Author: Berghout, S.
Title: Gibbs processes and applications
Issue Date: 2020-10-27

Gibbs Processes and Applications

Proefschrift

ter verkrijging
van de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 27 oktober 2020
klokke 13.45 uur

door

Steven Berghout

geboren te Krimpen aan den IJssel
in 1987.

Promotoren:

Prof. dr. E.A. Verbitskiy (Universiteit Leiden & Rijksniversiteit Groningen)
Prof. dr. W. Th. F. den Hollander

Samenstelling van de promotiecommissie

Prof. dr. P.D. Grunwald
Prof. dr. R.M. van Luijk
Prof. dr. A.C.D. van Enter (Rijksuniversiteit Groningen)
Prof. dr. C. Külske (Ruhr-Universität Bochum)
Prof. dr. A. Le Ny (Université Gustave-Eiffel, Paris)

This work was supported by the Netherlands Organisation for Scientific Research (NWO), grant 613.001.218.

Contents

Chapter 1

Introduction

In this thesis we will be primarily interested in discrete-time random processes – sequences of random variables indexed by $\mathbb{Z}$, and random fields – collections of random variables indexed by points of d-dimensional lattices $\mathbb{Z}^d$, $d \geq 2$. Furthermore, we will always assume that the random variables take values in some finite set (alphabet) $\mathscr{A}$.

In order to describe a random process or a random field one has to define the corresponding *probability model*, which we understand as a triple $(\Omega, \mathfrak{B}, \mu)$, where $\Omega = \mathscr{A}^{\mathbb{Z}} = \{\omega : \mathbb{Z} \to \mathscr{A}\}$ or $\Omega = \mathscr{A}^{\mathbb{Z}^d} = \{\omega : \mathbb{Z}^d \to \mathscr{A}\}$, the Borel σ-algebra $\mathfrak{B}$ is of subsets of Ω, and μ is some probability measure on the measurable space $(\Omega, \mathfrak{B})$.

The most intricate part of probabilistic modeling of random processes and random fields is the selection of an appropriate probability measure μ. One possible approach to the definition of an $\mathscr{A}$-valued ($|A| < \infty$) collection of random variables (process, field) indexed by $t \in T$, where T is a countable set, is by prescribing a *consistent* family of finite-dimensional marginal distributions

$$\left\{\mathbb{P}_{t_1,\ldots,t_k}(F_1 \times \ldots \times F_k) : \; t_1, \ldots, t_k \in T, \; F_1, \ldots, F_k \subseteq \mathscr{A}\right\}. \tag{1.1}$$

Here the family is called consistent if

$$\mathbb{P}_{t_1,\ldots,t_k} = \mathbb{P}_{\pi(t_1),\ldots,\pi(t_k)}$$

for any permutation π, and

$$\mathbb{P}_{t_1\ldots t_k}(F_1 \times \cdots \times F_k) = \mathbb{P}_{t_1\ldots t_k, t_{k+1},\ldots,t_{k+m}}(F_1 \times \cdots \times F_k \times A \times \cdots \times A)$$

for all $t_1, \ldots, t_{k+m} \in T$ and $F_1, \ldots, F_k \subseteq A$. Then, by the Kolmogorov extension theorem, there exists a probability space $(\Omega, \mathscr{F}, \mu)$ and a stochastic process $\{X_t :$

$\Omega \to \mathscr{A}\}$ such that

$$\mathbb{P}_{t_1 \dots t_k}(F_1 \times \cdots \times F_k) = \mu(X_{t_1} \in F_1, \dots, X_{t_k} \in F_k)$$

for all $t_1, \dots, t_k \in T$ and $F_1, \dots, F_k \subseteq A$.

An alternative approach to defining interesting classes of probability measures is based on prescribing the dependence structure of the underlying stochastic processes. Such models can be useful as approximations of 'real-life' or physical systems. Yet models with very wild dependence structures are neither very realistic nor very susceptible to analysis. Modern probability theory has identified a wide variety of useful probabilistic models of varying complexity.

1.1 Finite-range dependence models: Markov processes and fields

Markov chains have been introduced by Andrey Markov more than a hundred years ago as a model of the distribution of vowels and consonants in Pushkin's poem Eugene Onegin. Today, Markov processes are the most popular and the best studied examples of a probabilistic model of stochastic processes with an explicitly given dependence structure.

The characteristic property of Markov chains – the so-called Markov property – states that the conditional probability distribution of future values of the process, conditional on both past and present value, depends only on the present:

$$\mu(X_{n+1} = a_{n+1} \,|\, X_n = a_n, X_{n-1} = a_{n-1}, \dots) = \mu(X_{n+1} = a_{n+1} \,|\, X_n = a_n).$$

The beauty and simplicity of the model, as well as the intrinsic richness of the resulting class of stochastic processes, have led to both popularity of the model and a wide range of applications: from linguistics to bioinformatics, from queueing theory to Google's PageRank algorithm.

Similarly, a Markov random field is a random field where, for each finite domain $\Lambda \subset \mathbb{Z}^d$, the conditional probability of $a_\Lambda \in \mathscr{A}^\Lambda$ depends only on the values $a_{\partial\Lambda}$ on the boundary $\partial\Lambda = \{n \in \mathbb{Z}^d : \mathrm{dist}(n, \Lambda) = 1\}$:

$$\mu\big(X_\Lambda = a_\Lambda \,|\, X_{\mathbb{Z}^d \setminus \Lambda} = a_{\mathbb{Z}^d \setminus \Lambda}\big) = \mu\big(X_\Lambda = a_\Lambda \,|\, X_{\partial\Lambda} = a_{\partial\Lambda}\big).$$

Two famous examples of Markov random fields are the Ising and Potts models of Statistical Physics. Both will be discussed in this thesis.

1.2 Long-range dependence models

The finite-range models: Markov processes and random fields, already form a rich class of probabilistic models. However, there is a clear need to introduce and study more flexible models without implicit assumptions of bounded dependence. Various models of such kind have been proposed in Probability Theory, Statistics, Information Theory, and Statistical Mechanics. In this thesis we will focus on two particular classes:

- In dimension $d = 1$, the class of g-measures, also known as chains with infinite connections – a very natural generalization of Markov measures;
- In dimension $d \geq 1$, the class of Gibbs measures – probabilistic models originating in Statistical Mechanics.

1.2.1 g-measures.

For random processes, the natural extension of the Markov property is the requirement that the conditional probability of the present value depends continuously on infinitely many past values. The resulting process has *infinite memory*, but dependence on the past values of the process gets weaker as the distance to the origin increases. For historical reasons, we will consider conditional probabilities conditioned on future values. Let $\Omega_+ = \mathscr{A}^{\mathbb{Z}_+}$, where $\mathscr{A}$ is some finite set, and denote by $T : \Omega_+ \to \Omega_+$ the left shift on Ω_+.

Definition 1.1. Let $G(\Omega_+)$ be the set of all positive continuous functions

$$g : \Omega_+ \to (0,1)$$

that are *normalized* in the sense that

$$\sum_{a \in \mathscr{A}} g(ax) = 1$$

for all $x \in \Omega_+ = \mathscr{A}^{\mathbb{Z}_+}$.

Here, $ax \in \Omega_+$ denotes a configuration obtained by concatenation of the symbol a with an infinite string of letters x, i.e., $ax = (a, x_0, x_1, \ldots)$. Note that, since g is continuous, the n^{th} variation of g satisfies

$$\text{var}_n(g) \equiv \sup_{x,y \in \mathscr{A}^{\mathbb{Z}_+}} \left| g(x_0^\infty) - g(x_0^n y_{n+1}^\infty) \right| \to 0 \text{ as } n \to \infty,$$

Finally, define a g-measure as follows:

Definition 1.2. A translation invariant measure μ_+ on Ω_+ is called a g-measure for $g \in G(\Omega_+)$ if

$$\mu_+(x_0|x_1^\infty) = g(x_0^\infty)$$

for μ-a.e. $x \in \Omega_+$.

For any $g \in G(\Omega_+)$ the set of g-measures is not empty, and may contain several measures. Various conditions for uniqueness of g-measures have been established [5, 18, 31, 35, 45, 48, 83]. Moreover, since μ is translation invariant, we can use translation to uniquely extend the g-measure to $\Omega = \mathscr{A}^{\mathbb{Z}}$.

The above conditions typically relate to the continuity of the g-function. A simple but rather strong uniqueness condition is the summability of $\mathrm{var}_n(g)$:

$$\sum_{n=1}^{\infty} \mathrm{var}_n(g) < \infty.$$

Johansson and Öberg [48] established uniqueness under square summability of variations:

$$\sum_{n=1}^{\infty} \big(\mathrm{var}_n(g)\big)^2 < \infty.$$

Alternatively one can define a g-measure for a function g as an equilibrium state on Ω_+ for a potential $\phi = \log(g)$ [59]. An equilibrium state for a potential ϕ is a translation invariant probability measure that satisfies the variational principle:

$$h(\mu, S) + \int \phi \, d\mu = \sup_{\nu \in \mathscr{M}_S^1(\Omega_+)} \Big[h(\nu, S) + \int \phi \, d\nu \Big],$$

where $h(\lambda, S)$ denotes the Kolmogorov-Sinai entropy of the left-shift $S : \Omega+ \mapsto \Omega+$ and the S-invariant measure λ, and the supremum is taken over the set $\mathscr{M}_S^1(\Omega_+)$ of all translation invariant probability measures on Ω_+.

1.2.2 Gibbs states

The work by Dobrushin [22] and, independently, by Lanford and Ruelle [56] in the late 1960's allowed for a rigorous mathematical definition of Gibbs measures (states) for infinite particle systems. The central idea is to prescribe probabilities of configurations in finite volumes, conditioned on the configuration outside that volume. It turns out that if the collection of such conditional distributions for different finite volumes, called a specification, is *consistent*, then there exists at least one probability measure whose conditional probabilities coincide with those prescribed by the specification. It is possible that there are multiple probability measures consistent with a given specification. In this case we say that a phase transition occurs.

We will restrict ourselves to systems on the lattice $\mathbb{Z}^d$, $d \geq 1$, and to finite alphabets $|\mathscr{A}| < \infty$. Such systems cover many interesting physically relevant examples, including those with phase transitions.

We start with some definitions and notation. Let the configuration space $\Omega = \mathscr{A}^{\mathbb{Z}^d}$ be equipped with the product topology. We denote by $\mathfrak{B}(\Omega)$ the corresponding Borel $\sigma-$algebra. For a configuration $x \in \Omega$ we denote by x_n its value at site $n \in \mathbb{Z}^d$. Similarly, for a set $\Lambda \subset \mathbb{Z}^d$ we denote by $x_\Lambda = (x_n : n \in \Lambda)$ the restriction of x to Λ. For disjoint sets $V, W \subset \mathbb{Z}^d$, $V \cap W = \emptyset$, we denote the concatenation of $\tilde{x}_V$ and x_W by $\tilde{x}_V x_W$, i.e.,

$$(\tilde{x}_V x_W)_n = \begin{cases} \tilde{x}_n : n \in V \\ x_n : n \in W. \end{cases}$$

Finally, in $d = 1$, we use the shorthand notation $x_n^m = x_{n,n+1,\ldots,m}$.

We next turn to the definition of Gibbs states. The first important notion is that of an *interaction*.

Definition 1.3. Let $\{\Phi_\Lambda\}$ be a collection of functions on $\Omega = \mathscr{A}^{\mathbb{Z}^d}$, indexed by finite subsets $\Lambda \subset \mathbb{Z}^d$ (denoted $\Lambda \Subset \mathbb{Z}^d$), such that for all $\Lambda \Subset \mathbb{Z}^d$

$$\Phi_\Lambda(x) = \Phi_\Lambda(x_\Lambda),$$

i.e., $\Phi_\Lambda(x)$ depends only on x_Λ.

An interaction $\Phi = \{\Phi_\Lambda\}_{\Lambda \Subset \mathbb{Z}}$ represents contribution to the total energy from the particles or spins in Λ. The interaction Φ is called *uniformly absolutely convergent* (UAC) if

$$||\Phi|| := \sup_{n \in \mathbb{Z}^d} \sum_{n \in \Lambda \Subset \mathbb{Z}^d} \sup_{x \in \Omega} |\Phi_\Lambda(x)| = \sup_{n \in \mathbb{Z}^d} \sum_{n \in \Lambda \Subset \mathbb{Z}^d} ||\Phi_\Lambda||_\infty < \infty.$$

The requirement that the interaction Φ is UAC ensures that the total energy, or Hamiltonian, corresponding to a finite region $\Lambda \Subset \mathbb{Z}^d$, defined as

$$H_\Lambda(x) = \sum_{\Lambda \cap \Lambda \neq \emptyset} \Phi_\Lambda(x),$$

is well defined.

Definition 1.4. A probability measure μ on Ω is a called a Gibbs measure if for every $\Lambda \Subset \mathbb{Z}^d$

$$\mu(x_\Lambda | x_{\Lambda^c}) = \frac{e^{-H_\Lambda(x)}}{\sum_{\tilde{x}_\Lambda} e^{-H_\Lambda(\tilde{x}_\Lambda x_{\Lambda^c})}} =: \gamma_\Lambda^\Phi(x_\Lambda | x_{\Lambda^c}) \tag{1.2}$$

for μ—a.e. $x \in \Omega$.

This definition does not involve the inverse temperature β, which is commonly absorbed into the Hamiltonian. It can be shown that for any UAC interaction on Ω at least one Gibbs measure exists. Moreover, for many interesting examples a so-called phase transition occurs, i.e., there exist multiple Gibbs measures consistent with γ^Φ, c.f. (1.2).

The most famous examples are the Ising and Potts models.

- **Ising model**: $\mathscr{A} = \{-1, +1\}$ and $\Phi = \{\Phi_\Lambda : \Lambda \Subset \mathbb{Z}^d\}$ is given by:

$$\Phi_\Lambda(x) = \begin{cases} -Jx_n x_m, & \text{if } \Lambda = \{n, m\} \text{ and } ||n - m|| = 1, \\ 0, & \text{otherwise.} \end{cases}$$

- **Potts model** $\mathscr{A} = \{1, ..., q\}$ for some $q \geq 2$ and $\Phi = \{\Phi_\Lambda : \Lambda \Subset \mathbb{Z}^d\}$ is given by

$$\Phi_\Lambda(x) = \begin{cases} -2\beta \mathbb{I}[x_n = x_m] & \text{if } \Lambda = \{n, m\} \text{ and } ||n - m|| = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Both models exhibit a phase transition in dimension $d \geq 2$, depending on the parameters J and β, respectively.

Let us now return to the regularity properties of Gibbs measures mentioned earlier. Let $\Lambda_n = \{i \in \mathbb{Z}^d : ||i||_\infty \leq n\} = [-n, n]^d$. A function $f : \Omega \to \mathbb{R}$ is called continuous (or, quasilocal) if

$$\lim_{n \to \infty} \sup_{\tilde{x} \in \Omega} |f(x_{\Lambda_n} \tilde{x}_{\Lambda_n^c}) - f(x)| = 0.$$

For the UAC interaction Φ, the Hamiltonian H_Λ and the probability kernels γ_Λ^Φ, $\Lambda \Subset \mathbb{Z}^d$, are continuous on Ω. Moreover, again due to summability of the interaction Φ and hence the finiteness of H_Λ, for all $\Lambda \Subset \mathbb{Z}^d$:

$$\inf_x \gamma_\Lambda^\Phi(x_\Lambda | x_{\Lambda^c}) > 0.$$

Therefore we can conclude that Gibbs measures have positive continuous conditional probabilities. In fact, by the celebrated results of Kozlov and Sullivan [55, 84], this characterises the class of Gibbs measures.

In this thesis we will investigate the class of translation invariant Gibbs measures on one-dimensional symbolic spaces $\Omega = \mathscr{A}^{\mathbb{Z}}$. This class admits the following equivalent definition (for details see Chapter 2). Denote by $\mathscr{G}(\Omega)$ the class of continuous functions $\gamma : \Omega \to (0,1)$ that are normalized, i.e.,

$$\sum_{a\in\mathscr{A}} \gamma(\ldots, x_{-2}, x_{-1}, a, x_1, x_2, \ldots) = 1$$

for all $x \in \Omega$.

Definition 1.5. A translation invariant probability measure μ is called Gibbs for $\gamma \in \mathscr{G}(\Omega)$ if

$$\mu(x_0|x_{-\infty}^{-1}, x_1^{\infty}) = \gamma(\ldots, x_{-2}, x_{-1}, x_0, x_1, x_2, \ldots) = \gamma(x)$$

for μ-a.a. $x \in \Omega$.

1.3 Overview of the main results

This thesis can be divided in two parts. In the first part (Chapters 2 and 3) we study relations between one-sided and two-sided probabilistic models. In the second part (Chapters 4 and 5) we investigate the question whether the regularity properties of g- and Gibbs measures are preserved under so-called renormalisation transformations of the underlying probability spaces.

1.3.1 Summary of Chapter 2

The definitions of g-measures (Definition 1.2) and of translation invariant Gibbs measures in dimension $d = 1$ (Definition 1.5) show clear similarities: both classes of measures are defined by the requirement that conditional probabilities, either one-sided or two-sided, are given by a positive continuous function. A natural question is whether the two classes of measures are related. In fact, this question has been studied rather extensively.

Sinai [80] showed that Gibbs measures with Hölder-continuous functions γ are g-measures. Walters [90] extended this result to functions γ with *summable variation*, i.e.,

$$\sum_{n=1}^{\infty} \mathrm{var}_n(\gamma) < \infty, \quad \mathrm{var}_n(\gamma) = \sup_{x,y:x_{-n}^n = y_{-n}^n} |\gamma(x) - \gamma(y)|,$$

and later to functions γ satisfying the so-called Walters' condition.

Whether or not g-measures are always Gibbs, and vice versa, remained an open question until a few years before the start of this PhD project. In 2011, Gallo, Fernandez and Maillard [33] found an example of a g—measure that is not Gibbs. Shortly after the project started, a first example of a Gibbs measure that is not a g-measure was found in [6].

The main result of Chapter 2 is the necessary and sufficient condition for a g-measure to be a Gibbs measure.

Theorem 1.6. *Let μ be a g-measure on $\Omega_+ = \mathscr{A}^{\mathbb{Z}_+}$. Viewed as a measure on $\Omega = \mathscr{A}^{\mathbb{Z}}$, μ is Gibbs if and only if the sequence of functions $[\tilde{f}_n^{\sigma_0,\eta_0}]_{n\in\mathbb{N}}$, given by*

$$\tilde{f}_n^{\sigma_0,\eta_0}(x) = \prod_{i=-n}^{-1} \frac{g(x_i^{-1}\sigma_0 x_1^{\infty})}{g(x_i^{-1}\eta_0 x_1^{\infty})} \tag{1.3}$$

converges for all $\sigma_0, \eta_0 \in \mathscr{A}$ as $n \to \infty$, uniformly in $x \in \Omega$.

The condition in this theorem should be viewed as *regularity* requirement on the function g. Condition (1.3) is not easy to check. A sufficient condition for a g-measure to satisfy the condition of Theorem 1.6 is the so-called Good Future condition [29]: if

$$\partial_k(f) \equiv \sup_{x\in\Omega_+,\sigma_k,\eta_k\in\mathscr{A}} \left| f(x_0^{k-1}\sigma_k x_{k+1}^{\infty}) - f(x_0^{k-1}\eta_k x_{k+1}^{\infty}) \right|, \tag{1.4}$$

then $g \in G(\Omega_+)$ has Good Future if

$$\sum_{k=1}^{\infty} \partial_k(g) < \infty. \tag{1.5}$$

The novel and somewhat unexpected aspect of Theorem 1.6 is that condition (1.3) does not imply uniqueness of the corresponding g-measure. In particular, Hulse [46] showed that if $\lambda > 1$, then there exists a $g \in G(\Omega_+)$, with multiple g-measures, such that

$$\sum_{k=1}^{\infty} \partial_k(g) < \lambda < \infty.$$

For comparison, the one-sided one-dimensional analogue of the well known Dobrushin condition for g-measures states that there is a unique g-measure for functions $g \in G(\Omega_+)$ with

$$\sum_{k=1}^{\infty} \partial_k(g) < 1.$$

It follows that uniqueness is not required for a g-measure to be a Gibbs measure.

The question under which conditions a translation invariant Gibbs measure on the lattice $\mathbb{Z}$ is a g-measure remains open, with the exception of positive results due to Sinai and Walters for smooth potentials, and one recent negative example [6].

Another natural and similar question in this context is the following. Is the g-measure reversible in the following sense: the one-sided conditional probabilities in the opposite (time-reversed) direction are continuous? This question was first raised by Walters in [95]. A sufficient condition and regularity properties of the resulting reversed g-function were given. Under a weaker condition the regularity properties, given reversibility, were proven. Whether or not this weaker condition is sufficient for reversibility remained open. We show that this condition is not sufficient. However, finding a necessary and sufficient condition in terms of the g-function remains open.

1.3.2 Summary of Chapter 3

In this chapter we complete the first part of the thesis with a practical application of the relation between one-sided and two-sided models.

Various algorithms have been developed in Information Theory and Statistics to find approximations of stationary sources (measures) by Markov and, more generally, variable-length Markov models, which form a particular subclass of g-measures (see [61, 75]). The primary question addressed in Chapter 3 is: Given the fact that one-sided models can be converted into two-sided models, which one-sided algorithms produce good Gibbsian approximations of unknown sources?

In fact, all algorithms produce finite-range Markov approximations of the unknown source. For Markov measures, the correspondence between one-sided and two-sided models is a bijection. However, it is important to understand how well various one-sided, or unidirectional, algorithms perform when used for the estimation of two-sided conditional probabilities. This is particularly relevant as algorithms that produce direct two-sided estimates are less developed than their one-sided counterparts.

In this chapter we compared a number of one-sided algorithms using two metrics for the quality of the resulting two-sided model. The first quality metric originates Information Theory via the so-called denoising problem: namely, consider a finite sample $X_1^n = (X_1, \ldots, X_n)$ with $n \gg 1$ from an unknown stationary source. Some symbols in X_1^n have been randomly transformed into other symbols, resulting in noisy data Z_1^n. A denoiser uses the observed data Z_1^n and the knowledge of the noisy channel, to produce an estimate $\hat{X}_1^n$ of X_1^n. One celebrated algorithm – the Discrete Universal DEnoiser (DUDE) [97], requires a good estimate of the two-sided (bidirectional) conditional probabilities of $\{Z_n\}_{n\in\mathbb{Z}}$. We evaluate the quality of various algorithms indirectly by comparing the performance of the denoisers, which use the estimates of two-sided probabilities, computed from the one-sided estimates obtained by these algorithms. It was already shown in [100, 101] that such an approach can indeed lead to an improved (in comparison to original DUDE) denoising performance. In this thesis we consider several artificial sources as well as an English text.

As a second quality metric we consider the so-called erasure divergence – the two-sided variant of the Kullback-Leibler divergence. We use this metric to evaluate performance on a specific Gibbsian source, i.e., in a situation where we have a perfect knowledge of true two-sided conditional probabilities.

1.3.3 Summary of Chapter 4

In the second part of the thesis, we turn our attention to renormalisation of Markov and Gibbs measures.

We first address the Markov case. Suppose that $\mathscr{A}$ is a finite set, and $\{X_n\}_{n\in\mathbb{Z}_+}$ is a stationary $\mathscr{A}$-valued Markov chain with transition probability matrix P: $P \geq 0$ and $\sum_{b\in\mathscr{A}} P_{ab} = 1$ for all $a \in \mathscr{A}$. Suppose that $\pi : \mathscr{A} \to \mathscr{B}$ is a surjective map onto a second smaller alphabet $\mathscr{B}$, $|\mathscr{B}| < |\mathscr{A}|$. Define the corresponding *factor process* by $Y_n = \pi(X_n)$ for all n. If μ is the probability measure of $\{X_n\}$, then $\nu = \mu \circ \pi^{-1}$ is the probability measure on $\mathscr{B}^{\mathbb{Z}+}$ describing the process $\{Y_n\}$. Processes of such form – functions of Markov chains have been studied extensively in the past 60 years. For example, classical results [14, 52] provide necessary and sufficient conditions for the factor measure ν to be Markov. Similarly, there are various results providing sufficient conditions for ν to be a g-measure [45, 99]. The simplest example is that of a strictly positive transition matrix $P > 0$. In this case $\nu = \mu \circ \pi^{-1}$ is a g-measure for some Hölder-continuous function g:

$$\mathrm{var}_n(g) = \mathscr{O}(c^n)$$

for some $0 < c < 1$ and all $n \geq 1$.

Known results can informally be summarized as follows: the factor measure ν is typically not Markov (of any finite order), but, under relatively mild conditions, ν is a g-measure. The problem of identifying necessary and sufficient conditions for factors of Markov measures to be regular is still an open problem. The interesting case is when certain transitions are forbidden, i.e., the transition probability matrix has some zero elements. The support of the Markov measure μ is then a *subshift of finite type*:

$$\Omega_+ = \{x \in \mathscr{A}^{\mathbb{Z}_+} : P_{x_n,x_{n+1}} > 0, \forall n \in \mathbb{Z}_+\}.$$

If we extend π to a map from $\mathscr{A}^{\mathbb{Z}_+}$ to $\mathscr{B}^{\mathbb{Z}_+}$, then the support of the measure ν is $\Sigma_+ = \pi(\Omega_+)$ – a closed shift-invariant subset of $\mathscr{B}^{\mathbb{Z}_+}$, i.e., a certain subshift of $\mathscr{B}^{\mathbb{Z}_+}$. Throughout this chapter we will assume that $\Sigma_+ = \pi(\Omega_+)$ is also a subshift of finite type, even though in general it is only *sofic*. We establish a novel sufficient condition for regularity of ν, which supersedes all previous results. A similar condition has previously been applied to the question of regularity of factors of fully supported g-measures in [87]. Consider the fibres of the factor map $\pi : \Omega_+ \to \Sigma_+$:

$$\Omega_y = \pi^{-1}(y) = \{x \in \Omega : \quad \pi(x) = y\}, \quad y \in \Sigma_+.$$

We call $\{\mu_y\}_{y \in \Sigma_+}$ a family of conditional measures for μ on the fibres Ω_y if, for every $y \in \Sigma_+$, μ_y is a Borel probability measure on the fibre Ω_y, and

$$\mu = \int_{\Sigma_+} \mu_y \, \nu(dy),$$

meaning that, for any continuous function $f : \Omega_+ \to \mathbb{R}$,

$$\int_{\Omega_+} f(x)\mu(dx) = \int_{\Sigma_+} \left[\int_{\Omega_y} f(x)\mu_y \right] \nu(dy).$$

For any factor map π and every measure μ such a family, also called a *measure disintegration*, exists, but is not necessarily unique. The family of conditional measures can be thought of as a way of conditioning on the fibres, which are sets of measure zero. This disintegration can be used to describe the conditional probabilities of ν, namely, for ν-almost all $y \in \Sigma_+$,

$$\nu(y_0|y_1 y_2, \dots) = \sum_{a' \in \pi^{-1} y_1} \left[\sum_{a \in \pi^{-1} y_0} \frac{p_a P_{a,a'}}{p_{a'}} \right] \mu_{Ty}({}_0[a']), \tag{1.6}$$

where ${}_0[a'] = \left\{ x : x_0 = a' \right\}$, and $T : \Sigma_+ \to \Sigma_+$ is the left shift.

The measure ν is a g-measure on Σ_+ if the right-hand side of (1.6), which we denote by $\tilde{g}(y)$, defines a continuous function on Σ_+. In general, it is rather difficult to decide on continuity of $\tilde{g}(y)$. However, it can easily be shown that if the map

$$y \to \mu_y \tag{1.7}$$

is continuous in the weak topology, then $\tilde{g}(y)$ is indeed continuous, and hence, ν is a g-measure. Thus, existence of a *continuous measure disintegration* (CMD) $\{\mu_y\}$ is sufficient for ν to be a g-measure.

We discuss the literature on continuous measure disintegrations [85, 86]. We demonstrate that the existence of a continuous disintegration for Markov measures follows from the fibre-mixing condition [99], which was the weakest general sufficient condition for regularity of the factor measures, known prior to our work. In fact, we use two rather different techniques to show that fibre mixing implies CMD: one method originates in dynamical systems [27], the other uses results in statistical mechanics [81]. Moreover, we demonstrate with an example that one can have CMD without fibre mixing. Hence, our result is a substantial improvement of previously known results. However, in our final example we show that existence of a CMD is a sufficient, but not a necessary, condition for ν to be a g-measure.

1.3.4 Summary of Chapter 5

In the final chapter we consider factors of fully supported Gibbs measures on lattices $\mathbb{Z}^d$, $d \geq 1$. Again, let $\Omega = \mathscr{A}^{\mathbb{Z}^d}$, and define $\pi : \mathscr{A} \to \mathscr{B}$ to be a surjective map onto $\mathscr{B}$. We use the same symbol π to denote the coordinate-wise extension of π to a map $\pi : \mathscr{A}^V \to \mathscr{B}^V$ for any $V \subset \mathbb{Z}^d$. Suppose that μ is a Gibbs measure on Ω for some interaction Φ. Is the measure $\nu = \mu \circ \pi^{-1}$ on $\Sigma = \mathscr{B}^{\mathbb{Z}^d}$ Gibbs?

In Statistical Mechanics such factors appear in renormalisation of Gibbs measures. The behaviour of Gibbs measures under renormalisation is important in the study of the critical systems (renormalisation group method). It is paramount to controlling the occurrence of pathologies [39–41], that can appear due to a lack of regularity (quasi-locality) of conditional probabilities, i.e., non-Gibbsianness of the renormalized Gibbs states [47, 81].

The question of Gibbsianity of $\nu = \mu \circ \pi^{-1}$ distinguishes itself in an important way from the corresponding problem for Markov measures. For Markov measures, the interaction is relatively simple and the only potential source of singularities is the support of the measure or, to be more precise, the topological structure of the fibres. For fully supported Gibbs measures, singularities of the renormalized measures stem from the properties of the potential of the original Gibbs measure on the fibres.

Our main result is the following extension of the result for factors of Markov measures. Again, define fibres

$$\Omega_y = \pi^{-1}(y) = \{x \in \mathscr{A}^{\mathbb{Z}^d} : \quad \pi(x) = y\}, \quad y \in \mathscr{B}^{\mathbb{Z}^d}.$$

Theorem 1.7. *Suppose that μ is a Gibbs measure on $\mathscr{A}^{\mathbb{Z}^d}$ for the UAC interaction Φ and $\pi : \mathscr{A}^{\mathbb{Z}^d} \to \mathscr{B}^{\mathbb{Z}^d}$ is a factor map. Furthermore, suppose that μ admits a* *continuous family $\{\mu_y\}$ of conditional measures on fibres $\{\Omega_y\}$. Then $\nu = \mu \circ \pi^{-1}$ is a Gibbs state on $\mathscr{B}^{\mathbb{Z}^d}$ for some UAC interaction Ψ.*

This result has interesting implications for the long-standing van Enter-Fernández-Sokal hypothesis on the loss/preservation of Gibbsianity under renormalisation.

Conjecture 1.8. *The factor measure $\nu = \mu \circ \pi^{-1}$ is Gibbs if and only if for each $y \in \Sigma$ there exists a unique Gibbs measure on Ω_y for the original potential Φ.*

We obtain the following result.

Theorem 1.9. *If the interaction Φ is such that there is a unique Gibbs measure μ^y for Φ on the fibre $\mathscr{A}_y$ for all $y \in \mathscr{B}^{\mathbb{Z}^d}$, then the family of measures $\{\mu^y\}$ constitutes a continuous disintegration of μ, and hence $\nu = \mu \circ \pi^{-1}$ is Gibbs.*

Thus, we obtain the first proof in complete generality of the “easy” part of the van Enter-Fernández-Sokal conjecture.

Our proofs rely on the method of Tjur [85, 86] for construction of a continuous measure disintegration. In case of Gibbs measures, we show that any limiting measure in Tjur’s construction must be Gibbs for the original interaction.

The question of necessity (the “difficult” part of the van Enter-Fernández-Sokla conjecture) remains open, but we conjecture that the so-called non-Tjur points $y \in \Sigma$ are the primary candidates to be “bad points”. If this would indeed be the case, then this would lead to the complete proof of Conjecture 1.8.

Chapter 2

On the relation between Gibbs and g-measures

Thermodynamic formalism, the theory of equilibrium states, is studied both in dynamical systems and probability theory. Various closely related notions have been developed: e.g. Dobrushin–Lanford–Ruelle Gibbs, Bowen–Gibbs, and g-measures. We discuss the relation between Gibbs and g-measures in a one-dimensional context. Often g-measures are also Gibbs, but recently an example to the contrary has been presented. In this paper we discuss exactly when a g-measure is Gibbs and how this relates to notions such as uniqueness and reversibility of g-measures.

2.1 Introduction

Thermodynamic formalism, used in symbolic dynamics, has strong similarities to the study of DLR Gibbs measures [21,56] in statistical mechanics. For g—measures [51], or similar objects, such as chains of complete connections, variable length Markov chains and chains of infinite order these similarities are particularly pronounced. Via its natural extension a g—measure could be a one-dimensional Gibbs measure, or the corresponding counterpart in dynamical systems, a Bowen-Gibbs measure [11]. A recent example [33] shows that on $\mathbb{Z}$ the notions are not equivalent. The fundamental underlying question is the relation between one-sided conditional probabilities and their two-sided counterparts. In this respect these problems have a strong similarity to the reversibility question for g—measures,

[0]The chapter is based on S. Berghout, R. Fernández, E. Verbitskiy, *On the relation between Gibbs and g-measures*, Ergodic Theory & Dynamical Systems, Volume 39, Issue 12, December 2019, pp. 3224-3249.

i.e. when a projection of an extended $g-$measure on the negative integers is a $g-$measure in its own right. Both questions have been addressed in the literature. An early comparison between dynamical systems and statistical mechanics, by Sinai [80], uses Gibbs measures to study Anosov dynamical systems. Sufficient conditions for a $g-$measure to be Gibbs can be found in [29]. Besides the example in [33] it can easily be shown that an example constructed by Walters [95], used to show the existence of a non-reversible $g-$measure, is also non-Gibbsian.

We will extend these results in the literature by

- Presenting a necessary and sufficient condition, Theorem 2.11, for a g-measure to be Gibbs.
- Discuss when a $g-$measure is reversible.
- Discuss how well known classes of $g-$measures compare to the Gibbs condition.
- Show that there exist $g-$measures that are Gibbs measures in the non-uniqueness regime.
- Give an example demonstrating that there exists a $g-$measure with a potential in Bowen's class for which the reverse is not a $g-$measure.

In particular we will show that the non-Gibbsian example in [33] is a Bowen-Gibbs measure and that it is reversible. However, $g-$measures with a potential in Walters' class are all Gibbs measures. Furthermore, we will discuss an adaptation of Walters' example [95] to construct a reversible $g-$measure, for which the reverse $g-$measure has a slower decay of variation. As a preparation for these results and examples we will use the first sections of this paper to define the relevant classes of measures and recall some of their properties. In section 2.5 examples are given to highlight properties of some of the conditions mentioned throughout the paper. A table giving an comparison of some of these conditions is added as well. In the last section we review existing results on when Gibbs measures are $g-$measures.

2.2 Four classes of measures

2.2.1 General setting and notation

We will restrict ourselves to symbolic systems with a finite alphabet $\mathscr{A}$. The corresponding measurable spaces are the sets $X = \mathscr{A}^{\mathbb{Z}}$, $X_+ = \mathscr{A}^{\mathbb{Z}_+}$ and $X_- = \mathscr{A}^{\mathbb{Z}_-}$, equipped with the product topology and the corresponding σ-algebra of Borel sets. Elements of X will be referred to as two-sided sequences and elements of X_+ and X_- as one-sided sequences. We will use $\omega_i^j = \omega_i\omega_{i+1}...\omega_j$ as a shorthand notation for strings (words) over the alphabet $\mathscr{A}$. Another shorthand notation we will sometimes use is a^n, with $a \in \mathscr{A}$ and $n \in \mathbb{N}$, denoting a sequence of n subsequent identical elements a. Writing strings in order, for example $a^n b^m$, with $a, b \in \mathscr{A}$ and $n, m \in \mathbb{N}$, denotes the string that is a concatenation of the individual strings.

We write $\mu(\omega_i^j)$ for the measure of the cylinder set

$$[\omega_i^j] = \{\tilde{\omega} : \tilde{\omega}_i = \omega_i, \ldots, \tilde{\omega}_j = \omega_j\}.$$

These sets are of special importance, as they generate both the topologies and the $\sigma-$algebras of the spaces above. The shift operator $S : X_+ \to X_+$ is defined as $S(\omega_0\omega_1...) = (\omega_1\omega_2...)$. Similarly, we define a shift operator on X and a right shift, S_-, on X_-; note that the shift operator on X is invertible. In the present paper, we study $g-$measures and Gibbs measures that are translation-invariant: $\mu(S^{-1}A) = \mu(A)$ for all measurable sets A in the relevant $\sigma-$algebra. The set of translation-invariant measures will be denoted by $\mathscr{M}_S(X_+)$, $\mathscr{M}_S(X)$ or $\mathscr{M}_{S\_}(X_-)$.

The natural extension, as a dynamical system, uniquely maps a translation-invariant measure on X_+ (or X_-) to a translation-invariant measure on X. Let, for a bidirectional sequence, $\boldsymbol{\omega} \in X$, the projection $\pi : X \to X_+$ be defined by $\pi(\boldsymbol{\omega}) = \omega_0^\infty$. The corresponding projection for the measures on these spaces, $\pi^* : \mathscr{M}_S(X) \to \mathscr{M}_S(X_+)$, is given by $\mu_+(A) = (\pi^*\mu)(A) = \mu(\pi^{-1}(A))$, for $\mu \in \mathscr{M}_S(X)$ and $A \in \mathscr{B}_+$ a Borel measurable subset of X_+. Similarly a projection $\pi_- : X \to X_-$, given by $\pi_-(\boldsymbol{\omega}) = \omega_{-\infty}^0$, relates X_- to X. In this way we can identify the translation-invariant measures in $\mathscr{M}_S(X_+)$, $\mathscr{M}_S(X)$ and $\mathscr{M}_{S\_}(X_-)$ with each other.

2.2.2 The class of g-measures

Let $G(X_+)$ be the set of all positive continuous functions $g : X_+ \to (0, 1)$ which are *normalized*

$$\sum_{a\in\mathscr{A}} g(a\boldsymbol{\omega}) = 1 \text{ for all } \boldsymbol{\omega} \in X_+ = \mathscr{A}^{\mathbb{Z}_+}.$$

Definition 2.1. A translation-invariant measure μ_+ on X_+ is called a g-measure for $g \in G(X_+)$ if

$$\mu_+(\omega_0|\omega_1^\infty) = g(\omega_0^\infty)$$

for μ-a.e. $\boldsymbol{\omega} = (\omega_0, \omega_1, \ldots) \in X_+$.

Equivalently, one can say that μ_+ is a g-measure if, for any continuous function $f : X_+ \to \mathbb{R}$, one has

$$\int f(\boldsymbol{\omega})\mu_+(d\boldsymbol{\omega}) = \int \sum_{a\in\mathscr{A}} f(a\boldsymbol{\omega})g(a\boldsymbol{\omega})\mu_+(d\boldsymbol{\omega}).$$

g-measures on X_- are defined analogously. Finally, if μ_+ is a g-measure on X_+, we will also call the natural extension μ on X a g-measure. Using the projections defined above we can obtain the reverse of a g—measure μ_+ by extending it to X, resulting in a measure $\mu \in M_S(X)$, and then projecting onto X_- to obtain its reverse $\mu_- \in M_S(X_-)$. We call the g—measure reversible if μ_- is a g—measure. As not every g—measure is reversible, when we call $\mu \in \mathscr{M}_{\mathscr{S}}(X)$ a g—measure, we have to keep in mind with respect to which shift. The following result characterizes g-measures by uniform convergence of conditional probabilities [74].

Theorem 2.2 (Palmer, Parry, Walters [74])**.** *A fully supported translation-invariant probability measure μ_+ on X_+ is a g—measure if and only if the sequence of functions $[g_n]_{n\in\mathbb{N}}$, with*

$$g_n(\boldsymbol{\omega}) := \mu_+(\omega_0|\omega_1^n), \quad n \geq 1, \quad \boldsymbol{\omega} \in X_+,$$

converges uniformly in $\boldsymbol{\omega}$, as $n \to \infty$, to $g(\boldsymbol{\omega})$, for some $g \in G(X_+)$.

Note that the natural extension, μ, of μ_+ satisfies the same convergence of one-sided conditional probabilities if and only if μ_+ satisfies the conditions of the above theorem.

2.2.3 The class of DLR Gibbs measures

The classical definition of Gibbs measures for lattice systems $\mathscr{A}^{\mathbb{Z}^d}$, $d \geq 1$, involves the notions of interactions, Hamiltonians and specifications. We will use a novel equivalent definition, for translation-invariant measures in one dimension, that stresses the similarity with g-measures. Denote by $\mathscr{G}(X)$ the class of continuous functions $\gamma : X \to (0,1)$ that are normalised:

$$\sum_{a\in\mathscr{A}} \gamma(\ldots,\omega_{-2},\omega_{-1},a,\omega_1,\omega_2,\ldots) = 1,$$

for all $\boldsymbol{\omega} = (\ldots,\omega_{-2},\omega_{-1},\omega_0,\omega_1,\omega_2,\ldots) \in X$.

Definition 2.3. A translation-invariant measure μ is called Gibbs for $\gamma \in \mathscr{G}(X)$ if

$$\mu(\omega_0|\omega_{-\infty}^{-1},\omega_1^{\infty}) = \gamma(\ldots,\omega_{-2},\omega_{-1},\omega_0,\omega_1,\omega_2,\ldots) = \gamma(\boldsymbol{\omega})$$

for μ-a.a. $\boldsymbol{\omega} \in X$. Equivalently, for all continuous functions $f : X \to \mathbb{R}$ one has

$$\int f(\boldsymbol{\omega})\mu(d\boldsymbol{\omega}) = \int \sum_{a\in\mathscr{A}} f(\boldsymbol{\omega}^{(a)})\gamma(\boldsymbol{\omega}^{(a)})\mu(d\boldsymbol{\omega}),$$

where $\boldsymbol{\omega}^{(a)} = (\omega_n^{(a)})$, for $\boldsymbol{\omega} \in X$ and $a \in \mathscr{A}$, is defined as

$$\omega_n^{(a)} = \begin{cases} a, & n = 0, \\ \omega_n, & n \neq 0. \end{cases}$$

Theorem 2.2, for g-measures, has a counterpart for Gibbs measures in the following form:

Theorem 2.4 (Folklore). *A fully supported translation-invariant probability measure μ on X is Gibbs if and only if the double-indexed sequence of functions $[g_{n,m}]_{n,m\in\mathbb{N}}$, with*

$$g_{m,n}(\boldsymbol{\omega}) := \mu(\omega_0|\omega_1^n, \omega_{-m}^{-1}), \quad n, m \geq 1,$$

converges, as $n, m \to \infty$, to a function $\gamma \in \mathscr{G}(X)$, uniformly in $\boldsymbol{\omega}$.

Definition 2.3 seems to be new, we show the equivalence between the classical definition of Gibbs states and the one above in Section 2.3.1. Theorem 2.4 is a folklore result, for the convenience of readers we provide a proof.

