambda(d\eta|\sigma) = \int\int f(\tau_\Delta\eta_{\Lambda\setminus\Delta}\sigma_{\Lambda^c})\pi_\Delta(d\tau|\eta_{\Lambda\setminus\Delta}\sigma_{\Lambda^c})\pi_\Lambda(d\eta|\sigma_{\Lambda^c}) \quad (2.1)$$

We will often assume that by a specification Π we are indirectly also determining the whole system $(G,(S,\mathscr{T}_S),\Pi)$. Given a specification we are now in position to define what we mean by a consistent measure.[3]

Definition 2.1.4. [4] *A probability measure μ on Ω is said to be consistent with the specification $\Pi = \{\pi_\Lambda\}_{\Lambda\in\mathscr{L}}$ if its conditional probabilities for finite subsystems are given by the $\{\pi_\Lambda\}_{\Lambda\subset\mathbb{L}}$, that is, for each $\Lambda\in\mathscr{L}$ and $A\in\mathscr{F}$*

$$\mathbb{E}_\mu(1_A|\mathscr{F}_{\Lambda^c}) = \pi_\Lambda(A|\cdot)\ \mu\text{-}a.e. \quad (2.2)$$

We denote by $\mathscr{G}(\Pi)$ the set of all measures consistent with Π.

Thus, by a specification we assign the conditional expectations of the global measures, $\mathscr{G}(\Pi)$. Or said the other way around, a specification Π defines the conditional expectations, and we are searching for measures consistent with these conditional expectations. These consistent measures can further be characterized as follows.

[1] Definition 3.1 in [21]

[2] Definition 2.5 in [18]

[3] In the title we referred to Gibbs measures, however we will mainly be focusing on consistent measures. Gibbs measures are defined later as a subclass of the consistent measures, but all our results will deal with the general term of consistent measures. In the existing literature, however, these terms are often confused, for instance in [23].

[4] Definition 2.6 in [18]

Proposition 2.1.1. [5] *Let $\Pi = \{\pi_\Lambda\}_{\Lambda\subset\mathbb{L}}$ be a specification, let μ be a probability measure on Ω and let $\Lambda \subset \mathbb{L}$. Then the following are equivalent:*

a) *For each $A \in \mathscr{F}$, $\mathbb{E}_\mu(1_A|\mathscr{F}_{\Lambda^c}) = \pi_\Lambda(A,|\cdot)$ μ-a.e*

b) *There exists a measure ν_{Λ^c} such that $\mu = \nu_{\Lambda^c}\pi_\Lambda$*

c) $\mu = \mu\pi_\Lambda$

Condition c), the so-called DLR-equations, is written in operator form and should be understood similarly as condition iii) in definition 2.1.3. These will be very important for us in the proceeding chapters. To study the class of consistent measures we need to equip $\mathscr{G}(\Pi)$ with a topology. We will in this paper use the standard notion of weak convergence.

Definition 2.1.5. [6] *We say that a sequence of measures μ_n converges weakly to a measure μ if $\mu_n(f) \to \mu(f)$ for every bounded continuous function f.*

As discussed in [18] (p. 897-898) there are several other options for the choice of topology. For instance in [23] the so-called "topology of local convergence" is used which is to say $\mu_n(f) \to \mu(f)$ for all bounded quasilocal functions.

Definition 2.1.6. *A function is said to be quasilocal if it is the uniform convergent limit of some sequence of local functions, i.e. functions depending on spins in a finite volume only. That is, for each $\varepsilon > 0$ there exists a local function f_ε such that*

$$||f_\varepsilon - f||_\infty < \varepsilon$$

As further commented in [18], when S is separable and metrizable, then the weak quasilocal topology coincides with the weak topology. Our choice of the weak topology is not, however, without reason. There follow nice properties with this choice, one being that the set of probability measures on Ω, $M_{+1}(\Omega)$, is separable and metrizable (respectively complete metrizable, compact metrizable) if and only if Ω is.

Definition 2.1.7. [7] *A specification $\Pi = \{\pi_\Lambda\}_{\Lambda\subset\mathbb{L}}$ is said be quasilocal if, for each $\Lambda \in \mathscr{L}$ and bounded, measurable and quasilocal functions f implies $\pi_\Lambda f$ is bounded and quasilocal.*

A measure μ on Ω is said to be quasilocal if there exists a quasilocal specification with which μ is consistent.

There follow one important property when assuming a specification is quasilocal. As commented in in [18] (p. 910), if Π is a quasilocal specification, then we can weaken the DLR-equation above slightly. Under quasilocality it is sufficient to have $\mu = \mu\pi_\Lambda$ only for $\mathscr{F}_\Lambda$. Thus a consistent measure can be specified by all local event only, which will be an very important property for us later. Often it is also convenient to assume that Π is continuous in a certain sense.

Definition 2.1.8. [8]*[Feller property] A specification $\Pi = \{\pi_\Lambda\}_{\Lambda\subset\mathbb{L}}$ is said to be Feller if, for each $\Lambda \in \mathscr{L}$,*

$$f \in C(\Omega) \implies \pi_\Lambda f \in C(\Omega) \tag{2.3}$$

[5] Proposition 2.7 in [18]
[6] Definition in section 9.3 in [14]
[7] Definition 2.9 and 2.13 in [18]
[8] Definition 2.14 in [18]

Proposition 2.1.2. [9] *Let $\Pi = \{\pi_\Lambda\}_{\Lambda\subset\mathbb{L}}$ be a Feller specification. Let $\{\Lambda_n\}_{n\geq 1}$ be an increasing sequence of finite volumes whose union is $\mathbb{L}$, and let $(\nu_n)_{n\geq 1}$ be an arbitrary sequence of probability measures on Ω. Let μ be any limit point (in the weak topology) of the sequence $\{\nu_n\pi_{\Lambda_n}\}_{n\geq 1}$. Then μ is consistent with Π. In particular, $\mathscr{G}(\Pi)$ is a closed subset of $M_{+1}(\Omega)$.*

Actually, it can be proven that $\mathscr{G}(\Pi)$ is a convex set and that its extreme measures are mutually singular[10]. If we moreover assume S being a compact metric space then the set $\mathscr{G}(\Pi)$ have some additional properties.

Proposition 2.1.3. [11] *Let Ω be a compact metric space, let Π be a Feller specification, and let μ be an extreme point of $\mathscr{G}(\Pi)$. Then, for μ-a.e. ω*

$$\lim_{n\to\infty}\pi^{\omega}_{\Lambda_n} = \mu \tag{2.4}$$

in the weak topology, where Λ_n is a sequence of finite subsets converging to $\mathbb{L}$.

A direct consequence of this theorem is the existence of a consistent measures.

Corollary 2.1.1. [12] *Let Ω be a compact metric space, and let $\Pi = \{\pi_\Lambda\}_{\Lambda\subset\mathbb{L}}$ be a Feller specification. Then $\mathscr{G}(\Pi)$ is nonempty.*

In chapter 4, when studying the Dobrushin uniqueness condition, we only use properties of the single site specification $\{\pi_{\{x\}}\}_{x\in\mathbb{L}}$. The following theorem gives a sufficient condition for when the single site specification uniquely defines the whole specification.

Theorem 2.1.1. [13] *Let $\nu \in M(S,\mathscr{F}_S)$, and suppose Π is a specification which enjoys the following property: For each $x\in\mathbb{L}$ there exists a measurable function $\rho_{\{x\}} > 0$ on Ω such that $\pi_{\{x\}} = \rho_{\{x\}}\pi_{\{x\}}$. Then*

$$\mathscr{G}(\Pi) = \{\mu \in M_{+1}(\Omega,\mathscr{F}) : \mu\pi_{\{x\}} = \mu \textit{ for all } x\in\mathbb{L}\} \tag{2.5}$$

2.2 Examples

It is a saying that theory without examples is useless. Luckily the framework of mathematical statistical mechanics contains lots of examples. In this section we introduce some of the simplest ones, the Ising Model and the Hard-Core Model. Thereafter we look at some more general examples in the form of Hamiltonians, and discuss the border of quasilocal specifications.

Example. i) The Ising Model

The basic model referred to in Mathematical Statistical Mechanics is the so-called Ising Model ([34], page 3), which is a simplified model for describing certain properties of ferromagnetic metals. In the simplest model we consider $S = \{-1,1\}$ embedded on the integer lattice $\mathbb{L} = \mathbb{Z}^d$, with $\Pi = \{\frac{1}{Z_\Lambda(\cdot)}(e^{-\beta H_\Lambda(\cdot|\cdot)})\}_{\Lambda\subset\mathbb{Z}^d})$, where

$$H_\Lambda(\omega|\sigma) = \sum_{y\sim x:x,y\in\Lambda}\omega_x\omega_y + \sum_{y\sim x:x\in\Lambda,y\notin\lambda}\sigma_y\omega_x \tag{2.6}$$

[9] Proposition 2.22 in [18]

[10] See chapter 7 in [23]

[11] Proposition 2.23 in [18]

[12] Proposition 2.21 in [18]. See also the theory leading to Corollary 1.10 in [4] for an convincing deduction of the existence result.

[13] Theorem 1.33 in [23]. See also [20] for throughout study of how to construct a specification from its singletons.

Above, $x \sim y$ means the sites that are nearest neighbors. $Z_V(\sigma) = \sum_{\omega \in \Omega_\Lambda} e^{-\beta H_\Lambda(\omega|\sigma)}$ is the normalization constant (the partition function), and the spin space $S = \{-1, 1\}$ is assigned the discrete topology. This model was introduced by W. Lenz in 1920, and solved by his PhD student E. Ising in 1925 for $d = 1$. He discovered that for the model only possess one consistent measure, and claimed that this would also hold for general $d > 1$. In 1933, however, Peierls gave a proof of the following theorem.

Theorem 2.2.1 (Non-uniqueness in Ising Model)**.** [14] *In the Ising model in dimension $d \geq 2$, for all $\beta > 0$ sufficiently large, the system has more then one consistent measure.*

Thus Isings claim was proven to be wrong, and the question whether a mathematical model of statistical mechanics can exhibit several consistent measures had gained its prime example. Based on this example the study of characterizing when a model exhibit one or more consistent measure was established. This example serve as motivation for our study in the proceeding chapters for conditions implying uniqueness of the set of consistent measures.

ii) The Hard-core Model[15]

Let again $\mathbb{L} = \mathbb{Z}^d$, and let $S = \{0, 1\}$. Furthermore, define the potentials $U_{\{x,y\}}(1,1) = \infty$ and $U_{\{x,y\}}(s,t) = 0$ otherwise, and $U_x(s) = -s \ln \lambda$, where λ is the activity parameter. Then, similar as for the Ising model we let the specification be given by $\Pi = \{\frac{1}{Z_\Lambda}(e^{-\beta U_\Lambda(\cdot|\cdot)})\}_{\Lambda \subset \mathbb{Z}^d})$ where we now let

$$U_\Lambda(\sigma) = \sum_{x \sim y:\, \{x,y\} \cap \Lambda \neq \emptyset} U_{x,y}(\sigma_x, \sigma_y) + \sum_{x \in \Lambda} U_x(\sigma_x) \tag{2.7}$$

The interpretation here is that a spin of 1 stands for an occupied site so that a configuration, $\omega \in \Omega$, specifies a subset of occupied sites. The infinite energy that the potential asigns to a pair of occupied sites means that there is a hard constraint forbidding two neighboring sites from both being occupied. This model serves as an simple example of a model not captured by the generality given in iii) below.

iii) General Hamiltonian

The Ising model and the Hard-core model can be generalized further by the notions of interactions and Hamiltonians.

Definition 2.2.1. [16] *An interaction is a family $\Phi = (\phi_A)_{A \in \mathscr{L}}$ of functions $\phi_A : \Omega \mapsto \mathbb{R}$ such that for each $A \in \mathscr{L}$, the function ϕ_A is $\mathscr{F}_A$-measurable*

Note that the definition does not allow infinite values. Therefore, a "hard-core interaction", e.g. as in ii), is not included. On the other hand, the Ising model is obviously included in this framework.

Definition 2.2.2. [17] *Let Φ be an interaction. Then, for each $\Lambda \in \mathscr{L}$, the Hamilitonian $H^\phi_{\Lambda,\tau}$ with boundary conditions τ is the function*

$$H^\phi_{\Lambda,\tau}(\omega) = \sum_{\Delta \in \mathscr{L},\, \Delta \cap \Lambda \neq \emptyset} \phi_\Delta(\omega_\Lambda \tau_{\Lambda^c}) \tag{2.8}$$

[14] See for instance section 1.3.2 and in particular Theorem 1.20 in [4] for a modern presentation of the proof based on Peierls arguments.

[15] Example 2 in [42]

[16] Definitions 2.1 in [18]

[17] Definition 2.3 in [18]

provided that this sum converges to a finite limit for all $\omega \in \Omega$, in which case it is called convergent.

The main class of specifications are the so-called Gibbs-specification.

Definition 2.2.3. [18] *Let $\mu^0 = \prod_{x\in\mathbb{L}} \mu_x^0$ be a product probability measure, and let Φ be a convergent, μ^0-admissible interaction, that is $Z_\Lambda^\Phi(\omega_{\Lambda^c}) \in (0,\infty)$ where*

$$Z_\Lambda^\Phi(\omega_{\Lambda^c}) = \int \exp[-H_\Lambda^\Phi(\omega)] \prod_{x\in\Lambda} d\mu_x^0(\omega_x) \tag{2.9}$$

Then the Gibbs-specification $\{\pi_\Lambda^\Phi(\cdot,\cdot)\}_{\Lambda\in\mathscr{L}}$ on $\mathscr{F}\times\Omega$ is defined by

$$\pi_\Lambda^\Phi(A,\omega) = \frac{1}{Z_\Lambda^c}\Phi(\omega_{\Lambda^c}) \int 1_A(\omega) \exp(-H_\Lambda^\Phi(\omega)] \prod_{x\in\Lambda} d\mu_x^0(\omega_x) \tag{2.10}$$

A Gibbs measure is a measure consistent with a Gibbs-specification. This terminology is often mixed with our more general notion of a consistent measures, as for instance seen in the standard reference book [23]. In the title we referred to the uniqueness of Gibbs measures, rather than the consistent measures. However, all our result in the proceeding chapters will be presented for consistent measures. To see that a Gibbs-specification indeed forms a specification follows by showing that they satisfy the DLR-equations (see p. 906 in [18]).

Theorem 2.2.2. [19] *Let Φ be a uniformly convergent interaction and μ^0-admissible interaction. Then the Gibbs-specification Π^Φ is quasilocal.*

Thus a large class of Gibbs-specifications serve as examples satisfying the property of being quasilocal. As further comment in [18], by allowing for local constraints, as in the case of the Hard-Core model, the quasilocality is not interrupted. See the examples and comments on p. 908 in [18] for a throughout discussion.

---

[18] Definition 2.8 in [18]
[19] theorem 2.10 in [18]

Chapter 3

The Wasserstein Metric

In the previous chapter we introduced the concept of consistent measures. In the following chapter we will look deeper into how we can compare these measures under the weak topology. More formally we will introduce a distance, the Wasserstein distance, which metrizes the space $\mathscr{G}(\Pi)$. This will equip us with vital tools for the following chapters, when investigating the unicity of the consistent measures. Our presentation here is heavily influence by the St. Flour lecture notes [40] by C. Villani. Other references which have been of great importance are the book [14] by Dudley, the paper [7] by R.L. Dobrushin, and his St. Flour lecture notes [8].

3.1 Optimal Transportation

Before introducing the Wasserstein distance we will give a short introduction to the field of Optimal Transportation, to serve as a motivation for later, and to put the Wasserstein distance in a larger context. For a more in depth study we strongly recommend the book [39] by C.Villani and his lecture notes [40].

Definition 3.1.1 (Coupling)**.** [1] *Let $(\mathscr{X},\mu)$ and $(\mathscr{Y},\nu)$ be two probability spaces. Coupling μ and ν means constructing two random variables X and Y on some probability space (Ω,P) such that the law of X equals μ and the law of Y equals ν. The couple (X,Y) is called a coupling of (μ,ν).*

Equivalently, in a measure theoretic formulation, a coupling is a measure π on $\mathscr{X}\times\mathscr{Y}$ with marginals μ, and ν respectively.

We will mostly use the latter definition, and will denote $K(\mu,\nu)$ as the set of couplings of the measures μ and ν. Firstly, it is easy to see that this set is non-empty, as we always have the trivial coupling given by $\pi(A,B)=\mu(A)\nu(B)$. However, the importance of coupling is when one can extract non-trivial information from it, and thus for most purposes we are in need of finding a better one. In what sense a coupling can be better then another depends on the system one is studying. In the field of Optimal Transportation, one is given a cost function $c(x,y)$ and the better coupling is the one which minimizes the cost of transporting mass from μ to ν. That is, we are interested in finding the coupling $\pi\in K(\mu,\nu)$ such that

$$\int_{\mathscr{X}\times\mathscr{Y}} c(x,y)\pi(dx,dy) \tag{3.1}$$

is minimized. As a motivation for the definition of an optimal coupling, Villani introduced in [40] the following modern interpretation.

[1] Definition 1.1 in [40]

"The question is how to distribute all the bread from the bakeries in Paris to its numerous cafes most efficiently. Consider a large number of bakeries, producing bread, that should be transported each morning to cafes where costumers will eat them. The amount of bread that can be produced at each bakery and the amount that will be consumed at each cafe is known in advance, and can be modeled as a probability measure on a space S (which in Villanis example is Paris). Moreover S is equipped with a natural metric ρ such that the distance between two points is the length of the shortest path joining them. Thus, Optimal Transportation serves to answer the question of finding the optimal solution to distribute the breads from the bakeries to the cafes with respect to minimizing the total distance travelled under the delivery."[2]

In most cases there does even exist an optimal plan of transporting.

Theorem 3.1.1 (Existence of an optimal coupling)**.** [3] *Let $(\mathscr{X},\mu)$ and $(\mathscr{Y},\nu)$ be two Polish probability spaces; let $a:\mathscr{X}\to\mathbb{R}\cup\{-\infty\}$ and $b:\mathscr{Y}\to\mathbb{R}\cup\{-\infty\}$ be two upper semicontinuous functions such that $a\in L^1(\mu)$, $b\in L^1(\nu)$. Let $c:\mathscr{X}\times\mathscr{Y}\to\mathbb{R}\cup\{+\infty\}$ be a lower semicontinuous cost function, such that $c(x,y)\geq a(x)+b(y)$ for all x,y. Then there is a coupling of (μ,ν) which minimizes the total cost $\mathbb{E}c(X,Y)$ among all possible couplings (X,Y).*

When introducing the Wasserstein distance in the next section we will consider only cost functions $c(x,y)$ representing a distance function. Hence we may set $a\equiv b\equiv 0$ to assure that there exists an optimal coupling. Remark that the theorem does only give an existence result, and so we may have several optimal coupling (which is most often the case).

The next theorem states that given two couplings we can construct a new coupling (a coupling of two couplings) which maintains the coupling property.

Theorem 3.1.2 (Gluing Lemma)**.** [4] *Let (X_i,μ_i), $i=1,2,3$ be a Polish probability space. If π_1 is a coupling of μ_1 and μ_2, and π_2 is a coupling of μ_2 and μ_3, then one can construct a measure π defined on $\mathscr{X}_1\times\mathscr{X}_2\times\mathscr{X}_3$ such that $\pi(\cdot,\cdot,\mathscr{X}_3)=\pi_1(\cdot,\cdot)$ and $\pi(\mathscr{X}_1,\cdot,\cdot)=\pi_2(\cdot,\cdot)$.*

3.2 The Wasserstein Metric

As mentioned in the previous section we will only be focusing on optimal couplings related to a distance function ρ.

Definition 3.2.1 (Wasserstein Distance)**.** [5] *Let $(\mathscr{X},\rho)$ be a Polish metric space, and let $p\in[1,\infty)$. For any two probability measures μ, ν on $\mathscr{X}$, the Wasserstein distance of order p between μ and ν is defined by the formula*

$$W_\rho^p(\mu,\nu)=(\inf_{\pi\in K(\mu,\nu)}\int_{\mathscr{X}}\rho(x,y)^p\pi(dx,dy))^{1/p} \tag{3.2}$$

The special case when $p=1$ will be referred to as the Wasserstein distance only and will denoted by $W_\rho(\mu,\nu)$.

The Wasserstein distance has seemingly been rediscovered through the history by many mathematicians and can thus be found under various names, for instance the Kantorovich distance, the

[2] p. 42 in [40]
[3] Theorem 4.1 in [40]
[4] ref. page 22 in [40]
[5] Definition 6.1 in [40]

Kantorovich-Rubenstein distance, and so on.[6]. We have chosen to name it the Wasserstein distance as this was the name R. L. Dobrushin used when introducing it to the field of statistical mechanics.[7] Actually, it is more the word "distance" one should question. The main purpose of this section is, however, to show that the Wasserstein distance indeed is a distance, and that it in addition forms a metric for the consistent measures under the weak topology.

Theorem 3.2.1. [8] *Let $(\mathscr{X}, \rho)$ be a Polish space, and let $p \in [1, \infty)$. Then the Wasserstein distance W_ρ^p satisfies the properties for being a metric.*

Proof.

- $W_\rho^p(\mu, \nu) = 0 \iff \mu = \nu$:

  Assume that $W_\rho^p(\mu, \nu) = 0$, then there exists a transference plan which is entirely concentrated on the diagonal $(y = x)$ in $X \times X$. Thus $\nu = \mu$.

  Similarly, if $\nu = \mu$ implies $W_\rho^p(\mu, \nu) = 0$, by taking the coupling which is entirely concentrated on $(y = x)$.

- $W_\rho^p(\mu, \nu) = W_\rho^p(\nu, \mu)$:

  This follows simply from the definition of the Wasserstein distance as the coupling is reflexive when integrating over a distance function ρ.

- $W_\rho^p(\mu, \nu) \leq W_\rho^p(\mu, \gamma) + W_\rho^p(\gamma, \nu)$:

  This follows from the Gluing Lemma and the existence of an optimal coupling (Lemma 3.1.2 and 3.1.1). Let P_1 be the optimal coupling of μ and γ and let P_2 be the optimal coupling of γ and ν. By the Gluing Lemma there exists a coupling $P_{1,2}$ of μ, γ and ν which preserves P_1 and P_2. Let P_3 be the measure remaining when integrating out γ in $P_{1,2}$, which by the Gluing Lemma also is a coupling. Then we have that

$$
\begin{aligned}
W_\rho^p(\mu, \nu) &\leq [\int_X \rho(x_1, x_3)^p P_3(dx_1, dx_3)]^{1/p} \\
&= [\int_X \rho(x_1, x_3)^p P_{1,2}(dx_1, dx_2, dx_3)]^{1/p} \\
&\leq [\int_X (\rho(x_1, x_2) + \rho(x_2, x_3))^p P_{1,2}(dx_1, dx_2, dx_3)]^{1/p} \\
&\leq [\int_X \rho(x_1, x_2)^p P_{1,2}(dx_1, dx_2, dx_3)]^{1/p} + [\int_X (\rho(x_2, x_3)^p P_{1,2}(dx_1, dx_2, dx_3)]^{1/p} \\
&= W_\rho^p(\mu, \gamma) + W_\rho^p(\gamma, \nu)
\end{aligned}
$$

  where we have made used of the metric inequality and the Minkowski inequality[9].

□

To state that W_ρ^p indeed is a metric, we in addition need that it is finite. For this we need to restrict to measures within a certain subclass.

[6]See the notes after chapter 6 in [40] for further discussions and references)

[7]See [7]

[8]See p. 105–106 in [40]

[9]Theorem 4.39 in [36]

Definition 3.2.2 (Wasserstein space)**.** [10] *Consider the space $M_{+1}(\mathscr{X})$ of Borel-Probability Measures on $\mathscr{X}$. Then we define the Wasserstein space of order p, $W_\rho^p(\mathscr{X}) \subset M_1$ as the set*

$$W_p(\mathscr{X}) := \{\mu \in M_{+1}(\mathscr{X}) : \exists x_0 \in \mathscr{X} \text{ with } \int_{\mathscr{X}} \rho(x_0,x)^p \mu(dx) < +\infty\} \tag{3.3}$$

The definition above seems to rely on the choice of $x_0 \in \mathscr{X}$, but this is only artificial. For any $x_1 \in \mathscr{X}$ we have that $\rho(x,x_1) \le \rho(x,x_0) + \rho(x_0,x_1)$. Hence if $\int_{\mathscr{X}} \rho(x_0,x)^p \mu(dx) < \infty$, then

$$\int_{\mathscr{X}} \rho(x_1,x)^p \mu(dx) \le \int_{\mathscr{X}} (\rho(x_0,x) + \rho(x_1,x_0))^p \mu(dx) < \infty \tag{3.4}$$

Also note that in case the distance is bounded, for instance $||\rho||_\infty \le 1$, then the Wasserstein space contains all the Borel-Probability Measures on $\mathscr{X}$.

Theorem 3.2.2 (The Wasserstein distance metrizices the weak topology)**.** [11] *Assume that the metric space (X,ρ) is Polish and $\mathscr{F}$ is the Borel-σ-algebra. Then the distance $W_\rho(\cdot,\cdot)$ forms a metric in the weak topology over the set of measures in the Wasserstein space $W_\rho^1(X)$. Moreover, the metric space $(W_\rho^1(X), W_\rho)$ is complete.*

The proof that the Wasserstein distance with $p = 1$ metrizes the weak topology follows from the characterization of weak topology in theorem 11.3.3 in [14] and equation 14.6 in [8]. In [40] the theorem is given in more generality for every $p \in [1,\infty)$. However, it is done by introducing an stronger form of weak convergence on the space W_ρ^p. To our knowledge there has so far not been any strengthen of results for proving the uniqueness by use of these methods with $p > 1$. Actually, as commented by Villani (remark 6.6. in [40]), $p \le q \implies W_\rho^p \le W_\rho^q$. In particular, the Wasserstein distance W_ρ is the weakest of all. As we do not tend to follow the lines of general p any longer, we leave it to the reader to investigate this further.

Remark. There are also various other metrics which metrizes the weak topology, see chapter 11 in [14] for a throughout account, and chapter 6 in [40] for a discussion of why one should use the Wasserstein distance.

Theorem 3.2.3 (Topology of the Wasserstein space)**.** [12] *Let $\mathscr{X}$ be a complete separable metric space. Then the Wasserstein space $W_\rho^1(\mathscr{X})$, metrized by the Wasserstein distance W_ρ, is also a complete separable metric space. If $\mathscr{X}$ is compact, then $W_p(\mathscr{X})$ is also compact. However, if $\mathscr{X}$ is only locally compact, then $W_\rho^1(\mathscr{X})$ is not locally compact.*

Remark. As shown in [7], theorem 2, the above theorem can be extended to separable metric spaces, and many of our results in the proceedings hold even in this generality. However, we mainly restrict to Polish spaces for conveniency, to assure the existence of an optimal coupling.

A major question is the calculation of the Wasserstein distance for given metric ρ. As it is given by the optimal coupling we are also in position to give upper bounds. Simply calculate the transportation cost via any given coupling. Hence the problem is often addressed to finding a coupling which is close enough to the optimal coupling. For improvements of the uniqueness estimates considered later, explicit knowledge of the Wasserstein distance could be essential. For a couple of examples one may look at chapter 14 in [7].

[10] Definition 6.4 in [40]

[11] Proposition 14.1 in [8]. See also [40], chapter 6, for statements for general p

[12] Theorem 6.18 in [40]

3.3 The Dual Metric

Recall from the definition of weak topology that two probability measure $\mu, \nu \in M_{+1}(X)$ are considered to be equal if $\mu(f) = \nu(f)$ for every f which is bounded and continuous. Thus another possible metric for the space of probability measure is the one given by

$$D(\mu, \nu) = \sup_{f \in C_b(X)} \frac{|\mu(f) - \nu(f)|}{||f||_\infty}$$

From the definition of weak topology it follows that $D(\mu, \nu) = 0$ if and only if $\mu = \nu$. The definition is obviously also symmetric. Moreover, since for any $\mu, \nu, \gamma \in M_{+1}(X)$ and any $f \in C(X)$,

$$|\mu(f) - \nu(f)| \leq |\mu(f) - \gamma(f)| + |\gamma(f) - \nu(f)|$$

it follows that it also satisfies the metric inequality. To compare this metric to the Wasserstein distance is it convenient to restrict the metric D to a subclass of function, namely the bounded Lipschitz functions.

Definition 3.3.1 (Lipshitz functions and norm)**.** [13] *Let (S,d) be a metric space. A function f from S to $\mathbb{R}$ is called Lipschitzian, or Lipschitz, if for some $K < \infty$*

$$|f(x) - f(y)| \leq Kd(x,y) = \text{ for all } x, y \in S \tag{3.5}$$

Moreover, the Lipschitzian norm of f, $||f||_L$ is defined as the smallest of such K, i.e $||f||_L = \sup_{x \neq y} \frac{|f(x)-f(y)|}{d(x,y)}$.

We write $L(\Omega)$ for the space of Lipschitzian function from Ω to $\mathbb{R}$, and for $\Delta \subset S$ we write $L(\Omega)_\Delta$ for the set of Lipschitzian functions only dependent on values in Δ. In most case it will be clear what Ω is and thus we often only write L and L_Δ respectively. The reason for introducing the Lipschitzian functions is that they are perfectly adjusted for metric spaces, and possess many nice properties. For instance we have the following nice relation with the Wasserstein distance, which eventually proves that D is a finite metric, at least for the measures in the Wasserstein space.

Proposition 3.3.1. *Let μ and ν be any two functions in $M_{+1}(X)$, where X is some separable metric space (S, ρ). Then*

$$D_\rho(\mu, \nu) \leq W_\rho(\mu, \nu) \tag{3.6}$$

Proof. For any function $f \in L(X)$ we have that

$$\begin{aligned}
D_\rho(\mu, \nu) &= \sup_{f \in L(X)} \frac{|\mu(f) - \nu(f)|}{||f||_L} \\
&= \sup_{f \in L(X)} \frac{1}{||f||_L} \left| \int (f(x) - f(y)) P(dx, dy) \right| \qquad \text{for any } P \in K(\mu, \nu) \\
&= \sup_{f \in L(X)} \frac{1}{||f||_L} \left| \int \frac{f(x) - f(y)}{\rho(x,y)} \rho(x,y) P(dx, dy) \right| \\
&\leq \sup_{f \in L(X)} \frac{1}{||f||_L} \int \left| \frac{f(x) - f(y)}{\rho(x,y)} \right| \rho(x,y) P(dx, dy) \\
&\leq \sup_{f \in L(X)} \int \rho(x,y) P(dx, dy)
\end{aligned}$$

Since this holds for any $P \in K(\mu, \nu)$, it also holds for the optimal one, which implies the theorem. □

[13] Section 6.1 in [14]

Actually, in most cases the inequality above is not strict.

Theorem 3.3.1 (The Strong Dual theorem). [14] *For any separable metric space $(\mathscr{X}, \rho)$ and any two probability measure $\mu, \nu \in W_1(\mathscr{X})$*

$$W_\rho(\mu, \nu) = D_\rho(\mu, \nu)$$

Definition 3.3.2 (The Dual Distance). *For any two probability measures μ and ν on a separable metric space $(\mathscr{X}, \rho)$, the Dual distance D_ρ is given by*

$$D_\rho(\mu, \nu) = \sup_{f:\,||f||_L \leq 1} \{|\mu(f) - \nu(f)|\} \tag{3.7}$$

A slightly different version of the Strong Dual Theorem is given in [40], generalizing the notion to the setup of optimal transportation (see theorem 5.10 and thereafter). On the other hand this theorem is not so explicit as stated above (Villanis statement is about one and a half page long), so we omit it. However, the following part is of importance even in our setup.

Theorem 3.3.2. [15] *Let $(\mathscr{X}, \rho)$ be a Polish space, and consider two measures in $\mu, \nu \in W^1_\rho(\mathscr{X})$. Then there exist a function obtaining the supremum in the definition of the dual distance. That is*

$$D(\mu, \nu) = \max_{f \in L} \frac{|\mu(f) - \nu(f)|}{||f||_L} \tag{3.8}$$

Thus, similar as Theorem 3.1.1, saying that there exists a optimal coupling, we also in the dual situation have the existence of an "optimal" function.

3.4 The Variational Metric

Another classical notion of distance between probability measures is the Total Variation.

Definition 3.4.1. [16] *Let $(\mathscr{X}, \mathscr{F})$ be a measurable space and let $\mu, \nu \in M_1(\mathscr{X}, \mathscr{F})$. The total variation distance between μ and ν is then given by*

$$||\mu - \nu||_{TV} := \sup_{A \in \mathscr{F}} |\mu(A) - \nu(A)| \tag{3.9}$$

This distance is widely used in Probability theory, mostly since it often is easy to produce bounds. Here are some equivalent definitions of the Total Variation, as shown in [23].

Proposition 3.4.1. [17] *Given a measurable space $(\mathscr{X}, \mathscr{F})$ and two probability measures $\mu, \nu \in M_1(\mathscr{X}, \mathscr{F})$. Then the following are equivalent notions of their total variation distance*

i) $\sup_{A \in \mathscr{F}} |\mu(A) - \nu(A)|$

ii) $\sup\{\frac{|\mu(f) - \nu(f)|}{\delta(f)} \,|\, f \in \mathscr{F}$ *is bounded*$\}$, *where $f \in \mathscr{F}$ means that it is measurable, and* $\delta(f) = \sup_x f(x) - \inf_x f(x)$

[14] Theorem 11.8.2 in [14]

[15] See theorem 5.10iii) in [40]

[16] See page 141 in [23]. Note that some use $2TV(\mu, \nu)$ as the definition of the total variation, e.g [34] and [27].

[17] See page 141 in [23]

iii) *$P(|g_1 - g_2|)/2$, where P is an arbitrary measure with $\mu \ll P$ and $\nu \ll P$, and g_1 and g_2 are the P-densities of μ, ν respectively.*

When $\mathscr{X}$ is countable one have furthermore the following equivalent versions of the Total Variation distance[18]

$$\begin{aligned}
||\mu - \nu||_{TV} &= \max_{A \in \mathscr{F}} |\mu(A) - \nu(A)| \\
&= \frac{1}{2} \sum_{x \in \mathscr{X}} |\mu(x) - \nu(x)| \\
&= \sum_{x \in \mathscr{X}, \mu(x) \geq \nu(x)} |\mu(x) - \nu(x)| \\
&= \frac{1}{2} \sup\{ \sum_{x \in \mathscr{X}} f(x)(\mu(x) - \nu(x)) : ||f||_\infty \leq 1 \} \\
&= \inf_{P \in K(\mu,\nu)} \int \rho_1(x,y) P(dx,dy)
\end{aligned}$$

As the following proposition show, the total variation distance equals the Wasserstein distance corresponding to the $(0,1)$-metric ρ_1 given by

$$\rho_1(x,y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$$

Proposition 3.4.2. [19] *Let $\mathscr{X}$ be space metrized by the $(0,1)$-metric ρ_1, and let μ and ν be two probability measures on $\mathscr{X}$. Then*

$$||\mu - \nu||_{TV} = \inf_{P \in K(\mu,\nu)} \int \rho_1(x,y) P(dx,dy) \tag{3.10}$$

The proposition implies the following relation between the Wasserstein distance and the Total Variation distance

$$W_\rho(\mu,\nu) \leq \mathrm{Diam}(\mathscr{X}) \, ||\mu - \nu||_{TV} \tag{3.11}$$

where $\mathrm{Diam}(\mathscr{X}) = \sup_{s,t} \rho(s,t)$. In particular, for metrics bounded by 1, i.e. $\sup_{x,y \in \mathscr{X}} \rho(x,y) \leq 1$, we have that $\rho(x,y) \leq \rho_1(x,y)$, and so the inequality follows directly by integrating. An important question for what follows in the next chapter is whether by direct calculations such a strict inequality can be proven.

[18] See [25] section 4.1 and 4.2 for nice proofs and discussion.

[19] Proposition 4.2 and Proposition 4.4 in [24]. A complete proof in our setting is given in [27], proof of Theorem 5.2, where a optimal coupling is constructed.

Chapter 4

Dobrushins Uniqueness Condition

In this chapter we introduce ways to show uniqueness of the consistent measures, in particular the Dobrushin uniqueness condition, first introduced by Ronald Dobrushin in the paper [6] and [7]. Originally this condition was proven via an contracting coupling argument by use of optimal couplings. We present this method and compare it to a dual approach via averaging over observables. Moreover we introduce an even more general technique applicable for both approaches expanding the uniqueness condition. In the last sections we discuss how these techniques can be applied to show decay of correlation, and to compare measures consistent with different specifications. This chapter is mainly based on the the book [34] by B. Simon, the original papers [6] and [7] by R.L. Dobrushin, the lecture notes [19] by R. Fernandez and the paper [22] by H. Föllmer.

4.1 Uniqueness and Non-Uniquenss of Gibbs Measures

From this point on we will assume given a set of metrics $\{\rho_x\}$ on the single site spaces Ω_x. Furthermore, we assume also that for each $\Delta \in \mathscr{L}$, ρ_Δ is a metric on Ω_Δ. A natural choice would be to let $\rho_\Delta = \sum_{x\in\Delta}\rho_x$. However, we proceed in generality, and comment later on how ρ_Δ can be chosen.

As shown in the previous chapter, the Wasserstein distance $W_\rho(\cdot,\cdot)$, or equivalently, its dual distance, $D_\rho(\cdot,\cdot)$, metrizes the weak topology on the space of measures. This leads to our first uniqueness condition.

Condition 1. If the specification Π is quasilocal, S is a Polish space, and that for any two consistent measures, μ and ν, and any $\Delta \in \mathscr{L}$

$$D_{\rho_\Delta}(\mu,\nu) = 0 \tag{4.1}$$

or equivalently,

$$W_{\rho_\Delta}(\mu,\nu) = 0. \tag{4.2}$$

then there exist at most one measure consistent with Π in the Wasserstein space $W^1_\rho(\Omega)$.

In particular, if ρ is uniformly bounded, the condition above holds for every consistent measure. Moreover, if we further assume that the single site space S is compact and that Π satisfies the Feller property, then by Corollary 2.1.1 the condition even implies uniqueness. In what follows we also want to compare limits of specification, and are hence in need of a limit procedure of finite volumes.

Definition 4.1.1 (Covering limit of volumes)**.** *A sequence $\{\Lambda_n\}$ of finite subset of $\mathbb{L}$ is said to eventually cover $\mathbb{L}$ if for every $x \in \mathbb{L}$ there exists a $N \in \mathbb{N}$ such that $x \in \Lambda_n$ for each $n \geq N$.*

Under the assumption that the specification Π satisfies the Feller property, then by proposition 2.1.3 we also have the following conditions.

Condition 2 (Uniqueness condition by Wasserstein distance)**.** [1] If S is a compact Polish space, and Π is quasilocal specification satisfying the Feller property, then the set $\mathscr{G}(\Pi)$ contains exactly one consistent measure if and only if for each finite $\Lambda \subset \mathbb{L}$

$$\sup_{\tau,\gamma} W_{\rho_\Lambda}(\pi_{\Lambda_n}(\cdot|\tau), \pi_{\Lambda_n}(\cdot|\gamma)) \to 0 \text{ as } n \to \infty$$

where the sequence Λ_n eventually covers $\mathbb{L}$.

Proof. By theorem 2.1.1 there exists a consistent measure and by theorem 2.1.3 any extreme point of $\mathscr{G}(\Pi)$, μ, satisfies that $\lim_{n\to\infty} \pi^{\omega}_{\Lambda_n} = \mu$ μ-a.e. ω. Thus, if μ is the unique consistent measure, it follows immediately as any limit procedure $\pi^{\omega}_{\Lambda_n}$ is a consistent measure. For the other direction, recall that by quasilocality, any μ is characterized by its local events. □

We will also use an equivalent condition to show uniqueness of consistent measures. For this we need the definition of the oscillation of a function.

Definition 4.1.2 (Oscillation of a function)**.** *For any $\Delta \subset \mathbb{L}$ and function $f : \Omega \mapsto \mathbb{R}$, the oscillation of f on V, $\delta^{\rho}_{\Delta}(f)$, is defined as*

$$\delta^{\rho}_{\Delta}(f) = \sup_{\sigma=\omega \text{ off } \Delta} \frac{|f(\sigma) - f(\omega)|}{\rho_\Delta(\sigma,\omega)} \tag{4.3}$$

If $\Delta = \{x\}$ or $\Delta = \mathbb{L}$, then we will simply write $\delta^{\rho}_{x}(f)$ instead of $\delta^{\rho}_{\{x\}}(f)$ and $\delta^{\rho}(f)$ instead of $\delta^{\rho}_{\mathbb{L}}(f)$.

Similarly we introduce the notation $\delta_\Delta(f)$ as

$$\delta_\Delta(f) = \sup_{\sigma=\omega \text{ off } \Delta} |f(\sigma) - f(\omega)| \tag{4.4}$$

that is, we do not divide by $\rho_\Delta(\sigma,\omega)$ in this case.

Remark. Note that $\delta^{\rho}_{\Delta}(f)$ actually is the Lipschitz norm of f on the space Δ with metric ρ_Δ.

The definitions above will be used when showing uniqueness via averaging over observables. This concept follows naturally when realizing that for any function $f \in L_\Lambda$,

$$\delta(\pi_\Lambda f) \le \delta(f) \tag{4.5}$$

Here $\pi_\Lambda f(\omega) = \pi^{\omega}_{\Lambda} f = \pi_\Lambda(f|\omega) = \int_\Lambda f(\sigma_\Lambda \omega^c_\Lambda)\pi_\Lambda(d\sigma|\omega)$ is seen as a function on $\mathscr{F}_{\Lambda^c}$. Thus the inequalities above follows since for any $\omega \in \Omega$,

$$\inf_\sigma f(\sigma) \le \pi_\Lambda f(\omega) \le \sup_\sigma f(\sigma)$$

and implies that for any $\Delta \subset \Lambda$ and $f \in L_\Lambda$,

$$\delta(\pi_\Lambda f) = \delta(\pi_\Lambda \pi_\Delta f) \le \delta(\pi_\Delta f).$$

By the consistency property of the specification it follows that $\lim_{n\to\infty} \delta(\pi_{\Lambda_n} f)$ exists by the monotonicity and the upper bound $\delta(f)$. The relevant question is whether the above limit vanishes or not, which is summarized in the following condition.

[1] Proposition 2.2. in [42]

Condition 3 (Dual Uniqueness condition)**.** [2] If S is compact and Π is a quasilocal specification satisfying the Feller condition, then the set $\mathscr{G}(\Pi)$ contains exactly one measure if and only if for each finite $\Lambda \subset \mathbb{L}$

$$\sup_{f \in L_\Lambda, ||f||_L = 1} \{\sup_{\tau,\gamma} |\pi_{\Lambda_n}(f|\tau) - \pi_{\Lambda_n}(f|\gamma)|\} \to 0 \text{ as } n \to \infty \tag{4.6}$$

where Λ_n eventually covers $\mathbb{L}$.

Proof. If μ is a measure consistent with Π it follows that for any $\Lambda \in \mathscr{L}$, and local Lipschitz function $f \in L_\Delta$, $\mu(f) = \mu(\pi_\Lambda f)$. We prove in the following that

$$\sup_{\mu \in \mathscr{G}(\Pi)} \mu(f) = \lim_{n\to\infty}(\sup_\tau\{\pi_{\Lambda_n} f(\tau)\}) \tag{4.7}$$

and similarly

$$\inf_{\mu \in \mathscr{G}(\Pi)} \mu(f) = \lim_{n\to\infty}(\inf_\tau\{\pi_{\Lambda_n} f(\tau)\}) \tag{4.8}$$

which implies the condition.