Proof of Theorem 2.4. Let us start by showing that the convergence of finite-range conditional probabilities is sufficient for Gibbsianity. Since the sequence converges uniformly the limit will be continuous. Hence the measure will be consistent with a continuous uniformly non-null specification on single sites. This means that the measure satisfies the conditions of Definition 2.3 and therefore it is a Gibbs measure. We postpone the proof that the new definition is indeed equivalent to the classical definition to Theorem 2.10. In the other direction let μ be a translation-invariant Gibbs measure and let $\{\gamma_V : V \Subset \mathbb{Z}\}$ be the corresponding continuous specification. Assume $N > 0$ and $n, m > N$. Then

$$\mu(\omega_0|\omega_{-m}^{-1}\omega_1^n) = \frac{1}{\sum_{\sigma_0} \frac{\mu(\omega_{-m}^{-1}\sigma_0\omega_1^n)}{\mu(\omega_{-m}^{-1}\omega_0\omega_1^n)}}. \tag{2.1}$$

If the following quantity is uniformly convergent

$$\frac{\mu(\omega_{-m}^{-1}\sigma_0\omega_1^n)}{\mu(\omega_{-m}^{-1}\omega_0\omega_1^n)} = \frac{\int_X \gamma_{[-m,n]}(\omega_{-m}^{-1}\sigma_0\omega_1^n|\xi_{[-m,n]^c})\mu(d\xi)}{\int_X \gamma_{[-m,n]}(\omega_{-m}^{-1}\omega_0\omega_1^n|\xi_{[-m,n]^c})\mu(d\xi)}, \tag{2.2}$$

then it is bounded away from 0 and ∞ and therefore the denominator in (2.1) is uniformly convergent, bounded away from 1 and ∞. It then follows that the conditional probability (2.1) does converge uniformly. To show uniform convergence of (2.2) we will use the consistency relation for specifications, this is discussed later in Section 2.3.1. For $V \Subset \mathbb{L}$ and $V' \subset V$ the specification satisfies:

$$\gamma_V(\sigma_V|\omega_{V^c}) = \sum_{\eta_{V'}} \gamma_{V'}(\sigma_{V'}|\sigma_{V\setminus V'}\omega_{V^c})\gamma_V(\eta_{V'}\sigma_{V\setminus V'}|\omega_{V^c}).$$

Applying this relation for $V = \{-m, -m+1, ..., n-1, n\}$ and $V' = \{0\}$ results in:

$$\frac{\mu(\omega_{-m}^{-1}\omega_0\omega_1^n)}{\mu(\omega_{-m}^{-1}\sigma\omega_1^n)} = \frac{\int_X \gamma(\omega_0|\omega_{-m}^{-1}\omega_1^n\xi_{[-m,n]^c})\sum_{\eta_0\in\mathscr{A}}\gamma(\omega_{-m}^{-1}\eta_0\omega_1^n|\xi_{[-m,n]^c})\mu(d\xi)}{\int_X \gamma(\sigma_0|\omega_{-m}^{-1}\omega_1^n\xi_{[-m,n]^c})\sum_{\eta_0\in\mathscr{A}}\gamma(\omega_{-m}^{-1}\eta_0\omega_1^n|\xi_{[-m,n]^c})\mu(d\xi)}$$

As all factors under the integrals are positive, one can estimate this quantity from above by:

$$\begin{aligned}\frac{\mu(\omega_{-m}^{-1}\omega_0\omega_1^n)}{\mu(\omega_{-m}^{-1}\sigma_0\omega_1^n)} &\leq \frac{\int_X \sup_{\lambda\in X}\gamma(\omega_0|\omega_{-m}^{-1}\omega_1^n\lambda_{[-m,n]^c})\sum_{\eta_0\in\mathscr{A}}\gamma(\omega_{-m}^{-1}\eta_0\omega_1^n|\xi_{[-m,n]^c})\mu(d\xi)}{\int_X \inf_{\lambda\in X}\gamma(\sigma_0|\omega_{-m}^{-1}\omega_1^n\lambda_{[-m,n]^c})\sum_{\eta_0\in\mathscr{A}}\gamma(\omega_{-m}^{-1}\eta_0\omega_1^n|\xi_{[-m,n]^c})\mu(d\xi)}\\ &= \frac{\sup_{\lambda\in X}\gamma(\omega_0|\omega_{-m}^{-1}\omega_1^n\lambda_{[-m,n]^c})}{\inf_{\lambda\in X}\gamma(\sigma_0|\omega_{-m}^{-1}\omega_1^n\lambda_{[-m,n]^c})}\\ &\leq \frac{\sup_{\lambda\in X}\gamma(\omega_0|\omega_{-N}^{-1}\omega_1^N\lambda_{[-N,N]^c})}{\inf_{\lambda\in X}\gamma(\sigma_0|\omega_{-N}^{-1}\omega_1^N\lambda_{[-N,N]^c})}\end{aligned}$$

for any $N \leq \min\{m, n\}$. In the same way we get a lower bound:

$$\frac{\mu(\omega_{-m}^{-1}\omega_0\omega_1^n)}{\mu(\omega_{-m}^{-1}\sigma_0\omega_1^n)} \geq \frac{\inf_{\lambda\in X}\gamma(\omega_0|\omega_{-N}^{-1}\omega_1^N\lambda_{[-N,N]^c})}{\sup_{\lambda\in X}\gamma(\sigma_0|\omega_{-N}^{-1}\omega_1^N\lambda_{[-N,N]^c})}$$

By quasilocality of γ both the upper and lower bound converge uniformly to the same value, as $N \to \infty$, therefore $\mu(\omega_0|\omega_{-m}^{-1}\omega_1^n)$ converges uniformly. □

2.2.4 The class of Bowen-Gibbs measures

Sometimes measures with a slightly different definition, by Bowen [11], are also referred to as Gibbs measures. Bowen's definition is not equivalent to the original definition of Gibbs states due to Dobrushin, Lanford, and Ruelle. This non-equivalence was deliberate: as a definition, Bowen chose a particular result valid for one-dimensional Gibbs measures, giving a convenient expression for the probability of any cylindrical event.

Definition 2.5. A translation-invariant measure μ on X_+ or X is Bowen-Gibbs for a continuous potential ϕ, if there exist constants $c > 1$ and $P \in \mathbb{R}$ such that for all $\boldsymbol{\omega}$ and every $n \in \mathbb{N}$

$$\frac{1}{c} \leq \frac{\mu\big([\omega_0^{n-1}]\big)}{\exp\Big(\sum_{j=0}^{n-1} \phi(S^j \boldsymbol{\omega}) - nP\Big)} \leq c.$$

The constant $P = P(\phi)$ is called the *topological pressure* of ϕ, and can be defined independently [92]. We propose to use the name "Bowen-Gibbs measure", rather than "Gibbs measure" to avoid confusion. The naming problem also extends to the class of measures μ satisfying

$$\frac{1}{c_n} \leq \frac{\mu\big([\omega_0^{n-1}]\big)}{\exp\Big(\sum_{j=0}^{n-1} \phi(S^j \boldsymbol{\omega}) - nP\Big)} \leq c_n, \tag{2.3}$$

where the c_n grow at most sub-exponentially in n, i.e., $\frac{1}{n}\log c_n \to 0$. These measures were called weak Gibbs in [102]. However, there exists an independent notion of weak Gibbs states in Statistical Mechanics [23, 24].

2.2.5 Equilibrium states

Finally, let us recall the notion of equilibrium states. The three classes of invariant measures defined above turn out to be equilibrium states for appropriate potentials.

Definition 2.6. A translation-invariant measure μ on X_+ or X is called an *equilibrium state* for a continuous function (potential) ϕ, defined on X_+ or X, respectively, if

$$h(\mu, S) + \int \phi(\boldsymbol{\omega})\mu(d\boldsymbol{\omega}) = \sup_{\nu}\Big[h(\nu, S) + \int \phi(\boldsymbol{\omega})\nu(d\boldsymbol{\omega})\Big], \tag{2.4}$$

where the supremum is taken over $\mathscr{M}_S(X_+)$ or $\mathscr{M}_S(X)$, and $h(\mu, S)$ is the Kolmogorov-Sinai entropy.

By the variational principle [20, 38], the right-hand side in (2.4) coincides with the topological pressure $P(\phi)$ of ϕ. It is easy to check that any g-measure is also an equilibrium state for $\phi = \log g$ [59], and every Bowen-Gibbs measure for the potential ϕ is also an equilibrium state for the same potential. In the case of Gibbs measures the corresponding function ϕ can be expressed in terms of an underlying interaction Φ.

2.2.6 The relation between Gibbs and g-measures

In the present chapter we investigate the relation between the classes mentioned above. This problem has been addressed before. For example, for *sufficiently smooth* potentials, meaning potentials that have a sufficiently fast decay, such as Hölder continuous potentials or those that have summable variation, the corresponding g-measures are Gibbs, and vice versa.

The renewed interest in this problem was sparked by a recent example constructed by Gallo, Fernández, and Maillard [33] of a g-measure μ_+^{GFM} on X_+ such that its extension μ^{GFM} to X is not Gibbs. This example was found in the class of so-called Variable Length Markov Models. Even earlier, Walters [95] produced an example of a g-measure μ_+^W on X_+ such that its reversal μ_-^W is not a g-measure. We will show that this example is also an example of a g-measure that is not Gibbs. We will also show that the Gallo-Fernandez-Maillard measure μ_+^{GFM} is in fact more regular than Walters' example: its reversal μ_-^{GFM} is a g-measure. Unknown to the authors of [33], their g-function belongs to the so-called R-class, introduced earlier by Walters [96]. We will use this class to discuss the measure μ_+^{GFM} and generate other relevant examples. Our main result gives the necessary and sufficient condition for a g-measure to be Gibbs. The problem of finding necessary and sufficient conditions for Gibbs measures to be g-measures remains open and will be discussed briefly in Section 2.6.

2.3 Further properties of Gibbs measures

In the previous section we gave a definition of Gibbs measures adapted for a comparison with g—measures. Now we will discuss how it compares to the standard way of defining a Gibbs measure.

2.3.1 Gibbs measures

One option for defining Gibbs measures is to start with the notion of an interaction.

Definition 2.7. An interaction is a collection of functions, $\{\Phi_V\}$ on X, indexed by finite subsets $V \Subset \mathbb{Z}$ ($\Subset$ indicates that the subset is finite), such that

$$\Phi_V(\boldsymbol{\omega}) = \Phi_V(\boldsymbol{\omega}|_V),$$

i.e., $\Phi_V(\boldsymbol{\omega})$ depends only on the values of $\boldsymbol{\omega}$ in V. An interaction $\Phi = \{\Phi_V\}_{V\Subset\mathbb{Z}}$ is called *uniformly absolutely convergent* (UAC) if

$$\|\Phi\| := \sup_{n\in\mathbb{Z}} \sum_{n\in V\Subset\mathbb{Z}} \sup_{\boldsymbol{\omega}\in X} |\Phi_V(\boldsymbol{\omega})| = \sup_{n\in\mathbb{Z}} \sum_{n\in V\Subset\mathbb{Z}} \|\Phi_V\|_\infty < \infty.$$

If $\Phi = \{\Phi_V\}$ is a UAC interaction, for a finite set (volume) $\Lambda \Subset \mathbb{Z}$, the corresponding Hamiltonian is defined as

$$H_\Lambda(\boldsymbol{\omega}) = \sum_{V\cap\Lambda\neq\varnothing} \Phi_V(\boldsymbol{\omega}).$$

Finally, a specification γ^Ψ is a collection of probability kernels $\gamma_V^\Psi : \mathscr{B}_V \times X_{V^c} \to (0,1)$, indexed by $V \Subset \mathbb{Z}$ defined as

$$\gamma_V^\Phi(\boldsymbol{\sigma}_V|\boldsymbol{\omega}_{V^c}) = \frac{1}{Z^\Psi(\boldsymbol{\omega}_{V^c})} \exp\Big(-H_V(\boldsymbol{\sigma}_V\boldsymbol{\omega}_{V^c})\Big),$$

where $\boldsymbol{\sigma}_V\boldsymbol{\omega}_{V^c}$ is an element in X, equal to $\boldsymbol{\sigma}$ on V, and to $\boldsymbol{\omega}$ on V^c. The normalizing constant $Z^\Psi(\boldsymbol{\omega}_{V^c}) = \sum_{\boldsymbol{\sigma}_V\in\mathscr{A}^V} \exp\Big(-H_V(\boldsymbol{\sigma}_V\boldsymbol{\omega}_{V^c})\Big)$ is known as a partition function. We will refer to the specification density as a specification since the notions are equivalent in the context of this paper.

Definition 2.8. A probability measure μ on X is Gibbs for an interaction Φ (denoted by $\mu \in \mathscr{G}(\Phi)$) if it is *consistent* with the corresponding specification γ^Φ ($\mu \in \mathscr{G}(\gamma^\Phi)$), meaning that for every $V \Subset \mathbb{Z}$

$$\mu(\boldsymbol{\omega}_V|\boldsymbol{\omega}_{V^c}) = \gamma_V^\Phi(\boldsymbol{\omega}_V|\boldsymbol{\omega}_{V^c}) \quad \mu - a.s.$$

Equivalently, μ is Gibbs if for any continuous function f on X and every $V \Subset \mathbb{Z}$

$$\int f(\boldsymbol{\omega})\mu(d\boldsymbol{\omega}) = \int \sum_{\boldsymbol{\sigma}_V\in\mathscr{A}^V} f(\boldsymbol{\sigma}_V\boldsymbol{\omega}_{V^c})\gamma_V^\Phi(\boldsymbol{\sigma}_V|\boldsymbol{\omega}_{V^c})\mu(d\boldsymbol{\omega}).$$

The principal result due to Dobrushin, Lanford and Ruelle, is that for every UAC interaction Φ, there exists at least one Gibbs measure, i.e., $\mathscr{G}(\Phi) = \mathscr{G}(\gamma^\Phi) \neq \varnothing$. The specification $\gamma^\Phi = \{\gamma_V^\Phi\}$ has the following important properties:

- **Uniform non-nullness**: for every $V \Subset \mathbb{Z}$ there exist positive constants a_V, b_V such that
$$a_V \le \gamma_V^\Phi(\boldsymbol{\sigma}_V|\boldsymbol{\omega}_{V^c}) \le b_V$$
for all $\boldsymbol{\sigma}_V \in \mathscr{A}^V$ and every $\boldsymbol{\omega} \in X$.

- **Quasilocality**: for every $V \Subset \mathbb{Z}$,
$$v_{V,n} := \sup_{\boldsymbol{\sigma}_V\in\mathscr{A}^V} \sup_{\boldsymbol{\omega}\in X} \sup_{\boldsymbol{\eta},\boldsymbol{\chi}\in X} \Big|\gamma_V^\Phi(\boldsymbol{\sigma}_V|\boldsymbol{\omega}_{[-n,n]\setminus V}\boldsymbol{\eta}_{[-n,n]^c\setminus V}) - \gamma_V^\Phi(\boldsymbol{\sigma}_V|\boldsymbol{\omega}_{[-n,n]\setminus V}\boldsymbol{\xi}_{[-n,n]^c\setminus V})\Big| \to 0,$$
as $n \to \infty$.

A third important property is that the conditional probabilities are *consistent* between different sets $V \Subset \mathbb{Z}$, for all external configurations.

Based on these properties a specification can be defined without mentioning an interaction. Define a specification $\gamma = \{\gamma_V : V \Subset \mathbb{Z}\}$ as a family of probability kernels on $(X, \mathscr{B}(X))$ where the kernels $\gamma_V : \mathscr{B}(X) \times X \to (0,1)$ are such that for all $V \Subset \mathbb{Z}$ one has:

- For every $C \in \mathscr{B}(X)$, $\gamma_V(C|\cdot)$ is measurable w.r.t. $\mathscr{B}_{V^c} = \sigma\{[\boldsymbol{\omega}_W] : W \Subset V^c\}$
- For every $C \in \mathscr{B}_{V^c}(X)$, $\gamma_V(C|\boldsymbol{\omega}) = \mathbb{1}_C(\boldsymbol{\omega})$
- For every $W \subset V$, one has $\gamma_V \gamma_W = \gamma_V$, where the product $\gamma_W \gamma_V$ is given by

$$\gamma_V \gamma_W(C|\boldsymbol{\omega}) = \int_X \gamma_W(C|\tilde{\boldsymbol{\omega}})\gamma_V(d\tilde{\boldsymbol{\omega}}|\boldsymbol{\omega}).$$

A fundamental result of Kozlov and Sullivan [55, 84] states that a specification $\gamma = \{\gamma_V\}$ that is uniformly non-null and quasilocal is a Gibbsian specification for some UAC interaction Φ, i.e., $\gamma = \gamma^{\Phi}$. As g—measures are translation-invariant, we will only consider those Gibbs measures that are translation-invariant.

Definition 2.9. A specification is translation-invariant if for any $a \in \mathbb{Z}$ and $V \Subset \mathbb{Z}$

$$\gamma_V((S^a\boldsymbol{\omega})_V|(S^a\boldsymbol{\omega})_{V^c}) = \gamma_{V+a}(\boldsymbol{\omega}_{V+a}|\boldsymbol{\omega}_{(V+a)^c}).$$

For a translation-invariant quasilocal specification there exists a translation-invariant (Gibbs) measure consistent with this specification [77]. If a continuous energy function is constructed as $\varphi(\boldsymbol{\omega}) = \sum_{V \ni 0} \frac{1}{|V|}\Phi_V(\omega_0|\boldsymbol{\omega}_{\{0\}^c})$ from a UAC interaction Φ and μ is a translation-invariant measure, then μ is an equilibrium state for φ if and only if it is a Gibbs measure for Φ.

Specifications are defined for all finite subsets, for Definition 2.3 to make sense, however one needs to reconstruct a full specification from singletons.

Theorem 2.10. *Let μ be a translation-invariant measure on X, then μ is a Gibbs measure if and only if it satisfies the conditions of Definition 2.3, for some $\gamma \in \mathscr{G}(X)$.*

Proof. If the translation-invariant measure μ is Gibbs then it is consistent with a specification, therefore, for any finite $V \Subset \mathbb{Z}$

$$\mu(\boldsymbol{\omega}_V|\boldsymbol{\omega}_{V^c}) = \gamma_V(\boldsymbol{\omega}_V|\boldsymbol{\omega}_{V^c}),$$

for μ-a.e. $\boldsymbol{\omega}$. Thus, in particular, this holds for $V = \{0\}$.

In the opposite direction, suppose μ is a fully supported translation-invariant measure satisfying

$$\mu(\omega_0|\omega_{-\infty}^{-1},\omega_1^{\infty}) = \gamma(\ldots,\omega_{-2},\omega_{-1},\omega_0,\omega_1,\omega_2,\ldots) = \gamma(\boldsymbol{\omega}),$$

for μ-a.e. $\boldsymbol{\omega} \in X$ and some $\gamma \in \mathscr{G}(X)$. From theorem A.4 in [29], which also applies to non-invariant measures, it follows that a specification can be uniquely constructed from a family of non-null single-site kernels $\gamma_{\{i\}}$ if they satisfy the following consistency conditions: the expression

$$\frac{\gamma_{\{i\}}\big(\eta_i \,\big|\, \eta_j\, \omega_{\{i,j\}^c}\big)}{\sum_{\alpha_i} \frac{\gamma_{\{i\}}\big(\alpha_i \big| \eta_j\, \omega_{\{i,j\}^c}\big)}{\gamma_{\{j\}}\big(\eta_j \big| \alpha_i\, \omega_{\{i,j\}^c}\big)}}$$

must be invariant under the exchange $i \longleftrightarrow j$. The reason for this consistency condition is that the expression above is for μ—a.e. $\boldsymbol{\omega} \in X$ equal to

$$\mu\big(\eta_i \,\big|\, \eta_j\, \omega_{\{i,j\}^c}\big)\,\mu\big(\eta_j \,\big|\, \omega_{\{i,j\}^c}\big) \;=\; \mu\big(\eta_i\, \eta_j \,\big|\, \omega_{\{i,j\}^c}\big)\,,$$

which is symmetrical under the $i \longleftrightarrow j$ exchange. The construction in the proof of Theorem A.4 of [29] shows that the continuity of the single-site kernels extends to all finite-volume kernels. Alternatively the theorem can be proven using the conditions from [66]. □

2.4 Main results

2.4.1 Main theorems

Known conditions for g—measures to be Gibbs relate to rapid decay of variations of the corresponding g—functions. Likewise our main result is a regularity condition on the function g that is necessary and sufficient for the corresponding g—measure to be Gibbs.

Theorem 2.11. *Let μ be a g—measure on $X_+ = \mathscr{A}^{\mathbb{Z}_+}$. Viewed as a measure on $X = \mathscr{A}^{\mathbb{Z}}$, μ is Gibbs if and only if the sequence of functions $[\tilde{f}_n^{\sigma_0,\eta_0}]_{n\in\mathbb{N}}$, given by*

$$\tilde{f}_n^{\sigma_0,\eta_0}(\boldsymbol{\omega}) = \prod_{i=-n}^{-1} \frac{g(\omega_i^{-1}\sigma_0\omega_1^{\infty})}{g(\omega_i^{-1}\eta_0\omega_1^{\infty})} \tag{2.5}$$

converges, for all $\sigma_0,\eta_0 \in \mathscr{A}$, uniformly in $\boldsymbol{\omega} \in X$, as $n \to \infty$.

Proof. Let us start by showing that condition (2.5) is sufficient for μ to be Gibbs. We can express two-sided conditional probabilities $\mu(\omega_0|\omega_{-n}^{-1},\omega_1^m)$ as follows:

$$\mu(\omega_0|\omega_{-n}^{-1},\omega_1^m)=\frac{\mu(\omega_{-n}^{-1}\omega_0\omega_1^m)}{\sum_{\sigma_0\in\mathscr{A}}\mu(\omega_{-n}^{-1}\sigma_0\omega_1^m)}=\frac{1}{\displaystyle\sum_{\sigma_0\in\mathscr{A}}\frac{\mu(\omega_{-n}^{-1}\sigma_0\omega_1^m)}{\mu(\omega_{-n}^{-1}\omega_0\omega_1^m)}}. \tag{2.6}$$

Let us introduce the following functions on X: for fixed $\sigma_0,\eta_0\in\mathscr{A}$ put

$$f_{n,m}^{\sigma_0,\eta_0}(\boldsymbol{\omega})=\frac{\mu(\omega_{-n}^{-1}\sigma_0\omega_1^m)}{\mu(\omega_{-n}^{-1}\eta_0\omega_1^m)},$$

$$f_n^{\sigma_0,\eta_0}(\boldsymbol{\omega})=\tilde{f}_n^{\sigma_0,\eta_0}(\boldsymbol{\omega})\times\frac{g(\sigma_0\omega_1^\infty)}{g(\eta_0\omega_1^\infty)}=\prod_{i=-n}^{-1}\frac{g(\omega_i^{-1}\sigma_0\omega_1^\infty)}{g(\omega_i^{-1}\eta_0\omega_1^\infty)}\times\frac{g(\sigma_0\omega_1^\infty)}{g(\eta_0\omega_1^\infty)},$$

Since μ is a g—measure, by Theorem 2.2, for every $i\in\mathbb{Z}$, one has

$$\mu(\omega_i|\omega_{i+1}^m)\rightrightarrows g(\omega_i^\infty),$$

as $m\to\infty$, uniformly in $\boldsymbol{\omega}\in X$. Therefore, for each $n\in\mathbb{N}$,

$$\begin{aligned}f_{n,m}^{\sigma_0,\eta_0}(\boldsymbol{\omega})&=\frac{\mu(\omega_{-n}^{-1}\sigma_0\omega_1^m)}{\mu(\omega_{-n}^{-1}\eta_0\omega_1^m)}=\prod_{i=-n}^{-1}\frac{\mu(\omega_i|\omega_{i+1}^{-1}\sigma_0\omega_1^m)}{\mu(\omega_i|\omega_{i+1}^{-1}\eta_0\omega_1^m)}\times\frac{\mu(\sigma_0|\omega_1^\infty)}{\mu(\eta_0|\omega_1^\infty)}\\&\rightrightarrows\prod_{i=-n}^{-1}\frac{g(\omega_i^{-1}\sigma_0\omega_1^\infty)}{g(\omega_i^{-1}\eta_0\omega_1^\infty)}\times\frac{g(\sigma_0\omega_1^\infty)}{g(\eta_0\omega_1^\infty)}=f_n^{\sigma_0,\eta_0}(\boldsymbol{\omega})\end{aligned}$$

as $m\to\infty$ uniformly in $\boldsymbol{\omega}$. Since g is uniformly bounded away from 0, the functions $f_n^{\sigma_0,\eta_0}$, for fixed n, are continuous and bounded away from 0,

$$\inf_{\boldsymbol{\omega}\in X_+}f_n^{\sigma_0,\eta_0}(\boldsymbol{\omega})\geq\left(\frac{\inf_{\boldsymbol{\omega}}g(\boldsymbol{\omega})}{\sup_{\boldsymbol{\omega}}g(\boldsymbol{\omega})}\right)^{n+1}>0.$$

Uniform convergence of (2.5) immediately implies that

$$f_n^{\eta_0,\sigma_0}(\boldsymbol{\omega})=\tilde{f}_n^{\sigma_0,\eta_0}(\boldsymbol{\omega})\times\frac{g(\sigma_0\omega_1^\infty)}{g(\eta_0\omega_1^\infty)}$$

converges, uniformly in $\boldsymbol{\omega}$, as $n\to\infty$. Thus the limiting function,

$$f^{\sigma_0,\eta_0}(\boldsymbol{\omega})=\prod_{i=-\infty}^{-1}\frac{g(\omega_i^{-1}\sigma_0\omega_1^\infty)}{g(\omega_i^{-1}\eta_0\omega_1^\infty)}\times\frac{g(\sigma_0\omega_1^\infty)}{g(\eta_0\omega_1^\infty)},$$

is continuous and is bounded away from 0. Indeed, since for every n, one has $f_n^{\sigma_0,\eta_0}(\boldsymbol{\omega})f_n^{\eta_0,\sigma_0}(\boldsymbol{\omega})=1$, we also have $f^{\sigma_0,\eta_0}(\boldsymbol{\omega})f^{\eta_0,\sigma_0}(\boldsymbol{\omega})=1$; f^{σ_0,η_0} is bounded from above as a continuous function on the compact X, it is bounded below by $1/||f^{\sigma_0,\eta_0}||_\infty$, for all σ_0, η_0. Finally, Equation (2.6) implies that

$$\lim_{n\to\infty}\lim_{m\to\infty}\mu(\omega_0|\omega_{-n}^{-1},\omega_1^m) = \frac{1}{\sum_{\sigma_0\in\mathscr{A}} f^{\sigma_0,\omega_0}(\boldsymbol{\omega})} =: \gamma(\boldsymbol{\omega})$$

for all $\boldsymbol{\omega}$. Therefore μ is consistent with a continuous uniformly non-null specification; hence it is a Gibbs measure.

Conversely assume that μ is a g—measure such that $\tilde{f}_n^{\sigma,\eta}$ does not converge uniformly, furthermore assume μ is a Gibbs measure. Then $f_{n,m}^{\eta,\sigma}$, for any n, converges uniformly to $f_n^{\sigma,\eta}$ as $m \to \infty$. It follows that if $f_n^{\sigma_0,\eta_0}$ does not converge uniformly then $f_{n,m}^{\sigma,\eta}$ does not converge uniformly either. However, for a Gibbs measure μ

$$f_{n,m}^{\sigma_0,\eta_0}(\omega) = \frac{\mu(\omega_{-n}^{-1}\sigma_0\omega_1^m)}{\mu(\omega_{-n}^{-1}\eta_0\omega_1^m)} = \frac{\mu(\sigma_0|\omega_{-m}^{-1},\omega_1^n)}{\mu(\eta_0|\omega_{-m}^{-1},\omega_1^n)},$$

must converge uniformly. This follows from uniform non-nullness and Theorem 2.4 combined with uniform continuity of fractions on the relevant part of the domain. This leads to a contradiction, therefore a g—measure is a Gibbs measure if and only if the conditions of Theorem 2.11 are satisfied. □

Theorem 2.11 shows how the regularity of g—functions determines the continuity of two-sided conditional probabilities. A reversibility condition for g-measures should have a similar form.

Theorem 2.12. *A g—measure μ is reversible if and only if the sequence*

$$\hat{f}_n(\boldsymbol{\omega}) \equiv \left[\prod_{i=-n}^{-1} \frac{\mu(\omega_i|\omega_{i+1}^0)}{\mu(\omega_i|\omega_{i+1}^{-1})}\right]_{n\in\mathbb{N}} \tag{2.7}$$

converges uniformly in $\boldsymbol{\omega} \in X$, bounded away from 0, as $n \to \infty$.

Proof. The proof is almost immediate as $\hat{f}_n(\boldsymbol{\omega}) = \frac{\mu(\omega_{-n}\ldots\omega_{-1}|\omega_0)}{\mu(\omega_{-n}\ldots\omega_{-1})}$, hence, if $[\hat{f}_n]_{n\in\mathbb{N}}$ converges in $C(X)$ as $n \to \infty$, then $\mu(\omega_0)\hat{f}_n(\boldsymbol{\omega}) = \frac{\mu(\omega_{-n}\ldots\omega_{-1}\omega_0)}{\mu(\omega_{-n}\ldots\omega_{-1})}$ converges uniformly. Thus by Theorem 2.2 μ_-, the reverse of μ_+, is a g—measure. The corresponding g—function, $g_-(\boldsymbol{\omega})$, is given by $g_-(\boldsymbol{\omega}) = \lim_{n\to\infty}\mu(\omega_0)\hat{f}_n(\boldsymbol{\omega})$. □

Convergence of the sequence (2.7) is often more difficult to show than the convergence of (2.5). The reason is that it is not easily reduced to a property of the function g or the potential, the quantities one usually knows or controls.

2.4.2 Good future and uniqueness

In this section we will relate Theorem 2.11 to a known sufficient condition, called *Good Future* [29], introduced by Fernández and Maillard. This condition applies to processes that are not necessarily translation-invariant and in the case of g—measures it reduces to the following: for $f \in C(X_+)$ let

$$\partial_k(f) \equiv \sup_{\omega \in X_+, \sigma_k, \eta_k \in \mathscr{A}} \left| f(\omega_0^{k-1}\sigma_k\omega_{k+1}^\infty) - f(\omega_0^{k-1}\eta_k\omega_{k+1}^\infty) \right|, \tag{2.8}$$

then the g—function $g \in G(X_+)$ has Good Future if the following summability condition is satisfied:

$$\sum_{k=1}^{\infty} \partial_k(g) < \infty. \tag{2.9}$$

Furthermore we define $\mathrm{GF}(X_+)$ to be the set of g—functions satisfying this condition.

Remark 2.13. Summable variation of $g \in G(X_+)$ is equivalent to summable variation of $\log(g)$. Likewise, $g \in \mathrm{GF}(X_+)$ is equivalent to

$$\sum_{k=1}^{\infty} \partial_k(\log(g)) < \infty.$$

Theorem 2.14 (Fernández, Maillard [29]). *If μ is a g—measure and $g \in G(X_+)$, if $g \in \mathrm{GF}(X_+)$, then μ is Gibbs.*

Proof. As μ is a g—measure there exists $c > 0$ such that for all ω one has $c < g(\boldsymbol{\omega}) < 1 - c$. Thus, for any $\sigma_k \in \mathscr{A}$,

$$\frac{g(\omega) - \partial_k(g)}{g(\omega)} \leq \frac{g(\omega_0^{k-1}\sigma_k\omega_{k+1}^\infty)}{g(\omega)} \leq \frac{g(\omega) + \partial_k(g)}{g(\omega)}$$

and therefore

$$\max\left\{0, 1 - \frac{\partial_k(g)}{1-c}\right\} \leq \frac{g(\omega_0^{k-1}\sigma_k\omega_{k+1}^\infty)}{g(\omega)} \leq 1 + \frac{\partial_k(g)}{c}.$$

The function g is bounded away from 0 and 1 and continuous so we can choose an $N > 0$ such that $1 - \frac{\partial_k(g)}{c} > 0$, for $k \geq N$. Now notice that, due to the convergence of $\sum_{k=1}^{\infty} \partial_k(g)$, the following two products converge, for N as above, to a nonzero value:

$$\prod_{k=N}^{\infty}\left(1 - \frac{\partial_k(g)}{1-c}\right), \quad \prod_{k=N}^{\infty}\left(1 + \frac{\partial_k(g)}{c}\right).$$

And for sufficiently large $N > 0$:

$$\prod_{k=N+1}^{\infty}\left(1-\frac{\partial_k(g)}{1-c}\right) \leq \prod_{i=-\infty}^{-N-1}\frac{g(\omega_i\omega_{i+1}^{-1}\sigma_0\omega_1^{\infty})}{g(\omega_i^{\infty})} \leq \prod_{k=N+1}^{\infty}\left(1+\frac{\partial_k(g)}{c}\right).$$

It follows that for any $\varepsilon > 0$ there exists an N such that the upper and lower bounds above are ε—close to 1. This implies that the sequence

$$\prod_{i=-N}^{-1}\frac{g(\omega_i\omega_{i+1}^{-1}\sigma_0\omega_1^{\infty})}{g(\omega_i\omega_{i+1}^{\infty})}$$

is uniformly Cauchy and thus uniformly convergent. Therefore, by Theorem 2.11, a g—measure for which (2.9) converges uniformly is Gibbs. □

An interesting consequence of the above result is that the smoothness required for Gibbsianity does not imply uniqueness. Hulse [46] has shown that for any $\lambda > 1$ there exists a function $g \in G(X_+)$, with multiple g—measures, such that

$$\sum_{k=1}^{\infty}\partial_k(g) < \lambda < \infty.$$

An important distinction between the Good Future condition and Theorem 2.11 is that, in Eq. (2.5), the supremum is taken after the product, meaning single factors in the product in Eq. (2.5) are not always of the order of $\partial_k(g)$. Furthermore, different factors in the product might cancel, therefore g—measures with slowly decaying $\partial_k(g)$ could still be Gibbs states. We will give an example of a measure having such behaviour in Section 2.5.4.

2.4.3 Special classes of g-measures

There are several extensively studied classes of g—measures. We will discuss some of them from the point of view of being Gibbs. Hölder continuity was already known as a sufficient condition for being Gibbs [80]. Functions $g \in G(X_+)$ with $\sum_n \mathrm{var}_n(g) < \infty$ are said to have summable variation, the corresponding g—measures are known to be Gibbs [22, 56]. The set of functions with summable variation contains the Hölder continuous functions as a proper subset. In turn these are a subset of a set of functions which were first introduced by Walters [91]. Let

$$S_n\phi \equiv \sum_{i=0}^{n-1}\phi\circ S^i,$$

the set of functions $\mathrm{Wal}(X_+, S)$ is defined as:

$$\mathrm{Wal}(X_+, S) = \{\phi \in C(X_+) : \sup_{n\geq 1}\mathrm{var}_{n+p}(S_n\phi) \to 0 \text{ as } p \to \infty\}.$$

Remark 2.15. In order to compare Walters' class with Theorem 2.11 we point out that the uniform convergence in the theorem is equivalent to:

$$\sup_{n\geq 0} \partial_{n+p}(S_n \log g) \to 0 \quad \text{as } p \to \infty.$$

Given a potential $\phi \in \mathrm{Wal}(X_+, S)$, there exists a unique equilibrium state for ϕ that is also a $g-$measure [91]. Hölder continuity or summable variation of a potential $\phi \in C(X_+)$ implies $\phi \in \mathrm{Wal}(X_+, S)$.

Theorem 2.16 (Walters [95]). *If μ is (the natural extension of) a $g-$measure with* $\log(g) \in \mathrm{Wal}(X_+, S)$ *then it is also a $g-$measure in the reverse direction, with* $\log(g_-) \in \mathrm{Wal}(X_-, S_-)$*, and a Gibbs measure.*

Proof. The reversibility with $\log(g_-) \in \mathrm{Wal}(X_-, S_-)$ has been shown by Walters [95].

If $\log(g) \in \mathrm{Wal}(X_+, S)$ then the corresponding condition can be written as

$$\sup_{n\geq 1} \sup_{\omega_0^{n+p-1}=\eta_0^{n+p-1}} \left(\sum_{i=0}^{n-1} \log(g(S^i\omega)) - \sum_{i=0}^{n-1} \log(g(S^i\eta)) \right) \to 0, \text{ as } p \to \infty.$$

Which is equivalent to

$$\sup_{n\geq 1} \sup_{\omega_0^{n+p-1}=\eta_0^{n+p-1}} \log\left(\prod_{i=0}^{n-1} \frac{g(S^i\omega)}{g(S^i\eta)} \right) \to 0, \text{ as } p \to \infty.$$

Writing down the previous expression for the condition imposed on $\log(g)$ in a more explicit form one gets:

$$\sup_{n\geq 1} \sup_{\omega,\eta\in X_+} \log\left(\prod_{i=0}^{n-1} \frac{g(\omega_i \omega_{i+1}^{n-1} \omega_n^{n+p-1} \omega_{n+p}^{\infty})}{g(\omega_i \omega_{i+1}^{n-1} \omega_n^{n+p-1} \eta_{n+p}^{\infty})} \right) \to 0 \text{ as } p \to \infty.$$

We can, for $\omega \in X$, let g act on ω_i^{∞}, with $i \in \mathbb{Z}$. For every $\varepsilon > 0$ there exists a $\delta > 0$ such that if $|\log(1+x)| < \varepsilon$ then $|x| < \varepsilon$, therefore:

$$\sup_{n\geq 1} \sup_{\omega,\eta\in X} \left| \left(\prod_{i=-n-p}^{-p-1} \frac{g(\omega_i \omega_{i+1}^{-p-1} \omega_{-p}^{-1} \omega_0^{\infty})}{g(\omega_i \omega_{i+1}^{-p-1} \omega_{-p}^{-1} \eta_0^{\infty})} \right) - 1 \right| \to 0 \text{ as } p \to \infty.$$

Here the entire tail can be different in the fraction. We only need to allow them to differ at a single site. So, for all $\varepsilon > 0$ there exists an $N > 0$ such that if $p > 0$ then

$$\sup_{n\geq N} \sup_{\omega\in X, \sigma\in\mathscr{A}} \left| \left(\prod_{i=-n-p}^{-n-1} \frac{g(\omega_i \omega_{i+1}^{-1} \omega_0^{\infty})}{g(\omega_i \omega_{i+1}^{-1} \sigma \omega_1^{\infty})} \right) - 1 \right| < \varepsilon.$$

As the finite products are bounded away from 0 and ∞ the tail being arbitrarily close to 1 implies uniform convergence and boundedness, so that we can use theorem 2.11. Therefore the natural extension of a $g-$measure with $\log(g) \in \mathrm{Wal}(X_+, S)$ is Gibbs. $\square$

An even larger set of well behaved potentials is given by

$$\mathrm{Bow}(X_+, S) = \left\{ \phi \in C(X_+) : \sup_{n\geq 1} \mathrm{var}_n(S_n\phi) < \infty \right\}.$$

For any potential $\phi \in \mathrm{Bow}(X_+, S)$ a unique equilibrium state exists and this is a Bowen-Gibbs measure. However, there is no guarantee that an equilibrium state for a potential $\phi \in \mathrm{Bow}(X_+, S)$ is a $g-$measure.

2.4.4 Reversibility of g-measures

As was shown by Walters [95] and discussed above, if μ is a $g-$measure with $\log(g) \in \mathrm{Wal}(X_+, S)$, then its reverse μ_- is a $g-$measure and has $\log(g_-) \in \mathrm{Wal}(X_-)$. Walters proved that if $\log(g) \in \mathrm{Bow}(X_+, S)$ and if the measure is also a reverse $g-$measure, then the reverse $g-$function, g_-, satisfies $\log(g_-) \in \mathrm{Bow}(X_-)$. However it remained an open question whether μ_- is a $g-$measure if the potential of μ_+ satisfies $\log(g) \in \mathrm{Bow}(X_+, S)$. Borrowing elements from both non-Gibbsian examples [33, 95] we can show that this is not always the case.

Proposition 2.17. *There exists a $g-$measure μ_+ with* $\log(g) \in \mathrm{Bow}(X_+, S)$ *for which the reverse measure μ_- is not a $g-$measure.*

Proof. We will show this by constructing an example. First we define a $g-$function, then show that $\log(g) \in \mathrm{Bow}(X_+, S)$ and finally show that the reverse is not a $g-$measure.

Step 1. Take $\{\nu_k\}_{k\geq 0}$ a sequence of reals satisfying

(i) $\nu_k \to 0$, as $n \to \infty$,

(ii) there exists a $K > 0$ with $\left|\sum_{k=0}^{N} \nu_k\right| < K$ for all $N > 0$,

(iii) $\sum_{k=0}^{\infty} \nu_k$ does not converge,

(iv) $k \to \nu_k$ is injective.

The fourth requirement is not needed for the validity of the example, but it simplifies the combinatorial aspects of the proof. It is easy to see that such a sequence exists.

Let $\xi \in \left(1, 2^{\frac{1}{2K}}\right)$, $\mathscr{A} = \{0,1\}$, $n \geq 1$. Define the function g as follows

$$\begin{cases} g(0\,0^n10\boldsymbol{\eta}) = \frac{1}{2}\xi^{-\nu_n}, \\ g(1\,0^n10\,\boldsymbol{\eta}) = 1 - \frac{1}{2}\xi^{-\nu_n}, \end{cases} \\ \begin{cases} g(00^n11\boldsymbol{\eta}) = \frac{1}{2}\xi^{\nu_n}, \\ g(10^n11\boldsymbol{\eta}) = 1 - \frac{1}{2}\xi^{\nu_n}, \end{cases} \tag{2.10}$$

for all $\boldsymbol{\eta} \in \mathscr{A}^\infty$ and let $g(\boldsymbol{\omega}) = \frac{1}{2}$ otherwise. Then $0 < g < 1$ and $\sum_{\sigma\in\mathscr{A}} g(\sigma\omega_1^\infty) = 1$, for any $\boldsymbol{\omega} \in X_+$. Furthermore

$$\begin{aligned} \mathrm{var}_m(g) &= |g(\underbrace{00^n1}_{m}0\boldsymbol{\eta}) - g(\underbrace{00^n1}_{m}1\boldsymbol{\sigma})| = \left|\frac{1}{2}\xi^{\nu_{m-2}} - \frac{1}{2}\xi^{-\nu_{m-2}}\right| \\ &= \sinh(\log(\xi)\nu_{m-2}) \to 0 \quad \text{as} \quad m \to \infty, \end{aligned}$$

since $\nu_k \to 0$ as $k \to \infty$. Thus g is a continuous function.

Step 2. We will now show that $\log(g) \in \mathrm{Bow}(X_+, S)$. It sufficient to show that $\mathrm{var}_n(S_n \log g)$ is uniformly bounded for all $n \geq 1$. Note that

$$\mathrm{var}_n(S_n \log g) = \sup_{\boldsymbol{\omega},\boldsymbol{\eta},\boldsymbol{\sigma}\in X_+} \log\left(\prod_{i=0}^{n-1} \frac{g(\omega_i^{n-1}\eta_n^\infty)}{g(\omega_i^{n-1}\sigma_n^\infty)}\right). \tag{2.11}$$

Suppose $n \geq 1$ and let $\boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{\sigma} \in X$ be the points at which the supremum in (2.11) is attained. Suppose $0 \leq i < n-2$ is such that

$$\frac{g(\omega_i^{n-1}\eta_n^\infty)}{g(\omega_i^{n-1}\sigma_n^\infty)} \neq 1.$$

Then, by the definition of g, one necessarily has

$$\omega_i\omega_{i+1} = 00 \quad \text{or} \quad \omega_i\omega_{i+1} = 10.$$

Put

$$I_1^{(n)} = \left\{1 \leq i < n-1 : \omega_i\omega_{i+1} = 00 \text{ and } \frac{g(\omega_i^{n-1}\eta_n^\infty)}{g(\omega_i^{n-1}\sigma_n^\infty)} \neq 1\right\},$$
$$I_2^{(n)} = \left\{1 \leq i < n-1 : \omega_i\omega_{i+1} = 10 \text{ and } \frac{g(\omega_i^{n-1}\eta_n^\infty)}{g(\omega_i^{n-1}\sigma_n^\infty)} \neq 1\right\},$$

then

$$\prod_{i=0}^{n-1} \frac{g(\omega_i^{n-1}\eta_n^\infty)}{g(\omega_i^{n-1}\sigma_n^\infty)} = \prod_{i\in I_1^{(n)}\cup I_2^{(n)}\cup\{n-1\}} \frac{g(\omega_i^{n-1}\eta_n^\infty)}{g(\omega_i^{n-1}\sigma_n^\infty)}. \tag{2.12}$$

We proceed by evaluating the contribution of factors for indices in $I_1^{(n)}$. Note that since $\omega_i^{n-1}\boldsymbol{\eta}$ and $\omega_i^{n-1}\boldsymbol{\sigma}$ start with two 0's: $\omega_i\omega_{i+1} = 00$, one has

$$\frac{g(\omega_i^{n-1}\eta_n^\infty)}{g(\omega_i^{n-1}\sigma_n^\infty)} = \frac{\frac{1}{2}\xi^{\pm\nu_m}}{\frac{1}{2}\xi^{\pm\nu_k}} = \xi^{\pm\nu_m\mp\nu_k},$$

where $m, k \in \mathbb{N}\cup\{+\infty\}$ depend on the first occurrence of 10 or 11 in the corresponding sequences; with a minor abuse of notation we let $m = +\infty$ or $k = +\infty$ if $\omega_i^{n-1}\boldsymbol{\eta} = 0^\infty$ or $\omega_i^{n-1}\boldsymbol{\sigma} = 0^\infty$, respectively. Note that the first 1 in $\omega_i^{n-1}\boldsymbol{\eta}$ and $\omega_i^{n-1}\boldsymbol{\sigma}$ cannot occur before position $n-1$. Indeed, if 1 appears earlier, i.e., $\omega_j = 1$ for some $j \le n-1$, then necessarily

$$\frac{g(\omega_i^{n-1}\eta_n^\infty)}{g(\omega_i^{n-1}\sigma_n^\infty)} = 1.$$

Finally, since we concluded that the first 1 in $\omega_i^{n-1}\boldsymbol{\eta}$ and $\omega_i^{n-1}\boldsymbol{\sigma}$ occurs in position n or later, one has

$$\omega_i, \omega_{i+1}, \ldots, \omega_{n-2} = 0, 0, \ldots, 0.$$

Hence, if $I_1^{(n)} \neq \varnothing$ and $i_* = \min I_1^{(n)}$, then

$$\{i_*, i_*+1, \ldots, n-3\} \subseteq I_1^{(n)} \subseteq \{i_*, i_*+1, \ldots, n-2\}, \tag{2.13}$$

where we used the injectivity of the sequence. A similar argument applies to $I_2^{(n)}$: if $i \in I_2^{(n)}$, i.e., $\omega_i\omega_{i+1} = 10$, and

$$\frac{g(\omega_i^{n-1}\eta_n^\infty)}{g(\omega_i^{n-1}\sigma_n^\infty)} \neq 1,$$

then 1 does not appear in $\omega_{i+1}, \ldots, \omega_{n-2}$, and hence if $I_2^{(n)} \neq \varnothing$, $I_2^{(n)}$ is a singleton, say $I_2^{(n)} = \{i\}$ and $\omega_{i+1}, \ldots, \omega_{-2} = 0, \ldots, 0$.