Firstly, note that the limits above exists by the monotonicity. By the DLR equation $\mu(f) = \mu(\pi_{\Lambda_n} f)$, and hence $\mu(f) \le \sup_\tau(\pi_{\Lambda_n} f(\tau))$ for any Gibbs measure μ and any n. Therefore

$$\sup_{\mu \in \mathscr{G}(\Pi)} \mu(f) \le \lim_{n\to\infty}(\sup_\tau\{\pi_{\Lambda_n} f(\tau)\})$$

To show equality, let us consider an optimizing sequence of boundary condition, i.e. $\{\omega^n\}$ such that $\pi_{\Lambda_n} f(\omega^n) = \sup_\tau \pi_{\Lambda_n} f(\tau)$. Such boundary conditions exists because $\pi_{\Lambda_n}(\cdot)$ is a continuous function on the compact space of (boundary) configurations. The sequence of measures $\{\pi_{\Lambda_n}(\cdot|\sigma^n)\}$ has one or more accumulation points due to compactness, all of which are consistent measures. Any of these, selected for the given f, saturates the inequality. That proves the first statement. The second statement is proven analogously. □

The proof that the two conditions above are in fact dual uniqueness conditions follows from the next proposition.

Proposition 4.1.1. *Assume S is a Polish space, Π a specification and let Δ be a finite subset of $\mathbb{L}$. Then*

$$\sup_{\sigma,\omega} W_{\rho_\Delta}(\pi_\Delta(\cdot|\sigma), \pi_\Delta(\cdot|\omega)) = \sup_{f \in L_\Delta} \frac{\delta(\pi_\Delta f)}{\delta_\Delta^\rho(f)} = \sup_{f \in L_\Delta, ||f||_L = 1} \delta(\pi_\Delta f) \tag{4.9}$$

Proof. For any $\sigma, \omega \in \Omega$ we get from the strong Dual theorem (3.3.1) that

$$\begin{aligned} & \sup_{\sigma,\omega} W_{\rho_\Delta}(\pi_\Delta(\cdot|\sigma), \pi_\Delta(\cdot|\omega)) \\ = & \sup_{\sigma,\omega} \sup_{f \in L_\Delta} \frac{|\pi_\Delta(f|\sigma) - \pi_\Delta(f|\omega)|}{\delta_\Delta^\rho(f)} \\ = & \sup_{f \in L_\Delta} \sup_{\sigma,\omega} \frac{|\pi_\Delta(f|\sigma) - \pi_\Delta(f|\omega)|}{\delta_\Delta^\rho(f)} \\ = & \sup_{f \in L_\Delta} \frac{\delta(\pi_\Delta f)}{\delta_\Delta^\rho(f)} \end{aligned}$$

[2] Theorem 1.1 in [19]

Here the supremum is taken over all non-constant Lipschitz function. For the second line, by Theorem 3.3.2 there exist a function $f^{\sigma,\omega}$ obtaining the supremum for each $\sigma,\omega\in\Omega$. The swapping of supremums in the third line is thus allowed since Ω is a Polish space, and hence it contains a countable dense subset. $\square$

4.2 Dobrushin Uniqueness Condition

The conditions for uniqueness of consistent measures seen in the previous section are all based on an infinite volume analysis. In almost every system that one studies in mathematical statistical mechanics the complexity of the specifications makes it more or less impossible to study such limits. Thus the need for a criteria which is restricted to a finite volume analysis, and hence can be calculated in finite time, sees its importance. The Dobrushin uniqueness condition, introduced by R.L Dobrushin in 1968 in [6], is such a criteria. It shows that weakly dependent systems also admits only one consistent measure, by analyzing the single site specification only. The condition has historically been studied via coupling and via averaging over observables. We present both approaches and show how they are related.

Example. Let S be finite and consider the specifications given by the Hamiltonian $H_\Lambda^\omega(\sigma)=H_\Lambda(\sigma)=\sum_{x\in\Lambda}\phi_x(\sigma_x)$, then

$$\begin{aligned}\pi_\Lambda(\sigma|\omega) &= \frac{1}{Z_\Lambda}e^{-H_\Lambda(\sigma)}\\ &= \frac{1}{Z_\Lambda}\prod_{x\in\Lambda}e^{-\phi_x(\omega_x)}\\ &= \prod_{x\in\Lambda}\frac{1}{Z_x}e^{-\phi_x(\omega_x)}\end{aligned}$$

Since $Z_\Lambda=\prod_{x\in\Lambda}Z_x$. Thus, as each $\mu\in\mathscr{G}(\Pi)$ must satisfy the DLR equations, it follows by Theorem 5.11 in [36] that there exists only one such measure μ. Actually, the general theme is that independent specification implies uniqueness of the consistent measures.[3] The strategy of the Dobrushin uniqueness criteria is to show uniqueness when the specification is only slightly dependent, for instance in the Ising model when the temperature is high (small β). This we can estimate by having a control over how much a site x depend on a flipping of the spin at another site y.

To start our analysis we introduce the concept of an estimator.

Definition 4.2.1 (Dobrushin Estimator)**.** *The set $\{k_{x,y}\}_{x,y\in\mathbb{L}}$ is said to be an estimator for the specification Π if for every $\sigma=\omega$ off y*

$$W_{\rho_x}(\pi_x^\sigma,\pi_x^\omega)\le\rho_y(\sigma,\omega)k_{x,y} \tag{4.10}$$

In the dual setup, $\{\alpha_{x,y}\}_{x,y\in\mathbb{L}}$ is an estimator if for every $f\in L_x$,

$$\delta_y^\rho(\pi_x f)\le\delta_x^\rho(f)\alpha_{x,y} \tag{4.11}$$

[3] Theorem 5.1 in [36]

Thus, $k_{x,y}$ and $\alpha_{x,y}$ are estimates of how much a site x depend on a flipping of a spin at a site y. If $x = y$ the we can always set $k_{x,y} = \alpha_{x,y} = 0$, and we will assume so in the proceedings. We will in the following use the notation $k_{x,y}$ only when we want to emphasis that we are working in the coupling approach, but otherwise stick to $\alpha_{x,y}$. Furthermore, the optimal estimators are given by

$$k_{x.y} = \sup_{\sigma=\omega \text{ off } y} \frac{1}{\rho_y(\sigma,\omega)} W_{\rho_x}(\pi_x^\sigma, \pi_x^\omega) \tag{4.12}$$

and

$$\alpha_{x,y} = \sup_{f\in L_x,\, ||f||_L=1} \delta_y^\rho(\pi_x f) = \sup_{f\in L_x, ||f||_L=1} \sup_{\sigma=\omega \text{ off } y} \frac{|\pi_x^\sigma f - \pi_x^\omega f|}{\rho_y(\sigma,\omega)} \tag{4.13}$$

Hence $\alpha_{x,y} = k_{x,y}$, as for each $\sigma = \omega$ off y and $f \in L_x$,

$$\frac{|\pi_x^\sigma f - \pi_x^\omega f|}{\rho_y(\sigma,\omega)} \leq \frac{\delta_x^\rho(f) W_{\rho_x}(\pi_x^\sigma, \pi_x^\omega)}{\rho_y(\sigma,\omega)},$$

and we will in the following treat them as the same.

Condition 4 (Dobrushin Uniqueness Condition)**.** The Dobrushin Uniqueness condition is said to hold for a specification Π if there exists a Dobrushin estimator $\{\alpha_{x,y}\}_{x,y\in\mathbb{L}}$ and a constant $\gamma_1 \geq 0$ such that

$$\sup_{x\in\mathbb{L}} \sum_{y\in\mathbb{L}:\, y\neq x} \alpha_{x,y} = \gamma_1 < 1 \tag{4.14}$$

In words the above conditions is saying that the spin value at a site x is only weakly influenced by the spin values at other sites. In the following we will prove that it implies uniqueness of the consistent measure.

Theorem 4.2.1. *Given a system $(G,(S,\mathscr{T}_S,\mathscr{F}_S),\Pi)$ where S is a Polish metric space and Π is a quasilocal specification. If the Dobrushin Uniqueness Condition holds for a specification Π, then the Wasserstein space $W_\rho^1(\Omega)$ contains at most one measure μ consistent with Π.*

4.2.1 Proof by improving coupling

The following approach emphasis the main ideas how they were originally presented in [6] and [7].

Enumerate $\mathbb{L}$ by the natural numbers, let μ and ν be any two measures in $\mathscr{G}(\Pi)$, and consider any coupling $P \in K(\mu,\nu)$. Moreover, let $f_x = \int \rho_x(\alpha,\beta)P(d\alpha,d\beta)$, which we call the estimation function. Then we have the following important lemma.

Lemma 4.2.1 (Surgery Lemma)**.** *Assume we are given an estimator $\{k_{x,y}\}_{x,y\in\mathbb{L}}$. Then for any $x_0 \in \mathbb{L}$ there exists a coupling $\tilde{P}$ of μ and ν with corresponding estimation function $\tilde{f}$ such that*

$$\tilde{f}_x \begin{cases} = & f_x & \text{if } x \neq x_0 \\ \leq & \sum_{y\neq x} f_y k_{x,y} & \text{if } x = x_0 \end{cases} \tag{4.15}$$

Proof. For any $\tau,\eta \in \Omega$ there exists a coupling $P^{\tau,\eta} \in K(\pi_{x_0}(\cdot|\tau), \pi_{x_0}(\cdot|\eta))$ which satisfies

$$\int \rho_x(\alpha,\beta)P^{\tau,\eta}(d\alpha,d\beta) \leq \sum_{y\neq x_0} \rho_y(\tau,\eta)k_{x,y}$$

This follows by Theorem 3.1.1, saying that the optimal coupling exists, and as $W_{\rho_x}(\pi_x^\tau, \pi_x^\eta) \leq \sum_n W_{\rho_x}(\pi_x^{\tau_n}, \pi_x^{\tau_{n+1}})$ by the metric inequality, where $\{\tau^n\}$ is a set of telescoping configuration, that is, $\tau_x^n = \tau_x$ if $x \geq n$, and otherwise it is equal η_x.

Now, define our coupling $\tilde{P}$ to be the measure characterized by $\tilde{P} = P$ for all events outside Ω_{x_0}, and with $\tilde{P}_{x_0}(\cdot,\cdot|\tau,\eta) = P^{\tau,\eta}(\cdot,\cdot)$. That is, for every continuous function f we let

$$\tilde{P}(f) = \int\int f(\alpha,\beta)P^{\sigma,\omega}(d\alpha,d\beta)P(d\sigma,d\omega)$$

As μ and ν are consistent measure, and hence satisfies the DLR equations, $\tilde{P}$ is indeed a coupling of μ and ν. Moreover, we have that

1. For any $x \neq x_0$, $\tilde{f}(x) = \int \rho_x(\alpha,\beta)\tilde{P}(d\alpha,d\beta) = \int \rho_x(\alpha,\beta)P(d\alpha,d\beta) = f(x)$
2. $\tilde{f}(x_0) = \int\int \rho_x(\alpha,\beta)P^{\tau,\eta}(d\alpha,d\beta)P(d\tau,d\eta) \leq \int \sum_{y\neq x_0} k_{x_0,y}\rho_y(\tau,\eta)P(d\tau,d\eta) = \sum_{y\neq x_0} k_{x_0,y} f(y)$.

□

Given a coupling $P \in K(\mu,\nu)$ we can hence improve it by letting U_xP be the coupling which minimizes the estimation function of P and $\tilde{P}$ where $\tilde{P}$ is updated at x by use of the Surgery Lemma. That is, $(U_xf)_y = \min(f_y, \tilde{f}_y)$. Furthermore, we introduce the limit operator $UP := \lim_{n\to\infty} U_nU_{n-1}\cdot U_1(P)$. For any local event this definition indeed makes sense, and since every measure we are considering is characterized by local events by assuming that the specification is quasilocal, even the limit is well defined. Moreover, the operator U satisfies the following bound.

Proposition 4.2.1. *Let (Uf) be the estimation function of UP. Then* $\sup_{x\in\mathbb{L}}(Uf)_x \leq \gamma_1 \sup_{x\in\mathbb{L}} f_x$

Proof. By construction the estimation function (Uf) satisfies the inequality from the Surgery Lemma for every $x \in \mathbb{L}$. Hence, for any $x \in \mathbb{L}$

$$(Uf)_x \leq \sum_{y\neq x} f_y k_{x,y} \leq \sum_{y\neq x} k_{x,y} \sup_{x\in\mathbb{L}} f_x \leq \gamma_1 \sup_{x\in\mathbb{L}} f_x. \tag{4.16}$$

which implies the claim. □

Finally, by iterating U we get that $\sup_{x\in\mathbb{L}}(U^nf)_x \leq \gamma^n \sup_x f_x$, and hence, under the assumption that $\sup_x f_x < \infty$, which is the case for measures in the Wasserstein space, the estimation function converges to 0. Hence, this implies that the Wasserstein distance of μ and ν equals 0, and so $\mu = \nu$. As this holds for any $\mu,\nu \in G(\Pi)$, there can exists at most one measure consistent with Π.

4.2.2 Proof by averaging over observables

We continue by proving the same uniqueness result, only now by averaging over observables. The following approach is an adapted version of the proofs of Theorem V.1.3 and V.3.1 in B. Simons book [34]. We start of with introducing the concept of a spread estimator, which will be important also in the following chapters.

Definition 4.2.2. *Given two measure μ and ν, the set $\{\lambda_x\}_{x\in\mathbb{L}}$ is said to be a spread estimator for μ and ν if for each quasilocal Lipschitz function f*

$$|\mu(f) - \nu(f)| \leq \sum_{z\in\mathbb{L}} \delta_z^\rho(f)\lambda_z \tag{4.17}$$

It is important that the definition is restricted to quasilocal functions only. To see why this is crucial let $S=\{0,1\}$ and let $\mu(A)=1-\nu(A)=1$ where $\omega\in A$ if it contains only finitely many 1's. Thus, $|\mu(1_A)-\nu(1_A)|=1$, while $\delta_x^\rho(1_A)=0$ for each $x\in\mathbb{L}$. Hence there does not exist a spread estimator in this situation. As the consistent measures forms a convex set, and the extreme points are mutually singular, the above examples shows why the concept of quasilocality is important. Note that for any local function f, the definition indeed make sense. This can be seen by telescoping. Let f depend only on the sites $\{1,\cdots,n\}$. Then for any $\sigma=\omega$ off $\{1,\cdots,n\}$,

$$|f(\sigma)-f(\omega)|\le\sum_{i=1}^{n+1}|f(\sigma^{i-1})-f(\sigma^i)|$$

where $\sigma_x^i=\sigma_x$ if $x\ge i$, and otherwise equal to ω_x. Hence it follows that

$$\delta(f)\le\sum_{x\in\mathbb{L}}\delta_x(f)\tag{4.18}$$

As $\delta_x(f)=0$ if and only if $\delta_x^\rho(f)=0$, this proves that the definition of the spread estimator is indeed meaningful for local functions, hence also quasilocal functions.

We continue with a new Lemma, the Dusting Lemma, which shows how we can apply π_x as a "duster", where a function f is seen as the "dust".

Lemma 4.2.2 (Dusting Lemma). [4] *Given an estimator $\{\alpha_{x,y}\}_{x,y\in\mathbb{L}}$, then*

$$\delta_y^\rho(\pi_{\{x\}}f)\begin{cases}=0 & \text{if } y=x\\ \le\delta_y^\rho(f)+\delta_x^\rho(f)\alpha_{x,y} & \text{if } y\neq x\end{cases}\tag{4.19}$$

for every Lipschitz function f.

Proof. Let $\sigma=\omega$ off y and let $f\in L$. If $y=x$, then obviously $\delta_y^\rho(\pi_{\{x\}}f)=0$ since $\pi_{\{x\}}f$ does not depend on site x. Moreover, for $y\neq x$ we have that for any $\sigma=\omega$ off y,

$$\begin{aligned}
& \frac{|\pi_{\{x\}}(f|\sigma)-\pi_{\{x\}}(f|\omega)|}{\rho_y(\sigma,\omega)} & \\
= & \frac{|\pi_{\{x\}}(f_\sigma|\sigma)-\pi_{\{x\}}(f_\omega|\omega)|}{\rho_y(\sigma,\omega)} & \text{where } f_\tau(\eta)=f(\eta_x\tau_x^c)\\
\le & \frac{|\pi_{\{x\}}(f_\sigma|\sigma)-\pi_{\{x\}}(f_\sigma|\omega)|}{\rho_y(\sigma,\omega)}+\frac{|\pi_{\{x\}}(f_\sigma-f_\omega|\omega)|}{\rho_y(\sigma,\omega)} & \\
\le & \delta_y^\rho(\pi_{\{x\}}f_\sigma)+\delta_y^\rho(f) & \\
\le & \delta_x^\rho(f)\alpha_{x,y}+\delta_y^\rho(f) & \text{as } f_\sigma\in L_x
\end{aligned}$$

□

The name "Dusting Lemma" comes from the interpretation that by applying π_x we are dusting the function f at the site x. The Dusting Lemma tells us that when doing so we manage to clean the site x completely, while some of the dust spreads to the other sites. Hence, the dust at $y\neq x$ increases with the amount of dust at x, $\delta_x^\rho(f)$, times the dust parameter $\alpha_{x,y}$. The Dobrushin uniqueness condition

[4] Lemma 1.3 in [19]

then states that dusting at any site $x \in \mathbb{L}$ with the duster π_x reduces the total amount of dust since $\sum_{y\in\mathbb{L}} \alpha_{x,y} < 1$.

Now combining the Dusting Lemma with the concept of a spread estimator we are able to update the spread estimator, similar as in the Surgery Lemma 4.2.1.

Lemma 4.2.3 (Updating Lemma)**.** *Let $\{\lambda_z\}_{z\in\mathbb{L}}$ be a spread estimator for μ and ν. Then for any $x \in \mathbb{L}$ there exists a spread estimator $\{\tilde{\lambda}_z\}$ such that*

$$\tilde{\lambda}_z \begin{cases} = & \lambda_z \quad \text{if } z \neq x \\ \leq & \sum_{y\neq z} \alpha_{z,y}\lambda_y \quad \text{if } z = x \end{cases} \tag{4.20}$$

Proof. We have that for any local Lipschitz function f that

$$\begin{aligned} & |\mu(f) - \nu(f)| & \\ = & |\mu(\pi_x f) - \nu(\pi_x f)| & \text{by DLR equations} \\ \leq & \sum_{z\in\mathbb{L}} \delta_z^\rho(\pi_x f)\lambda_z & \\ \leq & \sum_{z\neq x}(\delta_z^\rho(f) + \delta_x^\rho(f)\alpha_{x,z})\lambda_z & \text{by Dusting Lemma} \\ = & \delta_x^\rho(f)(\sum_{z\neq x}\alpha_{x,z}\lambda_z) + \sum_{z\neq x}\delta_z^\rho(f)\lambda_z & \end{aligned}$$

As this holds for any such f completes the proof. □

The above Updating Lemma shows how to improve the spread estimator by dusting a site x. To prove uniqueness we need to show that by doing so iteratively over the whole lattice all the dust vanishes. This goes similar as in the coupling approach. Enumerate $\mathbb{L}$ by the natural numbers, and introduce the operator $T_n := \pi_1 \ldots \pi_n(f)$. Then by applying T_n we have the following proposition.

Proposition 4.2.2. *For any local function $f \in L$*

$$|\mu(T_n f) - \nu(T_n f)| \leq \sum_{z\notin\{1,\cdots,n\}} \delta_z^\rho(f)\lambda_z + \sum_{i=1}^{n} \delta_i^\rho(f)\gamma_1 \sup_{z\in\mathbb{L}} \lambda_z \tag{4.21}$$

Proof. This follows by induction. For $n = 0$, the statement simply follows by the definition of a spread estimator. Now, assume it holds for $n-1$. Then we have

$$\begin{aligned} & |\mu(T_n f) - \nu(T_n f)| & \\ = & |\mu(T_{n-1}\pi_n f) - \nu(T_{n-1}\pi_n f)| & \\ \leq & \sum_{z\notin\{1,\cdots,n-1\}} \delta_z^\rho(\pi_n f)\lambda_z + \sum_{i=1}^{n-1} \delta_i^\rho(\pi_n f)\gamma_1 \sup_{z\in\mathbb{L}}\lambda_z & \text{by induction hypothesis} \\ \leq & \sum_{z\notin\{1,\cdots,n\}} (\delta_z^\rho(f) + \delta_n^\rho(f)\alpha_{n,z})\lambda_z + \sum_{i=1}^{n-1}(\delta_i^\rho(f) + \delta_n(f)\alpha_{n,i})\gamma_1 \sup_{z\in\mathbb{L}}\lambda_z & \text{by Dusting Lemma} \\ = & \sum_{z\notin\{1,\cdots,n\}} \delta_z^\rho(f)\lambda_z + \sum_{z<n}\delta_z^\rho(f)\gamma_1\sup_{z\in\mathbb{L}}\lambda_z & \\ & +\delta_n^\rho(f)(\sum_{z\notin\{1,\cdots,n\}} \alpha_{n,z}\lambda_z + \sum_{i=1}^{n-1}\alpha_{n,i}\gamma_1\sup_{z\in\mathbb{L}}\lambda_z) & \\ \leq & \sum_{z\notin\{1,\cdots,n\}} \delta_z^\rho(f)\lambda_z + \sum_{i=1}^{n}\delta_i^\rho(f)\gamma_1\sup_{z\in\mathbb{L}}\lambda_z & \end{aligned}$$

□

Thus, by defining $T := \lim_{n\to\infty} T_n$, which is well defined for every local Lipschitz function, and since $\mu(Tf) = \mu(f)$ by the DLR equations, the proposition implies the following corollary.

Corollary 4.2.1. *For every local function $f \in L$*

$$|\mu(Tf) - \nu(Tf) \le \sum_{x\in\mathbb{L}} \delta_x^\rho(f)\gamma_1 \sup_{x\in\mathbb{L}} \lambda_x \tag{4.22}$$

Assuming that the Dobrushin uniqueness condition is satisfied and that $\sup_{x\in\mathbb{L}} \lambda_x = C < \infty$ it hence follows that

$$|\mu(f) - \nu(f)| = |\mu(Tf) - \nu(Tf)| \le \gamma_1 \sum_{z\in\mathbb{L}} \delta_z^\rho(f)C$$

for each local function $f \in L$. By iterating the operator T n-times it implies further that $|\mu(f) - \nu(f)| \le \gamma_1^n \sum_{x\in\mathbb{L}} \delta_z^\rho(f)C$ which converges to 0 as $n \to \infty$. Hence $|\mu(f) - \nu(f)| = 0$ for every local Lipschitz function f, which by condition 1 means that there can at most exists one measure consistent with Π.