We are now able to derive uniform bounds for (2.12). Firstly, since g is a continuous positive function, and thus uniformly bounded away from 0 and 1, by taking into account that $|I_2^{(n)}| \le 1$, one has

$$\prod_{i=0}^{n-1} \frac{g(\omega_i^{n-1}\eta_n^\infty)}{g(\omega_i^{n-1}\sigma_n^\infty)} = \prod_{i\in I_1^{(n)}\cup I_2^{(n)}\cup\{n-1\}} \frac{g(\omega_i^{n-1}\eta_n^\infty)}{g(\omega_i^{n-1}\sigma_n^\infty)} \le \left(\frac{\sup g}{\inf g}\right)^2 \prod_{i\in I_1^{(n)}} \frac{g(\omega_i^{n-1}\eta_n^\infty)}{g(\omega_i^{n-1}\sigma_n^\infty)}. \tag{2.14}$$

If $I_1^{(n)}$ is empty, the proof of the claim that $\log g$ is in the Bowen class is completed. If $I_1^{(n)}$ is not empty, taking (2.13) into account, one has

$$\prod_{i\in I_1^{(n)}} \frac{g(\omega_i^{n-1}\eta_n^{\infty})}{g(\omega_i^{n-1}\sigma_n^{\infty})} \le \left(\frac{\sup g}{\inf g}\right)\prod_{i=i_*}^{n-3} \frac{g(\omega_i^{n-1}\eta_n^{\infty})}{g(\omega_i^{n-1}\sigma_n^{\infty})}$$

Consider $(\omega_{i_*}\ldots\omega_{n-2}\omega_{n-1}\eta_n\eta_{n+1}\ldots) = (0,\ldots,0,\omega_{n-1}\eta_n\eta_{n+1}\ldots)$. Let k be the position of the first 1 in this string. Note $k \ge n-1$, we let $k=+\infty$ if 1 does not appear. Depending on the next symbol: 0 or 1, one has

$$\prod_{i=i_*}^{n-3} g(\omega_i^{n-1}\eta_n^{\infty}) = \frac{1}{2^{n-i_*-2}}\xi^{-\sum_{i=i_*}^{n-3}\nu_{k-i-1}} \text{ or } \prod_{i=i_*}^{n-3} g(\omega_i^{n-1}\eta_n^{\infty}) = \frac{1}{2^{n-i_*-2}}\xi^{+\sum_{i=i_*}^{n-3}\nu_{k-i-1}},$$

respectively. One has a similar expression for $(\omega_{i_*}\ldots\omega_{n-2}\omega_{n-1}\sigma_n\sigma_{n+1}\ldots)$. Since $\left|\sum_{j=m_1}^{m_2}\nu_j\right| = \left|\sum_{j=0}^{m_2}\nu_j - \sum_{j=0}^{m_1}\nu_j\right| \le 2K$ for all $m_1, m_2 \in \mathbb{N}$, we conclude that $\mathrm{var}_n(S_n \log g)$ is uniformly bounded.

Step 3. In the last part of the proof we show that the reverse of the unique g-measure μ_+ for the function g, given by (2.10), is not a g-measure. Let μ be the natural extension to X of μ_+. By Theorem 2.2 it suffices to present an element $\boldsymbol{\omega} \in X_-$ such that the sequence

$$a_n = \mu(\omega_0|\omega_{-1},\ldots,\omega_{-n}), \quad n\in\mathbb{N},$$

does not converge as $n\to\infty$. Let

$$\boldsymbol{\omega} = (\ldots,\omega_{-n},\ldots,\omega_{-2},\omega_{-1},\omega_0) = (\ldots,0,\ldots,0,1,0)\in X_-.$$

We use $0_{-\infty}^{-2}1_1 0_0$ as a shorthand notation for $\boldsymbol{\omega}$. Since μ is a fully supported g-measure, for $n\ge 2$, one has

$$a_n = \mu(0_0|1_{-1}0_{-n}^{-2}) = \frac{\mu(0_{-n}^{-2}1_{-1}0_0)}{\mu(0_{-n}^{-2}1_{-1}0_0)+\mu(0_{-n}^{-2}1_{-1}1_0)} = \frac{1}{1+\frac{\mu(0_{-n}^{-2}1_{-1}1_0)}{\mu(0_{-n}^{-2}1_{-1}0_0)}}.$$

Furthermore, note that

$$\begin{aligned} b_n = \frac{\mu(0_{-n}^{-2}1_{-1}1_0)}{\mu(0_{-n}^{-2}1_{-1}0_0)} &= \prod_{j=-n}^{-3}\frac{\mu(0_j|0_{j+1}^{-2}1_{-1}1_0)}{\mu(0_j|0_{j+1}^{-2}1_{-1}0_0)} \times \frac{\mu(0_{-2}1_{-1}1_0)}{\mu(0_{-2}1_{-1}0_0)} \\ &= \left(\prod_{j=-n}^{-3}\xi^{2\nu_{-j-2}}\right)\frac{\mu(0_{-2}1_{-1}1_0)}{\mu(0_{-2}1_{-1}0_0)} = C\,\xi^{2\sum_{j=1}^{n-2}\nu_j}, \end{aligned}$$

for some $C>0$. However, since the partial sums $\sum_{j=1}^{n}\nu_j$ are bounded and oscillate, the sequences $\{b_n\}$ and $\{a_n\}$ do not converge as $n\to\infty$. Thus we constructed an example μ_+ of a g—measure with $\log(g)\in \mathrm{Bow}(X_+,S)$, such that its reverse is not a g—measure. □

2.5 Examples and an Overview

As a large number of conditions are relevant to this paper, this section aims at providing some insight in how these conditions compare. First we will discuss a class of measures introduced by Walters [96], as it provides a convenient framework for discussing the example in [33], of a non-Gibbsian g—measure. Furthermore we will use this class to generate more examples. The first examples we will discuss are the non-Gibbsian g—measures that can be found in the literature. Then we will give two examples to clarify the distinction between Walters' condition and Good Future. A fifth example shows how the decay of variation of μ_+ can differ from the decay of variation of its reverse, μ_-, even if both are g—measures. Finally we will give a table showing how the classes differ, supported by examples.

2.5.1 Walters' natural space

In order to investigate the set of g—measures that are Gibbsian the space of functions, $R(X_+)$, introduced by Walters [96] is very useful. For example, the non-Gibbsian g—function from [33] is in this class. Consider the alphabet $\mathscr{A} = \{0, 1\}$. The set $R(X_+) \cap G(X_+)$ is parametrised by two sequences, conditions for being in $\text{Wal}(X_+, S)$ and $\text{Bow}(X_+, S)$ depend in an elegant way on these sequences. Walters showed that cylinder sets of the corresponding g—measures can be calculated explicitly in terms of these sequences and that the measures are unique.
To be precise, let $(\gamma_p)_2^\infty$ and $(\delta_p)_2^\infty$ be two sequences taking values in $(c, 1-c)$, with $c \in (0, \frac{1}{2})$, converging to γ and δ respectively. Define, for $p \geq 2$ and any $\boldsymbol{\eta} \in \mathscr{A}^\infty$:

$$g(0^p 1\boldsymbol{\eta}) = \gamma_p, \quad g(10^{p-1}1\boldsymbol{\eta}) = 1-\gamma_q,$$
$$g(1^p 0\boldsymbol{\eta}) = \delta_p, \quad g(01^{p-1}0\boldsymbol{\eta}) = 1-\delta_p.$$

For this to be consistent with $g \in G(X_+)$ the function must be continuous, we must then require that, for $q \geq 1$

$$g(0^\infty) = \gamma, \qquad g(1^\infty) = \delta,$$
$$g(10^\infty) = 1-\gamma, \qquad g(01^\infty) = 1-\delta.$$

This is what defines an element of $G(X_+) \cap R(X_+)$. For such a function Walters proved:

Theorem 2.18 (Walters [96]). *Let $g \in G(X_+) \cap R(X_+)$ be given in terms of $(\gamma_p)_2^\infty$ and $(\delta_p)_2^\infty$ as above. Then the following statements hold:*

1. $\log(g) \in \text{Bow}(X_+, S)$ *if and only if there exists $A > 1$ with*

$$A^{-1} \leq \gamma_2 \ldots \gamma_{n+1}/\gamma^n \leq A$$

and

$$A^{-1} \le \delta_2 ... \delta_{n+1}/\delta^n \le A$$

for all $n \ge 1$.

2. $\log(g) \in \mathrm{Wal}(X_+, S)$ *if and only if* $\sum_{n=2}^{\infty} \log(\gamma_n/\gamma)$ *and* $\sum_{n=2}^{\infty} \log(\delta_n/\delta)$ *are both convergent.*

All functions $g \in G(X_+) \cap R(X_+)$, define a unique g—measure that is reversible, even those for which $\log(g) \notin \mathrm{Bow}(X_+, S)$:

Theorem 2.19 (Walters [96]). *For $g \in G(X_+) \cap R(X_+)$ there exists a unique g—measure μ. Furthermore the measure of a cylinder set is equal to the measure of the reverse cylinder,*

$$\mu(\omega_0 \omega_1 ... \omega_{n-1}) = \mu(\omega_{n-1} ... \omega_1 \omega_0)$$

for all $\omega_0, \omega_1, ..., \omega_{n-1} \in \{0, 1\}$, *where* $n \ge 1$.

Due to the ease of construction of elements in $R(X_+)$ and good control of their properties this class constitutes a very nice source of examples.

2.5.2 A well behaved g-measure that is not Gibbs

Let us start with the example of a non-Gibbsian g—measure constructed in [33]. The measure is defined as follows. Let $p_\infty \in (0,1)$ and $\xi \in (1, (1-p_\infty)^{-2})$. For $k \ge 0$ put:

$$\nu_k = \frac{(-1)^{r_k}}{r_k}, \quad \text{with} \quad r_k = \inf\left\{ i \ge 1 : \sum_{j=1}^{i} j \ge k+1 \right\}.$$

The first few terms of this sequence are

$$-1, \frac{1}{2}, \frac{1}{2}, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, -\frac{1}{5}, -\frac{1}{5}, -\frac{1}{5}, -\frac{1}{5}, -\frac{1}{5}.$$

The function g is in the class $R(X_+) \cap G(X_+)$ and can thus be defined in terms of two sequences:

$$\gamma_p = (1-p_\infty)\xi^{\nu_{p-1}}$$
$$\delta_p = 1-(1-p_\infty)\xi^{-1},$$

In [33] the following was shown:

$$\mu^{GFM}\left(0_0 | 1_{-m} 0_{-m+1}^{-1}, 0_1^{n-1} 1_n\right) = \left(1 + c_{n,m} \xi^{\sum_{k=0}^{m-1} \nu_k - \nu_{k+n}}\right)^{-1},$$

where $c_{n,m}$ is a positive converging sequence as $m,n \to \infty$, bounded away from 0. It follows that the limiting behaviour is determined by a factor, $\xi^{\sum_{k=0}^{m-1} \nu_k - \nu_{k+n}}$. For any $M > 0$ one can choose $m,n > M$ such that this factor equals ξ^{-1}; likewise $m,n > M$ can be chosen such that the factor becomes 1. This oscillatory behaviour implies that the sequence does not converge, resulting in an essential discontinuity of the two-sided conditional probability at $\boldsymbol{\omega} = 0_{-\infty}^{+\infty}$.
Alternatively, it is easy to see that

$$\prod_{i=1}^{\infty} \frac{g(0^i 10^\infty)}{g(0^\infty)} = \prod_{i=1}^{\infty} \xi^{\nu_i}$$

does not converge and therefore we can use Theorem 2.11 to show that the measure is not Gibbs. Note that this example still satisfies the uniform non-nullness required for Gibbs measures. Additionally, $\sum_{n=2}^{\infty} \log(\gamma_n/\gamma)$ does not converge due to oscillating partial sums, while $\gamma_2 ... \gamma_{n+1}/\gamma^n$ and $\delta_2 ... \delta_{n+1}/\delta^n$ are bounded away from 0 and ∞, this implies, by Theorem 2.18, that $\log(g) \in \mathrm{Bow}(X_+, S) \setminus \mathrm{Wal}(X_+, S)$, showing that μ^{GFM} is a Bowen-Gibbs measure as well.

2.5.3 A more severely non-Gibbsian g-measure

Another example of a g—measure that is not Gibbsian, was, for different purposes, constructed by Walters [95]. This example was constructed to show the existence of a g—measure μ_+^W on X_+ such that its reversal on X_- is not a g—measure. It turns out that Walters' construction also provides a non-Gibbsian measure. While the example in Section 2.5.2 subtly breaks continuity, but still has the uniform non-nullness property, in the example by Walters both conditions are violated.
Let $(p_k)_{k=0}^{\infty}$, with $p_k \in [0,1)$ for all $k \in \mathbb{Z}_+$, $p_k \to 0$ as $k \to \infty$ and $\sum_{k=0}^{\infty} \frac{p_k}{1+p_k} = \infty$. For $k,l \geq 0$ let

$$g(000^k 1^l 101 \boldsymbol{\eta}_{k+l+4}^{\infty}) = \frac{1}{2}(1 - p_{k+l}) \quad \text{and} \quad g(100^k 1^l 101 \boldsymbol{\eta}_{k+l+4}^{\infty}) = \frac{1}{2}(1 + p_{k+l}),$$

$$g(000^k 1^l 100 \boldsymbol{\eta}_{k+l+4}^{\infty}) = \frac{1}{2}(1 + p_{k+l}) \quad \text{and} \quad g(100^k 1^l 100 \boldsymbol{\eta}_{k+l+4}^{\infty}) = \frac{1}{2}(1 - p_{k+l}),$$

for any $\boldsymbol{\eta} \in X_+$. For all other $\boldsymbol{\omega} \in X_+$ put $g(\boldsymbol{\omega}) = \frac{1}{2}$. Clearly, g is a continuous normalised function on X_+, for any such g at least one g—measure exists, let μ_+^W be such a measure and let μ^W be the natural extension of μ_+^W. Then we have the following recurrence relation: for $m \geq 2$

$$\frac{\mu^W(1^m 00)}{\mu^W(1^m 0)} = \frac{\mu^W(1|1^{m-1}00)}{\mu^W(1|1^{m-1}0)} \frac{\mu^W(1^{m-1}00)}{\mu^W(1^{m-1}0)} = \frac{\mu^W(1^{m-1}00)}{\mu^W(1^{m-1}0)},$$

where the last equality holds as $\mu^W(1|1^{m-1}00)$ and $\mu^W(1|1^{m-1}0)$ are both equal to $\frac{1}{2}$. This implies that $\frac{\mu^W(1^m00)}{\mu^W(1^m0)}$ is constant in m and due to μ^W being a g−measure this value is not 0 or 1.

$$\begin{aligned}\frac{\mu^W(0^n1^m00)}{\mu^W(0^n1^m0)} &= \frac{\mu^W(0^n1^m00)}{\mu^W(0^n1^m00)+\mu^W(0^n1^m01)} &&= \frac{1}{1+\frac{\mu^W(0^n1^m01)}{\mu^W(0^n1^m00)}} \\ &= \frac{1}{1+\left(\prod_{i=1}^{n-1}\frac{\mu^W(0|0^i1^m01)}{\mu^W(0|0^i1^m00)}\right)\frac{\mu^W(01^m01)}{\mu^W(01^m00)}} &&= \frac{1}{1+\left(\prod_{i=1}^{n-1}\frac{1-p_{i+m-2}}{1+p_{i+m-2}}\right)\frac{\mu^W(01^m01)}{\mu^W(01^m00)}}.\end{aligned}$$

For any $m \geq 1$ the product $\prod_{i=1}^{n-1}\frac{1-p_{i+m-2}}{1+p_{i+m-2}}$ tends, by the choice of $(p_k)_{k=0}^{\infty}$, to 0 as $n \to \infty$. Then it follows that

$$\lim_{n\to\infty}\frac{\mu^W(0^n1^m00)}{\mu^W(0^n1^m0)} = 1,$$

for any $m \geq 1$. It follows that there exists an $\varepsilon > 0$ such that for any $m \geq 1$ there exists an $n > 0$ such that $\left|\frac{\mu^W(1^m00)}{\mu^W(1^m0)} - \frac{\mu^W(0^n1^m00)}{\mu^W(0^n1^m0)}\right| > \varepsilon$. It follows that $\mu^W_-(0_0|.)$ is discontinuous at $0_{-1}1_{-\infty}^{-1}$. This shows that the reverse of μ^W_+ is not a g−measure. We now show that non-Gibbsianness follows from similar arguments. The conditional probability $\mu^W(0_0|1_{-n}^{-2}0_{-1}0_1^m)$ is, for a given $n \geq 2$, bounded away from 0 and 1, for all $m \geq 1$, as μ^W is a g−measure. Now for any $m \geq 1$ we have that $\mu^W(0_0|1_{-n}^{-2}0_{-1}0_1^m)$ is constant in $n \geq 2$, hence the conditional probability is bounded away from 0 and 1. On the other hand $\mu^W(0_0|0_{-n-p}^{-n-1}1_{-n}^{-2}0_{-1}0_1^m)$ tends, for any $n \geq 2$ and $m \geq 1$, to 1, as $p \to \infty$. Therefore a consistent specification would necessarily be discontinuous in $1_{-\infty}^{-2}0_{-1}^{\infty}$, furthermore uniform non-nullness is violated. It follows that μ^W is not a Gibbs measure.

2.5.4 Example: Walters' class, but no Good future

Neither Walters' condition nor Good Future implies the other. In this example we will construct a g−function with oscillating dependencies on far away spins. The requirements for $\log(g) \in \mathrm{Wal}(X_+, S)$ are such that a strong influence from individual spins far away can be allowed, by having individual contributions cancel. Using this we can let the decay of the dependencies fall off rather slowly and yet have $\log(g) \in \mathrm{Wal}(X_+, S)$. Let for $p \geq 2$

$$\gamma_p = \frac{1}{2} + \frac{(-1)^p}{2p}, \qquad \gamma = \frac{1}{2},$$
$$\delta_p = \frac{1}{2}, \qquad \delta = \frac{1}{2}.$$

Now let $g \in G(X_+, S) \cap R(X_+)$ be the function defined by these sequences, as explained in section 2.5.1. In this case $\sum_{n=2}^{\infty} \log(\frac{\gamma_n}{\gamma})$ and $\sum_{n=2}^{\infty} \log(\frac{\delta_n}{\delta})$ converge, hence $\log(g) \in \mathrm{Wal}(X_+, S)$. However

$$\partial_k(g) = \sup_{l>k}\{|\gamma_k - \gamma_l|\} = |\gamma_k - \gamma_{k+1}| = \frac{2k+1}{4k(k+1)},$$

hence

$$\sum_{k=2}^{\infty} \partial_k(g) = \infty.$$

Thus we conclude that there exists a $g \in G(X_+)$ such that $\log(g) \in \mathrm{Wal}(X_+, S) \setminus \mathrm{GF}(X_+)$.

2.5.5 Example: Good future outside of Walters' class

Conversely, potentials satisfying Walters' condition have to behave well even if the entire tail is changed, whereas Good Future depends on single symbol changes. We can construct an example satisfying the Good Future condition, while failing Walters' condition by exploiting this property. Define, on a two letter alphabet $\mathscr{A} = \{0, 1\}$, a g-function $g : X_+ \to (0, 1)$ as

$$g(0\omega_1\omega_2 ...) = \frac{1}{2} + \frac{2}{\pi^2}\sum_{i=1}^{\infty}\frac{\omega_i}{i^2}$$

$$g(1\omega_1\omega_2 ...) = \frac{1}{2} - \frac{2}{\pi^2}\sum_{i=1}^{\infty}\frac{\omega_i}{i^2}$$

As $\sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6}$ this function is an element of $G(X_+)$. Also $g \in \mathrm{GF}(X_+)$ because $\delta_k = \frac{2}{\pi^2 k^2}$, in fact it even satisfies Dobrushin's uniqueness criterion [31], which is stronger then Good Future. Thus the function defines a unique g-measure, which is also a Gibbs measure. However,

$$\begin{aligned}
\sup_{n\geq 1} \sup_{\eta_0^{n-1}=\omega_0^{n-1}} \left\{\sum_{i=0}^{n-1} \log\left(\frac{g(S^i\omega)}{g(S^i\eta)}\right)\right\} &\geq \sup_{n\geq 1}\left\{\log\left(\prod_{i=0}^{n-1}\frac{g(0_i^{n-1}1_n^{\infty})}{g(0_i^{\infty})}\right)\right\} \\
= \sup_{n\geq 1}\left\{\log\left(\prod_{i=0}^{n-1}\left(1+\frac{4}{\pi^2}\sum_{j=n-i}^{\infty}\frac{1}{j^2}\right)\right)\right\} &= \sup_{n\geq 1}\left\{\log\left(\prod_{i=1}^{n}\left(1+\frac{4}{\pi^2}\sum_{j=i}^{\infty}\frac{1}{j^2}\right)\right)\right\} \\
\geq \sup_{n\geq 1}\left\{\log\left(\prod_{i=1}^{n}\left(1+\frac{4}{\pi^2 i}\right)\right)\right\} &= +\infty
\end{aligned}$$

From this divergence it follows that $\log(g) \in \mathrm{GF}(X_+) \setminus \mathrm{Bow}(X_+, S)$.

2.5.6 Example: slow decay of variation of a reverse g-measure

We have seen that $\log(g) \in \mathrm{Wal}(X_+, S)$ is a sufficient condition for the corresponding g—measure to be reversible. Moreover the reverse measure μ_- is a g—measure which has $\log(g_-) \in \mathrm{Wal}(X_-)$. Similarly, if $\log(g) \in \mathrm{Bow}(X_+, S)$ and if the measure is also a reverse g—measure, then the reverse function satisfies $\log(g_-) \in \mathrm{Bow}(X_-)$ [95]. Suggesting that, to some extent, the regularity of the function g and its reverse counterpart g_- are comparable. However, we would like to point out that there exist g—measures for which the decay of variation of g_- is slower that the decay of variation of g. We demonstrate this by the following adaptation of Walters' example [95]:
Define the g—function by a sequence $(p_k)_{\geq 0}$, taking values in $(0,1)$, converging to $\frac{1}{2}$, as follows:

$$g(00^{k-2}101\boldsymbol{\eta}) = p_k, \qquad g(10^{k-2}101\boldsymbol{\eta}) = 1 - p_k,$$
$$g(00^{k-2}100\boldsymbol{\eta}) = 1 - p_k, \qquad g(10^{k-2}100\boldsymbol{\eta}) = p_k,$$

for $k \geq 3$ and any $\boldsymbol{\eta} \in X_+$. Let $g(\boldsymbol{\omega}) = \frac{1}{2}$, for all other $\boldsymbol{\omega} \in X_+$. Now let $p_k = \frac{1}{2}\xi^{\frac{1}{k^2}}$, with $\xi \in (0,1)$, then $\mathrm{var}_n(g) = \mathcal{O}\left(\frac{1}{n^2}\right)$, therefore $\log(g) \in \mathrm{Wal}(X_+, S)$. It follows that the reverse, μ_-, is a g—measure with $\log(g_-) \in \mathrm{Wal}(X_-, S_-)$. Let $a_k = \mu(0_0|0_{-1}1_{-2}0^{-3}_{-k+1})$, then a_k is bounded away from 0 and 1 by some constant $c > 0$. Then

$$\prod_{k=m}^{\infty} \frac{\mu(0_{-k}|0^{-3}_{-k+1}1_{-2}0_{-1}1_0)}{\mu(0_{-k}|0^{-3}_{-k+1}1_{-2}0_{-1})} = \prod_{k=m}^{\infty} \frac{p_k}{a_k(1-p_k) + (1-a_k)p_k}$$
$$= \prod_{k=m}^{\infty} \frac{1}{1 + a_k\left(\frac{1}{p_k} - 2\right)} \leq 1 + \frac{2c\log(\xi)}{m} + \mathcal{O}\left(\frac{1}{m^2}\right),$$

for $m > 3$. Note that $\log(\xi)$ is negative, therefore the above estimate establishes a bound away from 1. As we already established that μ_- is a g—measure Theorem 2.2 applies. Furthermore

$$\mu(\omega_0|\omega_1^n) = \sum_{\sigma_{n+1} \in \mathcal{A}} \mu(\omega_0|\omega_1^n\sigma_{n+1})\mu(\sigma_{n+1}|\omega_1^n) \leq \sup_{\sigma_{n+1}} \mu(\omega_0|\omega_1^n\sigma_{n+1}).$$

This can be used to determine the following lower bound to the variation of g_-, the reverse g—function:

$$\mathrm{var}_n(g_-) \geq \sup_{\omega \in X} \left|\mu(\omega_0|\omega^{-1}_{-n+1}) - g_-(\omega^0_{-\infty})\right|$$
$$\geq \left|\mu(1_0|0_{-1}1_{-2}0^{-3}_{-n+1})\left(1 - \prod_{k=n}^{\infty} \frac{\mu(0_{-k}|0^{-3}_{-k+1}1_{-2}0_{-1}1_0)}{\mu(0_{-k}|0^{-3}_{-k+1}1_{-2}0_{-1})}\right)\right|$$
$$\geq \frac{2c^2|\log(\xi)|}{n} + \mathcal{O}\left(\frac{1}{n^2}\right),$$

for sufficiently large n. Hence $\mathrm{var}_n(g) = \mathcal{O}\left(\frac{1}{n^2}\right)$ while the variation of g_- can be estimated from below by a term proportional to $\frac{1}{n}$.

2.5.7 The overview of the relevant classes

In this Section we will give an overview of a relation between various classes of measures discussed in this paper. Some classes of g—measures are not included in the overview. Let us only mention the following growing chain of classes of measures

- Markov measures,
- g-measures/Bowen-Gibbs measures with Hölder continuous potentials,
- g-measures/Bowen-Gibbs measures with summable variation potentials,
- g-measures/Bowen-Gibbs measures with $\log g \in \mathrm{Wal}(X_+, S)$.

Let us now turn to summarizing the properties of various classes.

Wal If $\log(g) \in \mathrm{Wal}(X_+, S)$, then there exists a unique g-measure $\mu = \mu_g$, which is also Gibbs and Bowen-Gibbs; its reversal μ_- is also a g-measure for some g_- with $\log g_- \in \mathrm{Wal}(X_-, S_-)$. However μ_g does not necessarily satisfy the good future condition, c.f., Section 2.5.4.

Bow If $\log(g) \in \mathrm{Bow}(X_+, S)$, then there exists a unique g-measure $\mu = \mu_g$, which is also Bowen-Gibbs, but not necessarily Gibbs. Moreover, μ does not necessarily have Good Future, nor does its reversal μ_- have to be a g-measure, c.f., Sections 2.5.5, 2.4.4 respectively.

GF If $g \in G(X_+)$ has satisfies the Good Future condition: $\sum_k \partial_k(g) < \infty$, then any g-measure is necessarily Gibbs. Note that there exists g-functions with Good Future that have several g-measures; and hence, $\log g \notin \mathrm{Bow}(X_+, S)$. It is not known whether the reverse of a g—measure with $g \in \mathrm{GF}(X_+)$ must be a g-measure.

Gibbs If μ is a Gibbs measure and a g—measure at the same time then it is not necessarily Bowen-Gibbs, nor does it have to have Good Future, this is highlighted in the examples in Sections 2.5.5 and 2.5.4. It is not known whether the reverse μ_- is a g-measure.

The table in this Section refers to examples of measures having properties discussed above. One example in this overview we have not mentioned before.

	Wal	Bow	GF	Gibbs	reversible	unique	R ∩ G
Wal		∅	2.5.4	∅	∅	∅	2.5.6
Bow	2.5.2		2.5.4	2.5.2	2.4.4	∅	2.5.6
GF	2.5.5	2.5.5		∅	?	2.4.2	2.5.5
Gibbs	2.5.5	2.5.5	2.5.4		?	2.4.2	2.5.5
reversible	2.5.2	RnB	2.5.4	2.5.2		?	2.5.6
unique	2.5.2	RnB	2.5.4	2.5.2	2.4.4		2.5.6
R ∩ G	2.5.2	RnB	2.5.2	2.5.2	∅	∅	

Table 2.1: The symbol in row A and column B represents an example which belongs to A, but not to B. The categories, in order, are g—measures with $\log(g) \in \mathrm{Wal}(X_+, S)$, $\log(g) \in \mathrm{Bow}(X_+, S)$, $g \in \mathrm{GF}(X_+)$, μ is a DLR Gibbs measure, the reverse of μ_+ is a g—measure, g has a unique corresponding g—measure and $g \in R(X_+) \cap G(X_+)$. We use the following symbols:

- "X.X" refers to a Section in this paper.
- "∅" means no example exists.
- "?" if no example is known.
- "RnB" An R class example with $\log(g) \notin \mathrm{Bow}(X_+, S)$, discussed below.

Namely a measure for a $g-$function in the R class, that does not belong to Bowen's class. Such a measure can easily be constructed. Take for example the sequences $\gamma_p = \frac{1}{2}\left(\frac{9}{4}\right)^{\frac{1}{p}}$, $\delta_p = \frac{1}{2}$. These sequences take values in the interval $(0,1)$ and converge to $\frac{1}{2}$ and therefore define a $g-$function in the space $R(X_+)$, as described in Section 2.5.1. Furthermore $\prod_{p=2}^{\infty}\left(\frac{9}{4}\right)^{\frac{1}{p}} = \left(\frac{9}{4}\right)^{\sum_{p=2}^{\infty}\frac{1}{p}} = \infty$, hence $\log(g) \notin \mathrm{Bow}(X_+, S)$.
The set of $g-$measures with $\log(g) \in \mathrm{Bow}(X_+,S) \setminus \mathrm{Wal}(X_+,S)$ is particularly interesting in the context of this paper. This set of measures contains examples of both reversible and non-reversible $g-$measures and an example of a $g-$measure that is not Gibbs. It would be interesting to know if there can be a $g-$measure with $\log(g) \in \mathrm{Bow}(X_+,S) \setminus \mathrm{Wal}(X_+,S)$ that is a Gibbs measure, we do not have a result on this.

2.6 When Gibbs measures are g-measures

Suppose μ is a translation-invariant Gibbs measure on $X = \mathscr{A}^{\mathbb{Z}}$, corresponding to a uniformly absolutely convergent interaction $\Phi = \{\Phi(\Lambda,\cdot), \Lambda \Subset \mathbb{Z}\}$. A natural question is to identify conditions under which μ (more precisely, a restriction μ_+ of μ to $X_+ = \mathscr{A}^{\mathbb{Z}_+}$) is a g-measure. For completeness we will discuss two known types of solutions to this problem.
The first approach to deciding whether a given Gibbs measure is a g-measure is based on the variational description of Gibbs states. Let μ be a Gibbs measure for an UAC interaction Φ. It is well known [77, Chapter 3], [37, Chapter15], that the functions

$$f_\Phi(\boldsymbol{\omega}) = \sum_{0\in V\Subset\mathbb{Z}} \frac{1}{|V|}\Phi_V(\omega_\Lambda),$$
$$\tilde{f}_\Phi(\boldsymbol{\omega}) = \sum_{0\in V\Subset\mathbb{Z}_+} \Phi_V(\omega_\Lambda),$$

are continuous on X. Note that $\tilde{f}_\Phi$ depends only on $\{\omega_k,\ k \geq 0\}$, and hence, can be viewed as a function on $X_+ = \mathscr{A}^{\mathbb{Z}_+}$. Moreover, μ is an equilibrium state for f_Φ and $\tilde{f}_\Phi$. If μ is a g-measure, then μ is an equilibrium state for $\log g \in C(X_+,\mathbb{R})$. Thus the question boils down to the problem of deciding whether for a given $f_\Phi \in C(X,\mathbb{R})$ or $\tilde{f}_\Phi \in C(X_+,\mathbb{R})$ there exists a positive continuous normalized function $g \in C(X_+,\mathbb{R})$ such that $\log g$ has the same set of equilibrium states. One typically distinguishes two types of equivalence conditions on continuous functions resulting in identical sets of equilibrium states:

- two continuous functions f_1, f_2 are called **cohomologous** if there exists a continuous function h and a constant c such that

$$f_1 - f_2 = h \circ \sigma - h + c,$$

where σ is the left shift. Equivalently, one says that $f_1 - f_2$ is a **coboundary**.

- by analogy with the standard notion in Statistical Mechanics, we say that two continuous functions f_1, f_2 are **physically equivalent** if

$$\int (f_1 - f_2)\, d\nu = \text{const}$$

for all translation-invariant probability measures ν.

Clearly, if f_1, f_2 are cohomologous, then they are physically equivalent. The opposite is not true in general, c.f., [10].
Walters [94] identified the necessary and sufficient conditions for a continuous function f_1 on the space of two-sided sequences X to be cohomologous to a continuous function f_2 which depends only on the values of $\{\omega_k : k \geq 0\}$, i.e., f_2 can be viewed as a function on X_+.

Theorem 2.20 (Walters [94]). *Denote by π the natural projection from $X = \mathscr{A}^{\mathbb{Z}}$ to $X_+ = \mathscr{A}^{\mathbb{Z}_+}$. Then $f_1 \in C(X, \mathbb{R})$ satisfies*

$$f_1(\boldsymbol{\omega}) = f_2 \circ \pi(\boldsymbol{\omega}) + h \circ \sigma(\boldsymbol{\omega}) - h(\boldsymbol{\omega}) + c$$

for some $f_2 \in C(X_+, \mathbb{R})$, $h \in C(X, \mathbb{R})$, and $c \in \mathbb{R}$ if and only if f_1 belongs to the class of functions satisfying the extended Walters condition, $W_\pi(X, S)$: namely, if for every $\varepsilon > 0$ there exists $N > 0$ such that whenever $\boldsymbol{\omega}, \tilde{\boldsymbol{\omega}}$ are such that $\pi(\boldsymbol{\omega}) = \pi(\tilde{\boldsymbol{\omega}})$ and $\boldsymbol{\omega}_{-N}^{N} = \tilde{\boldsymbol{\omega}}_{-N}^{N}$ one has

$$\left| \sum_{k=0}^{n-1} \left[f_1(\sigma^k \boldsymbol{\omega}) - f_1(\sigma^k \tilde{\boldsymbol{\omega}}) \right] \right| < \varepsilon.$$

for all $n \geq 1$.

This still leaves the question whether the one-sided potential can be chosen such that it is normalised. In the literature several results are known that will guarantee this. Sinai showed this for Hölder continuous potentials on X [80]. This result was later improved to potentials on X with summable variation [18] and then to potentials in the two-sided counterpart of Walters' class, $\mathrm{Wal}(X, S)$ [9]. Such a potential has been shown to be cohomologous to a one-sided potential in $\mathrm{Wal}(X_+, S)$, implying that the corresponding equilibrium state is unique and a $g-$measure. At

present, it is unknown whether potentials in Bowen's class, $\mathrm{Bow}(X_+, S)$, will always result in a continuous g—function.
The second approach is aimed at establishing continuity of versions of conditional probabilities. Suppose μ is a translation-invariant Gibbs measure on $X = \mathscr{A}^{\mathbb{Z}}$. Denote by γ the corresponding (translation-invariant) Gibbs specification. Denote by μ_+ the restriction of μ to $X_+ = \mathscr{A}^{\mathbb{Z}_+}$. The objective is to identify sufficient conditions on the Gibbsian specification γ (equivalently, on the potential Φ) such that μ_+ is a g-measure for some positive normalised continuous function $g : X_+ \to (0,1)$. Note that μ_+ is a g-measure if and only if for every continuous $f : X_+ \to \mathbb{R}$ the conditional expectation

$$\mathbb{E}_{\mu_+}(f|\mathscr{F}_1^{\infty}),$$

defined μ_+-almost everywhere, admits a continuous version. It is sufficient to check the above condition for $f(\omega) = \mathbb{1}_{[a_0]}(\omega)$, $a_0 \in \mathscr{A}$. The natural approach to construction of one-sided conditional expectation from the known two-sided conditional expectations (given by specification γ) is as follows. Consider again some $f \in C(X_+)$ and conditional expectations

$$\mathbb{E}_{\mu}\big(f|\mathscr{F}_1^{\infty} \cap \mathscr{F}_{-\infty}^{-n-1}\big), \quad n \geq 1.$$

If $f(\omega) = \mathbf{I}_{[a_0]}(\omega)$, $a_0 \in \mathscr{A}$, and $\omega \in \mathscr{A}^{\mathbb{Z}_+}$ and $\xi \in \mathscr{A}^{\mathbb{Z}_{<0}}$, then

$$\mathbb{E}_{\mu}\big(f|\mathscr{F}_1^{\infty} \cap \mathscr{F}_{-\infty}^{-n-1}\big)(\xi_{-\infty}^{-1}, \omega_0^{+\infty}) = \gamma_{[-n,0]}([a_0]|\omega_1^{\infty}, \xi_{-\infty}^{-n-1}).$$

Note that $\gamma_{[-n,0]}(\cdot|\omega_1^{\infty}, \xi_{-\infty}^{-n-1})$ is indeed $\mathscr{F}_1^{\infty} \cap F_{-\infty}^{-n-1}$-measurable and depends continuously on ω_1^{∞},$\xi_{-\infty}^{-n-1}$. Therefore, if we assume that $\gamma_{[-n,0]}(\cdot|\omega_1^{\infty}, \xi_{-\infty}^{-n-1})$ converges as $n \to \infty$, and the limit is **independent** of ξ, then the limit

$$\lim_{n\to\infty} \gamma_{[-n,0]}([a_0]|\omega_1^{\infty}, \xi_{-\infty}^{-n-1})$$

is a natural candidate for g-function of μ_+. The Hereditary Uniqueness Condition (HUC) [29] is a sufficient condition for $g(\boldsymbol{\omega}) = \lim_{n\to\infty} \gamma_{[-n,0]}([a_0]|\omega_1^{\infty}, \xi_{-\infty}^{-n-1})$ to exist, with g a continuous function that is a version of $\mathbb{E}_{\mu_+}(f|\mathscr{F}_1^{\infty})$. The HUC itself is a mostly technical condition related to the uniqueness of measures consistent with alterations of the specification γ. More interesting is that HUC is implied by two known uniqueness conditions for Gibbs measures. The first condition is Dobrushin's uniqueness condition [21, 57]. The second condition is Georgii's boundary uniformity [37].

Definition 2.21. A specification $\{\gamma_V\}$ satisfies boundary-uniformity if there exists a constant $K > 0$ so that for every cylinder set $A \in \mathscr{B}$ there exists a finite interval $W = [i, ..., j] \subset \mathbb{Z}$ such that $\gamma_W(A|\xi) \geq K\gamma_W(A|\eta)$ for all $\xi, \eta \in X$.

If μ satisfies the boundary uniformity condition then the resulting measure μ_+ will satisfy the one-sided boundary uniformity condition, meaning that, for the corresponding g—function g, there exists a $K > 0$ such that

$$\prod_{i=0}^{n-1} \frac{g(S^i \omega_0^n \eta_n^\infty)}{g(S^i \omega_0^n \xi_n^\infty)} < K,$$

for all $\boldsymbol{\omega}, \boldsymbol{\eta}, \xi \in X$. This means the resulting g—measure will have $\log(g) \in \mathrm{Bow}(X_+, S)$.

2.7 Concluding remarks

We discussed a condition and properties of g—measures that are Gibbs. Either by directly applying Theorem 2.11 or one of the sufficient conditions. The question how to determine for a given Gibbs measure whether it is a g—measure is largely open and we merely gave an overview of methods in the literature. Recently a first example of a Gibbs measure that is not a g-measure has been given in [6]. A particularly interesting question would be if a Gibbs measure with a potential for which $f_\Phi(\omega) = \sum_{X \ni_{\min} 0} \frac{1}{|X|} \Phi_X(\omega) \in \mathrm{Bow}(X_+, S) \setminus \mathrm{Wal}(X_+, S)$ can be a g—measure, as this precisely forms a boundary where the continuity of g via the Ruelle operator theorem becomes uncertain.

Chapter 3

Estimation of two-sided conditional probabilities

In the previous chapter we extensively used that one-sided conditional probabilities of a measure on $\mathscr{A}^{\mathbb{Z}}$ determine its two-sided conditional probabilities on a set of full measure. In this chapter we discuss applications of this relation to the problem of estimation of two-sided conditional probabilities. This problem gained significant interest in information theory with the introduction of the Discrete Universal DEnoiser (DUDE) [97]. A denoiser, and in particular the DUDE, attempts to estimate a stationary process $\{X_n\}_{n\in\mathbb{Z}}$, taking values in a finite set $\mathscr{A}$ given a noisy observation of this process. In this chapter we will be primarily interested in signals corrupted by a discrete memoryless channel, a construction that we will now describe. The noisy observation is modelled by a second process $\{Z_n\}_{n\in\mathbb{Z}}$, taking values in another finite alphabet $\mathscr{B}$. The process $\{Z_n\}_{n\in\mathbb{Z}}$ is related to $\{X_n\}_{n\in\mathbb{Z}}$ via a stochastic matrix Π: Given X_i, the random variable Z_i is determined by

$$\mathbb{P}(Z_n = b|X_n = a) = \Pi_{a,b}, \quad a \in \mathscr{A},\ b \in \mathscr{B}, \quad n \in \mathbb{Z},$$

independently from $\{Z_m : m \neq n\}$. The discrete memoryless channel is defined as the map defining Z_n from X_n and we refer to Π as the channel matrix. The DUDE motivates the modeling of two-sided conditional probabilities because it relies on a good approximation of the probability

$$\mathbb{P}(Z_0 = a_0 \,|\, Z_{-\infty}^{-1} = a_{-\infty}^{-1}, Z_1^{\infty} = a_1^{\infty}), \tag{3.1}$$

for $a \in \mathscr{A}^{\mathbb{Z}}$ and the stationary process $\{X_n\}_{n\in\mathbb{Z}}$. In this text we use Z_n^m to denote $(Z_i : n \leq i \leq m)$, for $n, m \in \mathbb{Z} \cup \{-\infty, \infty\}$. This problem is analogous to the much more extensively studied and better understood problem of finding approximations of one-sided conditional probabilities:

$$\mathbb{P}(X_0 = a_0 \,|\, X_{-\infty}^{-1} = a_{-\infty}^{-1}), \tag{3.2}$$

for any $x \in \mathscr{A}^{-\mathbb{Z}_+}$. There are many efficient algorithms for one-sided, or unidirectional, modeling [61, 75]. However, constructing a good bidirectional model turns out to be a much harder problem. The relation between one-sided and two-sided conditional probabilities suggests a natural solution; to estimate the one-sided model and use it to estimate the corresponding two-sided model. A solution along these lines has been suggested in [100, 101]. Even though this immediate *theoretical* relation between one-sided and two-sided models exists, from a *practical* point of view it is not clear how well various unidirectional models perform when they are used as two-sided models.

In this chapter we address this problem by comparing the quality of several unidirectional algorithms when used for two-sided modeling. We will consider two metrics. The first is a comparison of the quality of the resulting denoisers. As a secondary 'metric' we use the erasure divergence between the two-sided model and the original source. This divergence can only be used on artificial sources as it requires knowledge of the distribution.

The algorithms that we consider are implementations by Begleiter et al. [4] of LZ78, PPM-C, PST, LZ-MS, as well as two versions of context tree weighting. Interestingly PPM-C, PST and LZ-MS are primarily aimed at prediction, while both CTW algorithm and LZ78 are primarily compressors. We find that PPM-C, PST and LZ-MS result in a significant improvement over the DUDE on several sources, while being at least competitive on the remaining sources. On the other hand the versions of context tree weighting and LZ78 are not capable of consistently outperforming DUDE. Therefore, among the tested algorithms, the prediction algorithms clearly outperformed the compression algorithms.

3.1 Unidirectional modeling

Let us first consider the well understood one-sided models. Let $\{X_n\}_{n\in\mathbb{Z}}$ be a stationary stochastic process, taking values in a finite alphabet $\mathscr{A}$, referred to as the *information source*. Unidirectional estimators estimate conditional probabilities of the form

$$\mathbb{P}(X_{n+1} = a_{n+1} \,|\, X_0^n = a_0^n), \tag{3.3}$$

where $a_0^{n+1} \in \mathscr{A}^{n+2}$. A typical application of unidirectional models is loss-less compression. This is the process of storing data in such a way that the expected number of bits used is reduced while the original data can be fully recovered. This can be accomplished by assigning a variable number of bits to characters based on the one-sided model of the source. There is a theoretical limit to how much data can be compressed without loss of information. To be precise, the number of bits per character can not be reduced by more than a factor of the information entropy of the source. Here the information entropy $h_{\mathbb{P}}$ is given by

$$h_{\mathbb{P}} = -\lim_{n\to\infty} \frac{1}{n} \sum_{a_0^n \in \mathscr{A}^{n+1}} \mathbb{P}(X_0^n = a_0^n) \log_2(\mathbb{P}(X_0^n = a_0^n)).$$

The base 2 of the logarithm determines the unit of entropy, in this case bits. Many well-known algorithms are asymptotically optimal. Here asymptotically optimal means that if the length of the realisation of the source goes to infinity, then the average number of bits per character after compression converges to the information entropy. For finite n however an additional expected number of bits per character is needed, since we only have access to a one-sided model of the source, based on a finite amount of data, rather then $\mathbb{P}$ itself. Let $\hat{\mathbb{P}}$ be the measure corresponding to this one-sided model, then the additional expected number of bits per character is given by the Kullback-Leibler divergence between $\mathbb{P}$ and $\hat{\mathbb{P}}$:

$$D\left(\mathbb{P}\middle|\middle|\hat{\mathbb{P}}\right) = \lim_{n\to\infty} \frac{1}{n} \sum_{a_0^n \in \mathscr{A}^{n+1}} \mathbb{P}(X_0^n = a_0^n) \log_2\left(\frac{\mathbb{P}(X_0^n = a_0^n)}{\hat{\mathbb{P}}(\hat{X}_0^n = a_0^n)}\right).$$

Naively, a one-sided model can be obtained by constructing an order k Markov approximation of the source, for an appropriate choice of k. The problem with this approach is that the number of parameters that must be estimated in such a model scales as $|\mathscr{A}|^k$. In practice this often results in a model that either has too many parameters, or fails to capture relevant long-range dependencies.

Instead, a good unidirectional algorithm captures the relevant long-range interactions whilst avoiding an exponential scaling of parameters. A well known algorithm by Rissanen [75] accomplishes this by only selectively including long-range dependencies in its model. The corresponding probabilistic model was later formalized as a Variable-Length Markov Chain (VLMC) [13]. A VLMC is a process that generalizes the idea of an order k Markov process in the following way: rather than uniformly bounding the past of the sequence on which the one-sided conditional probabilities depend by a constant, the relevant part of the past is described by a function. This so-called context function depends on a finite-past realisation of the process $x_{-k}^{-1} \in \mathscr{A}^{\{-k,\dots,-1\}} = \{(a_{-k}, a_{-k+1}, \dots., a_{-1}) : a_i \in \mathscr{A} \text{ for } 1 \le i \le k\}$ of length $k > 0$. Then, either, $0 < l(x_{-k}^{-1}) \le k$, meaning that the relevant past symbols are given by $x_{-l(x_{-k}^{-1})}^{-1}$, or, $l(x_{-k}^{-1}) = +\infty$, meaning that the relevant context is longer then the sequence x_{-k}^{-1}. More formally we can define a context function as is done in [36]:

Definition 3.1. Let

$$A^* = \bigcup_{k=1}^{\infty} \mathscr{A}^{\{-k,\dots,-1\}}$$

and let $l : A^* \to \{0, 1, \dots\} \cup \{\infty\}$ be such that

- For $k \geq 1$ and any $a_{-k}^{-1} \in A^*$, we have $l\left(a_{-k}^{-1}\right) \in \{1, ..., k\} \cup \{\infty\}$.
- For any $a_{-\infty}^{-1} \in A^{-\mathbb{N}}$, if $l(a_{-k}^{-1}) = k$ for some $k \geq 1$ then $l(a_{-m}^{-1}) = \infty$, for all $m < k$ and $l(a_{-m}^{-1}) = k$, for all $m > k$.

then we call l a *context function*.

Subsequently, one can extend the context function to one-sided infinite sequences via $l(x_{-\infty}^{-1}) = \inf\left\{k : l(x_{-k}^{-1}) \neq \infty\right\}$ and then define a VLMC as

Definition 3.2. Let $\{X_n\}_{n\in\mathbb{Z}}$ be a *stationary* process and let $l : A^* \to \mathbb{N} \cup \{\infty\}$ be a context function such that

$$\mathbb{P}\left(X_0 = a_0 \middle| X_{-\infty}^{-1} = a_{-\infty}^{-1}\right) = \mathbb{P}\left(X_0 = a_0 \middle| X_{-l\left(a_{-\infty}^{-1}\right)}^{-1} = a_{-l\left(a_{-\infty}^{-1}\right)}^{-1}\right),$$

for all $a \in \mathscr{A}^{-\mathbb{N}}$, then we call $\{X_n\}_{n\in\mathbb{Z}}$ a Variable-Length Markov Chain.

Let l be the context function for a VLMC and let $a \in \Omega \equiv \mathscr{A}^{\mathbb{Z}}$, then we call $a_{n-l(a_{-\infty}^{n-1})}^{n-1}$ the context for the position $n \in \mathbb{Z}$, this is well defined if we view $a_{-\infty}^{n-1}$ as an element of $\mathscr{A}^{-\mathbb{N}}$. Note that a VLMC with a bounded context function is a Markov process of order $m = \sup_{a\in\mathscr{A}^{-\mathbb{N}}} l(a)$.

Another way to represent a VLMC is as a tree graph. This tree has the empty string as a root node and each vertex, x, is either a leaf (terminal node), or it has a child ax for each $a \in \mathscr{A}$. To each leaf, x, a probability distribution is assigned corresponding to the conditional distribution for the current symbol, given the context x.

3.1.1 Unidirectional algorithms

Even though VLMCs solve the problem of capturing long-range dependencies without an exponential growth in the number of parameters quite elegantly, not all unidirectional algorithms are most naturally described as VLMCs, see, for example [17, 61, 76, 98]. Nevertheless, any lossless compression algorithm can be used to create a probabilistic model of the source and vice versa [28]. In particular, the one-sided modeling algorithms by Begleiter et al. [4] were implemented so that they create VLMC's that approximate the information source. The algorithms that they implemented are LZ78 [61], LZ-MS [68], Prediction by Partial Match (PPM)[17], Probabilistic Suffix Trees (PST) [76] and two versions of the Context Tree Weighting algorithm (CTW) [98], namely Binary CTW (BI-CTW) and Decomposed CTW (DE-CTW) [89].

Below follows a brief description of these algorithms, for a more comprehensive overview we refer to [4] and the citations therein.

- The LZ78 algorithm is a compression algorithm that does not directly aim for a probabilistic description of the source. Instead, LZ78 constructs a tree of words, or dictionary, that can be used for coding without using explicit probability estimates. The first step in building this tree is to associate the empty string to the root node. In the second step the algorithm parses the text, collecting symbols until it forms a word w_1^n, $n \geq 1$, not yet contained in the tree. This word is then added underneath the node associated with w_1^{n-1} to the tree. By repeating the second step until the whole string is processed the tree is completed. By construction common patterns typically end up in long library entries, that are then stored efficiently by a reference to the tree. This algorithm was already modified to produce a VLMC describing the source by Langdon and Rissanen [58, 75].

- The LZ-MS algorithm is a variation of LZ78 aimed at prediction. This version of the algorithm has two parameters, s and m, that parses the input sequence s times, each time expanding the tree resulting from the last parse. Furthermore, after a word is added to the tree the parser moves back m steps, up to a number of positions equal to the length of the most recently added word.

- The two *Context Tree Weighting* algorithms, BI-CTW and DE-CTW are two implementations of CTW that extend to CTW algorithm from binary alphabets to larger alphabets. The CTW algorithm does not directly construct a VLMC approximating the source. Instead, the CTW, given a maximal depth d, computes a mixture of very simple VLMCs with maximal context length d. This mixture would a priori require keeping track of an exponentially growing number of models in the mixture. However, the strength of CTW is that the specific mixture can recursively be computed in linear time in the length of the source. The resulting algorithm is known to be an efficient compression algorithm, with theoretical performance guarantees. The BI-CTW is an adaptation of CTW to multiletter alphabets via the binary representation of elements in the alphabet. The DE-CTW finds a more sophisticated way to represent an alphabet in a binary way. This algorithm constructs a binary tree with $|\mathscr{A}|$ leaves, where, for each node in the tree we label its children by a 0 or a 1. Therefore each leaf, or letter in the alphabet is associated with a binary description. Using proper heuristics for the determination of this tree is reported to result in a better compression performance then the naive binary representation [89].

- Finally the PPM and PST algorithms do naturally construct a VLMC to describe the source as accurately as possible, given a maximal depth d. The PPM algorithm considers a maximal context tree of depth d for the given alphabet. In subsequent steps this maximal tree is then pruned. When a symbol is not observed in a a given context, the algorithm will instead use a shorter context using a so called escape mechanism. It is therefore quite typical, when the alphabet is large, that many symbols have not appeared in a given context. The way the algorithm deals with this essentially smaller alphabet is called an exclusion mechanism. The various escape and exclusion mechanisms define a range of PPM algorithms. The algorithm implemented by Begleiter et al. is called PPM-C, a version of PPM by Moffat and Neal [65].

- The PST algorithm starts by constructing a candidate tree consisting of strings that have occurred sufficiently often. In subsequent steps this tree will be pruned. One of the pruning steps is to remove contexts for which the estimated distributions are not sufficiently different from the distribution in the parent node. This is done by choosing a parameter $r > 0$ and removing a leaf a_0^n when
$$\frac{\hat{\Pr}(\hat{X}_{n+1} = a_{n+1}|\hat{X}_0^n = a_0^n)}{\hat{\Pr}(\hat{X}_{n+1} = a_{n+1}|\hat{X}_1^n = a_1^n)} < r,$$
for all a_{n+1}. Besides r, two other parameters, α and γ, are used in the determination of the relevance of a given context, pruning it when needed. Finally, probabilities are determined by empirical counts and additional smoothing; in particular no symbol will be assigned a probability smaller then a set parameter P_{min}. The conditional distributions are then normalized.

3.2 DUDE and bidirectional modeling

Let's recall the definition of a denoiser. Let $\{X_i\}_{i\in\mathbb{Z}}$ be a stationary process, taking values in a finite alphabet $\mathscr{A}$. A second process, $\{Z_i\}_{i\in\mathbb{Z}}$, which takes values in another finite alphabet $\mathscr{B}$, is constructed from $\{X_i\}_{i\in\mathbb{Z}}$ via a stochastic matrix $\Pi_{i,j}$: Given X_i the random variable Z_i is determined,by

$$\mathbb{P}(Z_n = j|X_n = i) = \Pi_{i,j}, \quad i \in \mathscr{A},\ j \in \mathscr{B}, \quad n \in \mathbb{Z},$$

independently from $\{Z_j : j \neq i\}$. We call the map defining Z_n from X_n a discrete memoryless channel, and we call Π the channel matrix. A denoiser and in particular the Discrete Universal DEnoiser (DUDE) [97] attempts to recover $\{X_i : 0 \leq i \leq n\}$ for $n \geq 0$, from $\{Z_i : 0 \leq i \leq n\}$ and the channel matrix $\Pi_{i,j}$.

Some denoisers also use the distribution of $\{X_i\}_{i\in\mathbb{Z}}$ as input. To determine what constitutes the best approximation of the source one can define a loss function $\{\Lambda_{i,l}\}_{i,l\in\mathscr{A}}$, where $\Lambda_{i,l}$ is the loss assigned to estimating the symbol $i \in \mathscr{A}$ by the symbol $l \in \mathscr{A}$. We will restrict ourselves to the so-called Hamming loss: $\Lambda_{i,j} = 1 - \delta_{i,j}$, i.e., we treat all errors equally. For the Hamming loss, when Π is invertible, the DUDE estimates X_t by:

$$\begin{aligned} \hat{X}_t &= \arg\min_{x_t \in \mathscr{A}} \sum_{\hat{x}_t \in \mathscr{A}} \Lambda_{x_t,\hat{x}_t} \mathbb{P}(X_t = \hat{x}_t | Z_0^n = z_0^n) \\ &= \arg\max_{x_t \in \mathscr{A}} \mathbb{P}(X_t = x_t | Z_0^n = z_0^n) \\ &= \arg\max_{x_t \in \mathscr{A}} \Pi^{-T}_{x_t,z_t} \mathbb{P}(X_t = z_t | Z_0^{t-1} = z_0^{t-1}, Z_{t+1}^n = z_{t+1}^n) \end{aligned} \tag{3.4}$$

Note that in the last expression the dependence of the DUDE on the two-sided conditional probabilities of $\{Z_t\}$ is introduced. The original DUDE replaces these probabilities by a count of how often z_t has occurred in the context z_1^{t-k}, z_{t+1}^k, for an appropriate choice of $k \geq 0$. We denote this count by $m(z_{t-k}^{t-1}, z_{t+1}^{t+k})[i] = |\{j : Z_j^{j+2k+1} = z_{t-k}^{t-1} i z_{t+1}^{t+k}\}|$. This count is not normalised to a probability distribution, as the normalising factor would only depend on the context and therefore not affect the argument minimizing equation 3.4. The resulting denoiser is given by:

$$\hat{X}(z_0^n, t) = \arg\max_{\hat{x} \in \mathscr{A}} \Pi^{-T}_{x_t,z_t} m(z_{t-k}^{t-1}, z_{t+1}^{t+k})[z_t]. \tag{3.5}$$

For a general loss function, Λ, the DUDE estimator is, instead, given by:

$$\hat{X}(z_0^n, t) = \arg\min_{\hat{x} \in \mathscr{A}} m(z_{t-k}^{t-1}, z_{t+1}^{t+k})[z_t] \sum_{x_t} \Pi^{-1}_{z_t,x_t} \Lambda_{x_t,\hat{x}} \Pi_{x_t,z_t}. \tag{3.6}$$

A further generalization to non-invertible channel matrices is possible by using a generalized inverse of Π, instead of Π^{-1}. Those generalizations are beyond the scope of this text. Now we are left with choosing the context length parameter k. The DUDE gives the guarantee that the denoising performance is asymptotically optimal when the choice of k satisfies certain properties. Namely, $k = k_n$ must depend on the length of the realization, n, such that $k \to \infty$, as $n \to \infty$ and

$$k_n |\mathscr{B}|^{2k_n} = o\left(\frac{n}{\log(n)}\right).$$

In practice the following criterion for choosing k turns out to be quite effective: denoise the output data for various values of $k > 0$ and select the value of k for which the denoised data is smallest after compression. The idea behind this criterion is that the data with the least amount of noise should be easier to predict and therefore to compress. Alternatively an estimator for k compatible with a rather general loss function was proposed [71].

The proposed estimate of $\mathbb{P}_{Z_t | z_0^{t-1}, z_{t+1}^n}$ is analogous to the Markov approximation of one-sided conditional probabilities and, in fact, despite the success of the DUDE, it has similar problems. This leaves space for improvement to the DUDE using good bidirectional context models.

First note that one can generalize the concept of a VLMC to the bidirectional setting. For example one can find a context function $m : \mathscr{A}^{-\mathbb{N}} \times \mathscr{A}^{\mathbb{N}} \to \mathbb{N}^2$, such that, for almost all $a \in \mathscr{A}^{\mathbb{Z}}$:

$$\begin{aligned}&\mathbb{P}\left(X_t = a_t \middle| X_{-\infty}^{t-1} = a_{-\infty}^{t-1}, X_{t+1}^{\infty} = a_{t+1}^{\infty}\right)\\ &= \mathbb{P}\left(X_t = a_t \middle| X_{t-m_1(a)}^{t-1} = a_{t-m_1(a)}^{t-1}, X_{t+1}^{t+m_2(a)} = a_{t+1}^{t+m_2(a)}\right).\end{aligned}$$

In unidirectional context modeling the VLMC, represented as a tree, is typically build using certain criteria to include or exclude nodes. In the bidirectional setting it is not clear how to do this in a unique way. Several solutions to the bidirectional context modeling problem can be found in the literature [32, 71–73]. When those solutions were applied to denoising a significant improvement over the DUDE was reported. Finally, Yu and Verdú [100] proposed a number of solutions for the bidirectional problem. One of those directly constructs the two-sided model from a one-sided model, an approach we will explore further in this chapter.

3.3 Gibbs measures

We now turn to the background necessary for using one-sided models to obtain two-sided models, using the literature on Gibbs measures. Note that the problem of estimating two-sided conditional probabilities is the problem of estimating the specification. First we recall the definition. Let γ be the one point specification as defined in the previous chapter, i.e., $\gamma : \Omega \to (0, 1)$, on $\Omega = \mathscr{A}^{\mathbb{Z}}$, that is *normalised*, namely:

$$\sum_{a \in \mathscr{A}} \gamma(\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots) = 1,$$

for all $a = (\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots) \in \Omega$.

Definition 3.3. A stationary process $\{X_n\}_{n\in\mathbb{Z}}$ with values in $\mathscr{A}$, or, equivalently, a translation invariant measure $\mathbb{P}$ on Ω, is called *Gibbs* if it is consistent with a continuous normalised specification $\gamma : \Omega \to (0, 1)$, namely:

$$\mathbb{P}(X_0 = a_0 \,|\, X_{-\infty}^{-1} = a_{-\infty}^{-1}, X_1^{+\infty} = a_1^{+\infty}) = \gamma(a), \tag{3.7}$$

for $\mathbb{P}$-a.a. $a \in \Omega$.

The positivity assumption, $\gamma > 0$, can be relaxed: Gibbs measures can be defined on subshifts of finite type [78, 79], meaning some strings of bounded length can be excluded. The following theorem shows that an important class of noisy processes is naturally Gibbs:

Theorem 3.4. *Suppose $\{X_n\}$ is a finite-state stationary Markov process with* strictly positive *transition probability matrix $P > 0$, and $\{Y_n\}$ is the (hidden Markov) process, obtained from $\{X_n\}$ via* strictly positive *channel matrix $\Pi > 0$. Then $\{X_n\}$ is Gibbs, and moreover, γ has an* exponentially decaying *continuity rate: for some $C > 0$ and $\theta \in (0,1)$,*

$$\mathrm{var}_n(\gamma) \equiv \sup_{\substack{a,\tilde{a}\in\Omega:\\ a_{-n}^{n}=\tilde{a}_{-n}^{n}}} \left|\gamma(a)-\gamma(\tilde{a})\right| \leq C\theta^n, \quad \forall n \in \mathbb{N}.$$

In the upcoming sections we will be discussing the approach by Yu and Verdú [100] to construct bidirectional models out of unidirectional models. A theoretical background is given by the results in [37] on Markov specifications. Note that these results follow the Gibbsian convention to consider only strictly positive transition matrices.

Definition 3.5. Let γ be a one-point specification on Ω, as above. We say that γ is a Markov one point specification if $\gamma(a_0|a_{-\infty}^{-1}, a_1^{\infty}) = g(a_{-1}, a_0, a_1)$ for some function $g : \mathscr{A}^3 \to [0,1]$.

It can be shown that such a specification admits a unique Gibbs measure and that this measure is a Markov measure. Given a stochastic transition matrix P denote the corresponding Markov measure by $\mathbb{P}_P$.

Theorem 3.6 ([37], Chapter 3)**.** *Let $\mathbb{P}$ be the unique Gibbs measure corresponding to a Markov specification γ and $\mathbb{P}_P$ the corresponding Markov chain for a positive stochastic matrix P. If we now identify $\mathbb{P}$ with $\mathbb{P}_P$, this establishes a one-to-one relation between the set of Markov specifications and the set of positive transition matrices:*

- $\gamma(a_0|a_{-1}, a_1) = \mathbb{P}_P(X_0 = a_0|X_{-1} = a_{-1}, X_1 = a_1)$
- $P(x,y) = Q(x,y)\frac{r(y)}{qr(x)}$,

where $Q(x,y) = \frac{\gamma(a,x,y)}{\gamma(a,a,y)}$, $a \in \mathscr{A}$ is arbitrary, q is the Perron-Frobenius eigenvalue of the matrix Q and r a corresponding right eigenvector.

This result also applies to any order k Markov chain for the finite alphabet $\mathscr{A}$ as they are in a one-to-one correspondence with Markov chains.

For an order k Markov chain, the bidirectional model can be obtained by direct computation, as is done by Yu and Verdú [100]. Assume $\{X_n\}_{n\in\mathbb{Z}}$ is a k-step

Markov process, furthermore, assume $n > 2k+1$. Then, for $j = k+1, \dots, n-k$, and $a_1^n \in \mathscr{A}^n$, such that $\mathbb{P}(a_1^{j-1} c_j a_{j+1}^n) > 0$ for some $c_j \in \mathscr{A}$, one has

$$\begin{aligned}
&\mathbb{P}(X_j = a_j \,|\, X_1^{j-1} = a_1^{j-1}, X_{j+1}^n = a_{j+1}^n) \\
&= \frac{\mathbb{P}\left(X_1^{j-1} = a_1^{j-1}, X_j = a_j, X_{j+1}^n = a_{j+1}^n\right)}{\sum_{\overline{c} \in A} \mathbb{P}\left(X_1^{j-1} = a_1^{j-1}, X_j = \overline{c}, X_{j+1}^n = a_{j+1}^n\right)} \\
&= \frac{\mathbb{P}\left(X_j = a_j, X_{j+1}^n = a_{j+1}^n \,|\, X_1^{j-1} = a_1^{j-1}\right)}{\sum_{\overline{c} \in A} \mathbb{P}\left(X_j = \overline{c}, X_{j+1}^n = a_{j+1}^n \,|\, X_1^{j-1} = a_1^{j-1}\right)} \\
&= \frac{\mathbb{P}\left(X_j = a_j, X_{j+1}^n = a_{j+1}^n \,|\, X_{j-k}^{j-1} = a_{j-k}^{j-1}\right)}{\sum_{\overline{c} \in A} \mathbb{P}\left(X_j = \overline{c}, X_{j+1}^n = a_{j+1}^n \,|\, X_{j-k}^{j-1} = a_{j-k}^{j-1}\right)} \\
&= \frac{\prod_{t=j}^{j+k} \mathbb{P}\left(X_t = a_t | X_{t-k}^{t-1} = a_{t-k}^{t-1}\right)}{\sum_{\overline{c} \in A} \prod_{t=j}^{j+k} \mathbb{P}\left(X_t = a_t^{(\overline{c},j)} | X_{t-k}^{t-1} = (a^{(\overline{c},j)})_{t-k}^{t-1}\right)},
\end{aligned} \tag{3.8}$$

where $a^{(\overline{c},j)}{}_1^n$ differs from a_1^n only in position j, with $a_j^{(\overline{c},j)} = \overline{c}$. It was noted that, under some constraints on the one-sided model, a theoretical performance guarantee for the the application to denoising can be given [100]. Moreover improvements over DUDE were shown when applied to specific denoising problems.

3.3.1 Processes used for testing

In order to test on sources with varying properties we vary the size of the alphabet, the decay of memory, as well as the length of the realisation. Since a distance between measures is one of the two performance metrics we will mainly use artificial sources for which we have full knowledge of the distribution.

Binary Symmetric Channel

The first process we consider is a Hidden Markov Model. Let P be the transition matrix of a homogeneous Markov chain, $P_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i)$ for all $n \in \mathbb{Z}$, given by

$$P = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}.$$

Let an emission matrix, $\Pi_{ij} = \mathbb{P}(Y_n = j | X_n = i)$, be given by

$$\Pi = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}.$$

Then the output process $\{Y_n\}_{n\in\mathbb{Z}}$ has infinite memory, however, by theorem 3.4 its continuity rate decays exponentially. Apart from its simplicity, the process $\{Y_n\}$ has the advantage that one can compute both its entropy as well as the probability of cylinder events easily [88].

Long memory in artificial sources

The process described above has infinite memory, but the decay of memory is still quick. Because of this we found that, for sufficiently long realisations, the k-step Markov approximation is often very accurate, even for small k. As a result the denoising performance of the best unidirectional algorithms and the original DUDE were all very similar. Moreover, the algorithms had a tendency to create range-k Markov approximations, using all available 2^k parameters, with the PST as the only exception. As sequences with alternating symbols are less likely than sequences with repeating symbols, the algorithms were not necessarily expected to behave like this.

Therefore, we also consider an artificial source for which the short-range Markov approximations are less accurate. In particular, we define a process as a variable-length Markov chain going through a noisy channel. Furthermore, we will use alphabets with more than two symbols.

We therefore test on two VLMCs. The first VLMC has a randomly generated context tree with 8 symbols. The context function is bounded by 2 and a lower bound of the transition probability is given by $4 \cdot 10^{-5}$.

The second VLMC was, likewise, randomly chosen. The context function of this VLMC is bounded by 2 and the transition probabilities are bounded from below by $4 \cdot 10^{-5}$, but the alphabet now consists of 26 symbols. To both VLMCs we add noise using the typewriter channel, i.e., for $i \in \{1, ..., n\}$, where $n = |\mathscr{A}|$ we have

$$\mathbf{\Pi}_{i,i} = 1 - \varepsilon, \quad \mathbf{\Pi}_{i,i+1} = \varepsilon, \quad \mathbf{\Pi}_{n,1} = \varepsilon$$

and $\mathbf{\Pi}_{ij} = 0$ for all other entries.

Noisy English text

Finally, we will also test the performance of a denoiser based on unidirectional algorithms on real text. We follow [97] and denoise an English version of 'Don

Quixote', corrupted by a so-called typewriter channel. The typewriter channel adds noise by replacing any symbol with a probability of .05, uniformly, by a neighbouring symbol on a QWERTY keyboard. Spaces are left in place, without noise, but we removed other symbols resulting in a 27 letter alphabet.

3.3.2 Metrics for the quality of a various algorithms

The first metric used will be denoising performance of the two-sided versions of the unidirectional algorithms. In this comparison we will treat correctness as equally important for all symbols, i.e., the relevant loss function is the Hamming loss, $\Lambda_{i,j} = 1 - \delta_{i,j}$. We will typically report the denoising quality as the fraction of symbols correct after denoising.

For the symmetric Markov chain through a binary symmetric channel we can compute the probability of any cylindric event. This allows us to use the Kullback-Leibler divergence between the information source and the VLMC constructed by the unidirectional algorithm. However, the Kullback-Leibler divergence is intrinsically one-sided. This is reflected in its direct relation to compression, but it also follows from a well known equality which we will recall now.

Suppose $\hat{\mathbb{P}}$ is a VLMC with maximal depth $k > 0$ and let, for $n \geq 1$,

$$\hat{H}_n = \sum_{a_0^n \in \mathscr{A}^{n+1}} \mathbb{P}(X_0^n = a_0^n) \log\left(\hat{\mathbb{P}}(\hat{X}_0^n = a_0^n)\right),$$

then the Kullback-Leibler divergence, denoted by D, between the information source $\mathbb{P}$ and the VLMC $\hat{\mathbb{P}}$ is given by:

$$\begin{aligned} D\left(\mathbb{P} \middle\| \hat{\mathbb{P}}\right) &= \lim_{n\to\infty} \frac{1}{n} \sum_{a_0^n \in \mathscr{A}^{n+1}} \mathbb{P}(X_0^n = a_0^n) \log\left(\frac{\mathbb{P}(X_0^n = a_0^n)}{\hat{\mathbb{P}}(\hat{X}_0^n = a_0^n)}\right) \\ &= -h_{\mathbb{P}} + \lim_{n\to\infty} \frac{\hat{H}_n}{n}, \end{aligned}$$

where $h_{\mathbb{P}}$ denotes the entropy of the measure $\mathbb{P}$. Writing

$$\lim_{n\to\infty} \frac{\hat{H}_n}{n} = \lim_{m\to\infty} \frac{1}{m} \sum_{n=2}^{m} \hat{H}_n - \hat{H}_{n-1},$$

one can then use translation invariance to write:

$$\begin{aligned}
&\lim_{m\to\infty}\frac{1}{m}\sum_{n=2}^{m}\hat{H}_n-\hat{H}_{n-1}\\
&\quad=\lim_{m\to\infty}\frac{1}{m}\sum_{n=2}^{m}\sum_{a_0^n\in\mathscr{A}^{n+1}}\mathbb{P}(X_0^n=a_0^n)\log\left(\hat{\mathbb{P}}(X_0=a_0|X_1^n=a_1^n)\right)\\
&\quad=\sum_{a_0^k\in\mathscr{A}^{k+1}}\mathbb{P}\left(X_0^k=a_0^k\right)\log\left(\hat{\mathbb{P}}\left(\hat{X}_0=a_0|\hat{X}_1^k=a_1^k\right)\right),
\end{aligned}$$

which is a finite sum. Hence knowing the volume of cylinder sets and the entropy of the source $\mathbb{P}$ allows one to compute the Kullback-Leibler divergence between $\mathbb{P}$ and any finite depth VLMC $\hat{\mathbb{P}}$. However, the last equality confirms that the Kullback-Leibler divergence is an intrinsically one-sided quantity. Instead we can consider a natural two-sided counterpart of the Kullback-Leibler divergence, known as erasure divergence [26]. First we recall the concept of erasure entropy, which can be defined as follows:

Definition 3.7. Let $\mathscr{X}=\{X_n\}_{n\in\mathbb{Z}}$ be a stationary process taking values in $\mathscr{A}$, and let $\mathbb{P}$ be the corresponding probability measure. The erasure entropy rate of $\mathscr{X}$, or, equivalently, of $\mathbb{P}$ is given by

$$h^-(\mathscr{X})=h^-(\mathbb{P})=-\lim_{k\to\infty}\sum_{a_{-k}^k\in\mathscr{A}^{2k+1}}\mathbb{P}\left(X_{-k}^k=a_{-k}^k\right)\log\left(\mathbb{P}\left(X_0=a_0\middle|X_1^k=a_1^k,X_{-k}^{-1}=a_{-k}^{-1}\right)\right).$$

Now erasure divergence is defined as follows:

Definition 3.8. Let $\mathbb{P}$ be a stationary measure on $\mathscr{A}^{\mathbb{Z}}$ and let $\{\hat{X}_n\}_{n\in\mathbb{Z}_+}$ be a VLMC, taking values in $\mathscr{A}$, with context length function $\sup_{x\in\mathscr{A}^{-\mathbb{N}}} l=k<\infty$, determined by the measure $\hat{\mathbb{P}}$. The erasure divergence between $\mathbb{P}$ and $\hat{\mathbb{P}}$ is given by

$$D^-\left(\mathbb{P}\middle|\middle|\hat{\mathbb{P}}\right)=-h_{\mathbb{P}}^-+\sum_{a_{-k}^k\in\mathscr{A}^{2k+1}}\mathbb{P}\left(X_{-k}^k=a_{-k}^k\right)\log\left(\hat{\mathbb{P}}\left(\hat{X}_0=a_0|\hat{X}_1^k=a_1^k,\hat{X}_{-k}^{-1}=a_{-k}^{-1}\right)\right).$$

For various VLMC's produced by the unidirectional algorithms we will compare the Kullback-Leibler divergence, the erasure divergence of the resulting two-sided model, and the denoising performance.

Selection of parameters

The dependency of some of the algorithms on additional parameters requires a way of selecting appropriate values. In [4] a selection of good values for the above parameters is discussed, with application to the compression problem in mind. For the parameters that do not correspond to the maximal depth of the created VLMC we perform measurements around the values suggested in [4] to select reasonable values. We discuss the parameter selection in section 3.3.6.

In this section we will address elimination of the maximal depth parameter. The algorithms LZ78 and LZ-MS do not have such a parameter. We introduce this parameter to LZ78 and LZ-MS to make working with the resulting VLMC more convenient. However, we found that this additional bound was often beneficial. For all the algorithms, including the DUDE, the optimal value of the interaction range parameter depends on the properties of the source as well as the length of the realisation. In the context of denoising we eliminate this parameter using the compression heuristic, i.e., we choose the value that make the denoised data most compressible.

3.3.3 Data

For the symmetrical Markov process on two symbols, with parameters $p = .1$ and $\varepsilon = .1$ and realizations of 100000 symbols we consider the erasure divergence, KL divergence and denoising performance of each of the unidirectional algorithms, without eliminating the parameter d.

We found that the two divergences behave very similar to each other. Moreover, the quality of denoising follows the divergences well. It is also clear that, around their optimal values of d, two algorithms do better then the others. PPMC and PST reach the lowest values for both entropies and moreover are very stable for high values of d. LZ-MS follows PPMC and PST closely and, interestingly, is optimal for a slightly smaller value of d than PPMC and PST. LZ78 is clearly worse than L-MS, PPMC and PST and finally both CTW algorithms perform quite poorly in these metrics. Interestingly they gain very little by varying d.

For the symmetrical Markov process on two symbols, with parameters $p = .1$ and $\varepsilon \in \{.05, .1, .3\}$ we tested realisations of lengths 100, 1000, 10000 and 100000 symbols. The maximal depth d of the tree, and k for the DUDE, is eliminated by the compression heuristic. The results are listed in the tables 3.1, 3.2, 3.3 and 3.4 below. We found that PPM-C, PST and LZ-MS are consistently comparable or better than the DUDE, whereas LZ78 is, with one exception, comparable or worse. Interestingly, the versions of context tree weighting, when applied to the longer realisations, were both worse then the classical DUDE, however, they were among the best algorithms on some of the shorter realisations, except that DE-CTW failed to denoise for $\varepsilon = .3$ regardless of the length of the realisation.

The fact that the good algorithms are comparable with the DUDE for long realisations and better on shorter realisations might be explained by the quick decay of variation in this process, making the simple Markov approximations efficient when enough data is available. Furthermore, we note that the performance of the CTW algorithms is somewhat surprising, as they are known as excellent denoisers.

Another aspect of note is that BI-CTW is systematically better then DE-CTW on the Binary Symmetric Channel. This could be explained by the fact that BI-CTW is developed for a two letter alphabet whilst DE-CTW is an adaptation to accommodate larger alphabets.

Algorithm	ε	Average k	Fraction of symbols correct	SEM
DUDE	.05	1.6	.946	.002
DUDE	.1	1.6	.915	.002
DUDE	.3	1.5	.709	.004
Algorithm	ε	Average d	Fraction of symbols correct	SEM
PPM-C	.05	1.2	.964	.002
PPM-C	.1	1.5	.928	.004
PPM-C	.3	1.7	.718	.007
PST	.05	1.2	.966	.002
PST	.1	1.5	**.940**	.004
PST	.3	1.8	.719	.008
LZ78	.05	1.8	.961	.003
LZ78	.1	1.8	.921	.005
LZ78	.3	1.6	.694	.009
LZ-MS	.05	2.7	.964	.003
LZ-MS	.1	2.7	.922	.004
LZ-MS	.3	1.8	.722	.007
BI-CTW	.05	1.4	.963	.003
BI-CTW	.1	1.7	.937	.003
BI-CTW	.3	2.3	**.750**	.006
DE-CTW	.05	1.2	**.967**	.003
DE-CTW	.1	1.4	.938	.004
DE-CTW	.3	1.2	.697	.006

Table 3.1: Binary symmetric channel with length 100.

Algorithm	ε	Average k	Fraction of symbols correct	SEM
DUDE	.05	1.3	.9679	.0005
DUDE	.1	1.8	.9381	.0008
DUDE	.3	1.6	.746	.002
Algorithm	ε	Average d	Fraction of symbols correct	SEM
PPM-C	.05	1.2	.969	.001
PPM-C	.1	2.0	.939	.002
PPM-C	.3	2.2	.756	.005
PST	.05	1.3	.967	.001
PST	.1	2.2	**.940**	.001
PST	.3	2.3	**.767**	.003
LZ78	.05	1.3	.967	.001
LZ78	.1	1.7	.938	.001
LZ78	.3	1.7	.732	.007
LZ-MS	.05	3.5	.966	.001
LZ-MS	.1	3.3	.936	.002
LZ-MS	.3	2.6	.740	.006
BI-CTW	.05	1.4	**.971**	.001
BI-CTW	.1	1.9	.938	.002
BI-CTW	.3	2.2	.751	.005
DE-CTW	.05	1.7	.968	.001
DE-CTW	.1	1.8	.934	.002
DE-CTW	.3	1.7	.66	.02

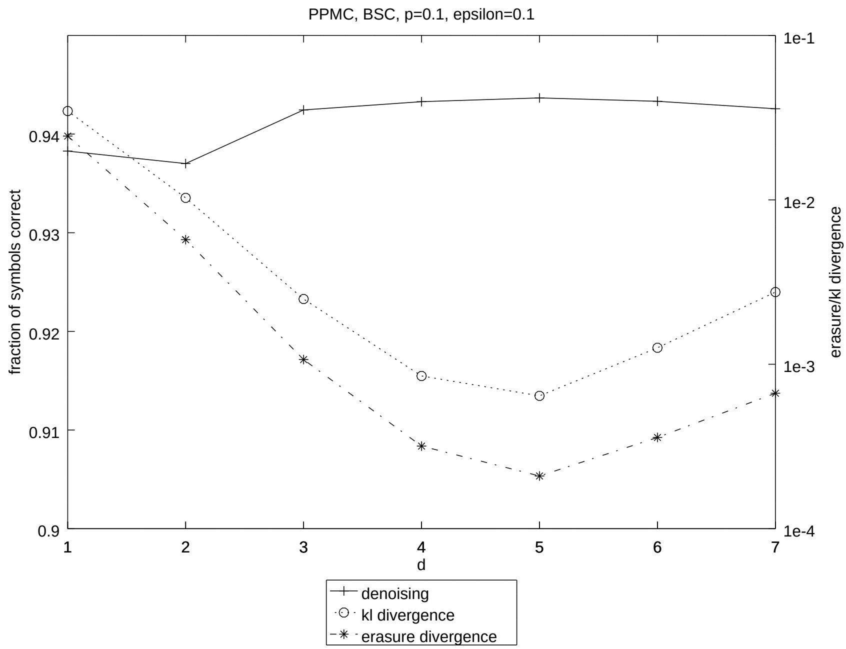
Table 3.2: Binary symmetric channel with length 1000.

Algorithm	ε	Average k	Fraction of symbols correct	SEM
DUDE	.05	1.5	.9689	.0002
DUDE	.1	2.2	**.9432**	.0001
DUDE	.3	2.5	.7674	.0007
Algorithm	ε	Average d	Fraction of symbols correct	SEM
PPM-C	.05	1.3	.9687	.0004
PPM-C	.1	3.9	.9419	.0005
PPM-C	.3	3.2	**.772**	.001
PST	.05	1.2	.9687	.0003
PST	.1	3.7	.9416	.0005
PST	.3	3	.771	.001
LZ78	.05	1.4	.9686	.0005
LZ78	.1	2.8	.9392	.0006
LZ78	.3	2.5	.763	.002
LZ-MS	.05	3.7	**.9706**	.0003
LZ-MS	.1	3.9	.9411	.0005
LZ-MS	.3	3.2	.770	.001
BI-CTW	.05	1.2	.9684	.0003
BI-CTW	.1	2.1	.9383	.0006
BI-CTW	.3	3.0	.753	.004
DE-CTW	.05	1.7	.9674	.0006
DE-CTW	.1	2.925	.9393	.0008
DE-CTW	.3	3.2	.70	.02

Table 3.3: Binary symmetric channel with length 10000.

Algorithm	ε	Average k	Fraction of symbols correct	SEM
DUDE	.05	1.6	.9696	.0001
DUDE	.1	2.2	.9432	.0001
DUDE	.3	3.5	.7797	.0003
Algorithm	ε	Average d	Fraction of symbols correct	SEM
PPM-C	.05	1	.9685	.0002
PPM-C	.1	4.86	.9429	.0001
PPM-C	.3	4.80	.7851	.0009
PST	.05	1.2	.9687	.0001
PST	.1	5.63	**.9436**	.0003
PST	.3	4.52	.7814	.0007
LZ78	.05	1.3	.9687	.0002
LZ78	.1	3.9	.9408	.0004
LZ78	.3	3.2	.776	.001
LZ-MS	.05	4.8	**.9706**	.0003
LZ-MS	.1	5.35	.9417	.0003
LZ-MS	.3	4.15	**.7830**	.0008
BI-CTW	.05	1.05	.9685	.0001
BI-CTW	.1	4.0	.934	.001
BI-CTW	.3	3.1	.751	.002
DE-CTW	.05	2.4	.9677	.0008
DE-CTW	.1	2.5	.9382	.0005
DE-CTW	.3	3.4	.71	.01

Table 3.4: Binary symmetric channel with length 100000.



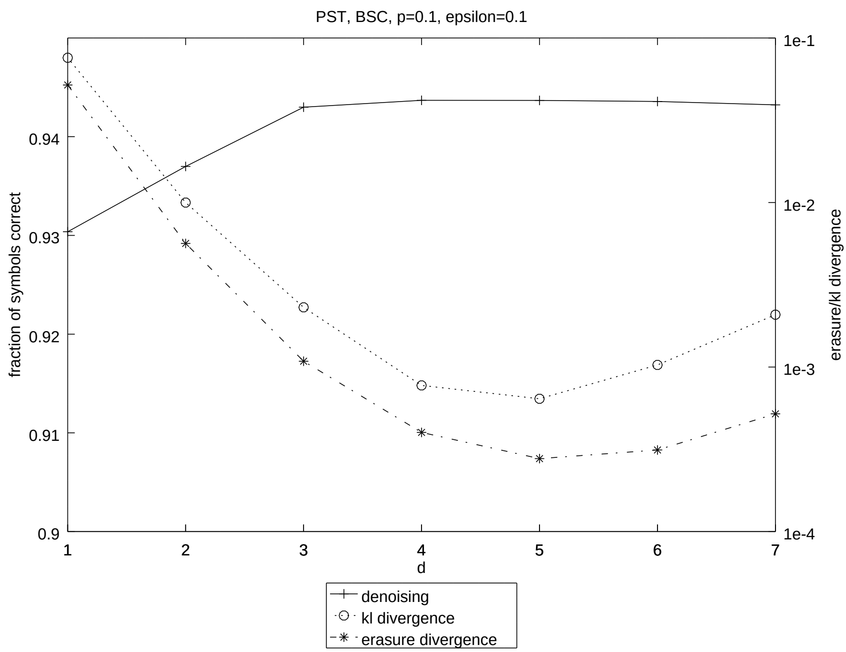


Figure 3.1: A comparison of the KL-divergence, erasure divergence and denoising performance for the unidirectional algorithms at different values of the parameter d. All realizations consist of 100000 symbols of a symmetrical Markov chain on two symbols going through a binary symmetric channel, with $p = .1$ and $\varepsilon = .1$.

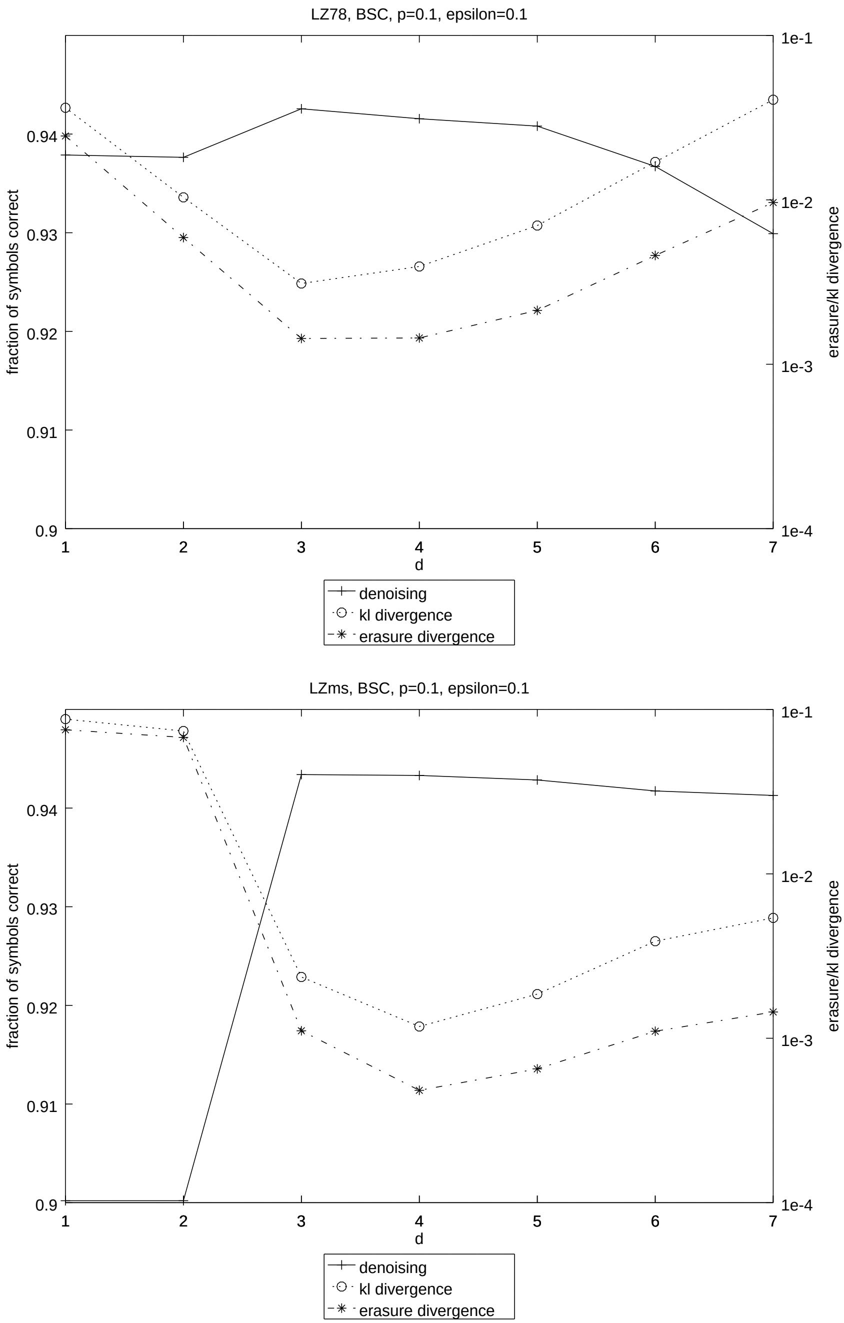
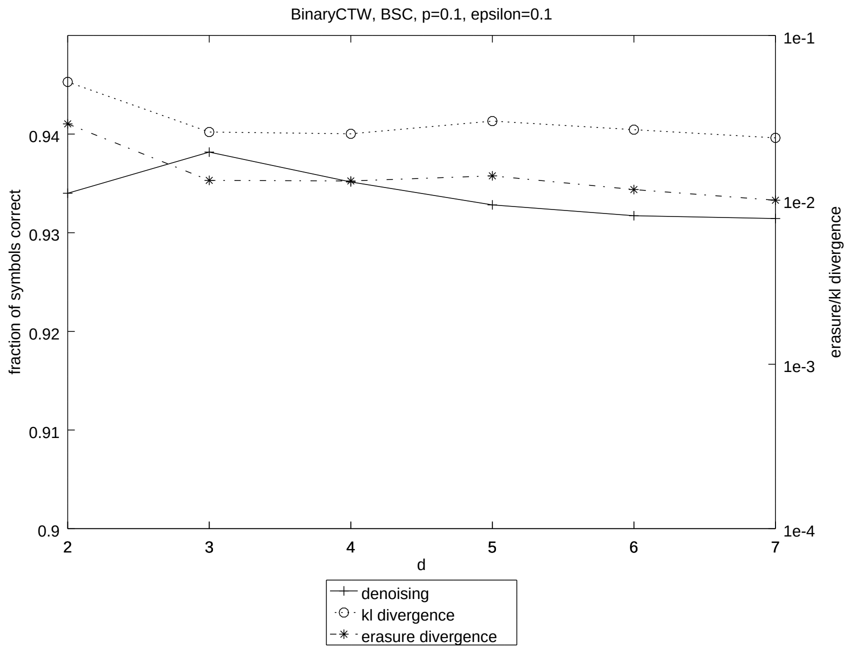


Figure 3.2: A comparison of the KL-divergence, erasure divergence and denoising performance for the unidirectional algorithms at different values of the parameter d. All realizations consist of 100000 symbols of a symmetrical Markov chain on two symbols going through a binary symmetric channel, with $p = .1$ and $\varepsilon = .1$.