4.2.3 Relation between the coupling approach and averaging over observables

We have now seen both the coupling approach and the approach by averaging over observables. There are many similarities between these approaches, and they can be seen to be more or less equivalent. In the coupling approach, given an estimator function f, we obtain from the Surgery Lemma a new estimator function $\tilde{f}$ given by

$$\tilde{f}_x \le \sum_{y\neq x} f_y k_{x,y}$$

On the other hand, by averaging over observables we obtain from the Updating Lemma the new spread estimator

$$\tilde{\lambda}_x \le \sum_{y\neq x} \alpha_{x,y}\lambda_y$$

Knowing these inequalities, the proofs were more or less identical, and did not essentially involve any application of methods based on coupling techniques or averaging operators.

In the coupling approach we needed $\sup_{x\in\mathbb{L}} f_x < \infty$, while in the averaging procedure we needed $\sup_{x\in\mathbb{L}} \lambda_x < \infty$. The estimator function f depended on a "good" choice of a coupling P of μ and ν, i.e. $f_x = \int \rho_x(\eta,\tau)P(d\eta,d\tau)$. For λ on the other hand we have that for any quasilocal function $f \in L$,

$$\begin{aligned} & |\mu(f) - \nu(f)| \\ \le\ & |\int f(\eta) - f(\tau)P(d\eta,d\tau)| \\ =\ & |\int \sum_{i=1}^{\infty} f(\eta_{x<i}\tau) - f(\eta_{x\le i}\tau)P(d\eta,d\tau)| \\ \le\ & \sum_{i=1}^{\infty} \int |f(\eta_{x<i}\tau) - f(\eta_{x\le i}\tau)|P(d\eta,d\tau) \\ \le\ & \sum_{i=1}^{\infty} \delta_i^\rho(f) \int \rho_x(\eta,\tau)P(d\eta,d\tau) \end{aligned}$$

Hence the bound for the estimator function f_x can also be used as a bound for the spread estimator λ_x. For instance, if we assume that the set of metrics ρ_x are uniformly bounded, say $\rho_x \leq 1$, then this gives us the trivial bounded $\lambda_x \equiv f_x \equiv 1$. However, this is not a necessary condition and may be weakened a bit. Even the requirement that the estimator function or the spread estimator are uniformly bounded can be weakened. This is best seen by how we proceeded in the proof by coupling where we after the Surgery Lemma obtained that $\tilde{f}_x = \sum_{y \neq x} k_{x,y} f_y$ for every $x \in \mathbb{L}$ (notice that we could have proceeded similarly in the dusting proof). Hence, rather than assuming that f_y is uniformly bounded, we may assume that it decays sufficiently such that $\sum_{y \neq x} k_{x,y} f_y < \infty$. There do exists results for models which does not need $\sup_{x \in \mathbb{L}} \rho_x < \infty$.[5] From the requirement that $\gamma_1 < 1$ it is immediate that $\alpha_{x,y}$ decays as $d(x,y)$ grows to infinite, hence for translation invariant systems[6] with $\mathbb{L} = \mathbb{Z}^d$ an exponential decay is sufficient.

4.3 A third approach to Dobrushin uniqueness condition

In the following we will consider an estimator $\alpha = \{\alpha_{x,y}\}_{x,y \in \mathbb{L}}$ as a $\mathbb{L} \times \mathbb{L}$ matrix, with each element on the diagonal equal to 0. Moreover, we use the notation $[\frac{1}{1-\alpha}]_{x,y}$ for the sum $\sum_{k \geq 0} [\alpha^k]_{x,y}$.

In the arguments in the previous section we did our analysis directly on the consistent measures. In this section we provide an similar analysis to compare finite volume specifications by using condition 3. Hence we have to limit ourself to quasilocal specifications on a compact Polish space having the Feller property. The following approach will go deeper into what is required of the matrix α, which in particular will give uniqueness under a dual condition. Another consequence of this approach is that we will be able to easily show several results showing decay of correlation under the Dobrushin uniqueness condition. The section is based extensively on the notes [19] by R. Fernandez, but many of the ideas originated already in the paper [22] by Föllmer.

We start of by showing that the uniqueness statement follows from the next theorem.

Theorem 4.3.1. [7] *Assume Ω is a compact Polish space. Given a quasilocal specification Π satisfying the Feller property, let $\Lambda \in \mathscr{L}$ and let $\{\alpha_{x,y}\}_{x,y \in \mathbb{L}}$ be an estimator for Π. Let $y \notin \Lambda$. If $[\alpha]^n \to 0$ as $n \to \infty$, then*

$$\delta_y(\pi_\Lambda f) \leq \sum_{x \in \Lambda} \delta_x^\rho(f) [\frac{\alpha}{1-\alpha}]_{x,y} \tag{4.23}$$

for every $f \in L_\Lambda$.

For the uniqueness it is thus sufficient that $[\frac{\alpha}{1-\alpha}]_{x,y} \to 0$ as $d(x,y)$ grows. For this to happen we may assume that the specification has finite range R, i.e. that $\pi_\Lambda f \in \mathscr{F}_{\partial_R \Lambda}$ meaning that it only depend on spins in a radius R from Λ. In this case $\alpha_{x,y} = 0$ whenever $d(x,y) > R$, and moreover $\alpha^n_{x,y} = 0$ whenever $d(x,y) > nR$. Furthermore, the Dobrushin uniqueness condition implies that $\alpha^n_{x,y} \to 0$. By assumption $\alpha_{x,y} \leq \gamma_1$. Assume $\alpha^n_{x,y} \leq \gamma_1^n$, then

$$\alpha^{n+1}_{x,y} = \sum_z \alpha_{x,z} \alpha^n_{z,y} \leq \sum_z \alpha_{x,z} \gamma_1^n \leq \gamma_1^{n+1}$$

[5] See for instance Theorem V.3.4 in [34] for an example were $S = \mathbb{R}$.

[6] In [10] instead of measures in the Wasserstein space one studies measures that has exponential growth. That is, all measures μ such that for some $s \in S$ and $g, G < \infty$,

$$\rho_x(P) = \int_\Omega \rho_x(\omega, s) P(d\omega) \leq G \exp(g\,||x||)$$

[7] Theorem 1.6 in [19]

Hence, when $d(x,y) > nR$, by induction $[\frac{\alpha}{1-\alpha}]_{x,y} \leq \frac{\gamma_1^n}{1-\gamma_1}$, which converges to 0 as the distance between x and y grows.

For the uniqueness statement consider now any local $f \in L_\Delta$, with $\Delta \subset \Lambda_n$. Then,

$$\delta_y(\pi_{\Lambda_n} f) \leq \sum_{x \in \Delta} \delta_x^\rho(f)[\frac{\alpha}{1-\alpha}]_{x,y} \tag{4.24}$$

Let $\{\Lambda_n\}$ be a sequence eventually covering $\mathbb{L}$, then this implies that $\delta_y(\pi_\Lambda) \to 0$ as n approaches infinity. As $\delta(\pi_\Lambda f) \leq \sum_{y \notin \Lambda} \delta_y(\pi_\Lambda f)$ condition 3 is satisfied implying the uniqueness statement.

Corollary 4.3.1. *Given a quasilocal specification Π of finite range R on a compact Polish space satisfying the Feller property. Let α be an estimator for Π. If α satisfies the Dobrushin uniqueness condition, then $|\mathscr{G}(\Pi)| = 1$.*

For uniqueness above we need that $[\frac{1}{1-\alpha}]_{x,y} \to 0$ as $d(x,y) \to \infty$. The Dobrushin uniqueness condition was shown to be a sufficient condition for this. Another sufficient condition for such a convergence is that $\sup_y \sum_{x \neq y} \alpha_{x,y} = \gamma_2 < 1$, which we will call the Dual-Dobrushin uniqueness condition[8]. Hence the above corollary can be seen as an extension of our previous uniqueness theorems. In words the condition says that if the effect of flipping any spin $y \in \mathbb{L}$ has a small effect on the other spins in $\mathbb{L}$, then we also have uniqueness.

Condition 5. We say that the Dual-Dobrushin uniqueness condition holds if

$$\gamma_2 = \sup_y \sum_{x \neq y} \alpha_{x,y} < 1 \tag{4.25}$$

Note that what was essential above was that $\alpha_{x,y}^n \leq \gamma_1^n$. Similar as for the Dobrushin uniqueness condition we have that $\alpha_{x,y}^n \leq \gamma_2^n$ under the Dual-Dobrushin uniqueness condition, and hence the uniqueness result follows.

We proceed with the the proof of the main theorem, Theorem 4.3.1. For this we re-introduce the concept of a spread estimator.

Definition 4.3.1. *Given $\Lambda \in \mathscr{L}$ and a specification Π, then a corresponding spread estimator is a set $\{\lambda_{x,y}\}_{x \in \Lambda y \notin \Lambda}$ of non-negative numbers satisfying*

$$\delta_y(\pi_\Lambda f) \leq \sum_{x \in \Lambda} \delta_x^\rho(f) \lambda_{x,y} \tag{4.26}$$

for any $f \in L_\Lambda$

Similar as for the Dusting Lemma we can then prove the following Lemma.

Lemma 4.3.1 (Spread Lemma)**.** *For any $g \in L(\Omega)$*

$$\delta_y(\pi_\Lambda f) \leq \delta_y(f) + \sum_{x \in \Lambda} \delta_x^\rho(f) \lambda_{x,y}$$

Proof. For any Lipschitz function g and configurations $\sigma = \omega$ off y,

$$\begin{aligned} & |\pi_\Lambda^\sigma f - \pi_\Lambda^\omega f| \\ \leq \quad & |\pi_\Lambda^\sigma f_\sigma - \pi_\Lambda^\omega f_\sigma| + |\pi_\Lambda^\omega (f_\sigma - f_\omega)| \\ \leq \quad & \delta_y(\pi_\Lambda f_\sigma) + \delta_y(f) \\ \leq \quad & \sum_{z \in \Lambda} \delta_z^\rho(f) \lambda_{z,y} + \delta_y(f) \end{aligned}$$

□

[8] In [42] this is referred to as Dobrushin-Shlosman condition

Assuming a bound for the spread estimator is given, which can be found in the same way as in the previous section, we continue with showing how the spread estimator can be updated.

Lemma 4.3.2 (Update of spread estimator)**.** *Given a spread estimator $\{\lambda_{x,y}\}$, then for any $x \in \Lambda$ one can construct another estimator $T_x\lambda$ defined by*

$$(T_x\lambda)_{z,y}\begin{cases} = & \lambda_{z,y} & \text{if } z \neq x \\ = & \min(\lambda_{z,y}, \alpha_{z,y} + \sum_{u\in\Lambda, u\neq z} \alpha_{z,u}\lambda_{u,y}) & \text{if } z = x \end{cases}$$

Proof. Consider any function $f \in C_\Lambda$. Then this follows by applying the Spread Lemma and the Dusting Lemma as follows

$$\begin{aligned} \delta_y(\pi_\Lambda f) &= \delta_y(\pi_\Lambda \pi_{\{x\}} f) && \text{by the DLR equations} \\ &\leq \delta_y(\pi_{\{x\}} f) + \sum_{x\in\Lambda\setminus\{x\}} \delta_z^\rho(\pi_{\{x\}} f)\lambda_{z,y} && \text{by the Spread Lemma} \\ &\leq \delta_y^\rho(f) + \delta_x^\rho(f)\alpha_{x,y} + \sum_{z\in\Lambda\setminus\{x\}} [\delta_z^\rho(f) + \delta_x^\rho(f)\alpha_{x,z}]\lambda_{z,y} && \text{by the Dusting Lemma} \\ &= \delta_x^\rho(f)[\alpha_{x,y} + \sum_{u\in\Lambda\setminus\{x\}} \alpha_{x,u}\lambda_{u,y}] + \sum_{z\in\Lambda\setminus\{x\}} \delta_z^\rho(f)\lambda_{z,y} \end{aligned}$$

Hence, $T_x\lambda$ is indeed a spread estimator. □

Thus, by applying the operator T_x to each site $x \in \Lambda$ and improving successively we get that $\{\lambda^1_{x,y}\}_{x\in\Lambda, y\notin\Lambda}$ given by

$$\lambda^1_{x,y} = \alpha_{x,y} + \sum_{z\in\Lambda\setminus\{x\}} \alpha_{x,z}\lambda_{z,y}$$

also forms a spread estimator. Similar as α, consider λ^1 as the $\mathbb{L}\times\mathbb{L}$ matrix where $\lambda_{x,y} = 0$ if $x \notin \Lambda$ or $y \in \Lambda$. Thus, written in matrix notation the above equation can be written as

$$\lambda^1 = P_\Lambda \alpha P_{\Lambda^c} + P_\Lambda \alpha P_\Lambda \lambda \tag{4.27}$$

Here P_Λ is the projection matrix, i.e. it equals the identity matrix inside Λ, and is 0 outside Λ. Iterating this process leads to the spread operator λ^n given by

$$\lambda^n \leq (\sum_{i=0}^{n-1} (P_\Lambda \alpha P_\Lambda)^i)(P_\Lambda \alpha P_{\Lambda^c}) + (P_\Lambda \alpha P \Lambda)^n \lambda \tag{4.28}$$

Hence, by passing to the limit and assuming that $\lim_{n\to\infty}(P_\Lambda \alpha P_\Lambda)^n = 0$ we get the spread estimator

$$\lambda^\infty = [\frac{P_\Lambda \alpha P_{\Lambda^c}}{1 - P_\Lambda \alpha P_\Lambda}].$$

This proves Theorem 4.3.1.

Remark. An obvious question is whether it is possible to improve our bound for the spread estimator. We will show in the next chapter that this indeed is possible.

Next we show how our new result implies a strong mixing condition.

Condition 6. The estimator α is said to satisfy the decay property if there exists some $m > 0$ such that

$$\gamma := \min(\sup_{x\in\mathbb{L}} \sum_{y\neq x} \alpha_{x,y} e^{md(x,y)}, \sup_{y\in\mathbb{L}} \sum_{x\neq y} \alpha_{x,y} e^{md(x,y)}) < 1 \tag{4.29}$$

Having a estimator satisfying the decay property implies the next lemma.

Lemma 4.3.3. [9] *If α is a matrix with $\alpha_{x,y} \geq 0$ for $x, y \in \mathbb{L}$, satisfying the decay property for some*

[9] Lemma 1.8 in [19]

$m \geq 0$ and some metric d on $\mathbb{L}$, then

$$[\frac{\alpha}{1-\alpha}]_{x,y} \leq \frac{\gamma}{1-\gamma} e^{-md(x,y)} \tag{4.30}$$

Proof. By the assumption either $\sup_x \sum_y \alpha_{x,y} < 1$ or $\sup_y \sum_x \alpha_{x,y} < 1$ (or both), and hence each $\alpha^k_{x,y}$ is well defined. Moreover, $\alpha_{x,y} e^{md(x,y)} < \gamma$ for any $x,y \in \mathbb{L}$. Assume $\alpha^k_{x,y} e^{md(x,y)} < \gamma^k$ for every $x,y \in \mathbb{L}$. Then we have for any $x,y \in \mathbb{L}$ that

$$\begin{aligned} \alpha^{k+1}_{x,y} e^{md(x,y)} &= \sum_z \alpha^k_{x,z} \alpha_{z,y} e^{md(x,y)} \\ &\leq \sum_z \alpha^k_{x,z} e^{md(x,z)} \alpha_{z,y} e^{md(z,y)} \\ &\leq \gamma^{k+1} \end{aligned}$$

Thus, for any $k \geq 0$, $\alpha^k_{x,y} \leq \gamma^k e^{-md(x,y)}$ which proves the claim. □

The decay lemma implies that for any $\Lambda \in \mathscr{L}$ and $f \in L_\Lambda$

$$\delta_y(\pi_\Lambda f) \leq \frac{\gamma}{1-\gamma} \sum_{x \in \Lambda} \delta^\rho_x(f) e^{-md(x,y)} \tag{4.31}$$

This inequality further implies that $\delta_y(\pi_\Lambda f) \leq C \delta^\rho_\Lambda(f) e^{-md(\Lambda,y)}$ for some constant C, which in [10] is referred to as "the strongest decay property" (equation 3.11 in [10]). We will discuss these notions in more depth after introducing the Dobrushin-Shlosman uniqueness condition considered in the next chapter.

4.4 Comparing specifications

More or less the same analysis as in the previous section can be used when comparing two specification, say Π and Γ, on the same configuration space Ω. In the following we compare two measures, rather than the specifications directly, to deduce more general statement from the section above (e.g. we do not need the Feller assumption, nor the compactness). Thus we follow the lines of Föllmer in [22].

Let μ and ν be two measures, μ consistent with Π and ν consistent with Γ. To compare μ and ν we assume that α is an estimator for Π. Given a spread estimator $\{\lambda_x\}_{x \in \mathbb{L}}$ for μ and ν, this can be updated in the following way.

Lemma 4.4.1. *Let λ be a spread estimator for μ and ν, and let f be any local function in L. Let x be any site in $\mathbb{L}$. Then there exist an spread estimator λ^1 given by*

$$\lambda^1_x = b_x + \sum_{z \neq x} \alpha_{x,z} \lambda_x \tag{4.32}$$

where $b_x = \int W_{\rho_x}(\pi^\omega_x, \gamma^\omega_x) d\nu(d\omega)$.

Proof. For updating the spread estimator, let f be any local function in L. Then for any $x \in \mathbb{L}$ the following holds

$$\begin{aligned} & |\mu(f)-\nu(f)| \\ = & |\mu(\pi_x f)-\nu(\gamma_x f)| \\ \leq & |\mu(\pi_x f)-\nu(\pi_x f)|+|\nu((\pi_x-\gamma_x)f) \\ \leq & \sum_{z\neq x}\delta_z^\rho(\pi_x f)\lambda_z+\delta_x^\rho(f)b_x \\ \leq & \sum_{z\neq x}\delta_z^\rho(f)\lambda_z+\delta_x^\rho(f)(\sum_{z\neq x}\alpha_{x,z}\lambda_z+b_x) \end{aligned}$$

Hence, by enumerating $\mathbb{L}$ and updating at each site by taking the minimum as seen in previous updating lemmas, the result follows. □

In matrix notation the lemma tells us that $\lambda^1 = \alpha\lambda + b$. Iterating this updating Lemma we thus obtain that $\lambda^n = \alpha^n\lambda + \sum_{i=0}^{n-1}\alpha^i b$. Hence we obtain the following.

Theorem 4.4.1. [10] *Given two measures μ and ν, consistent with the quasilocal specifications Π and Γ respectively, and defined on the same configuration space Ω, a Polish space. Let α be an estimator for Π. If the matrix $\alpha^n \to 0$, then for any $f \in L_\Lambda$, $\Lambda \in \mathscr{L}$,*

$$|\mu(f)-\nu(f)| \leq \sum_{x\in\Lambda}\sum_{z\in\mathbb{L}}\delta_x(f)[\frac{1}{1-\alpha}]_{x,z}b_z \tag{4.33}$$

Moreover, the Wasserstein space $W_\rho^1(\Omega)$ contains at most one measure μ consistent with Π.

The uniqueness result follows since $b_x \equiv 0$ whenever $\Pi = \Gamma$. The above theorem is hence the strongest uniqueness statement under a Dobrushin-like condition we have seen so far, only assuming Π to be quasilocal. As seen from the analysis, there are no particular assumption on the estimator α other than $\alpha^n \to 0$ and so the requirement that Π has finite range is not necessary. Moreover, $\alpha^n \to 0$ is satisfied both under the Dobrushin uniqueness condition and the dual-Dobrushin uniqueness condition. Actually the situation is even more general. In [30] and [15] this question is treated in more generality, and the uniqueness condition is shown to hold if $||\alpha||_M < 1$ for any norm $||\cdot||_M$ on countable infinite matrices.

In the paper, [22], Föllmer also proved the following general correlation inequality, which we include for completeness.

Theorem 4.4.2. [11] *Suppose that μ is consitent with the specification Π, with estimator α satisfying $\alpha^n \to 0$, and that*

$$\sup_k \int \rho_k(\sigma,\omega)^2\mu(d\sigma) < \infty$$

for some $\omega \in \Omega$. Then

$$|\mu(fg)-\mu(f)\mu(g)| \leq c^2\sum_{x,y}\delta_x(f)[\frac{\alpha}{1-\alpha}]_{x,y}\delta_y(g) \tag{4.34}$$

for any Lipschitz functions f and g, where $c = \sup_k \int[\inf_{s\in S}\int \rho_k(\sigma,s)^2\mu_k(d\sigma_k|\sigma)]\mu(d\sigma)$.

[10] Theorem 2.4 in [22]
[11] Theorem 3.7 in [22]

Chapter 5

Extensions of Dobrushins Uniqueness Condition

In the mid eighties Dobrushin and Shlosman introduced in a series of papers an improved technique for uniqueness of consistent measures, following a similar approach as the coupling earlier used by Dobrushin (see for instance [10], [11],[12] and [13]). These were however only stated for translation invariant systems on the underlying graph $\mathbb{Z}^d$. In the following years, several related conditions by averaging over observables appeared (for instance [37] and [1]), which were claimed to be equivalent to the one by Dobrushin and Shlosman. In the first two sections we go into the details of the techniques of Dobrushin and Shlosman and a similar condition by Lieb and Aizenman. We first present a detailed proof of the Dobrushin-Shlosman condition and discuss a bit about its properties, before we show that the two conditions are more or less equivalent. In the search for a rigorous proof of the Lieb-Aizenman we first obtain a condition which generalizes Dobrushins original condition to larger volumes. In the last section, however, we derive a new condition extending the one by Lieb-Aizenman beyond translation invariance and finite range, and improving.

The main references for the following chapter are the original paper [10] by Dobrushin and Shlosman, and the lecture notes [19] by R. Fernandez, were an introduction to the technique by Lieb and Aizenman is given. [1] The last section proving the Lieb-Aizenman condition from our extended version is also influenced by the paper [42] by D. Weitz.

5.1 Dobrushin-Shlosman Uniqueness Condition

In this section and the following one on Lieb-Aizenman uniqueness condition we assume that the underlying graph is the integer lattice $\mathbb{Z}^d$ for some constant $d > 0$. Moreover, we also assume that we are given a specification Π which is translation invariant and has finite range R. That is, we assume

$$\pi_\Delta(\cdot|\sigma) = \pi_{\Delta+t}(\cdot|\theta_t\sigma) \tag{5.1}$$

where t is any vector in $\mathbb{Z}^d$ and $(\theta_t\omega)_x = \omega_{x-t}$. Finite range means that for every $f \in L_\Lambda$, $\pi_\Lambda f \in \mathscr{F}_{\partial_R\Lambda}$, and is hence a local function. Lastly we also assume that the set of metrics $\{\rho_x\}_{x\in\mathbb{L}}$ is uniformly bounded, and that $\rho_\Lambda = \sum_{x\in\Lambda}\rho_x$ for any $\Lambda \in \mathscr{L}$.

[1] The technique by Lieb and Aizenman have apparently never been published, but been cited in several papers. In [1] a theorem claimed to follow from Lieb and Aizenmans approach is proven, though in the setting of Stochastic Ising model only.

The strategy of Dobrushin and Shlosman was to extend the original technique to larger volumes. In their paper, [10], two consistent meaures are compared and the result is valid for any Ω being a separable metric space. [2] However, to compare it to the Lieb-Aizenman condition as presented in [19] it is necessary to write the condition for measures seen as limits of specification.

Definition 5.1.1 (DS_V Estimator)**.** *Given a finite volume $V \subset \mathbb{L}$, the set $\{k_{V,y}\}_{y\in\mathbb{L}\setminus V}$ is said to be an DS_V estimator if for every $y \in \mathbb{L}\setminus V$ and for all $\sigma = \omega$ off y it holds that*

$$W_{\rho_V}(\pi_V(\cdot|\sigma), \pi_V(\cdot|\omega)) \le \rho_y(\sigma,\omega)k_{V,y} \tag{5.2}$$

where $\rho_V = \sum_{x\in V}\rho_x$.

Equivalently, in the dual setup, this is the same as saying that for every $y \in \mathbb{L}\setminus V$ and any functions $f \in L_V$ the DS_V estimator $\alpha_{V,y}$ satisfies

$$\delta_y^\rho(\pi_V f) \le \delta_V^\rho(f)\alpha_{V,y} \tag{5.3}$$

Remark. By translation invariance it follows that $\alpha_{V,x} = \alpha_{V+s,x+s}$ and $k_{V,x} = k_{V+s,x+s}$ for any translation $s \in \mathbb{Z}^d$. This is important for the original proof which we present in the following subsection. Also the proof of the Lieb-Aizenman condition as seen in [19] uses this relation extensively.

Assuming that we are given a DS_V estimator, the condition for uniqueness is the following.

Condition 7 (DS_V Uniqueness Condition)**.** The DS_V uniqueness condition is said to hold for a specification Π if there exist a DS_V estimator $\alpha_{V,y}$ such that

$$\frac{1}{|V|}\sum_{y\in\mathbb{L}\setminus V}\alpha_{V,y} = \gamma_3 < 1 \tag{5.4}$$

Hence, what we intend to show is that under this condition the following theorem holds.

Theorem 5.1.1. *Let Π be a quasilocal specification satisfying the Feller property, and Ω a compact Polish space. If the Dobrushin-Shlosman uniqueness condition holds for some volume $V \in \mathscr{L}$, then there exists exactly one measure μ consistent with Π.*

We will later comment on the improvements this new condition imposes. Actually, as discussed in [10] the above criteria is so strong that it for a large class of models converges to the boundary of phase transition as the volume V approaches $\mathbb{Z}^d$. However, the bad news are that calculating an estimator when V becomes large demand highly increasing computation power and the complexity is beyond reach. On the other hand, for simple cases, and for not too large volumes, the above criteria gives a deterministic way to exploit the region of uniqueness by doing a finite time calculation. This was for instance exemplified in [13] were they applied the above criteria to a 3×4 box for a certain Hamiltonian model, extending previously known uniqueness bounds.

5.1.1 Proof via updating coupling

The following proof is a rewriting of the proof given in [10] with some extended details. It was originally shown via the following two lemmas.

[2] As commented on earlier, they refer to a space slightly more general measure space than our Wasserstein space.

Lemma 5.1.1. *Let V be any finite volume in $\mathbb{Z}^d$ and suppose we are given a DS_V estimator $\{k_{V,y}\}$. Let Λ be any given volume, let $\sigma, \omega \in \Omega$, and let $T(\Lambda) := \{t \in \mathbb{Z}^d \,:\, (V \cup \partial_R V) + t \subset \Lambda\}$.*

Then for any $\delta > 0$ there exists a measure $P \in K(\pi_\Lambda(\cdot|\omega), \pi_\Lambda(\cdot|\sigma))$ such that for all $s \in T(\Lambda)$.

$$\sum_{t \in V} f_{t+s} \leq \sum_{t \in \partial_r V} k_{V,t} f_{t+s} + \delta \tag{5.5}$$

where f_t is the corresponding estimator function.

Remark. Recall from last section that the estimator function was defined for every $t \in \Lambda$ as

$$f_t := \int_\Lambda \rho_t(\sigma, \omega) P(d\sigma, d\omega).$$

Lemma 5.1.2. *Suppose the DS_V uniqueness condition holds. Let $\Lambda \subset \mathbb{Z}^d$ and $f_t \geq 0$ be given such that equation 5.5 above holds for all $s \in T(\Lambda)$. Let $\Delta \subset \Lambda$ be given and define $c(t) := \exp(-g_0 d(t, \Delta))$, defined for every $t \in \mathbb{Z}^d$. Then for some constants C and c (not depending on Λ) we have that*

$$\sum_{t \in \Lambda} f_t c(t) \leq C \sum_{\bar{\partial}_V \Lambda} f_t c(t) + \delta \frac{|\Lambda|}{c} \tag{5.6}$$

where $\bar{\partial}_V \Lambda := \{t \in \Lambda : d(t, \Lambda^c) \leq diam(V \cup \partial_r V)\}$, i.e. those points in distance $V \cup \partial_r V$ to the boundary of Λ.

Remark. Here g_0, c and C are some positive constant, depending on γ_3, and which will be specified in the proof.

We next proceed by first showing how these two lemmas implies the main theorem.

Proof of theorem. Let $\Delta \subset \Lambda$ for some finite $\Lambda \subset \mathbb{Z}^d$. Combining the two Lemmas we get that for every $\tau, \gamma \in \Omega$ there exist a measure $P \in K(\pi_\Lambda(\cdot|\tau), \pi_\Lambda(\cdot|\gamma))$ such that

$$\sum_{t \in \Lambda} f_t c(t) \leq C \sum_{\bar{\partial}_V \Lambda} f_t c(t) + \delta \frac{|\Lambda|}{c},$$

where $c(t)$ is defined for all $t \in \Lambda$ with respect to Δ. Furthermore, by construction we have that

$$\sum_{t \in \Lambda} f_t c(t) \geq \sum_{t \in \Delta} f_t \geq W_{\rho_\Delta}(\pi_\Lambda(\cdot|\tau), \pi_\Lambda(\cdot|\gamma))$$

Now, let $\Lambda = D_n$, the n-cube, and choose δ_n in such a way that $\lim_{n\to\infty} \delta_n |D_n| \to 0$. Moreover, as n goes to infinity, $c(t)$ becomes arbitrary small for $t \in \bar{\partial}_V(\Lambda)$ and hence $\sum_{t \in \bar{\partial}_V \Lambda} f_t c(t)$ converges to zero. This is so since f_t is bounded by $2\mathrm{Diam}(S)$ which is finite by the compactness assumption on S. This follows by applying the metric inequality, ie. for any coupling P and $s \in S$ we have that

$$f_t = \int \rho_t(\alpha, \beta) P(d\alpha, d\beta) \leq \int \rho_t(\alpha, s) + \rho_t(s, \beta) P(d\alpha, d\beta) \leq 2\mathrm{Diam}(S).$$

Thus, $\lim_{n\to\infty} W_{\rho_\Delta}(\pi_{\Lambda_n}(\cdot|\tau), \pi_{\Lambda_n}(\cdot|\gamma)) = 0$ for any sequence $\{\Lambda_n\}$ eventually covering $\mathbb{L}$. Since this holds for any Δ, τ and γ, by taking supremum over γ and τ this implies the uniqueness via condition 2. □

We now continue with the proofs of the two lemmas.

Proof of Lemma 5.1.1. The following proof is a simple extension of the Surgery Lemma for the Dobrushin uniqueness condition.

Let s be any element of $T(\Lambda)$, and let P be a coupling of $K(\pi_\Lambda(\cdot|\tau), \pi_\Lambda(\cdot|\tau))$ such that $\sum_{t\in\Lambda} f_t \leq W_{\rho_\Lambda}(\pi_\Lambda(\cdot|\tau), \pi_\Lambda(\cdot|\gamma)) + \delta$. Such a measure exists by definition of W_{ρ_Λ} as the optimal coupling, and since $\sum_{t\in\Lambda} f_t = \int_\Lambda \rho_V(\omega,\sigma)P(d\omega,d\sigma)$.

Remark. We have earlier seen that the optimal coupling exists. The δ allows us to take a coupling which is not necessarily the optimal, and hence extending the condition to hold for separable metric spaces.

By the possession of a DS_V estimator we can construct a measure $P^{\sigma,\omega}(\cdot,\cdot) \in K(\pi_{V+s}(\cdot|\sigma), \pi_{V+s}(\cdot|\omega))$ for any $\sigma,\omega \in \Omega$ such that

$$\int_{V+s} \rho_{V+s}(\tau,\gamma)P^{\sigma,\omega}(d\tau,d\gamma) \leq \sum_{t\in\partial_r V+s} k_{V,t}\rho_t(\sigma,\omega) + \delta$$

This can be seen by introducing a sequence of configurations $\{\sigma_i\}_{i\geq 0}$ which converges to ω, where $\sigma_0 = \sigma$, and $\sigma_i = \sigma_{i-1}$ on all sites exept at site i where we flip and let it equal ω_i. Then we have that $W_{\rho_{V+s}}(\pi_{V+s}(\cdot|\omega), \pi_{V+s}(\cdot|\sigma)) \leq \sum_i W_{\rho_{V+s}}(\pi_{V+s}(\cdot|\sigma_i), \pi_{V+s}(\cdot|\sigma_{i+1}))$ by telescoping and using the metric inequality.

We now perform what is often referred to as the surgery operation by applying the measures $P^{\sigma,\omega}$ as probability kernels. More formally, we define the measure $\tilde{P}$ by

$$\tilde{P}(f) = \int\int f(\eta,\tau)P^{\sigma,\omega}(d\eta,d\tau)P(d\sigma,d\omega) \tag{5.7}$$

Now, $\tilde{P}$ is in $K(\pi_\Lambda(\cdot|\tau), \pi_\Lambda(\cdot|\gamma))$ by the self consistency of the specification Π.

We define $\tilde{f}_t$ similarly as f_t, only now with respect to our new coupling $\tilde{P}$. Thus, $f_{t+s} = \tilde{f}_{t+s}$ for all $t \in \Lambda \setminus V$ and $\sum_{t\in V} \tilde{f}_{t+s} \leq \sum_{t\in\partial_r V} k_{V+s,t+s}\tilde{f}_{t+s} + \delta$. By translation invariance this is equivalent to

$$\sum_{t\in V} \tilde{f}_{t+s} \leq \sum_{t\in\partial_r V} k_{V,t}\tilde{f}_{t+s} + \delta$$

Hence, by considering the measure of P and $\tilde{P}$ which minimizes $\sum_{t\in\Lambda} f_t$ shows the existence of such a measure for a given $s \in T(\Lambda)$. Applying the same procedure to the minimal measure to each $s \in T(\Lambda)$ completes the proof.

□

We continue with the proof of the second lemma. Remark that in the proof the coupling does not play any role. What is essential is that we possess an inequality similar to the one obtained in Lemma 5.1.1. Moreover, it is the following argument which is in need of the finite range property.

Proof of Lemma 5.1.2. By assumption $\sum_{t\in V} f_{t+s} \leq \sum_{t\in\partial_r V} k_{V,t} f_{t+s} + \delta$. Now, multiply with $c(s)$ and sum over all $s \in T(\Lambda)$ to get

$$\sum_{s\in T(\Lambda)} c(s) \sum_{t\in V} f_{t+s} \leq \sum_{s\in T(\Lambda)} c(s) \sum_{t\in\partial_r V} k_{V,t} f_{t+s} + \sum_{s\in T(\Lambda)} \delta$$

By reordering the summation this is equivalent to

$$\begin{aligned} &\sum_{t\in\Lambda} f_t\Big(\sum_{s\in\mathbb{Z}^d : t-s\in V} c(s) - \sum_{s\in(T(\Lambda)^c) : t-s\in V} c(s)\Big) \\ \leq\ &\sum_{t\in\Lambda} f_t\Big(\sum_{s\in\mathbb{Z}^d : t-s\in\partial_r V} k_{V,t-s}c(s) - \sum_{s\in(T(\Lambda)^c) : t-s\in\partial_r V} k_{V,t-s}c(s)\Big) + \sum_{s\in T(\Lambda)} \delta \end{aligned}$$

Which implies the following inequality

$$\begin{aligned}&\sum_{t\in\Lambda} f_t\Big(\sum_{s\in\mathbb{Z}^d:t-s\in V} c(s)-\sum_{s\in\mathbb{Z}^d:t-s\in\partial_r V} k_{V,t-s}c(s)\Big)\\ \leq\ &\sum_{t\in\Lambda} f_t\Big(\sum_{s\in(T(\Lambda)^c):t-s\in V} c(s)-\sum_{s\in(T(\Lambda)^c):t-s\in\partial_r V} k_{V,t-s}c(s)\Big)\Big)+|\Lambda|\delta\end{aligned}$$

We want to use the last inequality to show the wanted bound. For this we introduce the quantities MAX and MIN given by

$$MIN := \exp(-g_0\mathrm{Diam}(V\cup\partial_r V)\leq \exp(g_0(\mathrm{dist}(t,\Delta)-\mathrm{dist}(s,\Delta)))\leq \exp(+g_0\mathrm{Diam}(V\cup\partial_r V) := MAX$$

For the expression on the left hand side of the inequality we have that for any $t\in\Lambda$

$$\begin{aligned}&\sum_{s\in\mathbb{Z}^d:t-s\in V} c(s)-\sum_{s\in\mathbb{Z}^d:t-s\in\partial_r V} k_{V,t-s}c(s)\\ &\geq c(t)|V|MIN-|V|\gamma MAX c(t)=c(t)|V|(MIN-\gamma_3 MAX)\end{aligned}$$

by applying the DS_V condition. By tuning g_0 we may let MIN become so close to MAX that $MIN-\gamma MAX>0$ (e.g. let $g_0\in(0,\frac{-\ln\gamma_3}{2\mathrm{Diam}(V\cup\partial_r V)})$). For the right hand side we have for every $t\in\Lambda$ that

$$\sum_{s\in(T(\Lambda)^c):t-s\in V} c(s)-\sum_{s\in(T(\Lambda)^c):t-s\in\partial_r V} k_{V,t-s}c(s)$$

is non-zero only for $t\in\bar{\partial}_V\Lambda$. Moreover, similar as for the left hand side we can bound this from above by $c(t)|V|(MAX-\gamma MIN)$ giving us the inequality

$$\sum_{t\in\Lambda} f_t c(t)|V|(MIN-\gamma_3 MAX)\leq\sum_{t\in\bar{\partial}_V\Lambda} f_t c(t)|V|(MAX-\gamma_3 MIN)+\delta|\Lambda|$$

By dividing by $|V|(MIN-\gamma MAX)$ we obtain the claimed inequality with $C=\frac{MAX-\gamma_3 MIN}{MIN-\gamma_3 MAX}$ and $c=\frac{1}{MIN-\gamma_3 MAX}$. □

5.1.2 Mixing conditions

An issue with the Dobrushin-Shlosman uniqueness condition, versus the Dobrushin uniqueness condition, is the concept of Mixing. Recall that in the Dobrushin regime we could prove that if the estimator α satisfied the decay property with constant γ, then for every $f\in L_\Lambda$

$$\delta_y(\pi_\Lambda f)\leq\frac{1}{1-\gamma}G_1 e^{-md(\Lambda,y)}$$

for some $m>0$ and constant G_1. However, as discussed in [10] this does not hold under the Dobrushin-Shlosman condition. There they prove that only a "very strong decay property" (see equation 3.10) is satisfied under the DS_V condition, that is,

$$\delta(\pi_\Lambda f)\leq G_2\sum_{t\in\bar{\partial}_V\Lambda} e^{-gd(t,\Lambda)} \tag{5.8}$$

for some $g, G_2>0$ (depending on Λ and γ_2), where $\bar{\partial}_{V\Lambda}=\{u\in\Lambda: d(u,\Lambda^c)\leq\mathrm{Diam}(V\cup\partial_R V)\}$, hence only those sites well inside the Λ, and depending on the range R and the volume V.

Originally we believed that we could improve this inequality substantively and give a relation more similar to the strong mixing relation, even in the Dobrushin-Shlosman regime. The idea was that the finite range property should not matter in the mixing condition, only the size of the volume V should play a role, giving a inequality similar to

$$\delta_y(\pi_\Lambda f) \leq \frac{1}{1-\gamma}\delta^\rho_{\Lambda^*}(f)G_3 e^{-md(\Lambda^*,y)} \tag{5.9}$$

for some $g, G_3 > 0$ with $\Lambda^* = \{z \in \Lambda \mid d(z,\Lambda^c) > |V|\}$ and $f \in L_{\Lambda^*}$. However, it appeared in the final writings that our proof of such a statement contained a vital mistake, and we have so far not been able to correct this mistake.

In a preliminary paper to the Dobrushin-Shlosman uniqueness condition, [9], the mixing condition we have seen were confused and assumed to be the same. The Czech models were introduced by the Czech PhD-student Navratil, in the context of criticizing this paper, and given as a counterexample. The prototype of such models have a unique measure in the full space, but non-uniqueness in a half space (for instance, if $\Omega = \mathbb{Z}^d$, the set of positive vectors forms a half-space, see [31]). These models also showed that the uniqueness criteria is not exhausting, as first claimed in [9]. This was however corrected in [10], were they prove that for a large class of models, the Dobrushin-Shlosman uniqueness criteria indeed is exhaustive, in that by taking larger volumes V, the range of models satisfying the uniqueness condition approach the phase transition boundary. Moreover, in [11] a large class of models are classified under the term complete analyticity which is highly related to the issue of strong mixing properties.

5.2 Lieb-Aizenman Uniqueness condition

In the dual approach the DS_V estimator for the Dobrushin-Shlosman uniqueness condition is written as $\delta^\rho_y(\pi_V f) \leq \delta^\rho_V(f)\alpha_{V,y}$. In [1] a similar uniqueness condition appear, however only shown for the Stochastic Ising model (see for instance [26] for a detailed introduction to this model), and to our knowledge never directly compared with the coupling approach originally taken by Dobrushin and Shlosman. The condition originates from unpublished work by Lieb and Aizenman, hence the name Lieb-Aizenman uniqueness condition. A throughout derivation of the Lieb-Aizenman condition for general models, i.e. not only restricted to the Stochastics Ising model, were in our hands by the unpublished lecture notes [19]. In this section we take a closer look at this condition and investigate its relation with the Dobrushin-Shlosman uniqueness condition. A proof of the validity of the condition follows from the derivation in the remaining sections were we even provide an extended version of the Lieb-Aizenman condition. The technique relies on the following variant of the estimator.

Definition 5.2.1 (Lieb/Aizenmann Estimator)**.** *Given a finite volume $V \subset \mathbb{L}$ the set $\{\alpha^V_{x,y}\}_{x\in V, y\in\mathbb{L}\setminus V}$ is said to be an A_V estimator if for every $y \in \mathbb{L}\setminus V$ and every function $f \in L_V$ it holds that*

$$\delta^\rho_y(\pi_V f) \leq \sum_{x\in V}\alpha_{x,y}\delta^\rho_x(f) \tag{5.10}$$

Given a A_V estimator, we similarly as in the Dobrushin-Shlosman case have the following condition which implies the uniqueness of the consistent measures.

Condition 8 (Lieb-Aizenmann Uniqueness Condition)**.** The Lieb-Aizenmann uniqueness condition is said to hold for a specification Π if there exist an A_V estimator $\{\alpha^V_{x,y}\}$ such that

$$\frac{1}{|V|}\sum_{x\in V\, y\in\mathbb{L}\setminus V}\alpha^V_{x,y} = \gamma_4 < 1 \tag{5.11}$$

Theorem 5.2.1. *Let Π be a quasilocal specification satisfying the Feller property, and Ω a compact Polish space. If the Lieb-Aizenman uniqueness condition holds for some volume $V \in \mathscr{L}$, then there exists exactly one measure μ consistent with Π.*

Seemingly, the Lieb-Aizenman uniqueness condition contain some more flexibility concerning the choice of estimator then the Dobrushin-Shlosman uniqueness condition as it may also depend on sites $x \in V$. However, we will see that this flexibility is mainly artificial.

Lemma 5.2.1. *Let $\Lambda \in \mathscr{L}$ and let $f \in L_\Lambda$. If $\rho_\Lambda \geq \max_{x\in\Lambda} \rho_x$, then*

$$\delta^\rho(f) \leq \sum_{x\in\mathbb{L}} \delta^\rho_x(f) \tag{5.12}$$

Moreover, if we let $\rho_\Lambda = \sum_{x\in\Lambda} \rho_x$, then

$$\delta^\rho_\Lambda(f) = \sup_{x\in\Lambda} \delta^\rho_x(f) \tag{5.13}$$

for any $f \in L_\Lambda$.