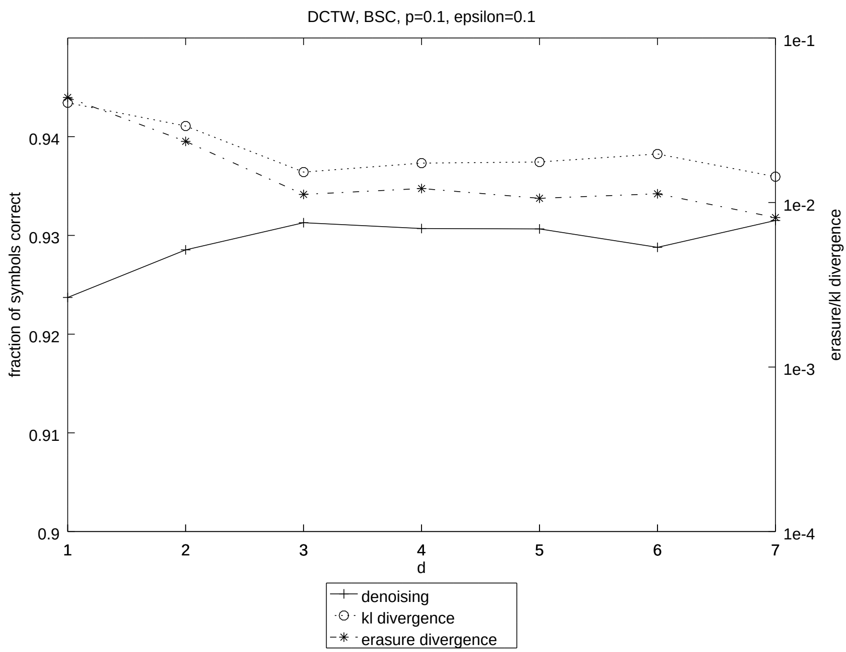


Figure 3.3: A comparison of the KL-divergence, erasure divergence and denoising performance for the unidirectional algorithms at different values of the parameter d. All realizations consist of 100000 symbols of a symmetrical Markov chain on two symbols going through a binary symmetric channel, with $p = .1$ and $\varepsilon = .1$.

3.3.4 Noisy VLMCs

The experiments on noisy VLMC's were performed for a realisation of 1000000 symbols of a VLMC on 26 symbols going through a typewriter channel with $\varepsilon =$ .05. This channel is given by

$$\Pi_{i,i} = 1 - \varepsilon, \quad \Pi_{i,i+1} = \varepsilon, \quad \Pi_{n,1} = \varepsilon,$$

for any $1 \leq i < n$, with n the size of the alphabet.
In this case we the performance of the DUDE and the unidirectional algorithms is no longer comparable. Instead, the classical DUDE only removes a very limited amount of noise, while the unidirectional algorithms, except the BI-CTW, performed better by a significant margin, as shown in table 3.5.

A second experiment on noisy VLMC's is performed on shorter realisations, of length 10000 and for an alphabet of 8 symbols. The noise is added using the typewriter channel for $\varepsilon = .05$. The results are given in table 3.6. Again, all unidirectional algorithms, except BI-CTW, outperform the DUDE. In both cases the PPM-C algorithm was the best performing algorithm, PST and DE-CTW are both close to the PPM-C, while LZ-MS falls behind on the shorter realisations. LZ78 is clearly worse then the best unidirectional algorithms in both cases.

Algorithm	Average k	Fraction of symbols correct	SEM
DUDE	1	.9510	$1 \cdot 10^{-4}$
Algorithm	Average d	Fraction of symbols correct	SEM
PPM-C	2	**.97449**	$3 \cdot 10^{-5}$
PST	2.6	.97291	$2 \cdot 10^{-5}$
LZ78	2	.96873	$4 \cdot 10^{-5}$
LZ-MS	2	.97285	$2 \cdot 10^{-5}$
BI-CTW	1	.94991	$3 \cdot 10^{-5}$
DE-CTW	2	.97260	$2 \cdot 10^{-5}$

Table 3.5: Denoising results of the realisations of a noisy VLMC of length 100000, with 26 letters in the alphabet.

Algorithm	Average d	Fraction of symbols correct	SEM
DUDE	1	.9519	.001
PPM-C	2	**.9727**	.0009
PST	1.7	.972	.001
LZ78	1.7	.9644	.0007
LZ-MS	1	.9665	.0006
BI-CTW	3	.949	.001
DE-CTW	2.1	.9721	.0008

Table 3.6: Denoising results of the realisations of a noisy VLMC of length 10000, with 8 letters in the alphabet.

3.3.5 Noisy English text

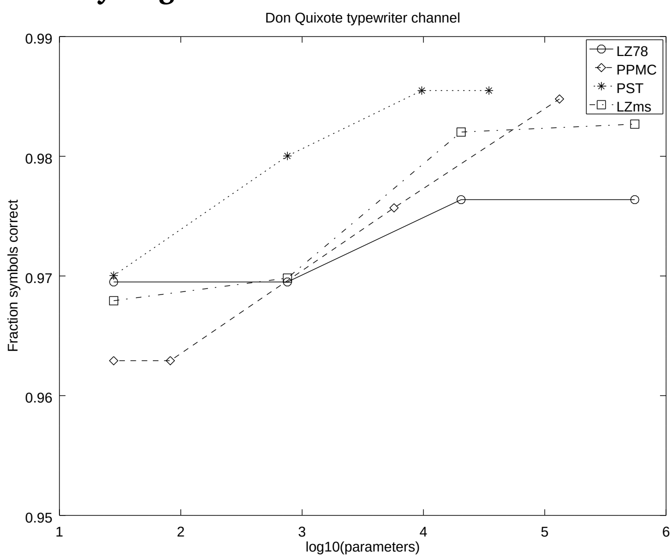


Figure 3.4: Parameters versus denoising for noisy English text.

An overview of the performance of the algorithms on the text by Don Quixote through a typewriter channel is given in table 3.7. The error probability, on the alphabet of 27 symbols, was set to .05 for non space characters. This resulted in 90110 errors before denoising, corresponding to a fraction of .040407 of the total number of symbols. In this case we also report the number of parameters in the resulting tree, except for the CTW algorithms as the way we extracted the context tree from the algorithms did not allow for that measurement.

As the optimal parameters on noisy text, for the LZ-MS algorithms, as reported by Begleiter et al. was somewhat different from the $m = 3, s = 3$ we used on the binary symmetric channel, we also tested $m = 2$, $s = 8$, close to the optimum as reported in [4] and used the compression heuristic to select the correct parameters. In figure 3.4 we plotted the values selected by the compression heuristic. It turned out that for $1 \leq d \leq 3$ the parameters $m = 3$, $s = 3$ were optimal whereas for $d = 4$ the settings with $m = 2, s = 8$ were slightly better. For PST we also changed the initial parameter r from 1.05 to 1.01 as the tree remained extremely small in the first case, which was reflected in the denoising performance.

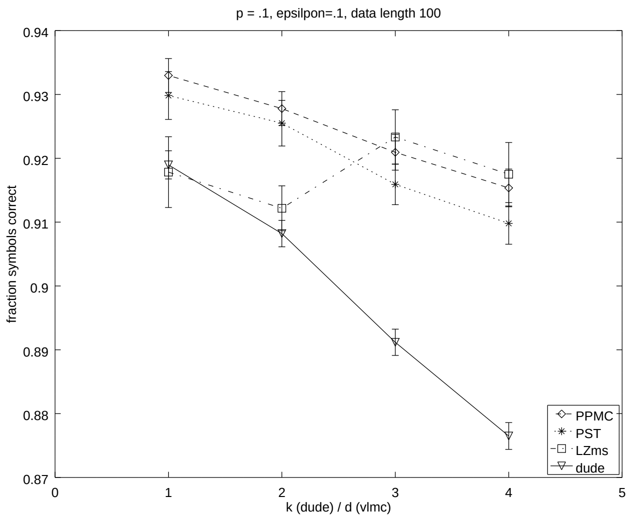
Algorihm	d / k	Parameters in model	Errors after	Fraction symbols correct
DUDE	2	-	48977	.978038
LZ78	3	20440	52685	.97637
PPM-C	4	132247	33926	.98479
PST	3	9694	32357	**.98549**
LZ-MS	4	551881	38607	.98269
BI-CTW	3	-	92672	.95844
DE-CTW	3	-	42295	.98103

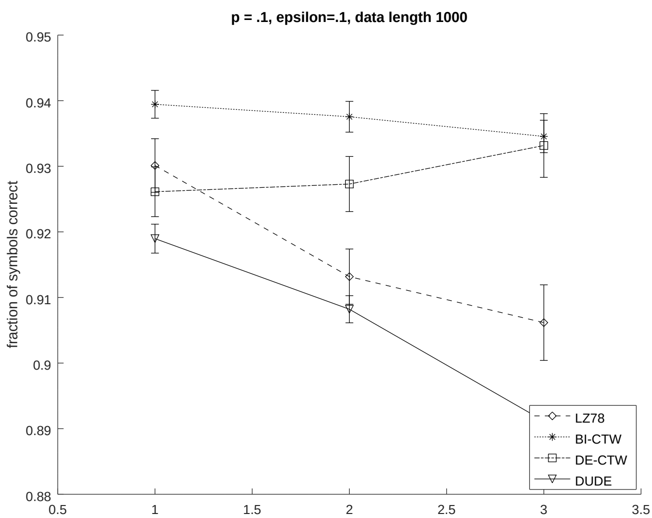
Table 3.7: Denoising results on a noisy English version of 'Don Quixote'.

We found that, except for the BI-CTW and LZ78 algorithms, the unidirectional algorithms outperform the DUDE, with PST and PPM-C being top performers by quite a large margin in the number of errors remaining. Moreover, the PST algorithm demonstrates that excellent performance can be achievement by a relatively small model.

3.3.6 Parameter selection

We now address the elimination of the algorithm parameters. In particular, all algorithms have a depth parameter, that was added to those unidirectional algorithms in which it is not already present. Subsequently we eliminate this parameter using the compression heuristic. We note that the compression heuristic correctly selected the best parameter value in most instances. A notable exception is shown in figure 3.5. The average number of bytes after compression is plotted against the parameter d, for the PPM-C and PST algorithms and next to it we show the corresponding quality of denoising. In this instance the heuristic incorrectly favours $d = 1$, while all other values of d result in a much better performance. How the unidirectional algorithms depend on the parameter d for denoising is shown in Figure 3.9, together with the dependence of the DUDE on the parameter k.





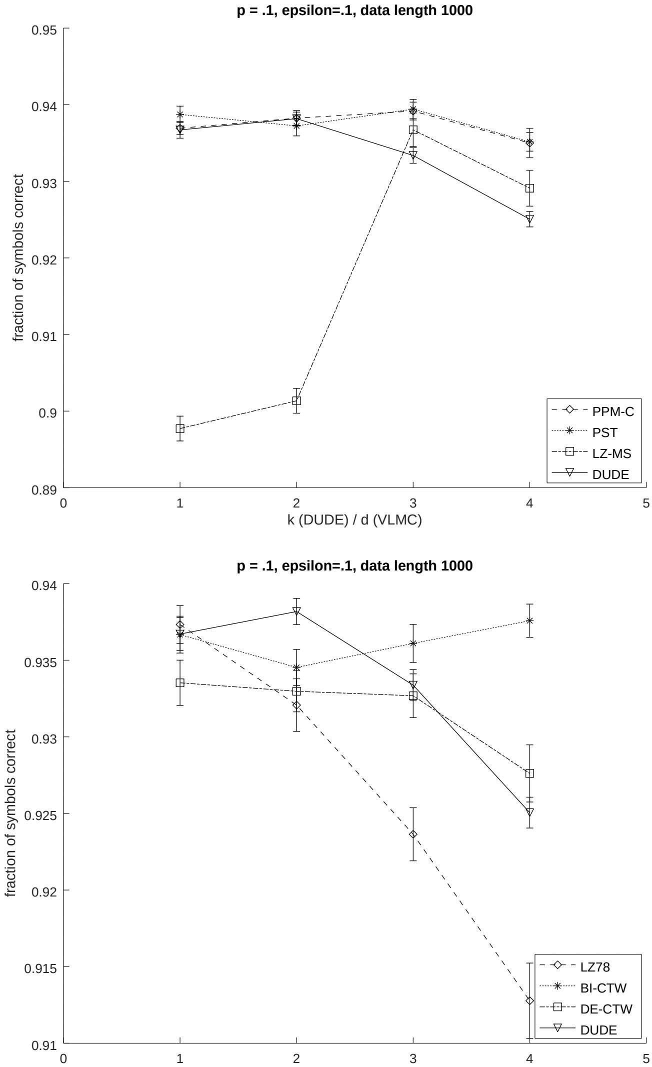


Figure 3.7: The dependence on d/k for the tested algorithms and the DUDE.

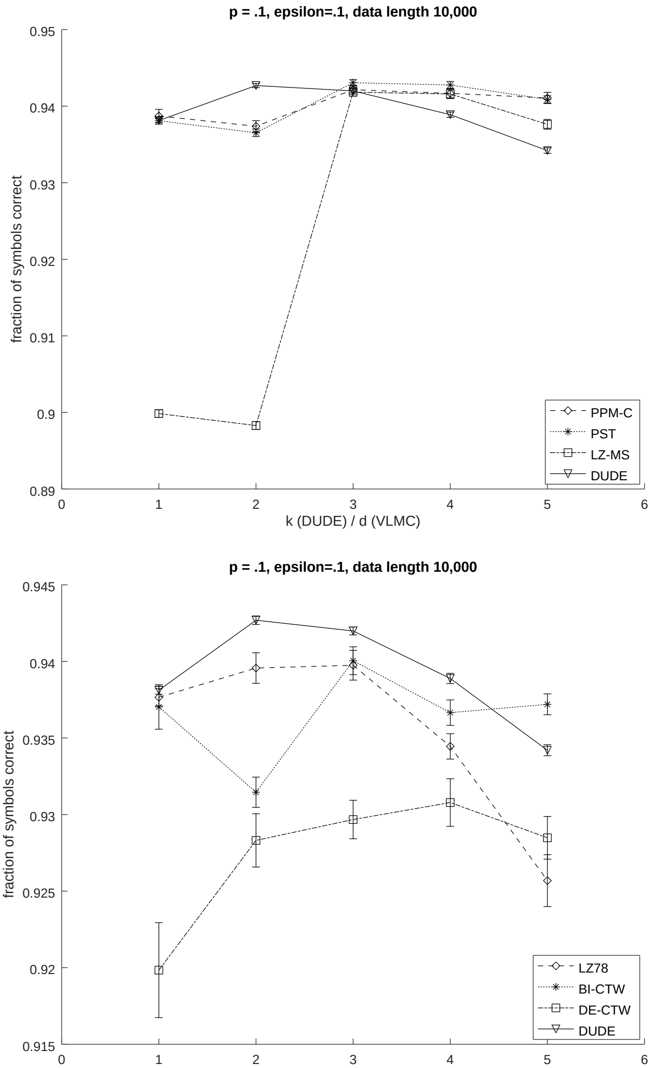


Figure 3.8: The dependence on d/k for the tested algorithms and the DUDE.

Besides the tree-depth d, the LZ-MS and PST algorithms have an additional set of parameters. For the experiments on the Binary Symmetric Channel we compared the performance for various values of the parameters on realisations with 100000 symbols. The tested values for PST were chosen around the values reported in [4] to be good for compression. In figure 3.10 we see that the resulting curves fall apart in two groups. That is, for a range of values the parameters result in similar denoising quality. The performance for the second group of parameters is significantly worse. For the remaining experiments we selected one of the well-performing parameter configurations. The values selected in [4] also fall in the range resulting in the upper curve.

Similarly the curves for different parameters m,s for LZ-MS are very similar, with the exception of some specific combinations that perform poorly for short d. We chose $m = 3$ and $s = 3$ for the remaining experiments involving artificial sources.

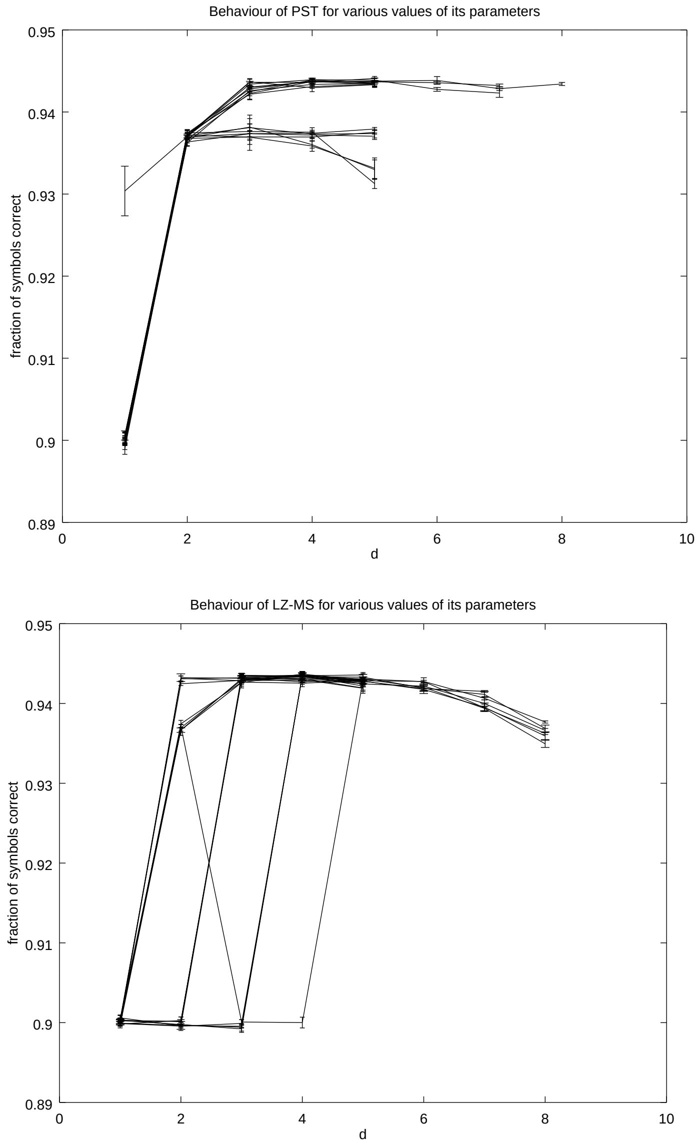


Figure 3.10: The dependence on their parameters for the PST and LZ-MS algorithms.

For English text going through a typewriter channel we performed one further test on the stability of the parameters and we made one adjustment for the PST algorithm. For the chosen parameters the PST algorithm generates a tree containing a very small number of nodes. Hence we also tested the PST for $r = 1.01$ instead of $r = 1.05$ and chose between those options based on the compression heuristic.

For LZ-MS the parameters chosen above were compared with $m = 2$, $s = 8$, parameters that performed well in compression [4]. A clear improvement in denoising was found. In both of the above cases the compression heuristic correctly selected the optimal denoiser.

3.4 Conclusions

We confirmed the finding in [26] that the unidirectional algorithm can result in comparable or better denoising performance. In particular the PST, PPM-C and LZ-MS algorithms were consistently as good as the DUDE or better. However, the versions of CTW and LZ78, in several instances, performed worse than the classical DUDE. Moreover, we saw that a unidirectional model that approximates the source well, as measured by the Kullback-Leibler divergence, tends to result in a good denoiser. However, some of the worse unidirectional models still resulted in good denoising performance. We also found that the introduction of a maximal depth parameter to the LZ78 and LZ-MS algorithms tended to improve the denoising performance. Finally, we note that the compression heuristic, that we used to eliminate the parameters of the algorithms, tended to select optimal values. We found one exception, where the compression heuristic significantly reduced the denoising of the PST and PPM-C algorithms, by selecting the wrong parameter value.

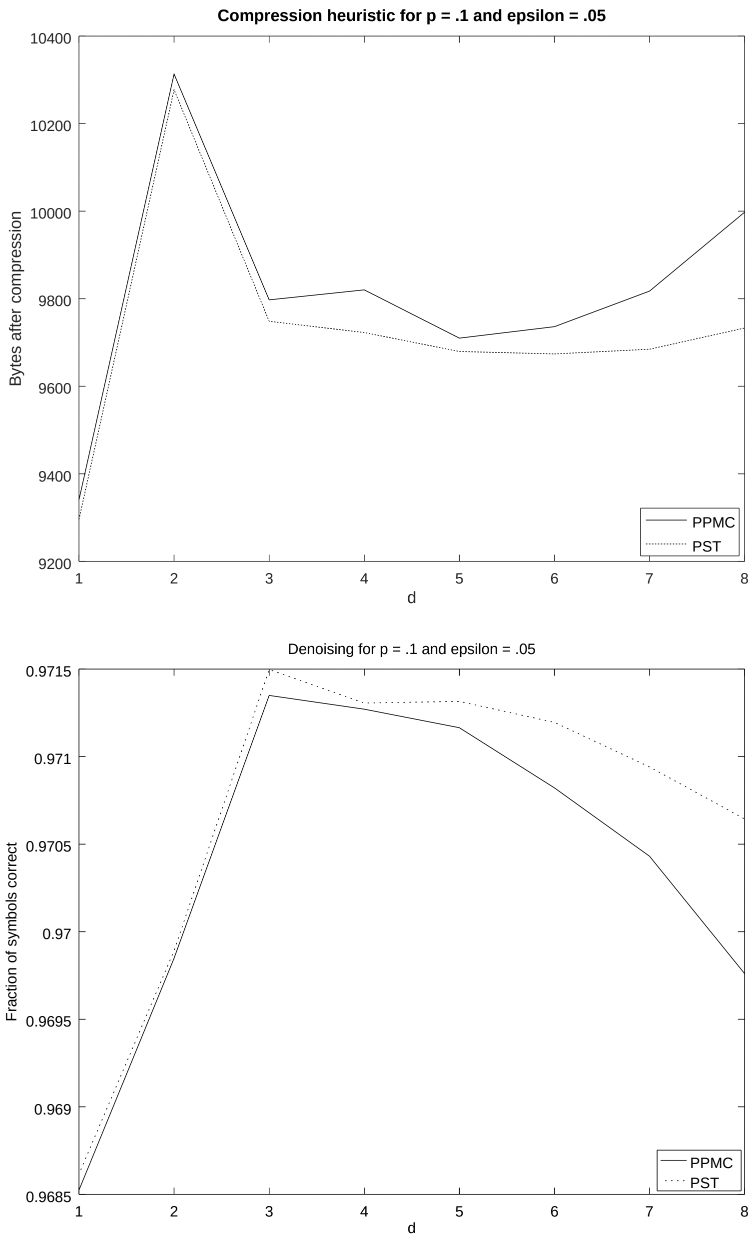


Figure 3.5: An instance where the compression heuristic for $p = .1$ and $\varepsilon = .05$ on a realization of 100000 symbols resulting a poor choice of d.

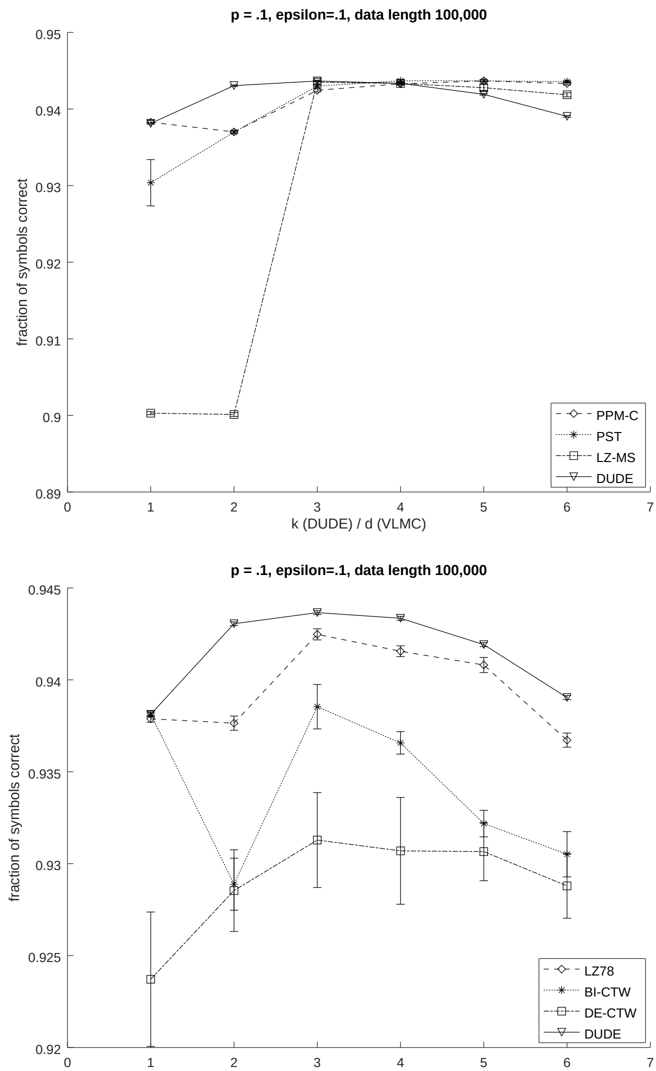


Figure 3.9: The dependence on d or k for the tested algorithms and the DUDE.

Chapter 4

On regularity of functions of Markov chains

4.1 Introduction

Suppose $\{X_n\}$ is a stationary Markov chain taking values in a finite set $\mathscr{A}$ and assume that we are not able to observe the values $\{X_n\}$ directly. Instead, we observe the values of some function of X_n that groups some elements in $\mathscr{A}$ together. To be precise, let $\pi : \mathscr{A} \to \mathscr{B}$, with $\mathscr{B}$ a smaller alphabet, and assume that we observe the process $\{Y_n\}$ given by

$$Y_n = \pi(X_n), \quad \text{for all } n. \tag{4.1}$$

Processes of this form have been studied extensively in the past 60 years and appear under a variety of different names in various fields: in Probability Theory, functions of Markov chains [14], grouped [45], lumped [52], or aggregated Markov chains [82]; one-block factors of Markov measures [64] or sofic measures [54] in Ergodic Theory. Note also that the Hidden Markov models [3] – very popular in Statistics, can be cast in the form (4.1) as well.

The factor process $\{Y_n\}$ is rarely Markov, the necessary and sufficient conditions have been found by Kemeny & Snell and Dynkin [25,52]. This raises the principal question; what is the dependence structure of the factor process?

It turns out that, under rather mild conditions on the underlying Markov chain and the coding map π, the resulting process can be seen as *approximately* or *nearly* Markov in the following sense: the conditional distribution of the next value Y_1 depends on the complete past $Y_{-\infty}^0 := (\dots, Y_{-2}, Y_{-1}, Y_0)$, but this dependence is *regular*, i.e., the distant past values $\{Y_{-n}\}$, for $n \gg 1$, have a diminishing effect on the distribution of Y_1. Stochastic processes with such properties occur naturally in many contexts, as a consequence many authors introduced concepts that formalize the notion of a measure that is approximately Markov. Among these concepts are the chains with complete connections [69], chains of infinite order [45], g-measures [51] and uniform martingales [50]. Although these concepts are very similar, they are not always equivalent, for a more detailed discussion see [30]. Among these notions g-measures are the most convenient for the purposes of this paper. Usually g-measures are defined on some subset of the product space $\mathscr{A}^{\mathbb{Z}_+}$. This space can be thought of as the collection of all allowed paths of a process starting at time 0. Like a Markov measure, a g-measure is introduced via its transition probabilities, the differences are that the transitions of a g-measure are described by a function $g : \mathscr{A}^{\mathbb{Z}_+} \to (0,1)$, rather than a matrix $P : \mathscr{A} \times \mathscr{A} \to [0,1]$ and that the time direction is reversed. That is, the vector $g(\cdot\, x_1^\infty)$ represents the distribution of the symbol in the origin, conditioned on the 'future' configuration x_1^∞. This time reversal is common in ergodic theory and is mostly inconsequential for our purposes, as a Markov measure satisfies the Markov property in both directions. A g-measure is approximately Markov due to the additional constraint that the function g is continuous. To clarify, continuity corresponds to a vanishing influence from far away symbols since, in the product topology, a function $g : \mathscr{A}^{\mathbb{Z}_+} \to \mathbb{R}$ is continuous if and only if:

$$\mathrm{var}_n(g) \equiv \sup_{x,y \in \mathscr{A}^{\mathbb{Z}_+}} \left| g(x_0^\infty) - g(x_0^n y_{n+1}^\infty) \right| \to 0, \text{ as } n \to \infty.$$

We can now use the language of g-measures to phrase the main result of this paper: we provide a novel sufficient condition for functions (factors) of Markov chains to belong to the class of g-measures. This condition is based on the application of the so-called fibre approach that originated in Ergodic Theory [60], but seems to be less known in Probability Theory.

Let us now describe this method briefly. Suppose μ is a Markov measure on $\Omega = \mathscr{A}^{\mathbb{Z}_+}$, $\pi : \Omega \to \Sigma$ a factor map and $\nu = \mu \circ \pi^{-1}$. Now define a fibre over $y \in \Sigma$ as the set $\Omega_y = \{x \in \mathscr{A}^{\mathbb{Z}_+} : \pi(x_i) = y_i\}$. Given a Markov measure μ and a factor map π, one can find a family of measures $\{\mu_y\}_{y \in \Sigma}$, called a disintegration of μ, such that μ_y is concentrated on Ω_y and $\mu = \int \mu_y \, d\nu$.

We will show in Theorem 4.9 that the factor measure ν is consistent with a function $\tilde{g} : \Sigma \to [0,1]$, i.e., $\nu(y_0 | y_1^\infty) = \tilde{g}(y)$ for ν-a.e. $y \in \Sigma$, where $\tilde{g}$ can be expressed in terms of the disintegration $\{\mu_y\}$. Based on this expression we identify two sufficient conditions for ν to be a g-measure. The first covers known results on lumpability, i.e., describes when the factor measure is Markov. The second condition is that the disintegration can be chosen in such a way that $y \to \mu_y$ is continuous. If $y \to \mu_y$ is continuous we call $\{\mu_y\}_{y \in \Sigma}$ a Continuous Measure

Disintegration (CMD). Our main result is that existence of a CMD implies that the factor measure is a g-measure.

Previously, this condition has been applied successfully to the analogous question: when is a factor of a fully supported g-measure itself a g-measure [49, 87]? In the context of factors of Markov measures, we show that the condition supersedes the currently known conditions.

These results are presented here in the following way: firstly, we introduce the necessary definitions, then we review known results in sections 4.2.1, 4.2.2 and 4.2.4. Subsequently we state our main theorem in section 4.3. In order to demonstrate that the condition in section 4.3 is more general than known results we apply the theory of non-homogeneous equilibrium states in section 4.4.2. We will also recall the constructive approach to continuous measure disintegrations by Tjur in section 4.4.3 to provide an interesting alternative to recover the known conditions in 4.4.5. Finally, in section 4.5, we discuss some examples to show that existence of a continuous measure disintegration is *strictly* weaker than the previously known conditions and that, unfortunately, it is not a necessary condition.

4.1.1 Notation

Suppose $\mathscr{A}$ is a finite set (alphabet) and M is $|\mathscr{A}| \times |\mathscr{A}|$ matrix with entries in $\{0,1\}$. The corresponding subshift of finite type (SFT) Ω_M is defined as

$$\Omega_M = \left\{ x = (x_n)_{n=0}^{\infty} \in \mathscr{A}^{\mathbb{Z}_+} : M(x_n, x_{n+1}) = 1 \quad \forall n \in \mathbb{Z}_+ \right\}.$$

We equip Ω_M with the product topology. We use the shorthand notation $a_n^m = (a_n, a_{n+1}, \ldots, a_m)$ for words in alphabet $\mathscr{A}$, and denote the corresponding cylinder sets as $[a_n^m] = \left\{ x \in \Omega_M : x_n^m = a_n^m \right\}$. Similarly, for a given finite set $\Lambda \subset \mathbb{Z}_+$, denote the configuration on the subset Λ by $a_\Lambda = (a_i)_{i \in \Lambda}$. A concatenation of two configurations a_Λ and b_Δ on disjoint sets $\Lambda, \Delta \subset \mathbb{Z}_+$ is denoted as $a_\Lambda b_\Delta$, to be precise:

$$(a_\Lambda b_\Delta)_i = \begin{cases} a_i & ,\text{if } i \in \Lambda, \\ b_i & ,\text{if } i \in \Delta. \end{cases}$$

For a given subshift of finite type Ω_M a Markov chain with probability transition matrix P is said to be *compatible* with Ω_M if $P_{ij} > 0 \iff M_{ij} = 1$ for all $i, j \in \mathscr{A}$. In complete analogy with the terminology for Markov chains, the subshift of finite type Ω_M is called

(a) ***irreducible*** if $\forall i, j \in \mathscr{A}$ there exists an $n = n(i,j) > 0$ such that $M^n(i,j) > 0$;

(b) ***aperiodic*** if $\forall i \in \mathscr{A}$, one has $\gcd\{m > 0 : M^m(i,i) > 0\} = 1$;

(c) ***primitive*** if there exists an $n > 0$ such that $M^n > 0$.

If the subshift of finite type Ω_M is irreducible, and P is a compatible probability transition matrix, then the (unique) stationary Markov measure μ has Ω_M as its support.

4.1.2 Single-block factor maps

Suppose $\mathscr{A}$ and $\mathscr{B}$ are finite sets, $|\mathscr{A}| > |\mathscr{B}|$, and $\pi : \mathscr{A} \to \mathscr{B}$ is a surjective map. We use the same symbol π to denote the map from $\mathscr{A}^{\mathbb{Z}_+}$ to $\mathscr{B}^{\mathbb{Z}_+}$ given by $\pi(x)_n = \pi(x_n)$ for all $n \in \mathbb{Z}_+$. Let μ be a stationary Markov measure corresponding to a Markov chain $\{X_n\}$, supported on an irreducible subshift of finite type $\Omega = \Omega_M \subset \mathscr{A}^{\mathbb{Z}_+}$, define the push-forward (or factor) measure ν as $\nu = \mu \circ \pi^{-1}$. The measure ν is supported on a subshift $\Sigma = \pi(\Omega) \subset \mathscr{B}^{\mathbb{Z}_+}$. In symbolic dynamics, Σ and ν are called the sofic shift and the sofic measure, respectively. Note that Σ is not necessarily a subshift of finite type. Throughout the paper we make the following standing assumptions on Ω and π:

$$\Omega = \Omega_M \text{ is an irreducible SFT,} \tag{A1}$$

$$\text{the one-block factor map } \pi : \mathscr{A}^{\mathbb{Z}_+} \to \mathscr{B}^{\mathbb{Z}_+} \text{ is such that } \Sigma = \pi(\Omega) \text{ is an SFT} \tag{A2}$$

i.e., $\Sigma = \Sigma_{M'}$ for some $\{0,1\}$ matrix M'. We note that using standard methods of symbolic dynamics (Fisher covers), it is possible to decide algorithmically whether for a given pair (Ω, π), the image Σ is indeed an SFT [78].

4.1.3 g-measures

As we will see below, factors of Markov measures are rarely Markov. Instead, it is far more common for factors of Markov measures to belong to the class of g-measures, i.e., measures having positive continuous conditional probabilities. Suppose $\Sigma \subseteq \mathscr{B}^{\mathbb{Z}_+}$ is a SFT and consider the following set of functions:

$$\mathscr{G}(\Sigma) = \left\{ g \in C(\Sigma, (0,1)) : \sum_{b \in \mathscr{B} : by \in \Sigma} g(by) = 1 \text{ for all } y \in \Sigma \right\}.$$

Definition 4.1. A translation invariant measure ν on Σ is called a g-measure for $g \in \mathscr{G}(\Sigma)$ if

$$\nu(y_0 | y_1^\infty) = g(y_0^\infty)$$

for ν-a.e. $x \in \Sigma$. Equivalently, ν is a g-measure if, for any continuous function $f : \Sigma \to \mathbb{R}$, one has

$$\int f(y)\nu(dy) = \int \sum_{b\in\mathscr{B}:\ by\in\Sigma} f(by)g(by)\nu(dy).$$

For any $g \in \mathscr{G}(\Sigma)$, at least one g-measure exists; however such a measure might not be unique [12]. A useful property of g-measures is that they characterized by the uniform convergence of finite one-sided conditional probabilities.

Proposition 4.2. [74] *A translation invariant probability measure ν on a SFT Σ is a g-measure if and only if the sequence of local functions on Σ*

$$g_n(y_0^n) := \nu(y_0|y_1^n),$$

converges uniformly to some function $g \in \mathscr{G}(\Sigma)$.

In the opposite direction, one can conclude that a given measure ν is not a g-measure if one is able to find a so-called *bad configuration* for ν.

Definition 4.3. A point $y \in \Sigma$ is called a bad configuration for ν if there exists an $\varepsilon > 0$ such that, for every $n \in \mathbb{N}$, one can find two points $\underline{y}, \overline{y} \in \Sigma$ and $m \in \mathbb{N}$ such that

$$y_0^n = \underline{y}_0^n = \overline{y}_0^n$$

and

$$\nu(y_0|y_1^n\overline{y}_{n+1}^{n+m}) - \nu(y_0|y_1^n\underline{y}_{n+1}^{n+m}) \geq \varepsilon > 0.$$

Existence of a bad configuration y implies that no version of the conditional probabilities $\nu(y_0|y_1^\infty)$ (defined ν-a.s.), can be continuous at y, and hence ν cannot be a g-measure for any continuous $g \in \mathscr{G}(\Sigma)$.

4.2 Properties of factors of Markov measures

Despite drawing significant interest in various fields, the problem of finding the necessary and sufficient conditions for factors of Markov measures to be regular is still open. On the other hand, the question under which conditions factors of Markov measures are Markov has been answered completely in 1960's.

4.2.1 Markov factors of Markov measures

Note that the Markovianity of the factor measure might depend on the initial distribution of the underlying Markov chain. The notion of *lumpability* was developed to address this question in a uniform fashion, i.e., independently of the initial distribution.

Let P be a stochastic matrix, indexed by $\mathscr{A} \times \mathscr{A}$ and $\pi : \mathscr{A} \to \mathscr{B}$ a factor map. Now let $\{X_n\}$ be a Markov chain with transition matrix P, then P is called *lumpable* for π if the process $Y_n = \pi(X_n)$ is Markov for all choices of the initial distribution p. The necessary and sufficient conditions for lumpability are quite restrictive, as demonstrated by the following result:

Theorem 4.4. [52] *Suppose P is an irreducible stochastic matrix, then P is lumpable with respect to $\pi : \mathscr{A} \to \mathscr{B}$ if and only if for any $y_1, y_2 \in \mathscr{B}$ we have*

$$\sum_{x_2 \in \pi^{-1}(y_2)} P_{x_1 x_2} = \sum_{x_2 \in \pi^{-1}(y_2)} P_{\tilde{x}_1 x_2} \tag{4.2}$$

for any $x_1, \tilde{x}_1 \in \pi^{-1}(y_1)$. The transition matrix of the factor chain $\{Y_n = \pi(X_n)\}$ is then given by

$$P^{(\pi)}_{y_1 y_2} = \sum_{x_2 \in \pi^{-1}(y_2)} P_{x_1 x_2}.$$

This condition is indeed very restrictive, in part due to a relatively strong requirement that the factor process $\{Y_n\}$ must be Markov for all initial distributions. Instead, one could require Markovianity only for a specific given initial distribution, this is a so-called weak lumpability property. It turns out that this question can be answered algorithmically in polynomial time [42]. Even though weak lumpability is a indeed a weaker condition than lumpability, it is still rather exceptional.

4.2.2 Fully supported Markov chains

Sufficient conditions for a factor measure to be a g-measure are substantially less restrictive than the conditions for (weak-) lumpability. We will discuss some positive and negative results, starting with the very basic positive result for Markov chains with strictly positive transition matrices P. This case was first considered in [45] and comes with an estimate of the continuity rate of the conditional probabilities (g-functions) of the factor measure:

Theorem 4.5 ([45]). *Let ν be a factor of a Markov measure μ with a positive transition matrix P, then ν is a g-measure satisfying*

$$\mathrm{var}_n(g) = \mathscr{O}(c^n),$$

for some $0 < c < 1$.

Let us only mention an intuitive, rough, argument for this result; suppose $P > 0$ is the transition matrix of the Markov process $\{X_n\}$. Suppose $y \in \Sigma$ then, ignoring some technicalities, we can consider the behaviour of μ on $\Omega_y = \pi^{-1}(y)$. In particular, the transition from X_{n+1} to X_n in Ω_y will be given by a positive rectangular matrix. It is well known that, if this matrix is square, then the corresponding map between the distributions of X_{n+1} and X_n is a contraction. For a rectangular matrix we can obtain the same result by using the Hilbert projective metric on the relevant distribution spaces. It is easy to show that this contraction will be uniform in n and therefore the result follows. A version of this argument can also be used to prove the more general results in [16, 99].

4.2.3 A highly non-regular factor measure

A factor measure ν of a Markov measure ν is not necessarily a g-measure. This situation can arise when any version of the conditional probabilities has an essential discontinuity in at least one point of Σ. In more extreme cases the conditional probabilities can be discontinuous everywhere. One such example was discussed by Blackwell [7], Furstenberg [34, Theorem IV.6], Walters [93] and Lorinzi et al [62]. Let $(X_n)_{n\in\mathbb{Z}_+}$ be a Bernoulli process taking values in $\{-1,1\}$ with

$$\mu(X_n = 1) = 1 - \mu(X_n = -1) = p,$$

for $0 < p < 1$, $p \neq \frac{1}{2}$. Then the process $(\tilde{X}_n)_{n\in\mathbb{Z}_+}$ with $\tilde{X}_n = (X_n, X_{n+1})$ is Markov. Consider the factor process $Y_n = \pi(\tilde{X}_n) = \pi(X_n, X_{n+1}) = X_n X_{n+1}$. Note that $\Sigma = \pi(\Omega)$ is the full shift on two symbols $\{-1,1\}$. In this example the conditional probabilities of the factor process $\{Y_n\}$ are discontinuous everywhere. Indeed, it is easy to see that every fibre over $y \in \Sigma$, i.e. $\Omega_y = \pi^{-1}(y) \subset \Omega$, consists of two points

$$x_y^+ = (1, y_0, y_0y_1, y_0y_1y_2, \ldots) \text{ and } x_y^- = (-1, -y_0, -y_0y_1, -y_0y_1y_2, \ldots).$$

We can now explicitly compute the conditional probabilities:

$$\nu(y_0|y_1^n) = \frac{\nu(y_0^n)}{\nu(y_1^n)} = \frac{\mu((x_y^+)_0^{n+1}) + \mu((x_y^-)_0^{n+1})}{\mu((x_y^+)_1^{n+1}) + \mu((x_y^-)_1^{n+1})}.$$

Since μ is the Bernoulli measure, it is easy to see that, with $S_n = \sum_{k=0}^n y_0 y_1 \ldots y_k$, one has

$$\begin{aligned}\nu(y_0^n) &= p^{1+\frac{n+1+S_n}{2}}(1-p)^{\frac{n+1-S_n}{2}} + p^{\frac{n+1-S_n}{2}}(1-p)^{1+\frac{n+1+S_n}{2}} \\ &= p^{\frac{n+1}{2}}(1-p)^{\frac{n+1}{2}}\left[p\left(\frac{p}{1-p}\right)^{\frac{S_n}{2}} + (1-p)\left(\frac{1-p}{p}\right)^{\frac{S_n}{2}}\right].\end{aligned}$$

Similarly,

$$\nu(y_1^n) = p^{\frac{n}{2}}(1-p)^{\frac{n}{2}}\left[p\left(\frac{p}{1-p}\right)^{\frac{\widetilde{S_n}}{2}} + (1-p)\left(\frac{1-p}{p}\right)^{\frac{\widetilde{S_n}}{2}}\right],$$

where $\widetilde{S_n} = \sum_{k=1}^{n} y_1 y_2 \cdots y_k$. Since, $S_n = y_0(1+\widetilde{S_n})$, using $\lambda = p/(1-p)$, one has

$$\nu(y_0 = 1|y_1^n) = \sqrt{p(1-p)}\left(\frac{p\lambda^{\frac{\widetilde{S_n}+1}{2}} + (1-p)\lambda^{-\frac{\widetilde{S_n}+1}{2}}}{p\lambda^{\frac{\widetilde{S_n}}{2}} + (1-p)\lambda^{-\frac{\widetilde{S_n}}{2}}}\right) = \sqrt{p(1-p)}\left(\frac{p\sqrt{\lambda}\lambda^{\widetilde{S_n}} + \frac{(1-p)}{\sqrt{\lambda}}}{p\lambda^{\widetilde{S_n}} + (1-p)}\right)$$

$$=: \frac{a\lambda^{\widetilde{S_n}} + b}{c\lambda^{\widetilde{S_n}} + d},$$

where

$$\frac{a}{c} = \sqrt{p(1-p)\lambda} = p \neq 1-p = \sqrt{\frac{p(1-p)}{\lambda}} = \frac{b}{d},$$

since $p \neq \frac{1}{2}$. Suppose for simplicity that $\lambda > 1$. For any y_1^n, one can choose a continuation z_{n+1}^{5n} such that $\widetilde{S_n} \gg 0$. Equally well, one can choose a continuation w_{n+1}^{5n} such that $\widetilde{S_n} \ll 0$. In the first case,

$$\nu(y_0 = 1|y_1^n z_{n+1}^{5n}) \simeq \frac{a}{c} = p$$

and in the second case,

$$\nu(y_0 = 1|y_1^n w_{n+1}^{5n}) \simeq \frac{b}{d} = 1-p.$$

Therefore, the conditional probabilities $\nu(y_0 = 1|y_1 y_2 \cdots)$ are everywhere *discontinuous*. In some sense this is the worst possible and most irregular behaviour possible. At the same time, when $p = \frac{1}{2}$, ν is a Bernoulli(½,½) product measure on $\{-1,1\}^{\mathbb{Z}_+}$. This example therefore highlights that regularity of the factor measure depends on both the properties of the coding map and the transition probabilities.