Proof. Enumerate $\Lambda = \{1,2,\dots,n\}$. For any $\sigma = \omega$ off Λ let $\{\sigma^j\}_{j=0}^{n+1}$ be a sequence of flippings from σ to ω such that $\sigma^0 = \sigma$ and $\sigma^j = \omega_{x<j}\sigma_{x\geq j}$. The inequality follows by telescoping as shown next

$$\begin{aligned}
\delta^\rho(f) = & \sup_{\sigma=\omega \text{ off } \Lambda} \frac{|f(\sigma)-f(\omega)|}{\rho_\Lambda(\sigma,\omega)} \\
\leq & \sup_{\sigma=\omega \text{ off } \Lambda} \sum_j \frac{|f(\sigma^j)-f(\sigma^{j+1})|}{\rho_\Lambda(\sigma,\omega)} \\
= & \sup_{\sigma=\omega \text{ off } \Lambda} \sum_j \frac{|f(\sigma^j)-f(\sigma^{j+1})|}{\rho_j(\sigma,\omega)} \frac{\rho_j(\sigma,\omega)}{\rho_\Lambda(\sigma,\omega)} \\
\leq & \sup_{\sigma=\omega \text{ off } \Lambda} \sum_j \delta^\rho_j(f) \frac{\rho_j(\sigma,\omega)}{\rho_\Lambda(\sigma,\omega)} \\
\leq & \sum_j \delta^\rho_j(f)
\end{aligned}$$

Hence $\delta^\rho(f) \leq \sum_{x\in\mathbb{L}} \delta^\rho_x(f)$. In case $\rho_\Lambda = \sum_{x\in\Lambda} \rho_x$ it also follows that $\delta^\rho_\Lambda(f) = \max_{x\in\Lambda} \delta^\rho_x(f)$. This can be seen from the second line from below, as $\sum_{i=1}^n \delta^\rho_i(f) \frac{\rho_i(\sigma,\omega)}{\rho_\Lambda(\sigma,\omega)}$ forms a convex combination. Taking the supremum thus yield $\delta^\rho_\Lambda(f) = \max_{x\in\Lambda} \delta^\rho_x(f)$. □

The lemma above essentially imply that the Lieb-Aizenman uniqueness condition is equivalent to the condition obtained by Dobrushin and Shlosman, as shown next.

Proposition 5.2.1. *For the optimal choice of $\{\alpha^V_y\}_{y\notin V}$ satisfying Dobrushin-Shlosman uniqueness condition and $\{\alpha^V_{x,y}\}_{x\in V y\notin V}$ satisfying Lieb-Aizenmans condition we have that for any $y \notin V$*

$$\alpha^V_y = \sum_{x\in V} \alpha^V_{x,y} \tag{5.14}$$

Moreover, given an estimator for Lieb-Aizenman condition, then we also have an estimator satisfying Dobrushin-Shlosman criteria

Proof. For any $f \in L_V$, let $x_0 \in V$ be a site such that $\delta^\rho_{x_0}(f)$ is maximized. Then

$$\begin{aligned} & \sum_{x\in V} \delta^\rho_x(f)\alpha^V_{x,y} \\ = \quad & \delta^\rho_{x_0}(f)(\alpha^V_{x_0,y} + \sum_{x\neq x_0, x\in V} \frac{\delta^\rho_x(f)}{\delta^\rho_{x_0}(f)}\alpha^V_{x,y}) \\ \leq \quad & \delta^\rho_V(f)(\sum_{x\in V}\alpha^V_{x,y}) \qquad \text{as } \delta^\rho_V(f) = \max_{x\in V}\delta^\rho_x(f) \end{aligned}$$

Furthermore, by considering a function $f \in L_V$ with the property that $\delta^\rho_x(f) = \delta^\rho_z(f)$ for all $x, z \in V$, we see that the above inequality actually is an equality, which proves the claim. □

Remark. Above we did not need $\{\alpha^V_{x,y}\}$ to be optimal. The lemma tells us that given an estimator $\alpha^V_{x,y}$, then $\sum_{x\in V}\alpha^V_{x,y} = \alpha^V_y$ is also an estimator in the setup of Dobrushin and Shlosman. Moreover, the optimal estimators satisfies this equality. Note that here it is essential that we use $\rho_V = \sum_{x\in V}\rho_x$. On the other hand, the setup by Lieb/Aizenman does only depend on the single site metrics and can thus be seen to be a generalization.

Remark. The optimal DS_V estimator is given by

$$\alpha_{V,y} = \sup_{\sigma=\omega \text{ off } y} \frac{1}{\rho_y(\sigma,\omega)} W_{\rho_V}(\pi^\sigma_V, \pi^\omega_V),$$

and is hence unique. The definition of the optimal A_V estimator on the other hand is not so straight forward, but an estimate can be found as follows. Enumerate $V = \{1,\dots,n\}$. Let $\sigma = \omega$ off y, then for every $f \in L_V$

$$\begin{aligned} & |\pi^\sigma_V f - \pi^\omega_V f| \\ \leq \quad & \int |f(\tau) - f(\eta)|\mu(d\tau, d\eta) \qquad \text{for any } \mu \in K(\pi^\sigma_V, \pi^\omega_V) \\ \leq \quad & \sum_{i=1}^n \int |f(\tau_{x\geq i}\eta) - f(\tau_{x>i}\eta)|\mu(d\tau, d\eta) \\ = \quad & \sum_{i=1}^n \int \frac{|f(\tau_{x\geq i}\eta) - f(\tau_{x>i}\eta)|}{\rho_i(\tau,\eta)}\rho_i(\tau,\eta)\mu(d\tau, d\eta) \\ \leq \quad & \sum_{i=1}^n \delta^\rho_i(f)\int \rho_i(\tau,\eta)\mu(d\tau, d\eta) \end{aligned}$$

Hence

$$\alpha^V_{x,y} = \sup_{\sigma=\omega \text{ off } y} \frac{1}{\rho_y(\sigma,\omega)} \inf_{P\in K(\pi^\sigma_V, pi^\omega_V)} \int \rho_x(\eta,\tau)P(d\eta, d\tau) \tag{5.15}$$

is more or less the best estimator one can have. Any approximation of $\alpha_{V,y}$ by applying a suitable coupling procedure immediately yields an A_V estimator too. However it is worth noticing that, unlike for the DS_V estimator, the optimal $\alpha^V_{x,y}$ is not unique as the optimal coupling in general is not.

In the proceeding we will continue in the formulation similar to Lieb-Aizenman. This will be convenient when generalizing the tools given in the former chapter for extending the results of Dobrushin and Shlosman beyond translation invariance and finite range. It should however be emphasized that the techniques we will use does not depend essentially on whether we use the coupling approach or by averaging over observables. The essential is that we obtain an inequality similar to the first Lemma above, a generalization of the Surgery Lemma.

5.3 Dobrushin Uniqueness Condition for Larger Volumes

We proceed by generalizing the results obtained in the previous chapter. To obtain the strongest results concerning uniqueness we assume given a quasilocal specifications, Π, on a Polish space Ω, with μ and ν two corresponding consistent measures. As earlier we assume that we are given a spread estimator $\{\lambda_z\}_{z\in\mathbb{L}}$ such that

$$|\mu(f) - \nu(f)| \leq \sum_{z\in\mathbb{L}} \delta_z^\rho(f)\lambda_z \tag{5.16}$$

Let $\Theta = \{\theta_n\}_{n\geq 1}$ be any partition of $\mathbb{L}$, consisting of non-overlapping sets such that each site $x \in \mathbb{L}$ is contained in exactly one element θ_n. Then we have the following generalized version of an estimator for Π.

Definition 5.3.1. *Given a partition Θ of $\mathbb{L}$, then we say that $\alpha^\Theta = \{\alpha^\Theta_{x,y}\}_{x,y\in\mathbb{L}}$ forms an estimator for Π if for each $\theta \in \Theta$*

$$\delta_y^\rho(\pi_\theta f) \leq \sum_{x\in\theta} \delta_x^\rho(f)\alpha^\Theta_{x,y} \tag{5.17}$$

for every $f \in L_\theta$.

Similar as the original Dobrushin estimator where each θ only contained a single site, the operator π_θ can be seen as a duster over all of θ.

Lemma 5.3.1 (Dusting Lemma)**.** *Let Π be a specification. Given a partition Θ of $\mathbb{L}$ and a corresponding estimator $\{\alpha^\Theta_{x,y}\}$, then for every $g \in L$ we have that for any $\theta \in \Theta$*

$$\delta_y^\rho(\pi_\theta g) = \begin{cases} = & 0, & \text{if } y \in \theta \\ \leq & \delta_y^\rho(g) + \sum_{x\in\theta} \delta_x^\rho(g)\alpha^\Theta_{x,y}, & \text{if } y \notin \theta \end{cases} \tag{5.18}$$

Proof. For $y \in \theta$, $\sup_{\omega=\sigma \text{ off } y} |\pi_\theta(g|\sigma) - \pi_V(g|\omega)| = 0$ since $\sigma = \omega$ on the boundary of θ.

For $y \notin \theta$ we have that for any $\sigma = \omega$ off y

$$\begin{aligned} & \frac{|\pi_\theta(g|\sigma) - \pi_\theta(g|\omega)|}{\rho_y(\sigma,\omega)} \\ = & \frac{|\pi_\theta(g_{\sigma_{\theta^c}}|\sigma) - \pi_\theta(g_{\omega_{\theta^c}}|\omega)|}{\rho_y(\sigma,\omega)} \qquad \text{where } g_{\sigma_{\theta^c}}(\omega) = g(\omega_\theta\sigma_{\theta^c}) \\ \leq \frac{|\pi_\theta(g_{\sigma_{\theta^c}} - g_{\omega_{\theta^c}}|\sigma)|}{\rho_y(\sigma,\omega)} + & \frac{|\pi_\theta(g_{\omega_{\theta^c}}|\sigma) - \pi_\theta(g_{\omega_{\theta^c}}|\omega)|}{\rho_y(\sigma,\omega)} \end{aligned}$$

The first term is bounded by $\delta_y^\rho(g)$ since, after integrating the functions only differs at y. The second term is bounded by $\delta_y^\rho(\pi_\theta g_{\omega_{\theta^c}})$, which, since $g_{\omega_{\theta^c}}$ only depends on values in θ, is bounded by $\sum_{x\in\theta} \delta_x^\rho(g)\alpha^\Theta_{x,y}$.

□

The interpretation is that by applying the duster π_θ we clean the sites inside θ. However, by doing so some dust is spread to the neighboring sites. $\alpha^\Theta_{x,y}$ is thus a measure of how much dust spreads from x to y when cleaning with the dusters π_θ, $\theta \in \Theta$. The next lemma shows how we can update the spread estimator by use of the estimator α^Θ, i.e., how to make it cleaner.

Lemma 5.3.2. *Let λ be a spread estimator for μ and ν, and let f be any quasilocal function in L. Let $\theta \in \Theta$. Then there exist an spread estimator $\tilde{\lambda}$ given by*

$$\tilde{\lambda}_x \begin{cases} = & \lambda_x & \text{if } x \notin \theta \\ = & \sum_{z\notin\theta} \alpha^{\Theta}_{x,z}\lambda_z & \text{if } x \in \theta \end{cases} \tag{5.19}$$

Proof. For updating the spread estimator λ, let f be a quasilocal function in L. Then for any $\theta \in \Theta$ it follows similar as in previous updating lemmas that

$$\begin{aligned} & |\mu(f) - \nu(f)| \\ = \; & |\mu(\pi_\theta f) - \nu(\pi_\theta f)| && \text{by the DLR-equations} \\ \leq \; & \sum_{z\notin\theta} \delta^{\rho}_z(\pi_\theta f)\lambda_z \\ \leq \; & \sum_{z\notin\theta} \delta^{\rho}_z(f)\lambda_z + \sum_{x\in\theta} \delta^{\rho}_x(f)(\sum_{z\notin\theta} \alpha_{x,z}\lambda_z) && \text{by the Dusting Lemma} \end{aligned}$$

□

For the single-site Dobrushin condition we could proceed by a taking the site-wise minimum spread estimator. For a larger volumes θ, however, we have no guaranty that this actually yield a spread estimator. Assume $\theta = \{x,y\}$, and let $\tilde{\lambda}$ be the updated spread estimator. Then it may be the case that $\tilde{\lambda}_x < \lambda_x$ and $\lambda_y < \tilde{\lambda}_y$, and hence the analysis tells us nothing about whether or not the site-wise minimum is a spread estimator. To overcome this problem we instead proceed as in the second proof of the Dobrushin uniqueness condition (see section 4.2.2), hence introducing the following condition.

Condition 9. Given a partition Θ of $\mathbb{L}$. The Dobrushin uniqueness condition is said to hold for Θ and a specification Π if there exists an corresponding estimator $\{\alpha^{\Theta}_{x,y}\}_{x,y\in\mathbb{L}}$ and a constant $\gamma_1 \geq 0$ such that

$$\sup_{x\in\mathbb{L}} \sum_{y\in\mathbb{L}:\, y\neq x} \alpha^{\Theta}_{x,y} = \gamma_1 < 1 \tag{5.20}$$

Assume further that the spread estimators are uniformly bounded, say by the constant c, which is the case for every measure in the Wasserstein space, and let $\lambda_x \equiv c$. By applying the updating lemma over a set θ we get the new estimator $\tilde{\lambda}$ which equals c for each $x \notin \theta$, and where

$$\tilde{\lambda}_x = \sum_{z\notin\theta} \alpha^{\Theta}_{x,z}\lambda_z \leq c\gamma_1 \tag{5.21}$$

Hence, this implies that the updated spread estimator also is the site-wise minimum, and so we can proceed as in the single-site Dobrushin condition. Enumerating the elements of the partition Θ and applying the updating rule to each of them we thus obtain the new spread estimator $T(\lambda)$, given by

$$(T(\lambda))_x \leq c\gamma_1$$

Iterating this procedure further thus provides us with a sequence of spread estimators converging uniformly to 0.

Theorem 5.3.1. *Given two measures μ and ν, consistent with the quasilocal specifications Π, defined on a Polish space Ω. Let Θ be any partition of $\mathbb{L}$ and let α^{Θ} be an estimator for Π. If the general Dobrushin uniqueness condition is satisfied, then the Wasserstein space $W^1_\rho(\Omega)$ contains at most one measure μ consistent with Π.*

Remark. For the single-site Dobrushin uniqueness condition we gave in theorem 4.4.1 a more general condition. Whether $[\alpha^{\Theta}]^n \to 0$ as $n \to \infty$ is a sufficient condition for absent of several consistent measures is still an open question.

As seen in the previous section, the Lieb-Aizenman uniqueness condition considers the average over all $x \in \theta$, and hence is still a better condition when it is applicable. The initial approach to show its validity, as seen in [19], was simply by taking a convex combination of such spread estimators.

Lemma 5.3.3. *Let $\{\lambda^j\}_{1\leq j\leq n}$ be a set of spread estimators. Then any convex combination $\sum_{i=1}^n \omega_i \lambda^j$ is also a spread estimator.*

Proof. Let $\sum_{i=1}^n \omega_i = 1$ with each $\omega_i \geq 0$ and let $\{\lambda_x^i\}$ be a set of n spread estimators for μ and ν. Then

$$\sum_{z\in\mathbb{L}} \delta_x(f) \sum_{i=1}^n \omega_i \lambda_x^i = \sum_{i=1}^n \omega_i \sum_{x\in\mathbb{L}} \delta_x(f)\lambda_x^i \geq \sum_{i=1}^n \omega_i |\mu(f) - \nu(f)| = |\mu(f) - \nu(f)| \tag{5.22}$$

□

However, one is once again put up against the dilemma whether or not one can update the spread estimator site-wise.[3] Hence we are lead to proceed by other methods, which actually will turn out to yield an improvement of the Lieb-Aizenman uniqueness condition.

5.4 Extension of Lieb-Aizenmans Condition

Let $\Theta = \{\theta_i\}_{i\geq 1}$ be a collection of finite subsets of $\mathbb{L}$ such that each site $x \in \mathbb{L}$ is contained in at least one of the elements. We call such a collection a cover of $\mathbb{L}$. Assign to each element θ_i a weight $\omega_i > 0$ such that $\sum_i \omega_i = c < \infty$. If not otherwise stated we will assume $c = 1$.

Definition 5.4.1. *Given a cover Θ of $\mathbb{L}$, then we say that $\alpha^{\Theta} = \{\alpha^{\Theta}_{x,y}\}_{x,y\in\mathbb{L}}$ forms an estimator for Π if for each $y \in \mathbb{L}$,*

$$\delta_y^{\rho}(\sum_{i\geq 1} \omega_i \pi_{\theta_i} f) \leq \sum_{x\neq y} \delta_x^{\rho}(f)\alpha^{\Theta}_{x,y} \tag{5.23}$$

for every $f \in L_{\partial_y\Theta}$, where $\partial_y\Theta = \{\theta \in \Theta : y \notin \theta\}$.

In the definition above $\delta_x^{\rho}(f) = 0$ for each $x \notin \partial_y\Theta$. Hence the corresponding $\alpha^{\Theta}_{x,y}$ may be chosen arbitrarily, and we will in the proceedings assume that it equals 0. The reason why they are included is only to lighten the notation.

Lemma 5.4.1 (Dusting Lemma)**.** *Let Π be a specification. Given a cover Θ of $\mathbb{L}$ and a corresponding estimator $\{\alpha^{\Theta}_{x,y}\}$, then for every $g \in L$ we have that for any $y \in \mathbb{L}$,*

$$\delta_y(\sum_{i\geq 1} \omega_i \theta_i g) \leq \delta_y^{\rho}(g)(1-\omega_y) + \sum_{x\neq y} \delta_x^{\rho}(f)\alpha^{\Theta}_{x,y} \tag{5.24}$$

where $\omega_y = \sum_{i:y\in\theta_i} \omega_i$.

[3]The proof in [19] of Lieb-Aizenman depends seemingly on taking a site-wise spread estimator

Proof. For any $\sigma = \tau$ off y we have

$$
\begin{aligned}
& \frac{1}{\rho_y(\sigma,\tau)}|\sum_{i\geq 1}\omega_i(\pi^{\sigma}_{\theta_i}(g)-\pi^{\tau}_{\theta_i}(g))| \\
= & \frac{1}{\rho_y(\sigma,\tau)}|\sum_{i\geq 1}\omega_i(\pi^{\sigma}_{\theta_i}(g_\sigma)-\pi^{\tau}_{\theta_i}(g_\tau))| \quad \text{where } g_\tau(\sigma)=g(\sigma_{\partial_y\Theta}\tau_{(\partial_y\Theta)^c}) \\
\leq & \frac{1}{\rho_y(\sigma,\tau)}|\sum_{i\geq 1}\omega_i(\pi^{\sigma}_{\theta_i}(g_\sigma)-\pi^{\tau}_{\theta_i}(g_\sigma))| + \frac{1}{\rho_y(\sigma,\tau)}|\sum_{i\geq 1}\omega_i\pi^{\tau}_{\theta_i}(g_\sigma-g_\tau)| \\
\leq & \delta^{\rho}_y(\sum_{i\geq 1}\omega_i\pi_{\theta_i}g_\sigma) + \frac{1}{\rho_y(\sigma,\tau)}|\sum_{i:\,\theta_i\in\partial_y\Theta}\omega_i\pi^{\tau}_{\theta_i}(g_\sigma-g_\tau)| \\
\leq & \sum_{x\neq y}\delta^{\rho}_x(g)\alpha^{\Theta}_{x,y} + \delta^{\rho}_y(g)(1-\omega_y)
\end{aligned}
$$

Where the last line follows as $\sum_{i:\,\theta_i\in\partial_y\Theta}\omega_i = 1-\omega_y$.

□

Assuming as earlier that we are given a spread estimator for μ and ν we get the following updating lemma.

Lemma 5.4.2. *Given a spread estimator $\{\lambda_x\}_{x\in\mathbb{L}}$ for μ and ν. Then $\{\lambda^1_x\}_{x\in\mathbb{L}}$ given by*

$$\lambda^1_x = (1-\omega_x)\lambda_x + \sum_{y\neq x}\alpha^{\Theta}_{x,y}\lambda_y \tag{5.25}$$

also forms a spread estimator.

Proof. Similar as for the other updating lemmas we have that for any quasilocal $f \in L$

$$
\begin{aligned}
& |\mu(f)-\nu(f)| \\
= & |\mu(\sum_{i\geq 1}\omega_i\pi_{\theta_i}f) - \nu(\sum_{i\geq 1}\omega_i\pi_{\theta_i}f)| \\
\leq & \sum_{z\in\mathbb{L}}\delta^{\rho}_z(\sum_{i\geq 1}\omega_i\pi_{\theta_i}f)\lambda_z \\
\leq & \sum_{z\in\mathbb{L}}[\delta^{\rho}_z(f)(1-\omega_z)\lambda_z + \sum_{x\neq z}\delta^{\rho}_x(f)\alpha^{\Theta}_{x,z}\lambda_z] \\
= & \sum_{z\in\mathbb{L}}\delta^{\rho}_z(f)[(1-\omega_z)\lambda_z + \sum_{y\neq z}\alpha^{\Theta}_{z,y}\lambda_y]
\end{aligned}
$$

□

Thus, if there exists a uniformly bounded for the spread estimator such that we may consider $\lambda_x = c < +\infty$ for each $x \in \mathbb{L}$, then similar as in the first approach in this section we get the following natural condition.

$$\sup_{x\in\mathbb{L}}\frac{1}{\omega_x}\sum_{y\in\mathbb{L}:\,y\neq x}\alpha^{\Theta}_{x,y} = \gamma_5 < 1 \tag{5.26}$$

Hence, under this condition, by implying the updated spread estimator we get that

$$\lambda^1_x = (1-\omega_x)\lambda_x + \gamma_5\omega_x c \tag{5.27}$$

for any initial spread estimator λ_x. We claim that by implying the updating lemma successively yield the spread estimator $\{\lambda_x^n\}_{x\in\mathbb{L}}$ given by

$$\lambda_x^n = (1-\omega_x)^n\lambda_x + \omega_x\gamma_5 c\sum_{j=0}^{n-1}(1-\omega_x)^j \tag{5.28}$$

This follows simply by induction. It clearly holds for $n = 0$. Assume it holds for $k-1$, then we have that

$$\begin{aligned}\lambda_x^k &= (1-\omega_x)(1-\omega_x)^{k-1}\lambda_x + \omega_x\gamma_5 c\sum_{j=0}^{k-2}(1-\omega_x)^j) + \gamma_5\omega_x c \\ &= (1-\omega_x)^k\lambda_x + \omega_x\gamma_5 c\sum_{j=0}^{k-1}(1-\omega_x)^j\end{aligned}$$

Hence, by taking the limit as $n\to\infty$ implies that $\lambda_x^\infty = \omega_x\gamma_5 c\frac{1}{\omega_x} = \gamma_5 c$. As this holds for any $x\in\mathbb{L}$, this shows that we can improve the uniform bound on the spread estimator by the factor γ_5. Iterating the whole process thus leads to a geometric convergence of the spread estimator, implying that the optimal spread estimator is the zero vector, and hence also that $\mu = \nu$.