4.2.4 Fibre mixing condition

In previous examples we saw that it is important to consider the behaviour of the Markov process $\{X_n\}$, given a realisation of the factor process $\{Y_n\}$. In particular the structure of the fibres $\pi^{-1}(y)$, $y \in \Sigma$, plays a crucial rule (c.f., Blackwell-Furstenberg example above). Theorem 4.5 can also be interpreted in this way.

To see this, recall that positivity of P implies that transitions between any letters, consistent with the fibre, are allowed. The regularity of the factor process is a consequence of the fact that each transition in this fibre, described by a positive rectangular matrix, acts as a contraction on distributions. The most general sufficient condition [99] for factors of Markov measures to be regular has a similar flavour. In particular, in [99] the above idea is generalised from positive matrices to the analogon of *primitive* matrices in the context of fibres; fibre mixing.

Definition 4.6 (Fibre mixing). Let Ω, Σ be subshifts of finite type and $\pi : \Omega \to \Sigma$ is a surjective 1-block factor. We say that π is ***fibre mixing*** if, for all $y \in \Sigma$, for all $x, \tilde{x} \in \Omega_y$ and every $n \in \mathbb{Z}_+$, there exists an $\hat{x} \in \Omega_y$, such that $x_0^n = \hat{x}_0^n$ and $\tilde{x}_{n+m}^{\infty} = \hat{x}_{n+m}^{\infty}$, for some $m \in \mathbb{Z}_+$.

Indeed, fibre mixing is a sufficient condition for the factor measure to be regular.

Theorem 4.7 (Yoo [99]). *Suppose*

(i) $\pi : \Omega \to \Sigma$ is a surjective 1-block factor map between irreducible subshifts of finite type Ω and Σ,

(ii) P is an irreducible stochastic matrix, compatible with the SFT Ω, and μ is the corresponding stationary Markov measure on Ω.

Suppose the factor π is fibre mixing. Then $\nu = \mu \circ \pi^{-1}$ is a g-measure on Σ, for a Hölder continuous g-function.

This result provides the most general set of sufficient conditions for regularity of factors of Markov chains known to date. Other sufficient conditions, e.g., found in [16, 53] imply fibre mixing and are strictly stronger. Let us reiterate that imposing conditions on fibres alone (i.e., the topological conditions on Ω, Σ, and π) is not optimal: the necessary and sufficient conditions must also take P into account, as demonstrated by the Blackwell-Furstenberg example discussed above.

4.3 Continuous measure disintegrations

In this section we will argue that imposing conditions on the behaviour of conditional measures on the fibres provides a more appropriate framework to study properties of the factor measures. As the first step, one has to properly define the conditional measures on the fibres. Fortunately, general results of measure theory provide the necessary tools.

Definition 4.8. We call $\mu_\Sigma = \{\mu_y\}_{y\in\Sigma}$ a family of conditional measures for μ on the fibres Ω_y if μ_y is a Borel probability measure on the fibre Ω_y,

$$\mu_y(\Omega_y) = 1,$$

for all $f \in L^1(\Omega,\mu)$ the map

$$y \to \int_{\Omega_y} f(x)\mu_y(dx)$$

is measurable and

$$\int_\Omega f(x)\mu(dx) = \int_\Sigma \int_{\Omega_y} f(x)\mu_y(dx)\nu(dy).$$

We will also refer to a family of conditional measures $\mu_\Sigma = \{\mu_y\}_{y\in\Sigma}$ for μ on fibres Ω_y as a disintegration of μ with respect to $\pi : \Omega \to \Sigma$.

By a celebrated theorem of von Neumann, for all subshifts Ω, Σ, a given continuous surjection $\pi : \Omega \to \Sigma$ and any Borel measure μ on Ω, there exists a disintegration $\mu_\Sigma = \{\mu_y\}_{y\in\Sigma}$ of μ with respect to π. Moreover, the disintegration is essentially unique in the sense that for any two disintegrations of μ, $\{\mu_y\}$ and $\{\tilde{\mu}_y\}$, we have $\nu(\{y : \tilde{\mu}_y(.) = \mu_y(.)\}) = 1$. We will be interested in continuous measure disintegrations (CMD): a measure disintegration $\mu_\Sigma = \{\mu_y\}$ is called continuous if for every continuous function $f : \Omega \to \mathbb{R}$, the function

$$y \to \int_{\Omega_y} f(x)\mu_y(dx)$$

is continuous. When a disintegration satisfies this constraint we call it a Continuous Measure Disintegration (CMD). Note that any measure μ admits at most one continuous disintegration.

As the conditional measures μ_y are not, in general, translation invariant, we introduce the following notation for cylinder sets in Ω_y:

$${}_n[a_k^m] = \left\{x \in \Omega_y : x_{n+k}^{n+m} = a_k^m\right\},$$

for $a \in \Sigma$ and $n, k, m \in \mathbb{Z}_+$. Using approach similar to that of [87], we will now show that a measure disintegration can be used to find an expression for the conditional probabilities of a factor measure.

Theorem 4.9. *Suppose*

(i) $\pi : \Omega \to \Sigma$ is a surjective 1-block factor map between irreducible subshifts of finite type Ω and Σ,

(ii) P is an irreducible stochastic matrix, compatible with the SFT Ω, and μ is the corresponding stationary Markov measure on Ω.

Suppose $\{\mu_y\}_{y\in\Sigma}$ is a disintegration of μ. Then $\nu=\mu\circ\pi^{-1}$ is consistent with the positive measurable normalized function $\tilde{g}:\Sigma\to(0,1)$, i.e.,

$$\nu(y_0|y_1,y_2,\dots)=\tilde{g}(y)\quad \nu-a.e.,$$

where

$$\tilde{g}(y)=\int_{\Omega_{Ty}}\left[\sum_{a\in\pi^{-1}y_0}\frac{p_a P_{a,x_0}}{p_{x_0}}\right]\mu_{Ty}(dx)=\sum_{a'\in\pi^{-1}y_1}\left[\sum_{a\in\pi^{-1}y_0}\frac{p_a P_{a,a'}}{p_{a'}}\right]\mu_{Ty}({}_0[a']), \tag{4.3}$$

and ${}_0[a']=\{w\in\Omega:w_0=a'\}$, $T:\Sigma\to\Sigma$ is the left shift on Σ, for $y=(y_0,y_1,\dots)$, $Ty=(y_1,y_2,\dots)\in\Sigma$.

Proof. The expression for $\tilde{g}$ originates from the following 'finite-dimensional' equality: denote by $\mathbb{P}$ the joint distribution of $(\{X_n\},\{Y_n\})$, where $\{X_n\}$ is the stationary Markov chain with the transition probability matrix P, and $Y_n=\pi(X_n)$ for all n. Then

$$\begin{aligned}\mathbb{P}(y_0|y_1^n)&=\frac{\mathbb{P}(y_0y_1^n)}{\mathbb{P}(y_1^n)}=\frac{\sum_{x_0^n\in\pi^{-1}y_0^n}\mathbb{P}(x_0x_1^n)}{\mathbb{P}(y_1^n)}=\sum_{x_1^n\in\pi^{-1}y_1^n}\left[\sum_{x_0\in\pi^{-1}y_0}\mathbb{P}(x_0|x_1^n)\right]\frac{\mathbb{P}(x_1^n)}{\mathbb{P}(y_1^n)}\\&=\sum_{x_1^n\in\pi^{-1}y_1^n}\left[\sum_{x_0\in\pi^{-1}y_0}\frac{p_{x_0}P_{x_0,x_1}}{p_{x_1}}\right]\frac{\mathbb{P}(x_1^n)}{\mathbb{P}(y_1^n)}\\&=\sum_{x_1\in\pi^{-1}y_1}\left[\sum_{x_0\in\pi^{-1}y_0}\frac{p_{x_0}P_{x_0,x_1}}{p_{x_1}}\right]\mathbb{P}\big(X_1=x_1|Y_1^n=y_1^n\big).\end{aligned}$$

The Markov measure μ, corresponding to $\{X_n\}$, is a g-measure for the function $g(x)=\frac{p_{x_0}P_{x_0,x_1}}{p_{x_1}}$, where p is the invariant distribution: $pP=p$. We will now show that $\tilde{g}$, given by (4.3), is positive and normalized and finally that $\nu=\mu\circ\pi^{-1}$ is consistent with $\tilde{g}$.

It is easy to check that $\tilde{g}$ is normalized. Indeed,

$$\sum_{y_0\in\mathscr{B}}\tilde{g}(y_0,y_1,y_2,\ldots)=\sum_{y_0\in\mathscr{B}}\left(\sum_{a'\in\pi^{-1}y_1}\left[\sum_{a\in\pi^{-1}y_0}\frac{p_aP_{a,a'}}{p_{a'}}\right]\mu_{Ty}({}_0[a'])\right)$$
$$=\sum_{a'\in\pi^{-1}y_1}\left[\sum_{y_0\in\mathscr{B}}\sum_{a\in\pi^{-1}y_0}\frac{p_aP_{a,a'}}{p_{a'}}\right]\mu_{Ty}({}_0[a'])=\sum_{a'\in\pi^{-1}y_1}\left[\sum_{a\in\mathscr{A}}\frac{p_aP_{a,a'}}{p_{a'}}\right]\mu_{Ty}({}_0[a'])$$
$$=\sum_{a'\in\pi^{-1}y_1}1\cdot\mu_{Ty}({}_0[a'])=1,$$

where we used that since p is the invariant distribution: $pP=p$, or $\sum_{a\in\mathscr{A}}p_aP_{a,a'}=p_{a'}$ for all $a'\in\mathscr{A}$, and hence $\tilde{g}$ is normalized.

The measurability of $\tilde{g}$ follows immediately from the measurability of the measure disintegration $\{\mu_y\}$. The positivity of $\tilde{g}$ is readily checked as well. Let $y=(y_0,y_1,\ldots)\in\Sigma$, then the transition from y_0 to y_1 is allowed in Σ. Since $\pi:\Omega\to\Sigma$ is surjective, it means that there is at least one pair (a,a') such that $\pi(a)=y_0$, $\pi(a')=y_1$ and $P_{aa'}>0$. Since the Markov chain is assumed to be irreducible it follows that the invariant distribution p is strictly positive, and hence

$$\varkappa=\min_{a,a':\,P_{aa'}>0}\frac{p_aP_{a,a'}}{p_{a'}}>0.$$

Therefore,

$$\tilde{g}(y)=\sum_{a'\in\pi^{-1}y_1}\left[\sum_{a\in\pi^{-1}y_0}\frac{p_aP_{a,a'}}{p_{a'}}\right]\mu_{Ty}({}_0[a'])\geq\sum_{a'\in\pi^{-1}y_1}\varkappa\,\mu_{Ty}({}_0[a'])=\varkappa>0. \tag{4.4}$$

Now we are going to show that $\nu=\mu\circ\pi^{-1}$ is consistent with $\tilde{g}$, or, equivalently, that for any continuous $h:\Sigma\to\mathbb{R}$, one has

$$\int_\Sigma h(y)\nu(dy)=\int_\Sigma\sum_{b\in\mathscr{B}:\ by\in\Sigma}h(by)\tilde{g}(by)\nu(dy).$$

Now we show consistency of ν with $\tilde{g}$ by using the fact that μ is a g-measure for

$$g(x)=\frac{p_{x_0}P_{x_0x_1}}{p_{x_1}}.$$

Consider an arbitrary $h\in C(\Sigma)$, and let $\{\mu_y\}$ be a measure disintegration for μ

and π, then

$$
\begin{aligned}
\int_\Sigma h(y)\nu(dy) &= \int_\Omega (h\circ\pi)(x)\mu(dx)\\
&= \int_\Omega \left[\sum_{a\in\mathscr{A}:P_{ax_0}>0} (h\circ\pi)\left(ax_0^\infty\right) g(ax_0^\infty)\right]\mu(dx)\\
&= \int_\Sigma\int_{\Omega_y}\left[\sum_{\substack{b\in\mathscr{B}:\\ b\pi(x)\in\Sigma}}\ \sum_{\substack{a\in\pi^{-1}b\\ P_{a,x_0}>0}} (h\circ\pi)(ax_0^\infty)g(ax_0)\right]\mu_y(dx)\nu(dy)\\
&= \int_\Sigma\int_{\Omega_y}\left[\sum_{\substack{b\in\mathscr{B}:\\ by\in\Sigma}}\ \sum_{\substack{a\in\pi^{-1}b\\ P_{a,x_0}>0}} (h\circ\pi)(ax_0^\infty)g(ax_0)\right]\mu_y(dx)\nu(dy)\\
&= \int_\Sigma\left(\sum_{\substack{b\in\mathscr{B}:\\ by\in\Sigma}} h(by)\int_{\Omega_y}\left[\sum_{a\in\pi^{-1}b} g(ax_0)\right]\mu_y(dx)\right)\nu(dy)\\
&= \int_\Sigma \sum_{\substack{b\in\mathscr{B}:\\ by\in\Sigma}} h(by)\tilde{g}(by)\nu(dy).
\end{aligned}
$$

Thus, ν is consistent with a positive normalized function $\tilde{g}:\Sigma\to(0,1)$. □

Therefore, if for some disintegration $\mu_\Sigma=\{\mu_y\}$, the function $\tilde{g}$, as defined in equation (4.3), is continuous, then ν is a g-measure. There are two obvious sets of sufficient conditions for continuity of $\tilde{g}$.

Corollary 4.10. *Under conditions of Theorem 4.9, the measure ν is a g-measure if there exists a disintegration $\mu_\Sigma=\{\mu_y\}$ such that $\tilde{g}(y)$, given by (4.3), is a continuous function on Σ.*

In particular, $\tilde{g}$ is continuous if one of the following conditions holds:

1) *matrix $Q=(Q_{a,a'})$ with $Q_{a,a'}=\frac{p_a P_{a,a'}}{p_{a'}}$ satisfies:*

$$\sum_{a\in\pi^{-1}(b)} Q_{aa'} = \sum_{a\in\pi^{-1}(b)} Q_{aa''}, \tag{4.5}$$

for any $b\in\mathscr{B}$ and any $a',a''\in\pi^{-1}(b')$, where $b'\in\mathscr{B}$.

2) *μ admits a* continuous measure disintegration *on the fibres* $\{\Omega_y = \pi^{-1}(y) : y \in \Sigma\}$*;*

Proof. If $\tilde{g}$ is indeed a continuous function, then ν is a g-measure by definition. We only have to show that conditions (1) and (2) imply continuity of $\tilde{g}$. Let us start with the first condition (4.5). Since

$$\tilde{g}(y) = \sum_{a' \in \pi^{-1} y_1} \left[\sum_{a \in \pi^{-1} y_0} \frac{p_a P_{a,a'}}{p_{a'}} \right] \mu_{Ty}({}_0[a']) = \sum_{a' \in \pi^{-1} y_1} \left[\sum_{a \in \pi^{-1} y_0} Q_{a,a'} \right] \mu_{Ty}({}_0[a']). \tag{4.6}$$

Condition (4.5) implies that for all $a' \in \pi^{-1} y_1$, the sums in the square-brackets have the same value. Let us denote the common value by S_{y_0,y_1}. Therefore,

$$\tilde{g}(y) = S_{y_0,y_1} \sum_{a' \in \pi^{-1} y_1} \mu_{Ty}({}_0[a']) = S_{y_0,y_1}$$

since μ_{Ty} is a Borel probability measure on the fibre $\Omega_{Ty} = \Omega_{(y_1,y_2,\ldots)}$.

Let us now consider the second assumption: suppose μ admits a continuous measure disintegration on the fibres $\{\Omega_y\}$, then for any $f \in C(\Omega)$, $y \to \mu_y(f) = \int f \, d\mu_y$ is continuous. In particular, since for any $b \in \mathscr{B}$, the function

$$G^b(x) = \sum_{a \in \pi^{-1} b} \frac{p_a P_{a,x_0}}{p_{x_0}},$$

is continuous on Ω as a function of x, we conclude that $\tilde{g}$ is continuous and hence ν is a g-measure. □

Remark 4.11. The first condition is simply a standard (strong) lumpability condition for the time-reversal of the original Markov chain. Note that lumpability conditions for the chain and its reversal are not equivalent in general. In this instance, however, we only consider the stationary chains, and hence, one should compare the weak lumpability conditions for the chain and its time reversal. It is somewhat surprising that we finish with the strong lumpability condition for the reversed chain, and not the weak lumpability condition.

Remark 4.12. The second sufficient condition requires existence of a continuous disintegration for μ: i.e., continuity of the map

$$y \mapsto \int_{\Sigma_y} f(x) \mu_y(dx) \tag{4.7}$$

for every continuous f on Ω. However, we only need continuity of integrals of rather 'simple' functions of the form

$$G^b(x) = \sum_{a\in\pi^{-1}b} \frac{p_a P_{a,x_0}}{p_{x_0}}, \quad b \in \mathscr{B}. \tag{4.8}$$

Thus the question is what is the relation between the requirements that there exists of a continuous measure disintegration for μ, and that there exists a disintegration such that for all $b \in \mathscr{B}$, the map $\Sigma \ni y \mapsto \int_{\Omega_y} G^b(x)\mu_y(dx) \in \mathbb{R}_+$ is continuous. The first condition of Corollary 4.10 then reads: for all $b \in \mathscr{B}$, $G^b(x) \equiv \text{const}$. In the last section we present an example of an irreducible Markov chain such that $G^b(x) \equiv \text{const}$, but μ does not admit a continuous disintegration. However, in a 'non-trivial' case $G^b(x) \not\equiv \text{const}$, we believe the difference between requiring continuity $y \to \int_{\mu_y} f(x)\mu_y(dx)$ for all continuous f, versus, only for simple functions depending only on the first coordinate $f(x) = f(x_0)$ is not substantial. The main reason is that we believe that the general hypothesis on regularity of factor measures proposed in Statistical Mechanics [81] applies to Markov chains as well.

We will proceed by investigating existence of a continuous measure disintegration using methods developed in thermodynamic formalism for fibred systems.

4.4 Thermodynamic formalism for fibred systems

There has been a lot of work done on thermodynamic formalism, equilibrium states and variational principles for fibred systems: starting from the celebrated work of Ledrappier and Walters [60] on relativized variational principles to the relatively comprehensive theory of Denker and Gordin [19], as well as extensive work on random subshifts of finite type [8]. We apply the methods developed in this field to provide sufficient conditions for the existence of continuous fibre disintegrations of Markov measures. Moreover, we apply, for the first time in a dynamical setting, a method originating in Mathematical Statistics, developed by Tjur [85, 86] in the 1970's, which provides a *constructive* approach to the construction of a continuous measure disintegration.

4.4.1 Fibres as non-homogeneous subshifts of finite type

The fibres of the factor map $\pi : \Omega \to \Sigma$ are not translation invariant. However, they admit a nice topological description: namely, as non-homogeneous or random subshifts of finite type.

Definition 4.13. Suppose $\mathbb{S} = \{S_n\}_{n\geq 0}$ is a collection of non-empty finite sets of bounded size. Let $\Omega^{\mathbb{S}} = \prod_{n\in\mathbb{Z}_+} S_n$ be the corresponding product space. Assume also that we are given a sequence of 0/1-matrices $\mathbb{M} = (M_n)_{n\in\mathbb{Z}_+}$, with M_n indexed by $S_n \times S_{n+1}$, such that for each n, M_n is *reduced*: it has no columns or rows with only 0 entries.

Then the set

$$\Omega_{\mathbb{M}} = \left\{x \in \Omega^{\mathbb{S}} : M_n(x_n, x_{n+1}) = 1, \text{ for all } n \in \mathbb{Z}_+\right\},$$

is called a non-homogeneous (random) subshift of finite type corresponding to the sequence $\mathbb{M}$.

It is easy to see that if $\Omega = \Omega_M$ is a SFT, $\pi : \Omega \to \Sigma$ is a 1-block factor map, then for any $y \in \Sigma$, the fibre Ω_y is a non-homogeneous SFT: indeed, let $S_n^y = \pi^{-1}(y_n)$, and put $M_n^y(x_n, x_{n+1}) = 0 \Leftrightarrow M(x_n, x_{n+1}) = 0$ for all $n \in \mathbb{Z}_+$ and $x_n \in S_n, x_{n+1} \in S_{n+1}$. In other words, $\Omega_y = \Omega_{\mathbb{M}^y}$, for $\mathbb{M}^y = \{M_n^y\}$, where M_n^y is a submatrix of M corresponding to rows $\pi^{-1}(y_n)$ and columns $\pi^{-1}(y_{n+1})$.
We recall the notion of a *transitive non-homogeneous SFT* introduced by Fan and Pollicott [27]:

Definition 4.14. A non-homogeneous SFT $\Omega_{\mathbb{M}}$, corresponding to a sequence of matrices $\mathbb{M} = (M_n)_{n\in\mathbb{Z}_+}$, is called *transitive*, i.e., there exists m such that

$$\prod_{j=n}^{n+m} M_j > 0$$

for all $n \geq 0$.

It turns out that the fibre mixing condition of Yoo is equivalent to the requirement that each fibre is a transitive non-homogeneous SFT. Moreover, the constant m can be chosen the same for all fibres.

Lemma 4.15. *The following conditions are equivalent:*

1. *the surjective 1-block factor map $\pi : \Omega \to \Sigma$ between irreducible SFTs, Ω and Σ, is fibre mixing;*

2. *for each $y \in \Sigma$, the fibre Ω_y is a transitive non-homogenous SFT.*

Proof. (1) $\Rightarrow$ (2): Assume that $\pi : \Omega \to \Sigma$ is fibre mixing, but there exists a $y \in \Sigma$ such that the corresponding sequence of matrices $\{M_n^{(y)}\}_{n\in\mathbb{Z}_+}$ is not transitive, i.e., there exists $n \in \mathbb{Z}_+$ such that for all $m \geq 0$, the matrix of size $|S_n| \times |S_{n+m+1}|$

$$M_{n,m}^{(y)} := \prod_{j=n}^{n+m} M_j^{(y)}$$

is not positive, i.e., some entries are equal to 0.

In fact there exists a specific row that contains a zero for all m. Indeed, if $M_{n,m}^{(y)} = \prod_{j=n}^{n+m} M_j^{(y)}$ has a zero for some $m \geq 1$, then $M_{n,m'}^{(y)} = \prod_{j=n}^{n+m'} M_n^{(y)}$ has a zero in exactly the same row for all $m' < m$. Similarly, if a certain row in $M_{n,m}^{(y)}$ is positive, it will remain positive in $M_{n,m'}^{(y)}$ for all $m' > m$, since $\{M_n^{(y)}\}$ is reduced: no column is identically zero.

Choose $x \in \Omega_y$ such that the row corresponding to x_n in $M_{n,m}^{(y)}$, contains 0 for all $m \geq 0$. We can do so by continuing x_n to the left and to the right using matrices $\{M_k^{(y)}\}$, to obtain the required point. It is indeed possible since the sequence $\{M_k^{(y)}\}$ is reduced. Similarly, for any $m \in \mathbb{N}$, choose $x^{(m)} \in \Omega_y$ such that $x_{n+m+1}^{(m)}$-column has a 0 in the x_n-row in $M_{n+m}^{(y)}$

$$\left(\prod_{j=n}^{n+m} M_j^{(y)}\right)_{x_n, x_{n+m+1}^{(m)}} = M_{n,m}^{(y)}(x_n, x_{n+m+1}^{(m)}) = 0.$$

Let $\bar{x} \in \Omega_y$ be some limit point of the sequence $\{x^{(m)}\}_{m\geq 0}$: $\bar{x} = \lim_k x^{(m_k)}$ (note that the fibre Ω_y is compact). We claim that points x and $\bar{x}$ cannot be 'connected' within the fibre Ω_y.

This is almost immediate, suppose that $\tilde{x} \in \Omega_y$ exists, such that $\tilde{x} = x_0^n \tilde{x}_{n+1}^{n+m} \bar{x}_{n+m+1}^{\infty}$ then this contradicts the zero entry in the matrix assumption as:

$$M_i^{(y)}(a, \hat{x}_{n+1}) \left(\prod_{j=n+1}^{n+m-1} M_j^{(y)}(\hat{x}_j, \hat{x}_{j+1})\right) M_{n+m}^{(y)}(\hat{x}_{n+m}, \bar{x}_{n+m+1}) > 0$$

It follows that if the factor map is fibre mixing, it must be primitive.

(2) $\Rightarrow$ (1): Conversely, let $x, \bar{x} \in \Omega_y$ and assume primitivity then, for any $i \in \mathbb{Z}$, we have $\prod_{n=i}^{i+m(i)} M_n^{(y)} > 0$ and therefore there exists an $\hat{x} \in \Omega_y$ such that

$$M_i^{(y)}(\hat{x}_i, \hat{x}_{i+1}) \cdots M_{i+m(i)}^{(y)}(\hat{x}_{i+m(i)}, \hat{x}_{i+m(i)+1}) > 0,$$

with $\hat{x}_i = x_i$, $\hat{x}_{i+m(i)+1} = \bar{x}_{i+m(i)+1}$. This means that $x_0^i \hat{x}_{i+1}^{i+m} \bar{x}_{i+m+1}^{\infty} \in \Omega_y$. Therefore Ω_y is fibre mixing. $\square$

Lemma 4.16. *Suppose $\pi : \Omega \to \Sigma$ is fibre mixing, and hence, for every $y \in \Sigma$, the sequence $\{M_n^{(y)}\}_{n\in\mathbb{Z}_+}$ is transitive: there exists m_y such that*

$$\prod_{j=n}^{n+m_y} M_j^{(y)} > 0 \tag{4.9}$$

for all n. Then $\sup_y m_y < \infty$, or, in other words, there exists one $m \in \mathbb{N}$, satisfying (4.9) for all y and n.

Proof. First we will show that any index of primitivity $m(n)$ for the fibre Ω_y is bounded from above by the index of primitivity $m(0)$ corresponding to $\Omega_{T^n y}$. Therefore it suffices to show that $m(0)$ is uniformly bounded in y.

Let $y \in \Sigma$ and let $m(0)$ be the index of primitivity corresponding to $\Omega_{T^n y}$. First note that, as any $x \in \Omega_y$ results in $T^m x \in \Omega_{T^m y}$, we have $\{x_{n+m} \in \mathscr{A} : x \in \pi^{-1}(y)\} \subset \{x_n \in \mathscr{A} : x \in \pi^{-1}(y)\}$, for all $n, m \in \mathbb{Z}_+$.

Now let $a \in \{x_n \in \mathscr{A} : x \in \pi^{-1}(y)\}$, $b \in \{x_{n+m(0)} \in \mathscr{A} : x \in \pi^{-1}(y)\}$, then

$$a \in \{x_0 : x \in \pi^{-1}(T^n y)\}, b \in \{x_{m(0)} : x \in \pi^{-1}(y)\}.$$

Therefore a word $\tilde{x}_0^{m(0)}$ exists such that $\tilde{x}_0 = a$, $\tilde{x}_{m(0)} = b$ and $\pi(\tilde{x}_0^{m(0)}) = y_n^{n+m(0)}$. It follows that the index of primitivity $m(0)$ for $\Omega_{T^n y}$ is an upper bound for the index of primitivity $m(n)$ for Ω_y.

Now assume for all $y \in Y$ that $(M_n)_{n\in\mathbb{Z}_+}$ is primitive and that the index of primitivity $m(0)$ is unbounded. We will show that these assumptions lead to a contradiction. Let $(y^{(i)})_{i\in\mathbb{Z}_+}$ be a sequence such that $y^{(i)} \in \{y : m(0) > i\}$. In fact we can choose this sequence, by compactness, in such a way that it converges, call the limit y.

Assume that $a^{(i)}, b^{(i)} \in \mathscr{A}$ are such that $\left(\prod_{n=0}^{i-1} M_n^{y^{(i)}}\right)_{a^{(i)},b^{(i)}} = 0$, then there exists an $\tilde{x}^{(i)} \in \Omega_{y^{(i)}}$, with $\tilde{x}_n^{(i)} = b$. Choose a converging subsequence of $\{\tilde{x}^{(i)}\}$, it must then be true that $(\prod_{n=0}^{i-1} M_n^y)_{a,\tilde{x}_i} = 0$, for any $i > 0$. It follows that if the matrix sequences is primitive, the bound on $m(n)$ is uniform in y and n. □

4.4.2 Non-homogeneous equilibrium states

Now we are ready to apply methods of thermodynamic formalism to construct directly the conditional measures on the fibres. The first and the most direct method is to use the approach of [87] which relies on the fundamental results of Fan and Pollicott [27] for transitive non-homogenous subshifts of finite type. Since the proof in the Markov case considered in the present paper is almost identical to (and, in fact, simpler than) the proof in the case of fully supported g-measures in [87], we will only sketch the necessary steps. We start by introducing the *averaging operators* acting on spaces of continuous functions on fibres Ω_y:

$$P_n^y f(x) = \sum_{\substack{a_0^n \in \pi^{-1} y_0^n: \\ a_0^n x_{n+1}^{+\infty} \in \Omega_y}} G_n^y(a_0 \dots a_n x_{n+1} \dots) f(a_0 \dots a_n x_{n+1} \dots),$$

where $G_n^y(x)$ is defined on Ω_y by

$$G_n^y(x) = \frac{Q(x_0, x_1) \dots Q(x_n, x_{n+1})}{\sum_{a_0^n:\ a_0^n x_{n+1}^{+\infty} \in \Omega_y} Q(a_0, a_1) \dots Q(a_n, x_{n+1})}, \quad Q(a, a') = \frac{p_a P_{a,a'}}{p_{a'}},\ a, a' \in \mathscr{A}. \tag{4.10}$$

Note that $\sum_{a_0^n:\ a_0^n x_{n+1}^{+\infty} \in \Omega_y} G_n^y(a_0 \dots a_n x_{n+1} \dots) = 1$ for all $x \in \Omega_y$, and hence $P_n^y \mathbf{1} = \mathbf{1}$. A probability measure μ^y on Ω_y is called a non-homogeneous equilibrium state associated to $\mathbb{G}^y = \{G_n^y\}$ if

$$\int_{\Omega_y} P_n^y f(x) \mu^y(dx) = \int_{\Omega_y} f(x) \mu^y(dx)$$

for all $f \in C(\Omega_y)$. Next we will show that the equilibrium states μ^y form a continuous measure disintegration, for now, we use a superscript distinguish from the notation for a disintegration.
The sequence of $\mathbb{G}^y = \{G_n^y\}$ given by (4.10), can easily be seen to satisfy the conditions of Theorem 1 of [27], and we immediately get the following corollary:

Corollary 4.17. *Suppose Ω, Σ are irreducible SFT's, and a 1-block surjective factor map $\pi : \Omega \to \Sigma$ is such that Ω_y is a transitive non-homogenous SFT for every $y \in \Sigma$. Then for each $y \in \Sigma$ there exists a unique non-homogeneous equilibrium state μ^y associated to $\mathbb{G}^y = \{G_n^y\}$. Moreover,*

$$P_n^y f(x) \rightrightarrows \int_{\Omega_y} f(x) \mu^y(dx) \tag{4.11}$$

uniformly on Ω_y, as $n \to \infty$.

Furthermore, the convergence in (4.11) turns out to be uniform in y as well. Using this rather strong property,we also immediately get the following corollary of Lemma 3.4 and 3.5 [87]:

Proposition 4.18. *Under the above conditions, the family $\{\mu^y\}$ of non-homogeneous equilibrium states on Ω_y associated to $\mathbb{G}^y$ forms a disintegration of μ, i.e., for every continuous function f one has*

$$\int_\Omega f(x)\mu(dx) = \int_\Sigma \int_{\Omega_y} f(x)\mu^y(dx)\nu(dy).$$

Moreover, the family $\{\mu^y\}$ is in fact continuous: for every continuous f,

$$y \mapsto \int_{\Omega_y} f(x)\mu^y(dx)$$

is a continuous function on ν.

Therefore, by Proposition 4.10, we conclude that ν is a g-measure.

Remark 4.19. The above method can be summarized as follows. The conditional measures on fibres are equilibrium states for the same potential as the starting measure μ. One needs to establish uniqueness of equilibrium states on the fibres first, and then prove continuity of the resulting family. In this particular case, one obtains continuity from the double uniform convergence of the averaging (transfer) operators. In the following section, we are going to show that uniqueness on each fibre is in fact sufficient, and one obtains continuity effectively for free.

4.4.3 Constructive approach to conditioning on fibres

General results on the existence of measure disintegrations are not constructive. To alleviate this problem, Tjur [85, 86] proposed a more direct method: the conditional measures μ_y on fibres can be obtained directly, in a unique way, as a limit of measures conditioned on sets with positive measure around y.

Suppose $y \in \Sigma$ and let D_y be the set of pairs (V, B), where V is an open neighbourhood of y and B is a subset of V with positive measure:

$$D_y = \{(B, V) : V \text{ open}, \ y \in V, \ B \subset V, \ \nu(B) > 0\}.$$

Now equip the collection D_y with a partial order given by $(V_1, B_1) \succcurlyeq (V_2, B_2)$, if $V_1 \subseteq V_2$. This partial order is upwards directed, as, for any $(V_1, B_1), (V_2, B_2) \in D_{y_0}$, there exists an element $(V_3, B_3) \in D_{y_0}$ such that $(V_3, B_3) \succcurlyeq (V_1, B_1)$ and $(V_3, B_3) \succcurlyeq (V_2, B_2)$. For each $(V, B) \in D_y$ we define a conditional measure μ^B:

$$\mu^B(\cdot) = \mu(\cdot \mid \pi^{-1}B).$$

Since D_y is upwards directed, the collection of conditional measures

$$\mathcal{N}_y = \left\{\mu^B(\cdot) : (V,B) \in D_y\right\},$$

is a ***net***, or a ***generalized sequence***, in the space of probability measures on Ω. We can now define the limit or accumulation points of this net as follows:

Definition 4.20. We call a measure $\tilde{\mu}$ on Ω an accumulation point of the net $\mathcal{N}_y$ if there exists a sequence $\{(V_n, B_n)\}_{n\geq 1} \subset D_y$, $n \geq 1$, such that

$$\mu^{B_n} = \mu(\cdot|\pi^{-1}B_n) \to \tilde{\mu}, \text{ as } n \to \infty,$$

weakly. Denote the set of all possible accumulation points by $\overline{\mathfrak{M}}_y$.

By standard compactness arguments we immediately conclude that $\overline{\mathfrak{M}}_y \neq \varnothing$, and for each $\lambda_y \in \overline{\mathfrak{M}}_y$, one has $\lambda_y(\Omega_y) = 1$.

Definition 4.21. The point $y \in \Sigma$ is called a Tjur point if $\overline{\mathfrak{M}}_y$ is a singleton, i.e., the net $\mathcal{N}_y$ has a limit, which we denote by μ^y.

Two basic theorems by Tjur provide sufficient conditions for the existence of continuous measure disintegrations. The first theorem states that, when conditional measures μ^y are defined ν-almost everywhere, they form a measure disintegration.

Theorem 4.22. [86, Theorem 5.1] *Suppose $\pi : \Omega \to \Sigma$ is a continuous surjection, as defined above, and $\nu = \mu \circ \pi^{-1}$. Assume, furthermore, that ν-almost all $y \in \Sigma$ are Tjur points. Then, for any $f \in L^1(\Omega, \mu)$, f is μ^y-integrable for ν-almost all y, and the function $y \mapsto \int f\, d\mu^y$ is ν-integrable and*

$$\int_\Omega f(x)\mu(dx) = \int_\Sigma \Big[\int_{\Omega_y} f(x)\mu^y(dx)\Big]\nu(dy).$$

The second theorem provides the desired continuity for the map $y \mapsto \mu^y$.

Theorem 4.23. [86, Theorem 4.1] *Denote by Σ_0 the set of all Tjur points in Σ. Then the map*

$$y \mapsto \mu^y$$

is continuous on Σ_0.

As a corollary, we immediately conclude

Corollary 4.24. *If $\nu = \mu \circ \pi^{-1}$ and for all $y \in \Sigma$ we have $|\overline{\mathfrak{M}}_y| = 1$, i.e., all points are Tjur, then μ admits a continuous disintegration, and hence ν is a g-measure.*

4.4.4 Gibbs measures on fibres

The main result of the previous section states that existence of a unique limit of the sequence of conditional measures $\mu(\cdot|\pi^{-1}B)$, $B \searrow y$, for all $y \in \Sigma$, is sufficient for the regularity of ν. However, this condition is not easy to validate directly. The general principle for renormalisation of Gibbs random fields formulated by van Enter, Fernandez, and Sokal, in the seminal paper [81], states that the conditional measures must be Gibbs for the *original* potential. Since the original measure μ is Markov, i.e., Gibbs for a two–point interaction, we have to study the Gibbs-Markov measures on the fibres. In the setting of this paper that means that the conditional measures are Markov. In fact, we have already seen this indirectly in the Fan-Pollicott construction on non-homogeneous equilibrium states on fibres. In this section we define Gibbs-Markov measures on fibres and show the absence of phase transitions, i.e., prove uniqueness on each fibre. In the following section we show that any limit measure in $\overline{\mathfrak{M}}_y$ must be Markov and, given that there is only one Markov measure on each fibre, we conclude that $|\overline{\mathfrak{M}}_y| = 1$ for all $y \in \Sigma$. Suppose Ω, Σ and $\pi : \Omega \to \Sigma$ are defined as above and μ is a stationary Markov measure with Ω as its support.

Definition 4.25. A Borel probability measure ρ on Ω_y is called Gibbs-Markov for the (irreducible) stochastic matrix P, if for all n and ρ-almost all $x = (x_0, x_1, \ldots) \in \Omega_y$

$$\rho(x_0^n|x_{n+1}^\infty) = \frac{Q(x_0,x_1)\ldots Q(x_n,x_{n+1})}{\sum\limits_{a_0^n:\ a_0^n x_{n+1}^{+\infty}\in\Omega_y} Q(a_0,a_1)\ldots Q(a_n,x_{n+1})}, \quad Q(a,a') = \frac{p_a P_{a,a'}}{p_{a'}},\ a,a' \in \mathscr{A}. \tag{4.12}$$

If we define the *interaction* $\Phi = \{\Phi_\Lambda(\cdot)\}$ – a collection of functions indexed by finite subsets Λ of $\mathbb{Z}_+$ –, by

$$\Phi_\Lambda(x) = \begin{cases} -\log Q(x_k, x_{k+1}), & \text{if } \Lambda = \{k, k+1\}, \\ 0, & \text{otherwise,} \end{cases}$$

then the expression (4.12) can be rewritten in a more traditional Gibbsian form:

$$\rho(x_0^n|x_{n+1}^\infty) = \frac{1}{Z_{[0,n]}(x_{n+1}^\infty)} \exp\left(-H_{[0,n]}(x)\right), \quad H_{[0,n]}(x) = \sum_{\Lambda\cap[0,n]\neq\varnothing} \Phi_\Lambda(x), \tag{4.13}$$

and $Z_{[0,n]}(x_{n+1}^\infty) = \sum_{a_0^n : a_0^n x_{n+1}^\infty \in \Sigma_y} \exp\left(-H_{[0,n]}(a_0^n x_{n+1}^\infty)\right)$ is the corresponding partition function. We denote by $\mathscr{G}_{\Omega_y}(\Phi)$ the set of all Gibbs probability measures for the interaction Φ. Since Ω_y is a non-homogeneous subshift of finite type, i.e., the *lattice system* as described in [77], the standard theory of Gibbs states implies

$\mathscr{G}_{\Omega_y}(\Phi)$ is a non-empty convex set of measures. Moreover, the extremal measures are tail-trivial. Thus two extremal measures in $\mathscr{G}_{\Omega_y}(\Phi)$ are either singular or equal.

To prove uniqueness of Gibbs-Markov measures on fibres we will use the classical boundary uniformity condition [31, 37]. Denote the right-hand side of (4.13) by $\gamma_{[0,n]}(x_0^n|x_{n+1}^\infty)$, and for a continuous function f, let

$$(\gamma_{[0,n]}f)(x) = \sum_{a_0^n : a_0^n x_{n+1}^\infty \in \Omega_y} f(a_0^n x_{n+1}^\infty)\gamma_{[0,n]}(a_0^n | x_{n+1}^\infty).$$

Then $\rho \in \mathscr{G}_{\Omega_y}(\Phi)$ if and only if for every continuous f on Ω_y the Dobrushin-Lanford-Ruelle equations are valid for every $n \geq 0$

$$\int_{\Omega_y} f(x)\rho(dx) = \int_{\Omega_y} (\gamma_{[0,n]}f)(x)\rho(dx).$$

Given the fact that the non-homogeneous subshift of finite type Ω_y is transitive, Φ is a *finite-range* potential, it is easy to check that the family of probability kernels $\gamma_{[0,n]}(\cdot|x_{n+1}^\infty)$ satisfies the so-called boundary uniformity condition: there exists $c > 0$ such that for any $a_0^m \in \pi^{-1}(y_0^m)$, and every $x, \tilde{x} \in \Omega_y$, for all sufficiently large n, one has

$$\big(\gamma_{[0,n]}\mathbb{1}_{[a_0^m]}\big)(x) \geq c\big(\gamma_{[0,n]}\mathbb{1}_{[a_0^m]}\big)(\tilde{x}). \tag{4.14}$$

Applying standard arguments for uniqueness of Gibbs measures under the boundary uniformity condition [31] one gets:

Lemma 4.26. *Suppose Ω_y is a transitive non-homogeneous subshift of finite type, and the potential Φ is such that the family of probability kernels $\{\gamma_{[0,n]}\}$ satisfies* (4.14). *Then there exists a unique Gibbs measure for Φ on Ω_y, i.e., $|\mathscr{G}_{\Omega_y}(\Phi)| = 1$.*

Proof. Consider two arbitrary extremal Gibbs measures $\rho, \tilde{\rho} \in \mathscr{G}_{\Omega_y}(\Phi)$. By integrating (4.14) first with respect to $\rho(dx)$, and then with respect to $\tilde{\rho}(d\tilde{x})$, one concludes that

$$\begin{aligned}\rho([a_0^m]) &= \iint \big(\gamma_{[0,n]}\mathbb{1}_{[a_0^m]}\big)(x)\rho(dx)\tilde{\rho}(d\tilde{x}) \\ &\geq \iint c\big(\gamma_{[0,n]}\mathbb{1}_{[a_0^m]}\big)(\tilde{x})\tilde{\rho}(d\tilde{x})\rho(dx) = c\tilde{\rho}([a_0^m]),\end{aligned}$$

and hence, $\rho \geq c\tilde{\rho}$. Similarly, $\tilde{\rho} \geq c\rho$. Since the distinct extremal measures in $\mathscr{G}_{\Omega_y}(\Phi)$ must be singular, we conclude that $\rho = \tilde{\rho}$. Hence, $\mathscr{G}_{\Omega_y}(\Phi)$ has a unique extremal element, and therefore $\mathscr{G}_{\Omega_y}(\Phi)$ is a singleton. □

4.4.5 Conditional measures are Markov

We are now going to show that any limit point of the net $\mathscr{N}_y$ must be a Gibbs-Markov measure on Ω_y, i.e., $\overline{\mathfrak{M}} \subseteq \mathscr{G}_{\Omega_y}(\Phi)$. Since we have already shown that $|\mathscr{G}_{\Omega_y}(\Phi)| = 1$ for all $y \in \Sigma$, we conclude that $|\overline{\mathfrak{M}}_y| = 1$ for all y, i.e., all points in ν are Tjur, and hence the ν is a g-measure.

Proposition 4.27. *Let μ be a stationary irreducible Markov measure for the interaction $\Phi = \{\phi_{i,i+1}\}$ and let $\overline{\mathfrak{M}}_y$ be defined as above. For all $y \in \Sigma$, one has $\overline{\mathfrak{M}}_y \subseteq \mathscr{G}_{\Omega_y}(\Phi)$.*

Proof. Suppose $\rho \in \overline{\mathfrak{M}}_y$:

$$\rho = \lim_{m\to\infty} \mu(\cdot|\pi^{-1}B_m),$$

for some sequence $(V_m, B_m) \in D_y$. Without loss of generality we may assume $V_m = [y_0^m]$. Moreover, since any measurable set B_m can be approximated arbitrarily well by cylinders, it is sufficient to consider only limit points of $\{\mu(\cdot|\pi^{-1}[y_0^m z_{m+1}^{m+n}])\}_{m,n\geq 0}$, provided $\nu([y_0^m z_{m+1}^{m+n}]) > 0$. Denote the set of all limits points of such conditional measures by $\mathfrak{M}_y$. We first prove the following lemma:

Lemma 4.28. *For all $y \in \Sigma$, any limit point in $\overline{\mathfrak{M}}_y$ is a linear combination of the limit points in $\mathfrak{M}_y$.*

Proof. Let $y \in \Sigma$, $\lambda \in \overline{\mathfrak{M}}_y$ and $(B_m, V_m) \in D_y$ is a sequence such that $\mu^{B_m} \to \lambda$. It suffices to show that each μ^{B_m} is a limit point of linear combinations in $\mathfrak{M}_y$. For any $m, n \in \mathbb{N}$ we can define a collection $C_{n,l}^{(m)}$, of disjoint cylinder sets in Σ, indexed by a finite set L_m, such that $\nu(B_m \Delta \cup_{l\in L_m} C_{n,l}^{(m)}) \to 0$, as $n \to \infty$.