Theorem 5.4.1. *Given two measures μ and ν consistent with the quasilocal specifications Π and defined on a Polish space Ω. Let Θ be a cover of $\mathbb{L}$ and let α^Θ be a corresponding estimator for Π. If $\sup_x \frac{1}{\omega_x}\sum_{y\neq x}\alpha_{x,y} < 1$, then the Wasserstein space $W_\rho^1(\Omega)$ contains at most one measure μ consistent with Π.*

Actually, we can get more out of this condition. Recall from the update lemma that

$$\lambda_x^1 = (1-\omega_x)\lambda_x + \sum_{y\neq x}\alpha_{x,y}^\Theta\lambda_y$$

In matrix notation this thus yields

$$\lambda^1 = \alpha^\Theta\lambda \tag{5.29}$$

where α is the interaction matrix with $\alpha_{x,x} = (1-\omega_x)$. Iterating thus yields $\lambda^n = \alpha^n\lambda$, given the following generalization.

Condition 10. Given a cover Θ of $\mathbb{L}$ and corresponding set of summable weights. The extended Lieb-Aizenman condition is said to hold for a specification Π if there exists an corresponding estimator $\{\alpha_{x,y}^\Theta\}_{x,y\in\mathbb{L}}$ such that

$$\lim_{n\to\infty}[\alpha]^n = 0. \tag{5.30}$$

Theorem 5.4.2. *Given two measures μ and ν consistent with the quasilocal specifications Π and defined on a Polish space Ω. If the extended Lieb-Aizenman condition is satisfied for some cover Θ with corresponding summable, then the Wasserstein space $W_\rho^1(\Omega)$ contains at most one measure μ consistent with Π.*

The extended Lieb-Aizenman put in the dusting interpretation tells us that it may also be sufficient to apply a duster which does not completely clean the sites which it is dusting. What is more important is that the total amount of dust decreases in the long term. It should also be noted that we could apply local dusters such as $\frac{1}{2}\pi_x + \frac{1}{2}\pi_z$ for some sites $x,z\in\mathbb{L}$ in the same way. However, one then needs several different dusters such that in the end one cover the whole space.

5.5 Derivation of the Lieb-Aizenman Condition

We claimed that the new condition is an extension of the Lieb-Aizenman uniqueness condition to systems which may not be translation invariant and are not restricted to have finite range only. For this we need an estimate of the optimal estimator under the extended version. For any $\sigma = \tau$ off y and $f \in L_{\partial_y \Theta}$ we have that

$$\begin{aligned}
& |\sum_{i\geq 1} \omega_i(\pi^{\sigma}_{\theta_i} f - \pi^{\tau}_{\theta_i} f)| \\
\leq & \sum_{i\geq 1} \omega_i |\pi^{\sigma}_{\theta_i} f - \pi^{\tau}_{\theta_i} f| \\
= & \sum_{i\geq 1} \omega_i |\int f(\alpha) - f(\beta) P^{\sigma,\tau}_{\theta_i}(d\alpha, d\beta)| \qquad \text{where } P^{\sigma,\tau}_{\theta_i} \in K(\pi^{\sigma}_{\theta_i}, \pi^{\tau}_{\theta_i}) \\
\leq & \sum_{i\geq 1} \omega_i \int |f(\alpha) - f(\beta)| P^{\sigma,\tau}_{\theta_i}(d\alpha, d\beta)
\end{aligned}$$

For each θ_i and $\alpha = \beta$ off θ_i consider a telescoping path from α to β, i.e $\{\alpha_j\}_{j=0}^{|\theta_i|}$ where $\alpha_0 = \alpha$ and $\alpha_{|\theta_i|} = \beta$. Then we have further that

$$\begin{aligned}
& \sum_{i\geq 1} \omega_i \int |f(\alpha) - f(\beta)| P^{\sigma,\tau}_{\theta_i}(d\alpha, d\beta) \\
\leq & \sum_{i\geq 1} \omega_i \int \sum_{j=1}^{|\theta_i|} |f(\alpha_{j-1}) - f(\alpha_j)| P^{\sigma,\tau}_{\theta_i}(d\alpha, d\beta) \\
= & \sum_{i\geq 1} \omega_i \sum_{j=1}^{|\theta_i|} \int \frac{|f(\alpha_{j-1} - f(\alpha_j)|}{\rho_j(\alpha,\beta)} \rho_j(\alpha,\beta) P^{\sigma,\tau}_{\theta_i}(d\alpha, d\beta) \\
\leq & \sum_{i\geq 1} \omega_i \sum_{j=1}^{|\theta_i|} \delta^{\rho}_j(f) \int \rho_j(\alpha,\beta) P^{\sigma,\tau}_{\theta_i}(d\alpha, d\beta).
\end{aligned}$$

Reordering the terms we see that this is equivalent to

$$= \sum_{z\in\mathbb{L}} \delta^{\rho}_z(f) \sum_{i\in B(z)} \omega_i \int \rho_z(\alpha,\beta) P^{\sigma,\tau}_{\theta_i}(d\alpha, d\beta)$$

where $B(z) = \{i : z \in \theta_i\}$. Hence by taking the supremum of $\sigma = \tau$ off y we see that an estimate of $\alpha^{\Theta}_{x,y}$ is to set

$$\alpha^{\Theta}_{x,y} = \sup_{\sigma=\tau \text{ off } y} \frac{1}{\rho_y(\sigma,\tau)} \sum_{i\in B(z)} \omega_i \int \rho_x(\alpha,\beta) P^{\sigma,\tau}_{\theta_i}(d\alpha, d\beta)$$

This estimate is less than $\sum_{i\in B(z)} \omega_i \sup_{\sigma=\tau \text{ off } y} \frac{1}{\rho_y(\sigma,\tau)} \int \rho_x(\alpha,\beta) P^{\sigma,\tau}_{\theta_i}(d\alpha, d\beta)$, which is seen to yield a convex combination of optimal estimators in the Lieb-Aizenman case (see equation 5.2). Hence our new condition is an improvement of the Lieb-Aizenman condition. On the other hand, we require that the weights are summable, while in the setting of Lieb-Aizenman the weights are constant. However, in the special case of $\mathbb{L}$ being the integer lattice we may overcome this problem by a simple approximation argument.[4]

[4] We learned the following argument from the derivation of Dobrushin-Shlosman uniqueness condition in [42]

Let $\mathbb{L} = \mathbb{Z}^d$ and let V be some finite volume in $\mathbb{L}$. Assume Π is translation invariant and consider Θ as the collection of all translations of V in $\mathbb{Z}^d$. From the derivation above $\alpha_{x,y}^{\Theta,1} \leq \sum_u \alpha_{x,y}^{V+u}$, where $\alpha_{x,y}^{V+u}$ is the the estimator from the Lieb-Aizenman condition and $\alpha_{x,y}^{\Theta,1}$ means that we are considering weights of 1 only. Hence, assuming that the Lieb-Aizenman condition holds, i.e. that $\sum_{x\in V}\sum_{y\notin V}\alpha_{x,y}^{V} < |V|$ we get that

$$\sum_{y\neq x}\alpha_{x,y}^{\Theta,1} \leq \sum_{y\neq x}\sum_{u}\alpha_{x,y}^{V+u} = \sum_{u\in V}\sum_{y\neq x}\alpha_{x-u,y-u}^{V} = \sum_{x\in V}\sum_{y\notin V}\alpha_{x,y}^{V} < |V|$$

Consider now the same systems, but assign to each $V_i \in \Theta$ the weight $\omega_i = (1+\varepsilon)^{-d(0,V_i)}$ for some $\varepsilon > 0$. The weights are clearly summable. Indeed, there are exactly $4n$ elements in $\mathbb{Z}^d$ of distance $n \geq 1$ from the origin, and so $\sum_{i\geq 1}\omega_i \leq |V|(1+\sum_{n=1}^{\infty}4n(1+\varepsilon)^{-n}) < \infty$. Moreover,

$$\begin{aligned} & \sum_{i\in B(x)}\omega_i\int \rho_x(\alpha,\beta)P_{\theta_i}^{\sigma,\tau}(d\alpha,d\beta) \\ \leq \quad & \frac{1}{|V|}\omega_x(1+\varepsilon)^r\sum_{i\in B(x)}\int \rho_x(\alpha,\beta)P_{\theta_i}^{\sigma,\tau}(d\alpha,d\beta) \end{aligned}$$

where $r = \mathrm{Diam}(V)$. Hence,

$$\sum_{y\neq x}\alpha_{x,y}^{\Theta,\omega} \quad \leq \quad \sum_{y\neq x}\frac{1}{|V|}\omega_x(1+\varepsilon)^r\alpha_{x,y}^{\Theta,1} \quad \leq \quad \omega_x \tag{5.31}$$

for some $\varepsilon > 0$ under the Lieb-Aizenman condition as $\sum_{y\neq x}\alpha_{x,y}^{\Theta,1} < |V|$. Thus we have proven that the Lieb-Aizenman condition implies our claimed extended version. Moreover, the proof above also holds beyond finite range specifications, extending the condition to what was previously known.

Remark. As seen from the derivation above, it seems convenient to compare the optimal estimators in a coupling perspective. This perspective will be made more explicit in the next chapter. Note also that our estimate $\alpha_{x,y}^{\Theta}$ is not the optimal one, since we have taken the supremum inside the summation of $\sum_{z\in\mathbb{L}}$.

Remark. It should be noted that what was essential above is that we can approximate the extended Lieb-Aizenman condition to hold beyond summable weights. Indeed, what we used was that $\sum_{y\neq x}\alpha_{x,y}^{\Theta,1} < |V|$, which is even stronger than the original Lieb-Aizenman condition. Moreover, for this we did not make use of the assumed translation invariance. It should be noted that a similar approach as above also could have been put forward for the column sum, that is under the condition $\sup_y\sum_{x\neq y}\alpha_{x,y}^{\Theta,\omega} < \omega_y$.

Chapter 6

Comparison of Uniqueness Conditions

In this chapter we introduce some other known uniqueness conditions and compare them to our various conditions seen in the previous chapters. First we introduce two uniqueness conditions given by D. Weitz in [42] following similar ideas as the initial one by Dobrushin, however written in a more dynamical approach. Next we discuss yet another uniqueness condition introduced in [3] by J. van den Berg and C. Maes which displays a relation between uniqueness of consistent measures and the critical probability for site-percolation. These "new" conditions are then compared to our "old" conditions seen in the previous chapter, and in particular the relation between a coupling approach and an averaging approach is highlighted. In the end section we review some of the unanswered question in this paper, and sketch areas which are in need of further investigation.

6.1 Dynamical Approach to Dobrushin Uniqueness Condition

In the original paper [42] by D. Weitz, the spin space S is assumed to be finite. Moreover, only specifications of nearest-neighbor type are considered, that is, specifications constructed via a Hamiltonian given by

$$H_\Lambda(\sigma) := \sum_{x\sim y:\,\{x,y\}\cap\Lambda\neq\emptyset} U_{x,y}(\sigma_x,\sigma_y) + \sum_{x\in\Lambda} U_x(\sigma_x) \tag{6.1}$$

for some pair potential $U_{x,y} : S\times S \mapsto \mathbb{R}\cup\{\infty\}$ and self-potential $U_x : S \mapsto \mathbb{R}\cup\{\infty\}$. Remark that the potential may be infinite, and so the hard-core model is for instance included in the set up. From the Hamiltonian the specification is then given by

$$\pi_\Lambda(\sigma|\tau) := \begin{cases} \frac{1}{Z_\Lambda^\tau}\exp(-H_\Lambda(\sigma)) & \text{if } \sigma=\tau \text{ off } \Lambda \\ 0 & \text{otherwise} \end{cases} \tag{6.2}$$

Moreover, these are only defined for so called feasible boundary conditions so that also Non-Gibbsian models, such as the Hard-core model, is within the range of models considered.

As we know, the given specification satisfies the DLR conditions $\pi_\Lambda\pi_\Delta = \pi_\Lambda$ for any $\Delta\subset\Lambda\in\mathscr{L}$. In Weitz dynamical approach this is seen as a stationary condition on the random process which at each instance of time picks a finite subset at random and updates the configuration. Thus, the DLR condition can be restated as saying that choosing a configuration according to π_Λ and then by π_Δ is the same as choosing a configuration according to π_Λ only.

Similar as for the extended Lieb-Aizenman condition we introduce a collection of finite blocks (regions) $\{\theta_i\}_{i\geq 1}$, which cover $\mathbb{L}$ finitely many times. Assign also a positive weight ω_i to each block θ_i. The update rule is then defined as follows.

Definition 6.1.1 (Local Update Rule)**.** *$\kappa = \{\kappa_i^\tau\}$ is said to be a local update rule for Π with respect to $\{\theta_i\}_{i\geq 1}$ if*

1. *For every τ and $i \in \mathbb{N}$, κ_i^τ is a probability distribution that agree with τ of θ_i.*
2. *For every feasible τ and i, $\pi_{\theta_i}^\tau$ is stationary under κ_i.*
3. *κ_i^τ and κ_i^σ projected on S^{θ_i} are the same whenever $\tau = \sigma$ on $\theta_i \cup \partial\theta_i$.*

A natural choice would be to let $\kappa_i^\tau = \pi_{\theta_i}^\tau$. However, by condition 3, κ_i^τ is also allowed to depend on the value of τ inside θ_i, and so other possibilities exists, such as the Metropolis update (see [42] page 5 for more details and section 5.3 for an example of a more sophisticated coupling).

From the definition of a Local Update Rule, the coupled update rule $\{K_i\}_{i\geq 1}$ is defined, where $K_i(\tau,\eta)$ is a coupling of κ_i^τ and κ_i^η (often taken as the optimal coupling). In the following we let $\rho_x(K_i(\eta,\xi))$ be the integral over ρ_x. Based on the Local Update Rule the following two definitions are given.

Definition 6.1.2 (Influence of a site)**.** *For a given coupled update rule K, the influence of a site y on site x, $I_{x\leftarrow y}$, is defined as the smallest constant for which, for all $\eta = \xi$ off y*

$$\sum_{i\in B(x)} \omega_i \rho_x(K_i(\eta,\xi)) \leq \rho_y(\eta,\xi) I_{x\leftarrow y} \tag{6.3}$$

where $B(x) = \{i \in \mathbb{N} : x \in \theta_i\}$.

Definition 6.1.3 (Total influence of a site)**.** *For a given coupled update rule K, the total influence of site y i, $I_{\leftarrow y}$, is defined as the smallest constant for which, for all $\eta = \xi$ off y*

$$\sum_{i\geq 1} \omega_i \rho_{\theta_i}(K_i(\eta,\xi)) \leq \rho_y(\eta,\xi) I_{\leftarrow y} \tag{6.4}$$

The first definition(6.1.2) combined with the following theorem is a rewriting or generalization of what we have called the Dobrushin uniqueness condition.

Theorem 6.1.1. *If a specification Π admits a coupled update rule K together with a bounded collection of metrics ρ, i.e. $\sup_{x\in\mathbb{L}} \sup_{\sigma=\tau \text{ off } x} \rho_x(\sigma,\tau) < \infty$, for which*

$$\sup_x \{\frac{1}{\omega_{B(x)}} \sum_y I_{x\leftarrow y}\} = \sup_x \frac{I_{x\leftarrow}}{\omega_{B(x)}} < 1 \tag{6.5}$$

where $\omega_{B(x)} = \sum_{i\in B(x)} \omega_i$ and $I_{x\leftarrow} = \sum_y I_{x\leftarrow y}$, then there exists exactly one Gibbs measure consistent with Π.

Similarly, the second definition (6.1.3) combined with the next theorem is a generalization of what we have called the dual-Dobrushin uniqueness condition, and in [42] referred to as a version of the Dobrushin-Shlosman condition.

Theorem 6.1.2. *If a specification Π admits a coupled update rule K together with a summable collection of metrics ρ, i.e. $\sum_{x\in\mathbb{L}} \sup_{\sigma=\tau \text{ off } x} \rho_x(\sigma,\tau) < \infty$, that satisfy $\sup_y \omega_{B(y)} < \infty$, $\inf \omega_{B(y)} > 0$ and*

$$\sup_y \{\frac{1}{\omega_{B(y)}} I_{\leftarrow y}\} < 1 \tag{6.6}$$

then there exists exactly one Gibbs measure consistent with Π.

Remark. As shown in section 5.2 of [42] the condition of the set of metrics being summable is essential. Thus the theorem does not apply in general when each distance are the same, e.g. $\rho_x = \rho$ for all x. However, by a similar argument as we used to show the Lieb-Aizenman condition in the previous chapter it can be shown to be valid for graphs with subexponential growth, e.g. like $\mathbb{Z}^d$. Another way to look at this statement for fixed site-metrics is that the weight most be summable, rather than the metrics. Lastly, in [43] the condition was improved to be correct without assuming $\inf_y \omega_{B(y)} > 0$. However, that $\sup_y \omega_{B(y)} < \infty$ is necessary, as shown in [42].

In proving the above stated theorems, Weitz uses the path coupling technique. The path coupling technique was formalized by M.E. Dyer and R. Bubley in [5] to construct couplings depending on two configurations, say σ and ω, based on a sequence of flippings (i.e., a path from σ to ω). It is essentially the same technique as originally used by Dobrushin in the Surgery Lemma, though put in a more general context. Weitz uses the path coupling technic to construct general couplings based on the single flipping coupling $K_i(\tau,\eta)$ (where $\tau = \eta$ off y for some $y \in \mathbb{L}$). Given a coupling, Q, he then constructs a new coupling, $F_S(Q)$ given by

$$F_S(Q) = \int K_S(\eta,\tau) Q(d\eta, d\tau),$$

where $K_S(\eta,\tau) = \sum_{i\in S} \frac{\omega_i K_i(\eta,\tau)}{\omega_S}$. Based on this he proofs

Lemma 6.1.1. *Given a coupled update rule K and a collection of metrics ρ. Let Q be any coupling, x any site and S any finite subset of $\mathbb{N}$ such that $B(x) \subset S$. Then, letting $\rho_x(Q) = \int \rho_x(\tau,\eta) Q(d\tau, d\eta)$,*

$$\rho_x(F_S(Q)) \leq \Big(1 - \frac{\omega_{B(x)}}{\omega_S}\Big)\rho_x(Q) + \frac{I_{x\leftarrow}}{\omega_S} \sup_{y\in\phi(B(x))} \rho_y(Q) \tag{6.7}$$

where $\phi(B(x)) = \cup_{i\in B(x)}(\theta_i \cup \partial\theta_i)$. Let Δ be any finite region, and S be such that $B(\Delta) \subset S$, then

$$\rho_\Delta(F_S(Q)) \leq \Big(1 - \frac{MAX + MIN}{2}\Big)\rho_\Delta(Q) + \frac{MAX}{\omega_S}\rho_{\phi(B(\Delta))}(Q) \tag{6.8}$$

where $MAX = \max_{y\in\phi(B(\Delta))}\{I_{\leftarrow y}\}$ and $MIN = \min_{y\in\phi(B(\Delta))}\{\omega_{B(y)} - I_{\leftarrow y}\}$.

The lemma can be seen as a version of what we have called the updating lemma adapted to Weitz settings. For instance, given a coupling Q with estimation function f_x, the first inequality tells that the updated coupling $F_S(Q)$ has estimation function $\tilde{f}_x$ satisfying

$$\tilde{f}_x \leq \Big(1 - \frac{\omega_{B(x)}}{\omega_S}\Big) f_x + \frac{I_{x\leftarrow}}{\omega_S} \sup_{y\in\phi(B(x))} f_y.$$

By applying the two lemmas Weitz shows, by use of induction, that for any finite $\Lambda \subset \mathbb{L}$ there exists for any $\sigma,\omega \in \Omega$ a coupling $Q_m \in K(\pi^\sigma_{\Lambda_n}, \pi^\omega_{\Lambda_n})$ such that $\rho_\Lambda(Q_m) \leq c|\Lambda|\alpha^m$, where $\alpha < 1$ and $c = \max_{x\in\Lambda_n} \sup_{s_1,s_2} \rho_x(s_1,s_2)$, which proves uniqueness under the Dobrushin uniqueness criteria (Total influence on a site). Moreover, for the dual uniqueness result (Total influence of a site) Weitz proves similarly that there exists a coupling $Q_m \in K(\pi^\tau_{\Lambda_{m+1}}, \pi^\eta_{\Lambda_{m+1}})$ such that $\rho_\Lambda(Q) \leq C\alpha^m$, where $C = \max_{\omega,\sigma} \rho_{\Lambda_m}(\sigma,\omega)$.

Remark. In [42] it is also commented that the proof should be valid for general finite range models. Moreover, by simply introducing the Wasserstein distance, rather then the Total Variation, the condition should easily be extendable beyond finite single spin space S.

6.2 Uniqueness Condition Related to Disagreement Percolation

In [3] another uniqueness condition was introduced, sort of similar to those we have seen, with connections to site-percolation.[1] For this we introduce the following estimator.

Definition 6.2.1. *Given a specification* Π*, then the set* $\{\alpha_x\}_{x\in\mathbb{L}}$ *is called a disagreement-estimator for* Π *if it satisfies*

$$\delta(\pi_x f) \leq \delta_x^\rho(f)\alpha_x \tag{6.9}$$

for every $f \in L_x$*. Equivalently, if for each* $\sigma, \tau \in \Omega$

$$||\pi_x^\sigma - \pi_x^\tau||_{TV} \leq \alpha_x \tag{6.10}$$

Thus, α_x represent the influence on x of flipping possibly all spins outside x. In [3], the following uniqueness condition is given

Theorem 6.2.1. [2] *Assume* S *finite and* Π *a specification of nearest neighbor type (i.e. only depend on the spin values at the boundary). Then, if*

$$\sup_{x\in\mathbb{L}} \alpha_x < p_c$$

where p_c *is the critical probability for the site-percolation on the underlying graph, then the exist exactly one measure consistent with* Π*.*

Thus the above condition depend on a good estimates of p_c, which in many situation can be very difficult to obtain itself. However, in [3] several examples are given, and in particular, for the Hard-Core Model the condition is shown to be better then the at that time known estimates by the Dobrushin-Shlosman condition.

6.3 Comparison of Uniqueness Conditions

In this section we summarize all the condition for uniqueness we have seen so far, and study which gives the better estimate. We do so by first looking at updates on single sites before looking at larger volumes. As we saw in the first section of this chapter, the coupling description by Weitz is more general then what we have studied so far. To be able to compare at this stage we thus restrict ourselves to so-called "heat-bath" updates, i.e. couplings of π_θ^σ and π_θ^τ. By assigning the weight 1 to each subset containing exactly one element, then the optimal conditions for only single-site updates can be summarized as shown in table 6.1.

The table contains four conditions. First is the original Dobrushin uniquness condition. Next is a condition by Weitz, the total influence of a site y on a site x. Third is the second condition by Weitz, the total influence of y. Last we have the condition by van den Berg and Maes. Comparing the optimal estimators, it is seen from the table that the first condition of Weitz is contained in the formulation of Dobrushin uniqueness condition. However, the second condition by Weitz is seen to be stronger then what we earlier called the dual-Dobrushin condition. This follows as

$$I_{\leftarrow y} = \sup_{\sigma=\tau \text{ off } y} \frac{1}{\rho_y(\sigma,\tau)} \sum_{y\neq x} W_{\rho_x}(\pi_x^\sigma, \pi_x^\tau) \leq \sum_{y\neq x} \sup_{\sigma=\tau \text{ off } y} \frac{1}{\rho_y(\sigma,\tau)} W_{\rho_x}(\pi_x^\sigma, \pi_x^\tau) = \sum_{x\neq y} \alpha_{x,y}$$

[1] A short introduction to site-percolation is given in their paper. [3]

[2] Corollary 2 in [3]

Optimal Estimator	Condition	Requirements
$\alpha_{x,y} = \sup_{f\in L_x} \sup_{\sigma=\tau \text{ off } y} \frac{\delta_y(\pi_x f)}{\delta_x^\rho(f)}$	$\alpha^n \to 0$	S Polish, Π quasilocal
$I_{x\leftarrow y} = \sup_{\sigma=\tau \text{ off } y} \frac{1}{\rho_y(\sigma,\tau)} W_{\rho_x}(\pi_x^\sigma, \pi_x^\tau)$	$\sup_x \sum_{y\neq x} I_{x\leftarrow y} < 1$	S discrete, Π finite range
$I_{\leftarrow y} = \sup_{\sigma=\tau \text{ off } y} \frac{1}{\rho_y(\sigma,\tau)} \sum_{y\neq x} W_{\rho_x}(\pi_x^\sigma, \pi_x^\tau)$	$\sup_y I_y < 1$	$\mathbb{L} \equiv \mathbb{Z}^d$, S discrete, Π finite range
$\alpha_x = \sup_{\sigma\neq\tau} W_{\rho_x}(\pi_x^\sigma, \pi_x^\tau)$	$\sup_x < p_c$	S finite, Π finite range

Table 6.1: Uniqueness conditions for single site

The extended Lieb-Aizenman condition should also be consider. It says that for $\mathbb{L} = \mathbb{Z}^d$ and any estimator satisfying

$$\delta_y(\sum_{x\in\mathbb{L}} \pi_x f) \leq \sum_{z\neq y} \delta_z^\rho(f)\alpha_{x,y}$$

implies uniqueness if $\lim_{n\to\infty}[\alpha]^n \to 0$. We saw earlier that for any $\sigma = \tau$ off y

$$|\sum_{x\in\mathbb{L}} (\pi_x^\sigma f - \pi_x^\tau f)| \leq \sum_{z\neq y} \delta_z^\rho(f) \int \rho_z(\alpha,\beta) P_x^{\sigma,\tau}(d\alpha, d\beta)$$

where $P_x^{\sigma,\tau} \in K(\pi_x^\sigma, \pi_x^\tau)$. Thus the requirement that $\sum_{y\neq x} \alpha_{x,y} \leq 1$ is better then the original Dobrushin condition, however difficult to write explicitly. On the other hand, $\sum_{x\neq y} \alpha_{x,y} = I_{\leftarrow y}$, and so the extended Lieb-Aizenman condition is a generalization of Weitz two conditions, at least for subexponential graphs.

Remark. The condition by van den Berg and Maes is somewhat different then the others and its hard to compare directly. An simple calculation shows that for the Ising model the Dobrushin uniqueness condition holds as long as $\beta < \frac{\ln 3}{4} \approx 0,275$, and by symmetry the conditions by Weitz yields the same. The condition by van den Berg and Maes on the other hand is satisfied whenever $\frac{e^{8\beta}-e^{-8\beta}}{2+e^{8\beta}+e^{-8\beta}} < p_c$. As commented in [3], p_c is expected to be less than $0,6$. Hence their condition applies only when for β values at least smaller then $0,175$ (exact estimates depends on knowledge of the critical probability for site percolation, which is an industry in itself). Hence, for this model, the Dobrushin-like conditions are seen to be better. However, as shown in [3] there are also models where their argument is better then the Dobrushin-like estimate, especially for models with hard-core constraints.

We continue to compare the conditions for general volumes. It is not known whether the condition by van den Berg and Maes can be extended to such cases, and it is therefore not included in the following. Assume given a finite cover $\Theta = \{\theta_i\}_{i\in\mathbb{N}}$ with corresponding weights ω_i. Moreover, we use the same notation, $\rho_x(\pi_\theta^\sigma, \pi_\theta^\tau)$, for the integral of ρ_x by the optimal coupling.

Optimal Estimator	Condition	Requirements
$\alpha_{x,y}^\Theta = \sup_{\sigma=\tau \text{ off } y} \frac{1}{\rho_y(\sigma,\tau)} \rho_x(\pi_{\theta_x}^\sigma, \pi_{\theta_x}^\tau)$	$\sup_x \sum_{y\neq x} \alpha_{x,y}^\Theta < 1$	S Polish, Π quasilocal
$\alpha_{V,y} = \sup_{\sigma=\tau \text{off} y} \frac{1}{\rho_y(\sigma,\tau)} W_{\rho_V}(\pi_V^\sigma, \pi_V^\tau)$	$\frac{1}{\lvert V\rvert} \sum_{y\notin V} \alpha_{V,y} < 1$	S Polish, Π finite range and TI.
$I_{x\leftarrow y} = \sup_{\sigma=\tau \text{ off } y} \frac{1}{\rho_y(\sigma,\tau)} \sum_{i\in B(x)} \omega_i \rho_x(\pi_{\theta_i}^\sigma, \pi_{\theta_i}^\tau)$	$\sup_x \frac{\sum_y I_{x\leftarrow y}}{\omega_{B(x)}} < 1$	S discrete, Π finite range
$I_{\leftarrow y} = \sup_{\sigma=\tau \text{ off } y} \frac{1}{\rho_y(\sigma,\tau)} \sum_{i:\, \theta_i\in\partial_R\{y\}} \omega_i W_{\rho_{\theta_i}}(\pi_{\theta_i^\sigma}, \pi_{\theta_i^\tau})$	$\sup_y \frac{I_y}{\omega_{B(y)}} < 1$	$\mathbb{L} \equiv \mathbb{Z}^d$, S discrete, Π finite range

Table 6.2: Uniqueness conditions for general volumes

In table 6.2 we first have the original Dobrushin condition for larger volumes, then the Dobrushin-Shlosman uniqueness condition, and then the two conditions by Weitz. It is seen that the Dobrushin condition holds for the largest class of models, for general Polish spaces and beyond finite range. The Dobrushin-Shlosman condition is known to hold under translation invariance and finite range, but with a Polish spin space. On the other hand, the conditions by Weitz is shown to hold only for finite spin space and nearest neighbors models, though claimed to be extendable to finite range and Polish spin space. Moreover, the second condition, the total influence of y, is seen to hold only for summable metrics, though an approximation argument extends beyond this for subexponential graphs. Under assumptions such that all the conditions are applicable, the Dobrushin condition is seen to be the weakest and then the Dobrushin-Shlosman. The two conditions conditions by Weitz are both better, though as they are dual in nature they are hard to compare with each other.

Lastly, we have the extended Lieb-Aizenman condition which is not included in the table as the optimal $\alpha^{\Theta}_{x,y}$ is difficult to write explicitly. As seen in section 5.4 the condition is valid for general quasilocal specifications on Polish spin space. Moreover, it extends the Dobrushin-Shlosman condition and when comparable with the Dobrushin condition, i.e. for subexponential graphs, it is also seen to be better. An upper bound of the optimal estimator was seen to be given by

$$\alpha^{\Theta}_{x,y} = \sup_{\sigma=\omega \text{ off } y} \frac{1}{\rho_y(\sigma,\tau)} \sum_{i\in B(x)} \omega_i \rho_x(\pi^{\sigma}_{\theta_i}, \pi^{\tau}_{\theta_i}) = I_{x\leftarrow y}.$$

Moreover for the column sum the optimal is seen to yield,

$$\sum_{x\neq y} \alpha_{x,y} = \sup_{\sigma=\tau \text{ off } y} \frac{1}{\rho_y(\sigma,\tau)} \sum_{x\neq y} \sum_{i\in B(x)} \omega_i \rho_x(\pi^{\sigma}_{\theta_i}, \pi^{\tau}_{\theta_i}) = I_{\leftarrow y}.$$

Hence the Lieb-Aizenman condition is seen to contain both of Weitz's condition, and moreover extend them to Polish spin spaces and beyond finite range. However, the condition is restricted to a summable weight set, which should be seen in relation to the summability condition of the metric in Weitz case, though the direct relation has not yet been put forward. As no such requirement is needed for Weitz's first condition, the total influence on x, this condition may yield better estimates in some cases (i.e. beyond subexponential graphs). On the other hand, the Lieb-Aizenman condition does not depend on a specific choice of the metrics on finite volume other than the single sites, as both Weitz's conditions and the Dobrushin-Shlosman condition does by assuming $\rho_V = \sum_{x\in V} \rho_x$.

6.4 Coupling vs Averaging

A central question in this paper has been the relation between the coupling approach and the averaging approach. In 2000, in the paper [35], A. Sokal raised the question whether one of the approaches is superior the other for showing Dobrushin-alike statements. Some years later Weitz presented his paper, [42], extending previous known results on the Dobrushin condition, moreover by using coupling arguments only. His approach was studied by several people, in particular by S. Winkler who in his PhD thesis, [43], conducted a comparison of the coupling approach and the averaging approach (or analytic approach as he calls it). It was there claimed that the question seemed to have a tendency towards the coupling method. With our results, in particular the extended Lieb-Aizenman condition, we claim to how balanced out this "competition". Indeed, it is our opinion that each approach by the coupling method can be described via averaging, and vice versa, as a result of the dual theorem for the Wasserstein metric. On the other hand we find it more natural to write the optimal estimators in the language of coupling, rather than the dual description. Moreover, as seen in the paper by Weitz, by the coupling approach one can extend the results beyond so-called ”heat-bath” coupling. How such an extension can be written in the averaging approach is not clear. Also for computational reasons the coupling approach seems to be favored, though this does not imply that it is better by theoretical means.

6.5 Future Work

The theory concerning uniqueness of consistent measures and the application of Dobrushins technique for uniqueness is far from being settled. There are still many open questions and connections which are ready to be explored. In this section we review some of the challenges which remains in our approach, and where we believe new discoveries possibly can be found. Moreover, we also give a short guide into topics in much relation with what is presented in this paper.

The main questions directly related to the research conducted while writing this paper are presented next.

1. Given a condition written by means of coupling, how to describe it by means of observables, and visa versa? Is there a natural way to translate such statements?

2. Given the interaction I of the elements $I_{x\leftarrow y}$ from Weitz first condition. Is the condition $\lim_{n\to\infty}[I]^n = 0$ sufficient for having uniqueness of the consistent measures?

3. What is the relation between the optimal estimators $\alpha_{x,y}$ for single site volumes, and $\alpha^V_{x,y}$ for a volume V? Can we estimate one of them by the other in an efficient way?

4. What is the relation between Total Variation distance and Wasserstein distance for general Polish spaces. Which one is preferable to use for computations? Can one improve the calculation of the conditions by using the Wasserstein distance for some models?

Another question which we have been dealing with is the mixing conditions the different uniqueness conditions impose. It was for a long time our belief that we could improve the mixing condition in the Dobrushin-Shlosman region, but our initial approach contained an vital mistake and is hence not included. For more on these issues one has to take into consideration the existence of exotic models, such as the Czech models seen in [31]. Moreover, the notion of complete analyticity will play a vital role, as seen in [11]. We next summarize other fields with connections to the Dobrushin uniqueness condition which would be of interest to study more deeply in the future.

1. There are deep connections between the Dobrushin uniqueness condition and so-called Logarithmic Sobolev Inequalities (LSI) as seen in [38] by D. Stroock and B. Zegarlinksi and [41] by B. Zegarlinski. Moreover, as for instance considered in [2] by F. Barthe and A. V. Kolesnikov, the LSI's are in much relation to newly developed theory in the field of Optimal Transportation.

2. Another issue is if one can improve the condition in specific cases by use of various cost functions instead of restricting to a metric as in the Wasserstein distance. In particular, the Wasserstein distance of order 2 is mainly used when studying Ricci curvature from differential geometry via Markov chains theory, see for instance [40] by C. Villani, [28] by Y. Ollivier and [29] by Y. Ollivier and A. Joulin.

3. A variant of the Dobrushin condition has also been written for interacting particle systems and probabilistic cellular automata, especially the ε-M criteria in [26] by Liggett, and the papers [32] and [33] by Shlosman and C. Maes.

4. The new conditions by Weitz is motivated by problems in Combinatorial Optimization and the search for improved mixing condition for various algorithms. There have been published many papers on these topics the last decade. We mention in particular the paper [16] by M. Dyer, L.A. Goldberg and M. Jerrum, and the paper [17] by M. Dyer, A. Sinclair, E. Vigoda and D. Weitz.

Bibliography

[1] Aizenman, M., and Holley, R. (1987) Rapid Convergence to Equilibrium of Stochastic Ising Models in the Dobrushin Shlosman regime. In: Percolation theory and ergodic theory of infinite particle systems, p. 1–11, Springer, New York.

[2] Barthe, F. and Kolesnikov, A. V. (2007) Mass transportation and variants of the logarithmic Sobolev inequality. In: J. Geom. Anal., vol. 18, p. 921–979.

[3] van den Berg, J. and Maes, C. (1994) Disagreement Percolation in the study of Markov Fields, The Annals of Probability, Vol. 22, p. 749–763.

[4] Bovier, A. (2006) Statistical Mechanics of disordered systems. A mathematical perspective. Cambrigde Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

[5] Bubley, R. and Dyer, M. E. (1997) Path coupling: A technique for proving rapid mixing in Markov chains. In: Proc. 38th Annual Symposium on Foundations of Computer Science, p. 223–231.

[6] Dobrushin, R. L. (1968), The description of a random field by means of conditional probabilities and conditions of its regularity. In: Theory of Probability and its applications, Volume XIII, no. 3, p. 197–224 (Translated from Russian).

[7] Dobrushin, R. L. (1970), Definition of a system of random variables by means of conditional distributions. In: Theor. Probability Appl., vol XV, no. 3, p. 458–486.

[8] Dobrushin, R. L. (1996), Perturbation methods of the theory of Gibbsian fields. In: Lectures on probability and statistics (Saint-Flour), Lecture notes in Math., 1648, Springer, Berlin, p. 1–66.

[9] Dobrushin, R. L. Pecherski, E. A. (1981) Uniqueness conditions for finitely dependent random fields. In ”Random fields”, vol. 1, North-Holland, Amsterdam-Oxford-N.Y, p 223–262.

[10] Dobrushin, R. L. and Shlosman, S. B. (1985) Constructive criteria for the uniqueness of Gibbs field. In: Statistical physics and dynamical systems, Birkhauser, Boston, p. 347–370.

[11] Dobrushin, R. L. and Shlosman, S. B. (1985) Completely analytical Gibbs fields. In: Statistical physics and dynamical systems, Birkhauser, Boston, p. 371–403.

[12] Dobrushin, R. L. and Shlosman, S. B. (1987) Completely analytical interactions: constructive description. In: J. Statist. Phys. 46, no. 5-6, p. 983–1014.

[13] Dobrushin, R. L., Kolafa, J., Shlosman, S. B. (1985) Phase diagram of the two-dimensional Ising antiferromagnet. Computer-assisted proof. In: Comm. Math. Phys. 102, no. 1, p. 89–103.

[14] Dudley, R. (1989) Probability and Real Analysis. The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.

[15] Dyer, M., Goldberg, L. A., Jerrum, M. (2009) Matrix norm and rapid mixing for spin systems. In: Ann. Appl. Probab. 19, no. 1, p. 71–107.

[16] Dyer, M., Goldberg, L. A., Jerrum, M. (2008) Dobrushin condition and systematic Scan. In: Combin. Probab. Comput. 17, no. 6, p. 761–779.

[17] Dyer, M., Sinclair, A., Vigoda, E., Weitz, D. (2004) Mixing in Time and Space for Lattice Spin Systems: A Combinatorial View, Random Structures Algorithms 24, no. 4, p. 461–479.

[18] van Enter, A. C. D., Fernandez, R., Sokal, A. D. (1993); Regularity Properties and Phatologies of Position-Space Renormalization-Group Transformations: Scope and Limitations of Gibbsian Theory. In: Journal of Statistical Physics, Vol 72, p. 879– 1167.

[19] Fernandez, R. (2000) High-temperature regime, unpublished notes.

[20] Fernandez, R. and Maillard, G. (2006) Construction of a specification from its singleton part. In: ALEA Lat. Am. J. Probab. Math. Stat. 2, p. 297–315.

[21] Fernandez, R. (2006), Gibbsianness and non-Gibbsiannes in lattice random fields. In: Mathematical Statistical Physics, Elsevier B. V., Amsterdam, p. 731–799.

[22] Föllmer, H. (1982), A covariance Estimate for Gibbs Measures. In: Journal of Functional Analysis, 46, no. 3, p. 387–395.

[23] Georgii, H. O. (1988), Gibbs measures and phase transitions. de Gruyter Studies in Mathematics, 9. Walter de Gruyter & Co., Berlin

[24] Georgii, H.O., Haggstrom, O., Maes, C. (2001), The random geometry of equilibrium phases. In: Phase transistions and critical phenomena, Vol. 18, p.1–142.

[25] Levin, D. A., Peres, Y., Wilmer, E. L. (2009) Markov Chains and Mixing Times. American Mathematical Society, Providence, RI.

[26] Liggett, T. M. (1985) Interacting Particle Systems. Grundleheren der Mathematischen Wissenschaften, 276. Springer-Verlag, New York.

[27] Lindvall, T. (1992) Lectures on the Coupling Method. Wiley Series in Probability and Mathematical Statistics John Wiley & Sons, Inc. New York.

[28] Ollivier, Y. (2009) Ricci curvature of Markov chains on metric spaces. In: J. Funct. Anal. 156, no. 3, p. 810–864.

[29] Ollivier, Y. and Joulin, A. (2010) Curvature, Concentration and Error Estimates for Markov Chain Monte Carlo. In: The Annals of Probability, Vol 38, No. 6, p. 2418–2442.

[30] T. de la Rue, R. Fernandez, A.D Sokal (2008); How to Clean a Dirty Floor: Probabilistic Potential Theory and the Dobrushin Uniqueness Theorem. In: Markov Processes and Related Fields, vol. 14, p. 1–78.

[31] S.B. Shlosman ; Uniqueness and half-space nonuniqueness of Gibbs states in Czech models. In: Theoretical and Mathematical Physics, Vol. 66, No. 3, p. 284–293 (Translated from Russian).

[32] Shlosman, S. B. and Maes, C. (1991) Ergodicity of probabilistic cellular automata: a constructive criterion. In: Comm. Math. Phys. 135, no. 2, p. 233–251.

[33] Shlosman, S. B. and Maes, C. (1993) When is an interacting particle system ergodic? In: Comm. Math. Phys. 151, no. 3, p. 447–466.

[34] Simon, B. (1993), The statistical mechanics of lattice gases. Vol. 1. Princeton Series in Physics. Princeton University Press, Princeton, NJ.

[35] Sokal, A. D. (2000) A personal List of Unsolved Problems Concerning Lattice Gases and Antiferromagnetic Potts Models, Inhomogeneous random systems. In: Markov Processes and Related Fields, vol. 7, no. 1, p. 21–38.

[36] Spreij, P. J. C. (2010), Measure Theoretic Probability. Lecture notes, Mastermath, Universiteit van Amsterdam, `http://staff.science.uva.nl/~spreij/onderwijs/master/mtp.pdf`

[37] Stroock, D. W., and Zegarlinski, B. (1992), The Logarithmic Sobolev Inequality for Discrete Spin Systems on a Lattice. In: Comm. Math. Phys. 149, no. 1, p. 175–193.

[38] Stroock, D. W. and Zegarlinski, B. (1992) The equivalence of the logarithmic Sobolev inequality and the Dobrushin-Shlosman mixing condition. In: Comm. Math. Phys. 144, no. 2, p. 303–323.

[39] Villani, C. (2003) Topics in optimal transportation. Graduate Studies in mathematics, 58. American Mathematical Society, Provdence, RI.

[40] Villani, C. (2009) Optimal Transportation - Old and new. Grundlehren der mathematischen Wissenschaften, 338. Springer-Verlag, Berlin.

[41] Zegarlinski, B. (1992) Dobrushin uniqueness theorem and logarithmic Sobolev inquealities. In: J. Funct. Anal. 105, no. 1, p. 77–111.

[42] Weitz, D. (2005), Combinatorial Criteria for Uniqueness of Gibbs Measures. In: Random Structures Algorithms, vol. 27, no. 4, p. 445–475.

[43] Winkler, S. N. (2007), Uniqueness of Gibbs Measures with Applications to Gibbs Sampling and the Sum-Product Algorithm, PhD thesis, Yale University. `http://gradworks.umi.com/3267392.pdf`