Given any $A \in \mathscr{F}(\Omega)$ we have that $\left|\mu^{B_m}(A) - \mu^{\cup_{l\in L_m} C_{n,l}^{(m)}}(A)\right| \to 0$ as $n \to \infty$. Also note that

$$\mu^{\cup_{l\in L_m} C_{n,l}^{(m)}}(A) = \frac{\mu(A \cap \pi^{-1} \cup_{l\in L_m} C_{n,l}^{(m)})}{\mu(\pi^{-1} \cup_{l\in L_m} C_{n,l}^{(m)})} = \sum_{l\in L_m} \mu(A|\pi^{-1}C_{n,l}^{(m)}) \frac{\mu(\pi^{-1}C_{n,l}^{(m)})}{\sum_{\tilde{l}\in L_m} \mu(\pi^{-1}C_{n,\tilde{l}}^{(m)})}.$$

In other words, each μ^{B_m} is a limit point of linear combinations of measures of the form $\mu^{C_{n,l}^{(m)}}$. Therefore λ is a limit point of linear combinations of measures in $\mathfrak{M}_y$. □

Hence, if we are able to prove that $\mathfrak{M}_y \subseteq \mathscr{G}_{\Omega_y}(\Phi)$, then we are able to conclude that $\overline{\mathfrak{M}}_y \subseteq \mathscr{G}_{\Omega_y}(\Phi)$ as well. Suppose

$$\rho = \lim_m \rho_m, \quad \rho_m = \mu(\cdot|\pi^{-1}[y_0^m z_{(m)}]),$$

where $z_{(m)}$ is some finite word in alphabet $\mathscr{B}$, such that $\nu([y_0^m z_{(m)}]) > 0$ for all m. We are going to show that ρ is a Markov measure on Ω_y, in other words

$$\rho(x_0^n | x_{n+1}^{n+\ell}) = \rho(x_0^n | x_{n+1}) \tag{4.15}$$

for all $n \geq 0$, $\ell \geq 1$, and $x \in \Omega_y$. Since ρ is the weak limit of ρ_m's, it is thus sufficient to establish (4.15) for ρ_m for all sufficiently large m.

Consider $x \in \Omega_y$, fix $n \geq 0$, $\ell \geq 1$. Choose m_0 such that for all $m \geq m_0$, K_m – the length of the word $y_0^m z_{(m)}$, satisfies $K_m > n + \ell$; e.g., $m_0 = n + \ell + 1$ suffices. Then

$$\begin{aligned}
\rho_m(x_0^n | x_{n+1}^{n+\ell}) &= \frac{\rho_m([x_0^n, x_{n+1}^{n+\ell}])}{\rho_m([x_{n+1}^{n+\ell}])} = \frac{\mu([x_0^n, x_{n+1}^{n+\ell}] \cap \pi^{-1}[y_0^m z_{(m)}])}{\mu([x_{n+1}^{n+\ell}] \cap \pi^{-1}[y_0^m z_{(m)}])} \\
&= \frac{\sum\limits_{a_0^{K_m} \in \pi^{-1}[y_0^m z_{(m)}]:\ a_0^{n+\ell} = x_0^{n+\ell}} \mu(a_0^{K_m})}{\sum\limits_{b_0^{K_m} \in \pi^{-1}[y_0^m z_{(m)}]:\ b_{n+1}^{n+\ell} = x_{n+1}^{n+\ell}} \mu(b_0^{K_m})} \\
&= \frac{\sum\limits_{a_0^{K_m} \in \pi^{-1}[y_0^m z_{(m)}]:\ a_0^{n+\ell} = x_0^{n+\ell}} \mu(a_0^n | a_{n+1}^{K_m}) \mu(a_{n+1}^{K_m})}{\sum\limits_{b_0^{K_m} \in \pi^{-1}[y_0^m z_{(m)}]:\ b_{n+1}^{n+\ell} = x_{n+1}^{n+\ell}} \mu(b_0^n | b_{n+1}^{K_m}) \mu(b_{n+1}^{K_m})} \\
&= \frac{\sum\limits_{a_0^{K_m} \in \pi^{-1}[y_0^m z_{(m)}]:\ a_0^{n+\ell} = x_0^{n+\ell}} \mu(x_0^n | x_{n+1}) \mu(a_{n+1}^{K_m})}{\sum\limits_{b_0^{K_m} \in \pi^{-1}[y_0^m z_{(m)}]:\ b_{n+1}^{n+\ell} = x_{n+1}^{n+\ell}} \mu(b_0^n | x_{n+1}) \mu(b_{n+1}^{K_m})} \quad \text{(since } \mu \text{ is Markov)} \\
&= \frac{\mu(x_0^n | x_{n+1}) \sum\limits_{a_0^{K_m} \in \pi^{-1}[y_0^m z_{(m)}]:\ a_0^{n+\ell} = x_0^{n+\ell}} \mu(a_{n+1}^{K_m})}{\sum\limits_{b_0^n : \pi(b_0^n x_{n+1}) = y_0^{n+1},\ P_{b_n x_{n+1}} > 0} \mu(b_0^n | x_{n+1}) \sum\limits_{b_{n+1}^{K_m} \in \pi^{-1}[y_{n+1}^m z_{(m)}]:\ b_{n+1}^{n+\ell} = x_{n+1}^{n+\ell}} \mu(b_{n+1}^{K_m})} \\
&= \frac{\mu(x_0^n | x_{n+1})}{\sum\limits_{b_0^n : \pi(b_0^n x_{n+1}) = y_0^{n+1},\ P_{b_n x_{n+1}} > 0} \mu(b_0^n | x_{n+1})},
\end{aligned}$$

is independent of m and of $x_{n+2}^{n+\ell}$. Hence, ρ, which is the weak limit of ρ_m's satisfies (4.15), and is thus a Markov measure on Ω_y.

□

These results can now be used to show that fibre mixing does indeed imply existence of a continuous measure disintegration, and hence by Corollary 4.24 regularity of the factor measure ν.

Corollary 4.29. *Let $\Omega \subset \mathscr{A}^{\mathbb{Z}_+}$ and $\Sigma \subset \mathscr{B}^{\mathbb{Z}_+}$ be mixing subshifts of finite type, and $\pi : \Omega \to \Sigma$ a 1-block factor map which is **fibre mixing**. Suppose μ is the stationary Markov measure consistent with Ω, then μ admits a continuous measure disintegration and hence $\nu = \mu \circ \pi^{-1}$ is a g-measure.*

Proof. By Lemma 4.26, for every $y \in \Sigma$, there is a unique Gibbs-Markov measure on Ω_y: $|\mathscr{G}_{\Omega_y}(\Phi)| = 1$. By Proposition 4.28, $\varnothing \neq \overline{\mathfrak{M}}_y \subset \mathscr{G}_{\Omega_y}(\Phi)$, and hence $|\overline{\mathfrak{M}}_y| = 1$ for all $y \in \Sigma$. Thus all points in Σ are Tjur, and hence by Corollary 4.24, μ admits a continuous disintegration, which allows us to conclude that ν is a g-measure. □

4.5 Examples

Existence of a continuous measure disintegrations of Markov measures thus follows from the fibre-mixing condition. In fact, it is a weaker condition: it implies regularity of the Furstenberg example (see Section 4.2.3) for the exceptional parameter value $p = \frac{1}{2}$, which is not fibre mixing. Recall, $\{X_n\}_{n\in\mathbb{Z}_+}$ is a Bernoulli process, with a parameter $p \in (0,1)$ taking values in $\mathscr{A} = \{-1, 1\}$ and $\{Y_n\}_{n\in\mathbb{Z}_+}$ is defined by $Y_n = X_n X_{n+1}$. The fibres in this example are $\Omega_y = \pi^{-1}(y) = \left\{x_y^+, x_y^-\right\}$, where

$$x_y^+ = (1,\ y_0,\ y_0 \cdot y_1,\ y_0 \cdot y_1 \cdot y_2, ...), \quad x_y^- = (-1,\ -y_0,\ -y_0 \cdot y_1,\ -y_0 \cdot y_1 \cdot y_2, \ldots).$$

If $p = \frac{1}{2}$, then $\{Y_n\}$ are independent, and $\nu = \mu \circ \pi^{-1}$ is the Bernoulli measure . We now show that $\{\mu_y\}_{y\in\Sigma}$ defined by

$$\mu_y = \frac{1}{2}\left(\delta_{x_y^+} + \delta_{x_y^-}\right).$$

is a continuous measure disintegration of μ. It is clear that, given y, the measure μ_y is a Borel measure supported on Ω_y. Moreover, one has

$$\mu_y(f) - \mu_{\tilde{y}}(f) = \frac{1}{2}\left(f(x_y^+) + f(x_y^-) - f(x_{\tilde{y}}^+) - f(x_{\tilde{y}}^-)\right)$$

and since $y \to x_y^+$ and $y \to x_y^-$ are a continuous maps, for any continuous function f and any $\varepsilon > 0$, one can choose $\delta > 0$, such that $d(y, \tilde{y}) < \delta$ implies $|\mu_y(f) - \mu_{\tilde{y}}(f)| < \varepsilon$.

We now show that $\{\mu_y\}$ is indeed a disintegration of μ. For $x = (x_i)_{i\geq 0}$, let $\bar{x} = (\bar{x}_i)$ with $\bar{x}_i = -x_i$ for all $i \geq 0$; note that $x_y^- = \overline{x_y^+}$. It is sufficient to validate consistency of disintegration $\{\nu_y\}$ for indicators of cylindric sets:

$$\int_\Sigma \int_{\Omega_y} \mathbb{1}_{[a_0^n]}(x)\mu_y(dx)\nu(dy) = \frac{1}{2}\int_\Sigma \Big(\mathbb{1}_{[a_0^n]}(x_y^+) + \mathbb{1}_{[a_0^n]}(x_y^-)\Big)\nu(dy)$$
$$= \frac{1}{2}\int_\Omega \Big(\mathbb{1}_{[a_0^n]}(x^+_{\pi(\tilde{x})}) + \mathbb{1}_{[\bar{a}_0^n]}(x^+_{\pi(\tilde{x})})\Big)\mu(d\tilde{x})$$
$$= \frac{1}{2}\int_\Omega \mathbb{1}_{[a_0^n]\cup[\bar{a}_0^n]}(\tilde{x})\,\mu(d\tilde{x}) = \int_\Omega \mathbb{1}_{[a_0^n]}(\tilde{x})\mu(d\tilde{x}).$$

Hence the μ admits for a continuous disintegration. This example only works for a very specific parameter value $p = 1/2$. Interestingly, there exists another example that has exactly the same continuous measure disintegration. Let $p \in (0,1)$ and $\{X_n\}_{n\in\mathbb{Z}_+}$ be a Markov chain taking values in $\{-1,1\}$, with the transition probability matrix

$$P = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}.$$

The stationary distribution is distribution $\rho = \left(\frac{1}{2}, \frac{1}{2}\right)$. Then the factor process

$$Y_n = \pi(X_n, X_{n+1}) = X_n \cdot X_{n+1}.$$

is Bernoulli for all values of $p \in (0,1)$. Let $M_+(w_n^m) = \sum_{i=n}^m \mathbb{1}_{+1}(w_i)$ and $M_-(w_n^m) = \sum_{i=n}^m \mathbb{1}_{-1}(w_i)$, then

$$\nu(Y_0 = y_0 | Y_1^n = y_1^n) = \frac{\nu(Y_0^n = y_0^n)}{\sum_{w\in\{-1,1\}} \nu(Y_0^n = wy_1^n)}$$
$$= \frac{\mu(X_0^{n+1} = (x_y^+)_0^{n+1}) + \mu(X_0^{n+1} = (x_y^-)_0^{n+1})}{\sum_{w\in\{-1,1\}} \mu(X_0^{n+1} = (x^+_{wy_1^\infty})_0^{n+1}) + \mu(X_0^{n+1} = (x^-_{wy_1^\infty})_0^{n+1})}$$
$$= \frac{2p^{M_+(y_0^n)}(1-p)^{M_-(y_0^n)}}{2p^{M_+(y_0^n)}(1-p)^{M_-(y_0^n)} + 2p^{M_+(\bar{y}_0 y_1^n)}(1-p)^{M_-(\bar{y}_0 y_1^n)}}$$
$$= \begin{cases} p & : y_0 = +1 \\ 1-p & : y_0 = -1, \end{cases}$$

for any $n \geq 1$, where we again used the notation $\bar{y}_0 = -y_0$. It follows that the process $\{Y_n\}_{n\in\mathbb{Z}_+}$ is Bernoulli with the parameter p. Note that this example has exactly the same fibre structure as the last example: $\Omega_y = \pi^{-1}(y) = \left\{x_y^+, x_y^-\right\}$, where

$$x_y^+ = (1, y_0, y_0 \cdot y_1, \ldots), \quad x_y^- = (-1, -y_0, -y_0 \cdot y_1, \ldots).$$

Moreover, the same continuous measure disintegration exists: $\{\mu_y\}_{y\in\Sigma}$ with

$$\mu_y = \frac{1}{2}\left(\delta_{x_y^+} + \delta_{x_y^-}\right).$$

Continuity and consistency follow by an identical computation as for the Furstenberg example above. Therefore, we have another example of a factor measure with a continuous measure disintegration, but without fibre mixing conditions.

4.5.1 Markov factor without continuous measure disintegration

We now show by an example that existence of a continuous measure disintegration is not necessary. In this example, the factor measure ν is Markov. Let $\{X_n\}_{n\in\mathbb{Z}_+}$ be a stationary Markov chain taking values in $\mathscr{A} = \{1,2,3,4\}$ defined by the probability transition matrix:

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

Define the factor map π as follows: let $\mathscr{B} = \{a,b,c\}$ and put $\pi : \mathscr{A} \to \mathscr{B}$, by $\pi(1) = \pi(3) = a$, $\pi(2) = b$, $\pi(4) = c$. Then the space $\Sigma \subset \mathscr{B}^{\mathbb{Z}_+}$ is a subshift of finite type with forbidden words $\{bb, cc, bc, cb\}$.

This example is not lumpable as $(1,0,0,0)$ is an initial distribution for which the factor process is not Markov; the transition from state a to state c in the output process has probability 0 until the first occurrence of the word ba. However, direct application of the result in [52], shows that the stationary chain $\{X_n\}$ is weakly lumpable with respect π, i.e., $\{Y_n = \pi(X_n)\}$ is a Markov process, and one can easily compute the corresponding transition probability matrix $\tilde{P}$. The stationary invariant distribution of $\{X_n\}$ is $p = \left(\frac{1}{3}, \frac{1}{6}, \frac{1}{3}, \frac{1}{6}\right)$. Hence, ν is a Markov measure with the probability transition matrix

$$\tilde{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

We proceed by showing that no continuous measure disintegration exists. In this particular case, the map π is *finite to one factor map*, meaning the fibres have a bounded number of elements.

Since $\pi^{-1}(b) = 2$ and $\pi^{-1}(c) = 4$, but $\pi^{-1}(a) = \{1,3\}$. If we assume that $y_n = a$ and $m = \min\{m > n : y_m \in \{b,c\}\}$ is finite, then $\pi^{-1}(y)_n = 1$ if $y_m = b$, and $\pi^{-1}(y)_n = 3$ if $y_m = c$. Therefore elements $y \in \Sigma$ are uniquely decodable if $y \neq y_0^{n-1} a_n^\infty$ for any $n \geq 1$, i.e., if y does not end with infinite string of a's. Otherwise, if $y = y_0^{n-1} a_n^\infty$ for some $n \geq 1$, then the fibre contains exactly two points, corresponding to one of the two possible tails: an infinite number of 1's or 3's. This fibre structure makes a continuous measure disintegration of μ impossible: Suppose $\{\mu_y\}_{\mathbb{Z}_+}$ is a measure disintegration of μ for the factor map π. Then μ_y is supported on Ω_y for each $y \in \Sigma$. Furthermore, consider the point $z = a_0^\infty$. Then any open neighbourhood of z contains, for some $n > 0$ the cylinder sets $[a_0^n b]$ and $[a_0^n c]$. For each $y \in [a_0^n b]$ we have $\Omega_y \subset [1_0]$, while for each $y' \in [a_0^n c]$ we have $\Omega_{y'} \subset [3_0]$. Hence

$$\left| \int_{\Omega_y} \mathbb{1}_{[1_0]}(x) \mu_y(dx) - \int_{\Omega_{y'}} \mathbb{1}_{[1_0]}(x) \mu_{y'}(dx) \right| = |1 - 0| = 1.$$

Since both cylinders $[a_0^n b]$ and $[a_0^n c]$ have positive ν-measure, and $\mathbb{1}_{[1_0]}$ is a continuous function we conclude that no measure disintegration μ can be continuous at $z = a_0^\infty$.

One can also use entropy methods to conclude that the factor measure is well-behaved. Suppose a subshift of finite type is defined by a primitive 0/1-matrix M. Then the *topological entropy* of the SFT X_M is equal to $\log(\lambda)$, where λ is the largest eigenvalue of M. In our case,

$$M = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

has characteristic polynomial $\lambda^4 - 2\lambda^3 - \lambda^2 + 2\lambda$, with the largest eigenvalue equal to 2. The topological entropy is an upper bound for the entropy of measures on the SFT, $h_\mu \leq h_{top}(\Omega) = \log(2)$. For Markov measures we have

$$h_\mu = -\sum_{x_0, x_1 \in \mathscr{A}} p_{x_0} P_{x_0,x_1} \log(P_{x_0,x_1}),$$

which in our case is $h_\mu = \log(2)$. For an irreducible SFT the measure of maximal entropy, also known as Parry measure, is unique and is Markov. Moreover, a finite-to-one factor map between two SFTs sends the measure of maximal entropy to the measure of maximal entropy. Thus since ν is a measure of maximal entropy on Ω, then so is $\nu = \mu \circ \pi^{-1}$, and hence ν is also Markov.

Chapter 5

On preservation of Gibbsianity under renormalisation

In this chapter we establish sufficient conditions for the preservation of the Gibbs property under renormalisation transformations on lattices $\mathbb{Z}^d$. Our result can be viewed as a first proof in complete generality of the easy part of the well-known van Enter-Fernández-Sokal hypothesis [81], which informally can be stated as that the renormalized Gibbs states remains Gibbs if and only if there are no *hidden phase transitions*.

5.1 Introduction

We start by recalling the necessary theory of Gibbs states for lattice systems; $\Omega = \prod_{n\in\mathbb{Z}^d} \mathscr{A}_n$, where the alphabets $\mathscr{A}_n$ are finite with $|\mathscr{A}_n| \leq M < \infty$ for all n. The dependency of the alphabets $\mathscr{A}_n$ on $n \in \mathbb{Z}^d$ is atypical, however the part of the theory on Gibbs measures that we need applies directly. We will use the following notation: $\mathscr{A}^\Lambda = \prod_{n\in\Lambda} \mathscr{A}_n$, and $x_\Lambda = x|_\Lambda$ for the restriction of $x \in \Omega$ to Λ.

Gibbs measures on Ω are defined via an *interaction*:

Definition 5.1. An interaction is a collection of functions, $\{\Phi_\Lambda\}$ on Ω, indexed by finite subsets $\Lambda \Subset \mathbb{Z}^d$ ($\Subset$ indicates that the subset is finite), such that

$$\Phi_\Lambda(x) = \Phi_\Lambda(x|_\Lambda),$$

i.e., $\Phi_\Lambda(x)$ depends only on the values of x on Λ. An interaction $\Phi = \{\Phi_\Lambda\}_{\Lambda\Subset\mathbb{Z}^d}$ is called *uniformly absolutely convergent* (UAC) if

$$\|\Phi\| := \sup_{n\in\mathbb{Z}^d} \sum_{n\in\Lambda\Subset\mathbb{Z}^d} \sup_{x\in\Omega} |\Phi_\Lambda(x)| = \sup_{n\in\mathbb{Z}^d} \sum_{n\in\Lambda\Subset\mathbb{Z}^d} \|\Phi_\Lambda\|_\infty < \infty.$$

We denote the space of UAC interactions by $\mathscr{B}^1(\Omega)$. For $\Phi = \{\Phi_\Lambda\} \in \mathscr{B}^1(\Omega)$ and a finite set (volume) $V \Subset \mathbb{Z}^d$, the corresponding Hamiltonian is defined as

$$H_V^\Phi(x) = \sum_{\Lambda \cap V \neq \varnothing} \Phi_\Lambda(x).$$

The Hamiltonian H_V^Φ is continuous (quasilocal) on Ω. Meaning that, if $\mathbb{L}_n = [-n, n]^d$, and, for every $\Lambda \Subset \mathbb{Z}^d$,

$$\sup_{\bar{x}_\Lambda \in \mathscr{A}^\Lambda} \sup_{x \in \Omega} \sup_{y,z \in \Omega} \left| H_\Lambda^\Phi(\bar{x}_\Lambda x_{\mathbb{L}_n \setminus \Lambda} y_{\mathbb{L}_n{}^c \setminus \Lambda}) - H_\Lambda^\Phi(\bar{x}_\Lambda x_{\mathbb{L}_n \setminus \Lambda} z_{\mathbb{L}_n{}^c \setminus \Lambda}) \right| \to 0,$$

as $n \to \infty$. Finally, given $\Phi \in \mathscr{B}^1(\Omega)$ and $\Lambda \Subset \mathbb{Z}^d$, define the corresponding specification density as

$$\gamma_\Lambda^\Phi(\bar{x}_\Lambda | x_{\Lambda^c}) = \frac{1}{Z^\Phi(x_{\Lambda^c})} \exp\Big(-H_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c}) \Big),$$

where $\bar{x}_\Lambda x_{\Lambda^c}$ is an element in Ω, equal to $\bar{x}$ on Λ, and to x on Λ^c. The normalizing constant $Z^\Phi(x_{\Lambda^c}) = \sum_{\bar{x}_\Lambda \in \mathscr{A}^\Lambda} \exp\Big(-H_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c}) \Big)$ is called a partition function. The specification for Φ, $\tilde{\gamma}^\Phi$, is a collection of probability kernels $\tilde{\gamma}_\Lambda^\Phi : \mathscr{B}_\Lambda \times \Omega_{\Lambda^c} \to (0,1)$, indexed by $\Lambda \Subset \mathbb{Z}^d$, where $\mathfrak{B}_V = \sigma\{[a_{V'}] : V' \Subset V\}$ for any $V \subset \mathbb{Z}^d$. The specification is then determined by the equality

$$\tilde{\gamma}_\Lambda^\Phi(f | x_{\Lambda^c}) = \sum_{\bar{x}_\Lambda \in \mathscr{A}^\Lambda} f(\bar{x}_\Lambda x_{\Lambda^c}) \gamma_\Lambda^\Phi(\bar{x}_\Lambda | x_{\Lambda^c}),$$

for any bounded measurable $f : \Omega \to \mathbb{R}$. In the remainder of the text we will not explicitly distinguish between specifications and their densities.

Definition 5.2. A probability measure μ on Ω is Gibbs for the interaction Φ (denoted by $\mu \in \mathscr{G}_\Omega(\Phi)$) if it is *consistent* with the corresponding specification γ^Φ ($\mu \in \mathscr{G}_\Omega(\gamma^\Phi)$), meaning that for every $\Lambda \Subset \mathbb{Z}^d$

$$\mu(x_\Lambda | x_{\Lambda^c}) = \gamma_\Lambda^\Phi(x_\Lambda | x_{\Lambda^c}) \quad \text{for } \mu - a.a.\ x \in \Omega.$$

Equivalently, μ is Gibbs if for any continuous function f on Ω and every $\Lambda \Subset \mathbb{Z}^d$

$$\int f(x) \mu(dx) = \int \sum_{\bar{x}_\Lambda \in \mathscr{A}^\Lambda} f(\bar{x}_\Lambda x_{\Lambda^c}) \gamma_\Lambda^\Phi(\bar{x}_\Lambda | x_{\Lambda^c}) \mu(dx). \tag{5.1}$$

A principal result due to Dobrushin, Lanford and Ruelle is that for every UAC interaction Φ, there exists at least one Gibbs measure, i.e., $\mathscr{G}_\Omega(\Phi) = \mathscr{G}_\Omega(\gamma^\Phi) \neq \varnothing$. The inhomogeneous space Ω we consider in this chapter is covered in the work by Ruelle.

When more than one Gibbs measure exists for a given interaction Φ, we say that the system has a phase transition.

The UAC property of Φ implies that $\gamma^{\Phi} = \{\gamma_{\Lambda}^{\Phi}\}$ has the following important properties:

- **Uniform non-nullness**: for every $\Lambda \Subset \mathbb{Z}^d$ there exist positive constants a_{Λ}^{Φ}, $b_{\Lambda}^{\Phi} \in (0,1)$ such that
$$a_{\Lambda}^{\Phi} \leq \gamma_{\Lambda}^{\Phi}(\bar{x}_{\Lambda}|x_{\Lambda^c}) \leq b_{\Lambda}^{\Phi}$$
for all $\bar{x}_{\Lambda} \in \mathscr{A}^{\Lambda}$ and every $x \in \Omega$.

- **Quasilocality**: let $\mathbb{L}_n = [-n,n]^d$, $n \geq 1$, for every $\Lambda \Subset \mathbb{Z}^d$,
$$v_{\Lambda,n} := \sup_{x,\bar{x}\in\Omega} \sup_{y,z\in\Omega} \left|\gamma_{\Lambda}^{\Phi}(\bar{x}_{\Lambda}|x_{\mathbb{L}_n\setminus\Lambda}y_{\mathbb{L}_n{}^c\setminus\Lambda}) - \gamma_{\Lambda}^{\Phi}(\bar{x}_{\Lambda}|x_{\mathbb{L}_n\setminus\Lambda}z_{\mathbb{L}_n{}^c\setminus\Lambda})\right| \to 0,$$
as $n \to \infty$.

Note that in the present context quasilocality is equivalent to continuity in the product topology.

It turns out that there is a second way to introduce Gibbs measures: a fundamental result of Kozlov and Sullivan [2, 55, 84] states that, for a specification $\gamma = \{\gamma_{\Lambda}\}$ on Ω that is uniformly non-null and quasilocal, there exists a UAC interaction Φ such that $\gamma = \gamma^{\Phi}$. First we need to define a specification. In particular we require some additional properties that are immediate when γ^{Φ} is obtained from an interaction Φ:

Definition 5.3. A specification for the lattice $\mathbb{Z}^d$ is a collection of probability kernels $\gamma = \{\gamma_{\Lambda}\}_{\Lambda \Subset \mathbb{Z}^d}$, such that

- $x \to \gamma(A|x)$ is a $\mathfrak{B}_{\Lambda^c}$−measurable map for each $\Lambda \Subset \mathbb{Z}^d$ and measurable set A.

- For $\Lambda \Subset \mathbb{Z}^d$ we have $\gamma_{\Lambda}(A|x) = \mathbb{1}_A(x)$, when $x \in \Omega$ and $A \in \mathfrak{B}_{\Lambda^c}$.

- For $\Lambda' \subset \Lambda \Subset \mathbb{Z}^d$
$$\int_{\Omega} \gamma_{\Lambda'}(A|x')\gamma_{\Lambda}(dx'|x) = \gamma_{\Lambda}(A|x).$$

The second property above is commonly referred to as properness and the third property is called consistency. The Kozlov-Sullivan theorem can now be stated as follows:

Theorem 5.4. *Suppose μ is a (fully supported) Borel probability measure on $\Omega = \mathscr{A}^{\mathbb{Z}^d}$ and $\gamma = \{\gamma_\Lambda : \Lambda \Subset \mathbb{Z}^d\}$ is a uniformly non-null quasilocal specification on Ω such that*

$$\int_\Omega f(x)\mu(dx) = \int_\Omega \sum_{\bar{x}_\Lambda \in \mathscr{A}^\Lambda} f(\bar{x}_\Lambda x_{\Lambda^c})\gamma_\Lambda(\bar{x}_\Lambda x_{\Lambda^c})\mu(dx) \tag{5.2}$$

for all continuous functions $f \in C(\Omega)$. Then μ is a Gibbs state for some UAC potential Φ.

From this one also gets a well known convenient characterisation of DLR Gibbs measures:

Theorem 5.5. *A fully supported probability measure on $\Omega = \mathscr{A}^{\mathbb{Z}^d}$, or $\Omega = \prod_{n\in\mathbb{Z}^d} \mathscr{A}_n$, is a Gibbs measure if and only if for every $\Lambda \Subset \mathbb{Z}^d$ there exists a continuous function $\gamma_\Lambda > 0$ on Ω such that*

$$\sup_x \left|\mu(x_\Lambda | x_{\mathbb{L}_n \setminus \Lambda}) - \gamma_\Lambda(x)\right| \to 0, \quad \text{as } n \to \infty$$

5.2 Fuzzy Gibbs states

Now restrict to Gibbs measures on homogeneous spaces: $\Omega = \mathscr{A}^{\mathbb{Z}^d}, \Sigma = \mathscr{B}^{\mathbb{Z}^d}$, with $|\mathscr{B}| \leq |\mathscr{A}| < \infty$. Suppose $\pi : \mathscr{A} \to \mathscr{B}$ is a surjective map, which we will refer to as a fuzzy map or a (single-block) factor map. We use the same letter π to denote the (componentwise) extension of π to a mapping from $\mathscr{A}^V$ onto $\mathscr{B}^V$ for any subset $V \subseteq \mathbb{Z}^d$. Then the map $\pi : \Omega \to \Sigma = \mathscr{B}^{\mathbb{Z}^d}$ defines the factor measure $\nu = \mu \circ \pi^{-1}$. If μ is a Gibbs state (measure), we will refer to the measure ν as a fuzzy Gibbs state (measure).

The measure ν is not necessarily Gibbs. That is, there does not exist a quasilocal and non-null specification consistent with ν. In the present context (spaces $\Omega = \mathscr{A}^{\mathbb{Z}^d}, \Sigma = \mathscr{B}^{\mathbb{Z}^d}$, and a single-block factor map $\pi : \Omega \to \Sigma$) non-nullness is preserved, hence potential non-Gibbsianness of the factor measure ν is due to non-quasilocality of conditional probabilities. The regularity of the conditional probabilities of factor measures has also been studied in other areas of mathematics. For example, regularity of factors of Markov measures has been studied extensively [14, 45, 52, 99]. Another example is regularity of factors of g—measures on a full shift [49, 87]. In the context of Gibbs measures, factors are specific examples of the more general renormalisation group transformations, used in the study of critical phenomena, see for example [15]. It was noted that some of these transformations resulted in pathological behaviour of the transformed measures [39–41]. Examples of this behaviour were suggested in [47] and proven in [81].

It turns out that the problem is non-Gibbsianness of the renormalized measure. In [81] it was shown that the non-Gibbsianness is typically caused by, so-called, hidden phase transitions. The central idea can be summarized as follows; given a Gibbs measure μ consistent with an interaction Φ and a factor map π, one finds a configuration $\tilde{y} \in \Sigma$, in which ν is not quasilocal. A good candidate configuration would be an $\tilde{y} \in \Sigma$ such that on the space $\Omega_{\tilde{y}} = \pi^{-1}(\tilde{y})$ there exists a phase transition for the interaction Φ. The reason is as follows: selecting a configuration far from the origin in Σ corresponds to selecting an element $\bar{y}$ in a small neighbourhood around $\tilde{y}$. If, for any two sufficiently far away configurations, $\bar{y}_1$ and $\bar{y}_2$, two different phases are selected on $\Omega_{\bar{y}_1}$ and $\Omega_{\bar{y}_1}$ then the distributions in $\Omega_{\bar{y}_1}$ and $\Omega_{\bar{y}_1}$ near the origin can be different. This might then be used to show that $\tilde{y}$ is a point of discontinuity for $\nu(\tilde{y}_0|\tilde{y}_{\{0\}^c})$. When, instead, for all $y \in \Sigma$ a unique Gibbs measure exists on Ω_y, for Φ, then the measure ν is Gibbs. We refer to this condition as the *van Enter-Fernández-Sokal hypothesis*. Furthermore, when the Gibbs measures on the fibres have a phase transition it is called a *hidden phase transition*.

The problem of finding a necessary and sufficient condition for the Gibbsianness of a factor measure is still open. Moreover, in certain examples, such as the decimation of the Ising model and the fuzzy Potts model, the Gibbsianness of the fuzzy measure is still an open problem for certain values of the (inverse) temperature.

The condition we discuss in this chapter is based on the sufficient conditions for regularity of factors of $g-$measures discussed in [87] and the previous chapter. The condition is based on the fact that, for factors such as above, a disintegration of the measure exists, i.e., a collection of measures on the fibres $\{\mu_y\}_{y\in\Sigma}$ exists such that $\mu(\cdot) = \int \mu_y(\cdot)\nu(dy)$. One can find an explicit expression for the conditional probabilities of the factor measures in terms of this integration. Now, if a version of $\{\mu_y\}$ can be chosen such that $y \to \mu_y$ is continuous in the weak topology, then regularity of the factor measure follows. In the original example [87] and the previous chapter this is formulated for a one-dimensional lattice., but, as we will show, it applies equally well to $\mathbb{Z}^d$. We give an expression for the specification of ν in terms of the measure disintegration. Furthermore, we show that existence of a continuous measure disintegration is a sufficient condition for the factor measure to be Gibbs. Moreover, when a continuous measure disintegration exists, the measures μ_y can be defined constructively [85, 86]. Using this approach we show that the absence of phase transitions on the sets Ω_y, for the interaction of μ, Φ, implies existence of a continuous measure disintegration. That is, we provide a proof for the easy direction van Enter-Fernández-Sokal hypothesis in full generality.

5.2.1 The decimation of the Ising model

We now turn to the discussion of known examples of fuzzy Gibbs measures that are not fully understood. The first example is the decimation of the Ising model. Let $d = 2$, $\mathscr{A} = \{-1, 1\}$ and define an interaction $\Phi_{i,j}(x) = -Jx_i x_j$, for i and j nearest neighbours on the lattice $\mathbb{Z}^d$ (denoted by $i \sim j$). Here J plays the role of an inverse temperature. This interaction defines the Ising model. This model has a critical temperature $J_c = \frac{1}{2}\log(1+\sqrt{2})$ [70]. That is, for $J < J_c$, there is a unique Gibbs measure and for $J > J_c$ there is a phase transition, i.e., multiple Gibbs states coexist.

An important example of a fuzzy map for the Ising model is the decimation transformation. Let $b \geq 2$, then $\pi : \mathscr{A}^{\mathbb{Z}^d} \to \mathscr{A}^{\mathbb{Z}^d}$ via $\pi(x)_i = \pi(x_{bi})$ is the decimation transformation. This map is visualized for $d = 2$, $b = 2$ in Figure 5.1. To see this as a fuzzy transformation consider, for $n \in \mathbb{Z}^d$, the fundamental domain of $\mathbb{Z}^d$ defined by $B_n = \{n' \in \mathbb{Z}^d : n_i \leq n'_i < n_i + b, \text{ for } 1 \leq i \leq d\}$. Now define $\mathscr{A}' = \mathscr{A}^{B_0}$, then we can identify $\Omega = \mathscr{A}^{\mathbb{Z}^d}$ with $\Omega' = \mathscr{A}'^{\mathbb{Z}^d}$ by assigning to $x \in \Omega$ an element $x' \in \Omega'$ defined by $x'_n = x_{B_{nb}}$ for all $n \in \mathbb{Z}^d$. The map π is now a fuzzy map from $\Omega' = \mathscr{A}'^{\mathbb{Z}^d}$ to $\Omega = \mathscr{A}^{\mathbb{Z}^d}$.

It was proposed in [47] and rigorously proven in [81] that this transformation, when applied to the Ising model for $d = 2$ at a sufficiently low temperature, results in a non-Gibbsian factor measure. Moreover, the non-Gibbsianness in this model is related to the phase transition in the Ising model.

It was shown in [81] that the decimation transformation for $b = 2$ and $J > \frac{1}{2}\cosh^{-1}(1+\sqrt{2}) \approx 1.73J_c$ results in factor measures which are not Gibbs measures. On the other hand, Haller and Kennedy [44] showed that for $J < 1.36J_c$, the 2-decimation of any Gibbs state for the Ising model in two dimensions is Gibbs.

5.2.2 Possible loss of Gibbsianity under renormalisation

We will now give a rough outline of the method by which non-Gibbsianness of the decimated Ising model was proven in [81]. First note that a specification defines a Gibbs measure if it satisfies both non-nullness and quasilocality. In the case of fuzzy transformations, the non-nullness is preserved, thus a loss of Gibbsianness corresponds to a loss of quasilocality. Hence, to show that a measure is not Gibbs we can proceed by finding a point where the measure fails to be quasilocal, a so-called *bad configuration*:

Definition 5.6. A point $y \in \Sigma$ is called a bad configuration for ν if there exists a

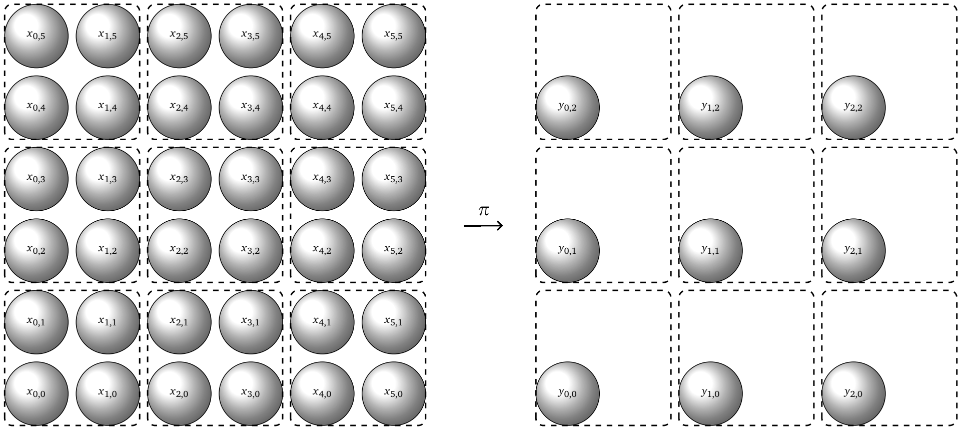


Figure 5.1: An example of a 2-decimation. The transformation is a fuzzy transformation when one identifies the configuration in a dashed box as an individual lattice site.

$\Lambda \Subset \mathbb{Z}^d$ and an $\varepsilon > 0$ such that for every $V \Subset \mathbb{Z}^d$ one can find $W \Subset \mathbb{Z}^d$, $V \subset W$, and two points $\underline{y}, \overline{y} \in \Sigma$ such that

$$\nu(y_\Lambda | y_{V\setminus\Lambda}\overline{y}_{W\setminus V\setminus\Lambda}) - \nu(y_\Lambda | y_V \underline{y}_{W\setminus V\setminus\Lambda}) \geq \varepsilon > 0.$$

Existence of a bad configuration y implies that no version of the conditional probabilities $\nu(y_\Lambda | y_{\Lambda^c})$, which is defined ν-a.s., can be continuous at y and hence no quasilocal specification can be consistent with ν.

In [81] it is shown that the alternating configuration $\tilde{y} \in \Sigma$, where $\tilde{y}_{i,j} = (-1)^{i+j}$, is a bad configuration for the 2–decimation of the Ising model in two dimensions. To understand why this is a bad configuration we can analyse the possible configurations in Ω leading to the alternating configurations: $\Omega_{\tilde{y}} = \pi^{-1}(\tilde{y}) \subset \Omega$. This fibre can be represented as a lattice model, as depicted in Figure 5.2. For the positions that are removed by the decimation we then have the possible spin values $\{-1, +1\}$, those spins are referred to as internal spins. We distinguish two different types of internal spins, those that have four other internal spins as their neighbours, these are denoted by $\tilde{x}$ in Figure 5.2. Additionally, we have internal spins with two other internal spins and two spins given by $\tilde{y}$ as their neighbours, those spins we label by $x_{i,j}$. The first important observation is that the fixed neighbours of the spins $x_{i,j}$ have *opposite* spins. Hence, the internal spins experience no net interaction external neighbours. The internal spins form a decorated lattice as depicted in Figure 5.3. This system can be reduced further as the $\tilde{x}_{i,j}$ spins interact with each other via the internal spins between them. The system of spins $\tilde{x}_{i,j}$ can be seen as an Ising model on $\mathbb{Z}^2$, for an inverse temperature $\tilde{J} = \frac{1}{2}\log(\cosh(2J))$.

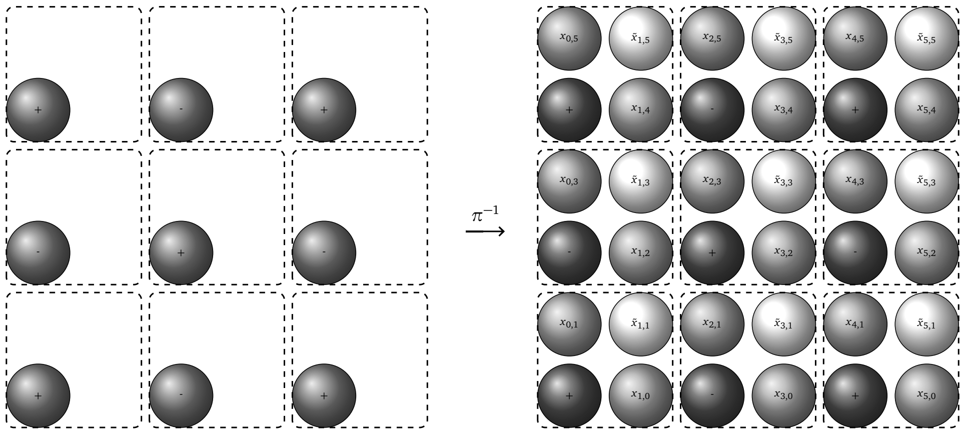


Figure 5.2: The preimage of the alternating configuration has two type of internal spin nodes. Those neighbouring two alternating spins and those neighbouring no spins determined by the alternating configuration. The latter form an Ising model amongst themselves.

The main observation is that the internal spins experience a phase transition for $J > \frac{1}{2}\cosh^{-1}\left(1+\sqrt{2}\right) \approx 1.73J_c$, where J_c is the critical temperature for the original Ising model. It can be shown that, for the alternating configuration $\tilde{y}$ on an arbitrarily large finite set Λ around the origin, the internal spins around the origin will still experience a strong influence from the internal spins corresponding to the configuration outside Λ. We can now choose either a $+1$ or -1 configuration outside Λ. Such a configuration acts as a magnetic field on the internal spins far from the origin. Because of the phase transition, this affects the expected internal spin values near the origin. The final step is choosing the alternating configuration only for $\Lambda - \{(0,0)\}$ and proving that the internal spins around the origin affect the distribution for y_0. To summarize, if we choose the alternating configuration $\tilde{y}$ in a arbitrarily large area around the origin, then there exists a phase transition in the internal spin model. This phase transition results in a strong long range interaction between the spin particle at the origin and the far away configurations. This can then be used to show that $\tilde{y}$ is a bad configuration for the $2-$decimation of the Ising model in $\mathbb{Z}^2$. This phase transition is commonly referred to as a hidden phase transition. The mechanism is used in a similar way to show loss of Gibbsianity for a number of renormalisation transformations applied to various Gibbs measures [81].

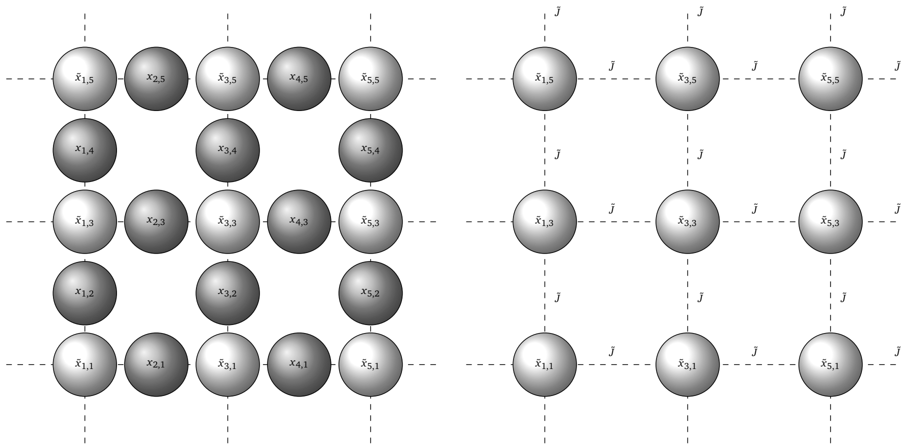


Figure 5.3: The decorated Ising model on the left, for the interaction J results in an Ising model for $\tilde{J} = \frac{1}{2}\log(\cosh(2J))$ on the right when integrating out the $x_{i,j}$ particles. Hence the $\tilde{x}_{i,j}$ particles have a phase transition for $J > \frac{1}{2}\cosh^{-1}\left(1+\sqrt{2}\right)$.

5.2.3 Fuzzy Potts model

We now consider the fuzzy Potts model [43, 63]. Let $\mathscr{A} = \{1, ..., q\}$, $q \geq 2$ be a finite alphabet and define a nearest-neighbour interaction $\{\Phi_{i,j}\}_{i\sim j}$, via $\Phi_{i,j}(x) = 2\beta\delta(x_i, x_j)$, with $\beta \geq 0$ the inverse temperature. Now any factor $\nu = \mu \circ \pi^{-1}$ of the Potts model is referred to as a fuzzy Potts model.

It is well known that the Potts model has a critical temperature. More specifically, for any $d \geq 2$, there exists a critical temperature $\beta_c = \beta_c(q, d) \in (0, +\infty)$ [1], such that for $\beta < \beta_c$ there is a unique Gibbs measure, and for $\beta > \beta_c$ there are multiple Gibbs measures, i.e., there is a phase transition. Consider now the following transformation: let $\mathscr{B} = \{1, ..., m\}$, $m < q$, and define a factor map $\pi : \mathscr{A} \mapsto \mathscr{B}$, by

$$\pi(x) = \begin{cases} 1: & x \in \{1, \ldots r_1\} \\ 2: & x \in \{r_1+1, \ldots, r_1+r_2\} \\ \cdots & \\ m: & x \in \{q-r_m+1, \ldots, q\}, \end{cases}$$

where $1 \leq r_i < q$ for all $1 \leq i \leq m$ and, without loss of generality, we may assume that $r_1 = \min\{r_i : r_i > 1\}$. Then the following holds:

Theorem 5.7 ([43]). *Let $\nu = \mu \circ \pi^{-1}$ be a fuzzy Potts measure as above. Then:*

- *ν is Gibbs if $\beta < \beta_c(r_1, d)$;*

- *ν is not Gibbs if $\beta > \frac{1}{2}\log\frac{1+(r_1-1)p_c(d)}{1-p_c(d)}$,*

where $p_c(d) \in (0,1)$ is the critical value for independent bond percolation on $\mathbb{Z}^d$.

It seems that the critical temperature is related to the (non-)Gibbsianness of the fuzzy measure. However, this result leaves a gap in the temperature range where Gibbsianness of the factor measure remains undetermined. Moreover, the bound of the temperature for which the factor is Gibbs is the critical temperature of a smaller Potts model with r_1 particles.

5.3 Conditional measures on fibres

5.3.1 Measure disintegrations

We now turn to the discussion of a novel method we propose to study the Gibbs properties of fuzzy Gibbs states. Previously this method has been used to study the regularity of factors of g—measure [87] and in the previous chapter we applied it to the problem of regularity of factors of Markov measures. The main idea is to use a conditional measure disintegration:

Definition 5.8. A family of measures $\mu_\Sigma = \{\mu_y\}_{y\in\Sigma}$ is called a family of conditional measures for μ on fibres Ω_y if

(a) $\mu_y(\Omega_y) = 1$;

(b) for all $f \in L^1(\Omega,\mu)$, the map

$$y \to \int_{\Omega_y} f(x)\mu_y(dx)$$

is measurable and

$$\int_\Omega f(x)\mu(dx) = \int_\Sigma \int_{\Omega_y} f(x)\mu_y(dx)\nu(dy).$$

By a celebrated theorem of von Neumann [67], for all product spaces over finite alphabets Ω, Σ and a continuous surjection $\pi : \Omega \to \Sigma$, a disintegration $\mu_\Sigma = \{\mu_y\}_{y\in\Sigma}$ is exists for any Borel measure μ on Ω.

Note also that

$$\int_{\Omega_y} f(x)\mu_y(dx) = \mathbb{E}_\mu\Big(f \,\Big|\, \pi^{-1}\mathfrak{B}(\Sigma)\Big),$$

where $\pi^{-1}\mathfrak{B}(\Sigma)$ is the sub-σ-algebra of $\mathfrak{B}(\Omega)$ given by

$$\left\{\pi^{-1}(C): \ C \in \mathfrak{B}(\Sigma)\right\}$$

If $\boldsymbol{\mu}_\Sigma = \{\mu_y\}$ and $\tilde{\boldsymbol{\mu}}_\Sigma = \{\tilde{\mu}_y\}$ are two families of conditional measures of μ on fibres $\{\Omega_y\}$, then

$$\nu\Big(\Big\{y : \mu_y \neq \tilde{\mu}_y\Big\}\Big) = 0.$$

5.3.2 Continuous measure disintegrations

Definition 5.9. A family of conditional measures $\{\mu_y\}_{y\in\Sigma}$ for μ on fibres Ω_y is called ***continuous*** if for every continuous $f : \Omega \to \mathbb{R}$, the map

$$y \mapsto \int_{\Omega_y} f(x)\mu_y(dx)$$

is continuous on Σ.

The principal question is whether, for a given measure μ and a factor map $\pi : \Omega \to \Sigma$, there exists a continuous disintegration of μ. First however, we show that existence of a continuous measure disintegration is related to the Gibbsianness of $\nu = \mu \circ \pi^{-1}$.

Theorem 5.10. *Suppose $\mu \in \mathcal{G}_\Omega(\Phi)$ with $\Phi \in \mathscr{B}_1(\Omega)$, $\pi : \Omega \to \Sigma$ is a 1-block factor map. Suppose μ admits a continuous family $\{\mu_y\}$ of conditional measures on fibres $\{\Omega_y\}$. Then $\nu = \mu \circ \pi^{-1}$ is a Gibbs state on Σ.*

First we introduce some notation and a simple lemma. Denote by ϕ_{Λ^c} the projection from Ω to Ω_{Λ^c}; we will use the same map to denote projection from Ω_y to $\Omega_{y_{\Lambda^c}}$. Now let the projection (restriction) of the measure μ_y on Ω_y to $\Omega_{y_{\Lambda^c}}$ be denoted by $\mu_{y_{\Lambda^c}}$. That is, $\mu_{y_{\Lambda^c}} = \mu_y \circ \phi_{\Lambda^c}^{-1}$.

Lemma 5.11. *Suppose $f : \Omega \to \mathbb{R}$ is a continuous function. Suppose furthermore that $f(x)$ depends only on the values on Λ^c, i.e., $f(x) = f(x_{\Lambda^c})$, where Λ is some finite subset of $\mathbb{Z}^d$. Consider a Borel probability measure ρ on Ω, and denote by ρ_{Λ^c} the restriction of ρ to Ω_{Λ^c}, i.e., $\rho_{\Lambda^c} = \rho \circ \phi_{\Lambda^c}^{-1}$. Then*

$$\int_\Omega f(x)\rho(dx) = \int_{\Omega_{\Lambda^c}} f(x_{\Lambda^c})\rho_{\Lambda^c}(dx_{\Lambda^c}).$$

Proof. This follows from a simple computation:

$$\int_{\Omega_{\Lambda^c}} f(x_{\Lambda^c})\rho_{\Lambda^c}(dx_{\Lambda^c}) = \int_\Omega f(\phi_{\Lambda^c}x)\rho(dx) = \int_\Omega f(x)\rho(dx).$$

□

Proof of Theorem 5.10. We are going to show that ν is consistent with a quasi-local non-null specification $\gamma = \{\gamma_\Lambda\}$, given by

$$\gamma_\Lambda(y_\Lambda y_{\Lambda^c}) = \int_{\Omega_{y_{\Lambda^c}}} \Big[\sum_{\bar{x}_\Lambda \in \pi^{-1} y_\Lambda} \gamma_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c}) \Big] \mu_{y_{\Lambda^c}}(dx). \tag{5.3}$$

Here

$$\Omega_{y_{\Lambda^c}} = \prod_{n \in \Lambda^c} \mathscr{A}_{y_n}, \quad \mathscr{A}_{y_n} = \{x_n \in \mathscr{A} : \pi(x_n) = y_n\}$$

is the fibre over y_{Λ^c} for the single-site factor map $\pi_{\Lambda^c} : \mathscr{A}^{\Lambda^c} \to \mathscr{B}^{\Lambda^c}$ and $\mu_{y_{\Lambda^c}}$ is the restriction of μ_y to $\Omega_{y_{\Lambda^c}}$. We will proceed in two main steps: first we show that the continuity and non-nullness requirements are satisfied. Then we show consistency with the factor measure ν.

Let us start by establishing the non-nullness and continuity of the functions $\{\gamma_\Lambda : \Lambda \Subset \mathbb{Z}^d\}$. Firstly, the fact that γ_Λ's are uniformly non-null follows immediately from the non-nullness of a Gibbsian specification $\{\gamma_\Lambda^\Phi\}$ of μ. Indeed, since

$$\sum_{\bar{x}_\Lambda \in \mathscr{A}^\Lambda} \gamma_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c}) = \sum_{y_\Lambda \in \mathscr{B}^\Lambda} \sum_{\bar{x}_\Lambda \in \pi^{-1} y_\Lambda} \gamma_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c}) = 1,$$

and

$$0 < a_\Lambda^\Phi \le \gamma_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c}) \le b_\Lambda^\Phi < 1$$

for all $\bar{x}_\Lambda$ and x_{Λ^c}, one has expression (5.3) implies that that for all $y \in \Sigma$ and every $\Lambda \Subset \mathbb{Z}^d$

$$\gamma_\Lambda(y_\Lambda y_{\Lambda^c}) \in \Big[\inf_{x_{\Lambda^c}} \sum_{\bar{x}_\Lambda \in \pi^{-1} y_\Lambda} \gamma_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c}), 1 - \inf_{x_{\Lambda^c}} \sum_{\bar{x}_\Lambda \in \pi^{-1} y_\Lambda} \gamma_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c}) \Big]$$

and hence bounded away from zero and one, i.e., $\gamma = \{\gamma_\Lambda\}$ is uniformly non-null.

Secondly, $\gamma_\Lambda(y) = \gamma_\Lambda(y_\Lambda y_{\Lambda^c})$ depends continuously on $y \in \Sigma$. Let us show that the map

$$y_{\Lambda^c} \mapsto \int_{\Omega_{y_{\Lambda^c}}} f(x) \mu_{y_{\Lambda^c}}(dx) =: \mu_{y_{\Lambda^c}}(f).$$

is continuous on Σ_{Λ^c} for all $f \in C(\Omega_{\Lambda^c})$. It suffices to check this for indicator functions of cylinder sets. Suppose $f(x) = \mathbb{1}_{[a_W]}(x)$ for some finite set $W \subset \Lambda^c$. Then

$$\mu_{y_{\Lambda^c}}(f) = \mu_{y_{\Lambda^c}}([a_W]) = \mu_y(\phi_{\Lambda^c}^{-1}[a_W]).$$

The preimage of $[a_W]$ under $\phi_{\Lambda^c}^{-1}$ in Ω_y is the union of disjoint cylinders

$$\phi_{\Lambda^c}^{-1}[a_W] = \bigsqcup_{z_\Lambda \in \pi^{-1} y_\Lambda} [z_\Lambda a_W].$$

Hence,

$$\mu_{y_{\Lambda^c}}(f) = \sum_{z_\Lambda \in \pi^{-1} y_\Lambda} \int_{\Omega_y} \mathbb{1}_{[z_\Lambda a_W]}(x)\mu_y(dx)$$

and thus depends continuously on y_{Λ^c}, since μ_y depends continuously on y and therefore it is a finite sum of continuous functions. Finally, the function

$$f(x_{\Lambda^c}) = \sum_{\bar{x}_\Lambda \in \pi^{-1} y_\Lambda} \gamma_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c})$$

is continuous on $\Omega_{y_{\Lambda^c}}$, and hence $\gamma_\Lambda(y) = \gamma_\Lambda(y_\Lambda y_{\Lambda^c})$, given by (5.3), is continuous on Σ.

We now turn to showing that that for every $\Lambda \Subset \mathbb{Z}^d$, ν satisfies the corresponding DLR equations:

$$\int_\Sigma g(y)\nu(dy) = \int_\Sigma \Big[\sum_{\bar{y}_\Lambda \in \mathscr{B}^\Lambda} g(\bar{y}_\Lambda y_{\Lambda^c})\gamma_\Lambda(\bar{y}_\Lambda y_{\Lambda^c})\Big]\nu(dy).$$

for every $g \in C(\Sigma)$ and the function $\gamma_\Lambda : \Sigma \to (0,1)$ given above. Consider $g \in C(\Sigma)$, then

$$\begin{aligned}
\int_\Sigma g(y)\nu(dy) = \int_\Omega g \circ \pi(x)\mu(dx) &\overset{(1)}{=} \int_\Omega \Big[\sum_{\bar{x}_\Lambda \in \mathscr{A}^\Lambda} g \circ \pi(\bar{x}_\Lambda x_{\Lambda^c})\gamma_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c})\Big]\mu(dx) \\
&\overset{(2)}{=} \int_\Sigma \left(\int_{\Omega_y} \Big[\sum_{\bar{x}_\Lambda \in \mathscr{A}^\Lambda} g \circ \pi(\bar{x}_\Lambda x_{\Lambda^c})\gamma_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c})\Big]\mu_y(dx)\right)\nu(dy) \\
&= \int_\Sigma \left(\int_{\Omega_y} \Big[\sum_{\bar{y}_\Lambda \in \mathscr{B}^\Lambda} \sum_{\bar{x}_\Lambda \in \pi^{-1}\bar{y}_\Lambda} g \circ \pi(\bar{x}_\Lambda x_{\Lambda^c})\gamma_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c})\Big]\mu_y(dx)\right)\nu(dy) \\
&= \int_\Sigma \left(\sum_{\bar{y}_\Lambda \in \mathscr{B}^\Lambda} g(\bar{y}_\Lambda y_{\Lambda^c}) \int_{\Omega_y} \Big[\sum_{\bar{x}_\Lambda \in \pi^{-1}\bar{y}_\Lambda} \gamma_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c})\Big]\mu_y(dx)\right)\nu(dy) \\
&= \int_\Sigma \left(\sum_{\bar{y}_\Lambda \in \mathscr{B}^\Lambda} g(\bar{y}_\Lambda y_{\Lambda^c}) \int_{\Omega_{y_{\Lambda^c}}} \Big[\sum_{\bar{x}_\Lambda \in \pi^{-1}\bar{y}_\Lambda} \gamma_\Lambda^\Phi(\bar{x}_\Lambda x_{\Lambda^c})\Big]\mu_{y_{\Lambda^c}}(dx)\right)\nu(dy) \\
&= \int_\Sigma \sum_{\bar{y}_\Lambda \in \mathscr{B}^\Lambda} g(\bar{y}_\Lambda y_{\Lambda^c})\gamma_\Lambda(\bar{y}_\Lambda y_{\Lambda^c})\nu(dy).
\end{aligned}$$

where in (1) we have used that μ is Gibbs for Φ, and in (2) that $\{\mu_y\}$ is a family of conditional probabilities for μ on fibres $\{\Omega_y\}$. Hence, we proved that ν satisfies the DLR equations with quasilocal and non-null specifications $\{\gamma_\Lambda\}$, and hence by Theorem 5.4, the measure ν is Gibbs. □

Remark 5.12. In principle we need to know that the collection $\{\gamma_\Lambda\}_{\Lambda\Subset\mathbb{Z}^d}$ satisfies the requirements for a specification. However, those requirements are chosen to reflect properties of conditional expectations and are in fact immediate from consistency. In particular, as $\gamma_\Lambda(A|x) = \mathbb{E}_\nu(A|\mathfrak{B}_{\Lambda^c})(x)$ ν—a.e., measurability with respect to $\mathfrak{B}_{\Lambda^c}$, as well as properness, follow form the definition of a conditional expectation, while consistency corresponds to the tower property for conditional expectations.

5.4 Tjur points

For factors of Gibbs measures conditional measure disintegrations are guaranteed to exist, however, as the approach is not constructive, it is difficult to obtain properties of the corresponding conditional measures. In order to relate Theorem 5.10 to the concept of hidden phase transitions we will first recall a constructive approach to measure disintegrations developed by Tjur [85, 86].

Suppose $y_0 \in \Sigma$. Denote by D_{y_0} the set of pairs (V, B), where V is an open neighbourhood of y_0 and B is a measurable subset of V such that $\nu(B) > 0$. A pair (V_1, B_1) is said to be ***closer*** to y_0 than (V_2, B_2), denoted by $(V_1, B_1) \succcurlyeq (V_2, B_2)$ if $V_1 \subseteq V_2$. This relation gives a partial order on D_{y_0}. Moreover, $(D_{y_0}, \succcurlyeq)$ is upwards directed: for any two elements in D_{y_0} there exists a third element, which is closer to y_0 than both of them. For $(V, B) \in D_{y_0}$ define a measure μ^B on Ω as the conditional measure on $\pi^{-1}B$:

$$\mu^B(\cdot) = \mu(\cdot \,|\, \pi^{-1}B).$$

The set $\{\mu^B(\cdot) \,|\, (V, B) \in D_{y_0}\}$ is a ***net***, or a ***generalized sequence***, in the space of probability measures on Ω.

Definition 5.13. If the limit (in the sense of net convergence described below)

$$\mu^{y_0} = \lim_{D_{y_0} \ni (V,B) \uparrow \infty} \mu^B \tag{5.4}$$

exists and belongs to a set of probability measures on Ω, then μ^{y_0} is called the conditional distribution of x, given $\pi(x) = y_0$; and we say that y_0 is a Tjur point.

The limit of a generalized sequence is understood in the net convergence sense, in the weak topology: More specifically, the limit μ^{y_0} exists if, for any $\varepsilon > 0$ and

$f \in C(\Omega)$, there exists an open neighbourhood V of y_0 such that for any $B \subseteq V$ with $\nu(B) > 0$ one has

$$\left|\int f(x)\mu^B(dx) - \int f(x)\mu^{y_0}(dx)\right| < \varepsilon.$$

Definition 5.13 requires that the limit, if it exists, is a probability measure. For compact spaces Ω this is automatic.

The limiting distributions $\{\mu^y\}$, when they exist, are constructed with the explicit hope to be the fibre measures in a conditional measure disintegration of μ. As mentioned earlier, any disintegration $\{\mu_y\}$ of μ is defined ν-almost everywhere. Therefore, if the limiting distributions $\{\mu^y\}$ are defined ν-a.e., i.e., the set of Tjur points has full ν-measure, one should hope that $\{\mu^y\}$ could constitute a valid disintegration of μ. Indeed, this is true, as the following result shows.

Theorem 5.14. [86, Theorem 5.1] *Suppose the measures $\{\mu^y\}$, as defined in (5.4), exist for almost all $y \in \Sigma$. Then for any integrable $f \in L^1(\Omega, \mu)$, f is μ^y-integrable for almost all y, and the function $y \mapsto \int_{\Omega_y} f\, d\mu^y$ is ν-integrable; furthermore*

$$\int_\Omega f(x)\mu(dx) = \int_\Sigma \Big[\int_\Omega f(x)\mu^y(dx)\Big]\nu(dy).$$

Continuity will be guaranteed by

Theorem 5.15. [86, Theorem 4.1] *Denote by Σ_0 the set of all Tjur points in Σ. Then the map*

$$y \mapsto \mu^y$$

is continuous on Σ_0.

The key result is the following criterion for existence of a continuous measure disintegration, adapted to our situation.

Theorem 5.16. [86, Theorem 7.1] *Suppose $\pi : \Omega \to \Sigma$ is a continuous, surjective map and μ is a Radon probability measure on Ω. The following conditions are equivalent:*

(i) the family of measures $\{\mu_y\}$ on fibres $\{\Omega_y\}$ constitute a continuous disintegration of μ (c.f., Definition 5.9);

(ii) conditional distributions μ^y are defined for all $y \in \Sigma$ and $\mu^y = \mu_y$, for all $y \in \Sigma$.

Proof. Since [86] is not readily available, for convenience we reproduce the proof of the statement. Firstly, assume that $\{\mu_y\}$ is a continuous disintegration of μ. Suppose $B \subseteq \Sigma$ has positive measure, $\nu(B) > 0$, and $f \in C(\Omega)$. Then

$$\begin{aligned}\int f(x)\mu^B(dx) &= \frac{1}{\nu(B)}\int_{\pi^{-1}B} f(x)\mu(dx) = \frac{1}{\nu(B)}\int_\Omega f(x)\mathbb{1}_{\pi^{-1}B}(x)\mu(dx)\\ &= \frac{1}{\nu(B)}\int_\Sigma\Big[\int_{\Omega_y} f(x)\mathbb{1}_{\pi^{-1}B}(x)\mu_y(dx)\Big]\nu(dy)\\ &== \frac{1}{\nu(B)}\int_\Sigma \mathbb{1}_B(y)\Big[\int_{\Omega_y} f(x)\mu_y(dx)\Big]\nu(dy)\\ &= \frac{1}{\nu(B)}\int_B\Big[\int_{\Omega_y} f(x)\mu_y(dx)\Big]\nu(dy).\end{aligned}$$

Note that the function $y \mapsto \int_{\Omega_y} f(x)\mu_y(dx) =: \mu_y(f)$ is assumed to be continuous. Furthermore, one can use standard arguments to show that if h is a continuous function on Σ, then

$$\frac{1}{\nu(B)}\int_B h(y)\nu(dy) \to h(y_0)$$

for any sequence of positive ν-measure sets B tending to y_0.

In the opposite direction, assume that μ^y exists for every $y \in \Sigma$. By Theorem 5.14, for any μ-integrable function f

$$\int_\Omega f(x)\mu(dx) = \int_\Sigma\Big[\int_{\Omega_y} f(x)\mu^y(dx)\Big]\nu(dy).$$

Furthermore, by Theorem 5.15, the map $y \mapsto \mu^y$ is continuous. It remains to show that μ^y is supported on Ω_y. Suppose $g \in C(\Sigma)$ and $\nu(B) > 0$. Then, since $\nu = \mu \circ \pi^{-1}$, one has

$$\frac{1}{\nu(B)}\int_{\pi^{-1}B} g(\pi(x))\mu(dx) = \frac{1}{\nu(B)}\int_B g(y)\nu(dy).$$

For B's tending to y_0, we obtain from the above equality that

$$\mu^{y_0}(g\circ\pi) := \int_\Omega g\circ\pi(x)\mu^{y_0}(dx) = g(y_0)\int_\Omega \mu^{y_0}(dx) = g(y_0).$$

Therefore, $\mu^{y_0} \circ \pi^{-1} = \delta_{y_0}$, and hence, μ^{y_0} is supported on $\Omega_{y_0} = \pi^{-1}(y_0)$. We have established that that $\{\mu^y\}$ is a valid disintegration of μ; $\mu^y(\Omega_y) = 1$ for all y, and the map $y \mapsto \int_{\Omega_y} f\, d\mu^y$ is continuous. Hence, $\{\mu^y\}$ is a continuous measure disintegration of μ. Since any measure admits at most one continuous disintegration, we immediately conclude that $\mu^y = \mu_y$ for all $y \in \Sigma$.

□

5.4.1 Lattice systems

The notion of Tjur points is useful in rather general settings. For lattice systems like $\Omega = \mathscr{A}^{\mathbb{Z}^d}$ one can derive a number of additional properties. Of particular importance are the cylinder sets, as they form a basis of the product topology in lattice systems, they are are clopen (both closed and open) and often convenient to work with. For our purposes, it is sufficient to consider only symmetric cylinder sets: namely, let $\mathbb{L}_n = [-n, n]^d \subset \mathbb{Z}^d$, $n \geq 1$, and denote by C_n^y the corresponding cylinder

$$C_n^y = \left\{\tilde{y} \in \Sigma : \quad \tilde{y}_{\mathbb{L}_n} = y_{\mathbb{L}_n}\right\}.$$

We will call n the size of cylinder C_n^y.

As we have seen above y_0 is a Tjur point, equivalently, the limit $\mu^{y_0} \in \mathscr{M}(\Omega)$ is defined, if for any $\varepsilon > 0$ and every $f \in C(\Omega)$, there exists an open neighbourhood V of y_0 such that for any $B \subseteq V$ with $\nu(B) > 0$ one has

$$\left|\int f(x)\mu^B(dx) - \int f(x)\mu^{y_0}(dx)\right| < \varepsilon.$$

Without loss of generality, one can substitute “there exists an open neighbourhood V” with a condition “there exists a cylinder set $V = C_n^{y_0}$ for some $n \geq 1$”. Ideally we would like to relate Tjur points to the convergence of conditional measures of the form $\mu^{C_n^y}$, however, it turns out an additional requirement of the convergence being uniform in $y \in \Sigma$ is needed:

Theorem 5.17. *For $\Omega = \mathscr{A}^{\mathbb{Z}^d}, \Sigma = \mathscr{B}^{\mathbb{Z}^d}$ and $\pi : \Omega \to \Sigma$ a 1-block factor and μ a Gibbs measure and $\nu = \mu \circ \pi^{-1}$, then the following conditions are equivalent*

(1) Every point $y \in \Sigma$ is a Tjur point; i.e., the limiting conditional distribution μ^y exists for all y;

(2) for all $y \in \Sigma$, the sequence of measures $\mu^{C_n^y}$ converges as $n \to \infty$ (to the limit μ^y), and the convergence is uniform in y: for every $\varepsilon > 0$ and $f \in C(\Omega)$ there exists $N \geq 1$ such that for all $n \geq N$,

$$\left|\int f(x)\mu^{C_n^y}(dx) - \int f(x)\mu^y(dx)\right| < \varepsilon \tag{5.5}$$

for all $y \in \Sigma$.

Proof. Fix $\varepsilon > 0$ and $f \in C(\Omega)$. An open neighbourhood V of y is called (ε, f)-good if for every $B \subseteq V$ with $\nu(B) > 0$ one has

$$\left| \int f(x)\mu^B(dx) - \int f(x)\mu^y(dx) \right| < \varepsilon.$$

Clearly, if V is an (ε, f)-good open neighbourhood of y, and V_1 is another open neighbourhood of y such that $V_1 \subset V$, then V_1 is (ε, f)-good as well. Note that a point $y \in \Sigma$ is a Tjur point if and only if for all $\varepsilon > 0$ and $f \in C(\Omega)$ there exists an (ε, f)-good neighbourhood of y.

Now suppose all $y \in \Sigma$ are Tjur points then, for every $y \in \Sigma$, there exists an (ε, f)-good open neighbourhood V_y of y. Thus for some $n_y \in \mathbb{N}$, the cylinder $C^y_{n_y} \subset V_y$ is (ε, f)-good as well. Therefore, we have a cover of Ω by cylinder sets $\{C^y_{n_y} | \ y \in \Sigma\}$. Since Ω is compact one can select some finite subcover, say $\{C^y_{n_y} | \ y \in E\}$, where $|E| < \infty$. Let $N = \max_{y \in E} n_y \in \mathbb{N}$. For those $y \in E$ with $n_y < N$, we can refine the corresponding cylinder $C^y_{n_y}$, and substitute it by a partition into a number of disjoint N-cylinders in our finite cover. This way we obtain a finite cover of Ω by cylinders of the form C^y_N, $y \in E'$, with $|E'| < \infty$. Note that all cylinders $\{C^y_N | \ y \in E'\}$ remain (ε, f)-good, since any refinement of an (ε, f)-good cylinder is an (ε, f)-good cylinder. In fact, by removing duplicate cylinders, we conclude that all cylinders of size N form a partition of Ω into (ε, f)-good sets. Consider an arbitrary $y \in \Sigma$, since the corresponding cylinder of size N, C^y_N is (ε, f)-good, we conclude that for all $n \geq N$ (5.5) holds, and hence the convergence $\mu^{C^y_n} \to \mu_y$ is uniform.

In the opposite direction, suppose for every y, the limit

$$\lim_{n \to \infty} \mu^{C^y_n} := \mu^y$$

exists, and the convergence is uniform in y. Fix $\varepsilon > 0$ and $f \in C(\Omega)$. Let $N \geq 1$ be such that (5.5) holds. We will show that every cylinder is $(2\varepsilon, f)$-good. Consider an arbitrary $y \in \Sigma$, and an arbitrary set $B \subset C^y_N$ with $\nu(B) > 0$. For any $\delta > 0$ (to be specified latter), the set B can be approximated by a disjoint union of a finite number of cylinder sets

$$C = \bigsqcup_{k=1}^{K} C^{y_k}_{m_k}, \quad y_k \in \Sigma, m_k \geq N, \ k = 1, \ldots, K.$$

such that

$$\nu(B \triangle C) < \delta. \tag{5.6}$$

Note that without loss of generality, we may assume that $C_{m_k}^{y_k} \subseteq C_N^y$. Indeed, if this is not the case, then $y|_{\Lambda_N} \neq y_k|_{\Lambda_N}$. But that means that $C_N^{y_k} \cap C_N^y = \varnothing$, and since $B \subset C_N^y$, cylinder $C_{m_k}^{y_k} \subseteq C_N^{y_k}$ can be removed, without jeopardizing the quality of approximation in (5.6).

Furthermore, one has

$$\left| \int_{\pi^{-1}B} f(x)\mu(dx) - \int_{\pi^{-1}C} f(x)\mu(dx) \right| \leq ||f||\mu(\pi^{-1}B \triangle \pi^{-1}C) < ||f||\nu(B\triangle C) < \delta ||f||,$$

and hence

$$\left| \frac{1}{\nu(B)} \int_{\pi^{-1}B} f(x)\mu(dx) - \frac{1}{\nu(B)} \int_{\pi^{-1}C} f(x)\mu(dx) \right| \leq \delta \frac{||f||}{\nu(B)}.$$

Moreover,

$$\left| \frac{1}{\nu(B)} \int_{\pi^{-1}C} f(x)\mu(dx) - \frac{1}{\nu(C)} \int_{\pi^{-1}C} f(x)\mu(dx) \right| \leq \left| \frac{1}{\nu(B)} - \frac{1}{\nu(C)} \right| \nu(C)||f|| \leq \delta \frac{||f||}{\nu(B)}.$$

Therefore,

$$\left| \frac{1}{\nu(B)} \int_{\pi^{-1}B} f(x)\mu(dx) - \frac{1}{\nu(C)} \int_{\pi^{-1}C} f(x)\mu(dx) \right| \leq 2\delta \frac{||f||}{\nu(B)}.$$

Introduce the shorthand notation $C_k = C_{m_k}^{y_k}$, $k = 1, \ldots, K$. Note that we argued above that $C_k \subseteq C_N^y$ for all k. One has

$$\begin{aligned}
&\left| \frac{1}{\nu(C)} \int_{\pi^{-1}C} f(x)\mu(dx) - \int f(x)\mu^y(dx) \right| \\
&\quad = \left| \sum_{k=1}^{N} \frac{\nu(C_k)}{\nu(C)} \left[\frac{1}{\nu(C_k)} \int_{\pi^{-1}C_k} f(x)\mu(dx) - \int f(x)\mu^y(dx) \right] \right| < \sum_{k=1}^{N} \frac{\nu(C_k)}{\nu(C)} \varepsilon = \varepsilon.
\end{aligned}$$

Therefore, combining all inequalities we conclude that

$$\begin{aligned}
&\left|\frac{1}{\nu(B)}\int\limits_{\pi^{-1}B} f(x)\mu(dx)-\int f(x)\mu^y(dx)\right| \\
&\quad\leq\left|\frac{1}{\nu(B)}\int\limits_{\pi^{-1}B} f(x)\mu(dx)-\frac{1}{\nu(C)}\int\limits_{\pi^{-1}C} f(x)\mu(dx)\right| \\
&\quad\quad+\left|\frac{1}{\nu(C)}\int\limits_{\pi^{-1}C} f(x)\mu(dx)-\int f(x)\mu^y(dx)\right| < 2\delta\frac{||f||}{\nu(B)}+\varepsilon.
\end{aligned}$$

If we let $\delta = \frac{1}{2}\frac{\nu(B)\varepsilon}{||f||+1}$, we conclude that C_N^y is $(2\varepsilon, f)$-good, and hence the set of all Tjur points is Σ. □

Remark 5.18. It is interesting to investigate whether (and under which conditions) one is able to drop the a priori requirement that the conditional measures $\mu^{C_n^y}$ converge uniformly. For example, the following two requirements are also sufficient:

- for every $y \in \Sigma$, $\mu^{C_n^y} \to \mu^y$;
- the family of measures $\{\mu^y : y \in \Sigma\}$ is continuous.

It is well known that pointwise convergence of continuous functions is not sufficient for the limit to be continuous; uniform convergence does imply continuity of the limit. A little less known is the condition of ***quasi-uniform convergence*** which, by the Arzelá–Aleksandrov theorem, is equivalent to the requirement that a limit of continuous functions is continuous.

To be precise, let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of functions from a topological space Ω into a metric space Σ converging pointwise to f, we call it quasi-uniformly convergent if for any $\varepsilon > 0$ and any integer $N > 0$ there exists an open cover $\{V_i\}_{i\in I}$, $I \subset \mathbb{N}$ of Ω and $n_i > N$, for $i \in I$, such that $d(f(x), f_{n_i}(x)) < \varepsilon$ for every $x \in V_i$.

5.5 Limiting conditional distributions and hidden phase transitions

Combining the main results of the previous section: Theorems 5.16 and 5.17, allows us to formulate the necessary and sufficient conditions for existence of a continuous measure disintegration in terms of convergence of measures without the use of Tjur points, i.e., by considering conditional probabilities, conditioned on cylindric events.

Theorem 5.19. *The measure μ on Ω admits a continuous disintegration $\{\mu_y\}$ under $\pi:\Omega\to\Sigma$ if and only if conditional measures $\{\mu^{C_n^y}\}_{n\geq 1}$ converge as $n\to\infty$ uniformly.*

On the other hand, validating uniform convergence in specific examples is not always a straightforward task. To come closer to the van Enter-Fernandez-Sokal criterion on preservation of Gibbs property under renormalisation in the absence of hidden phase transitions, we will study sets of all possible limiting distributions.

As in the previous section, fix $y\in\Sigma$ and consider the net of conditional measures

$$\mathcal{N}_y=\Big\{\mu^B(\cdot)=\mu(\cdot|\pi^{-1}B):\ (V,B)\in D_y\Big\}.$$

Definition 5.20. A measure $\tilde{\mu}$ is an accumulation point of the net $\mathcal{N}_y$ if for all $f\in C(\Omega)$, $\varepsilon>0$, and for every open set V containing y, there exists a set $B\subseteq V$, $\nu(B)>0$, such that

$$\left|\int f(x)\mu^B(dx)-\int f(x)\tilde{\mu}(dx)\right|<\varepsilon.$$

Denote by $\overline{\mathfrak{M}}_y$ the set of all possible accumulation points of $\mathcal{N}_y$. Clearly, since Ω is compact, $\overline{\mathfrak{M}}_y$ is not empty. It turns out that all accumulation points are in fact Gibbs states:

Theorem 5.21. *For every $y\in\Sigma$ the following holds:*

(a) $\overline{\mathfrak{M}}_y\neq\varnothing$ and for every $\lambda_y\in\overline{\mathfrak{M}}_y$, $\lambda_y(\Omega_y)=1$.

(b) Suppose μ is a Gibbs measure on Ω for potential Φ, then $\overline{\mathfrak{M}}_y\subseteq\mathcal{G}_{\Omega_y}(\Phi)$, where $\mathcal{G}_{\Omega_y}(\Phi)$ is the set of Gibbs states on Ω_y for potential Φ.

Proof. First we show (a). From compactness it follows that $\overline{\mathfrak{M}}_y\neq\varnothing$. Now, let $\lambda_y\in\overline{\mathfrak{M}}_y$ then, for any cylinder set $C\subset\Omega$ such that $C\cap\Omega_y=\emptyset$ there exists an $n>0$ such that $y_{\mathbb{L}_n}\cap C=\emptyset$. As $y_{\mathbb{L}_n}$ is an open set containing y, it follows that $\lambda_y(C)=0$ and therefore λ_y is concentrated on Ω_y. Moreover, by compactness, $\lambda_y(\Omega)=1$, hence $\lambda_y(\Omega_y)=1$.

We now turn to part (b). Consider an arbitrary measure $\lambda\in\overline{\mathfrak{M}}_y$. As before, let C_n^y be a sequence of cylinders. Since λ is an accumulation point, there exists a sequence of measurable sets $\{B_n\}$, $B_n\subset C_n^y$ and has positive ν-measure, such that

$$\mu_n=\mu(\cdot|\,\pi^{-1}B_n)\to\lambda$$

weakly in $\mathcal{M}^1(\Omega)$, as $n\to\infty$. Let $\Lambda\Subset\mathbb{Z}^d$, and consider $f(x)=\mathbb{1}_{[a_\Lambda]}(x)=\mathbb{1}(x_\Lambda=a_\Lambda)$ for some $a_\Lambda\in\mathscr{A}^\Lambda$. Since $\mu_n\to\lambda$ weakly, we have

$$\mu_n([a_\Lambda])\to\lambda([a_\Lambda]),\quad \text{as } n\to\infty.$$

Note that if $\pi(a_\Lambda)\neq y_\Lambda$, then $\lambda([a_\Lambda])=0$ since for all sufficiently large n (i.e., such that $\Lambda\subset\Lambda_n$) one has

$$[a_\Lambda]\cap\pi^{-1}B_n\subset[a_\Lambda]\cap\pi^{-1}[y_\Lambda]=\varnothing.$$

Otherwise, one has

$$\begin{aligned}\mu_n([a_\Lambda])=\frac{\mu([a_\Lambda]\cap\pi^{-1}B_n)}{\nu(B_n)}&=\frac{1}{\nu(B_n)}\int\mathbb{1}_{a_\Lambda}(x)\mathbb{1}_{\pi^{-1}B_n}(x)\mu(dx)\\&=\frac{1}{\nu(B_n)}\int\gamma_\Lambda(\mathbb{1}_{a_\Lambda}\cdot\mathbb{1}_{\pi^{-1}B_n}|x)\mu(dx)\qquad\text{(DLR eq's)}\\&=\frac{1}{\nu(B_n)}\int\gamma_\Lambda(a_\Lambda|x_{\Lambda^c})\mathbb{1}_{\pi^{-1}B_n}(a_\Lambda x_{\Lambda^c})\mu(dx)\\&\geq\frac{\alpha_\Lambda}{\nu(B_n)}\int\mathbb{1}_{\pi^{-1}B_n}(a_\Lambda x_{\Lambda^c})\mu(dx).\end{aligned}$$

Here α_Λ is the lower bound of γ_Λ, which exists by non-nullness.

If $x\in\pi^{-1}B_n$ and n is large enough (i.e., $\Lambda\subset\Lambda_n$), then $a_\Lambda x_{\Lambda^c}\in\pi^{-1}B_n$ as well. Hence we can continue as

$$\mu_n([a_\Lambda])\geq\frac{\alpha_\Lambda}{\nu(B_n)}\int_{\pi^{-1}B_n}\mu(dx)=\frac{\alpha_\Lambda}{\nu(B_n)}\mu(\pi^{-1}B_n)=\alpha_\Lambda>0. \tag{5.7}$$

Thus $\lambda([a_\Lambda])>0$ for all $\Lambda\Subset\mathbb{Z}^d$ and $a_\Lambda\in\pi^{-1}y_\Lambda$. Now fix an arbitrary $\Lambda\Subset\mathbb{Z}^d$ and let k be such that $\Lambda\subsetneq\mathbb{L}_k$, and an arbitrary $a\in\pi^{-1}y$. Our goal is to estimate the conditional probability $\lambda(a_\Lambda|a_{\mathbb{L}_k\setminus\Lambda})$ (well-defined by the above argument) and to show that

$$\sup_{a\in\Omega}\left|\lambda(a_\Lambda|a_{\mathbb{L}_k\setminus\Lambda})-\frac{\gamma^\Phi_\Lambda(a_\Lambda|a_{\Lambda^c})}{\sum\limits_{\bar a_\Lambda\in\pi^{-1}y_\Lambda}\gamma^\Phi_\Lambda(\bar a_\Lambda|a_{\Lambda^c})}\right|\to 0 \tag{5.8}$$

as $k\to\infty$. This implies that the measure λ on Ω_y is Gibbs with the corresponding specification

$$\gamma^y_\Lambda(a_\Lambda|a_{\Lambda^c}):=\frac{\gamma^\Phi_\Lambda(a_\Lambda|a_{\Lambda^c})}{\sum\limits_{\bar a_\Lambda\in\pi^{-1}y_\Lambda}\gamma^\Phi_\Lambda(\bar a_\Lambda|a_{\Lambda^c})},$$

i.e., λ is Gibbs on Ω_y for the original potential Φ. Consider

$$\mu_n(a_\Lambda\,|\,a_{\mathbb{L}_k\setminus\Lambda})=\frac{\mu_n([a_\Lambda a_{\mathbb{L}_k}])}{\mu_n([a_{\mathbb{L}_k\setminus\Lambda}])}=\frac{\mu([a_\Lambda a_{\mathbb{L}_k}]\cap\pi^{-1}B_n)}{\mu([a_{\mathbb{L}_k\setminus\Lambda}]\cap\pi^{-1}B_n)}.$$

We are going to show that for all sufficiently large n

$$\sup_{a\in\Omega_y} |\mu_n(a_\Lambda \,|\, a_{\mathbb{L}_k\setminus\Lambda}) - \gamma^y_\Lambda(a_\Lambda|a_{\Lambda^c})| = u_{\Lambda,k} + v_n,$$

where $v_n \to 0$ and $u_{\Lambda,k} \to 0$ as $k \to \infty$. This bound, together with the weak convergence $\mu_n \to \lambda$ will imply the desired conclusion (5.8). We can approximate the measurable set $B_n \subset \Sigma$ by a disjoint union of cylindric events $\sqcup_{j=1}^{M_n} C_j^{(n)}$ such that

$$\nu\left(B_n \,\triangle \bigsqcup_{j=1}^{M_n} C_j^{(n)}\right) \le \delta_n \nu(B_n), \tag{5.9}$$

where $\{\delta_n\}$ is a sequence converging to 0. Since $B_n \subset [y_{\mathbb{L}_n}]$, without loss of generality we may also assume that $C_j^{(n)} \subset [y_{\mathbb{L}_n}]$ for all j. Inequality (5.9) implies that

$$\nu\left(\bigsqcup_{j=1}^{M_n} C_j^{(n)}\right) \le \nu\left(\bigsqcup_{j=1}^{M_n} C_j^{(n)} \setminus B_n\right) + \nu(B_n) \le (1+\delta_n)\nu(B_n).$$

Similarly we get

$$\nu\left(\bigsqcup_{j=1}^{M_n} C_j^{(n)}\right) \ge (1-\delta_n)\nu(B_n),$$

and hence

$$\nu\left(B_n \,\triangle \bigsqcup_{j=1}^{M_n} C_j^{(n)}\right) \le \delta_n \nu(B_n) \le \frac{\delta_n}{1-\delta_n}\nu\left(\bigsqcup_{j=1}^{M_n} C_j^{(n)}\right) =: \delta'_n \nu\left(\bigsqcup_{j=1}^{M_n} C_j^{(n)}\right)$$

A well-known inequality states that for a probability measure $\mathbb{P}$ and any measurable events A, B, C, and D, one has

$$\mathbb{P}(A\cap B) \le \mathbb{P}(C\cap D) + \mathbb{P}(A \,\triangle\, C) + \mathbb{P}(B \,\triangle\, D).$$

Applying this inequality and using that $\nu = \mu \circ \pi^{-1}$, we conclude that

$$\begin{aligned}\mu([a_{\mathbb{L}_k}]\cap \pi^{-1}B_n) &\le \mu\left([a_{\mathbb{L}_k}]\cap \pi^{-1}\sqcup_{j=1}^{M_n} C_j^{(n)}\right) + \nu\left(B_n \,\triangle\, \sqcup_{j=1}^{M_n} C_j^{(n)}\right)\\ &\le \mu\left([a_{\mathbb{L}_k}]\cap \pi^{-1}\sqcup_{j=1}^{M_n} C_j^{(n)}\right) + \delta'_n \mu\left(\pi^{-1}\sqcup_{j=1}^{M_n} C_j^{(n)}\right)\\ &= \sum_{j=1}^{M_n}\left\{\mu\left([a_{\mathbb{L}_k}] \mid \pi^{-1}C_j^{(n)}\right) + \delta'_n\right\}\mu\left(\pi^{-1}C_j^{(n)}\right).\end{aligned}$$

Similarly, we conclude that

$$\mu([a_{\mathbb{L}_k\setminus\Lambda}\cap \pi^{-1}B_n]) \ge \sum_{j=1}^{M_n}\left\{\mu\left([a_{\mathbb{L}_k\setminus\Lambda}] \mid \pi^{-1}C_j^{(n)}\right) - \delta'_n\right\}\mu\left(\pi^{-1}C_j^{(n)}\right). \tag{5.10}$$

In the first part of the proof (c.f, (5.7)) we have shown that for any finite Λ and all n such that $\Lambda_n \subset \Lambda$, the conditional probability $\mu([a_\Lambda]|\pi^{-1}B_n) \geq \alpha_\Lambda > 0$, i.e., it is bounded away from zero, uniformly in B_n. Therefore, $\mu\left([a_{\mathbb{L}_k}] \mid \pi^{-1}C_j^{(n)}\right) \geq \alpha_{\mathbb{L}_k} > 0$ and hence all the terms in (5.10) are positive. Thus

$$\begin{aligned}\frac{\mu([a_{\mathbb{L}_k}] \cap \pi^{-1}B_n)}{\mu([a_{\mathbb{L}_k\setminus\Lambda}] \cap \pi^{-1}B_n)} &\leq \frac{\sum_{j=1}^{M_n}\left\{\mu\left([a_{\mathbb{L}_k}] \mid \pi^{-1}C_j^{(n)}\right)+\delta_n'\right\}\mu\left(\pi^{-1}C_j^{(n)}\right)}{\sum_{j=1}^{M_n}\left\{\mu\left([a_{\mathbb{L}_k\setminus\Lambda}] \mid \pi^{-1}C_j^{(n)}\right)-\delta_n'\right\}\mu\left(\pi^{-1}C_j^{(n)}\right)} \\ &\leq \max_j \frac{\mu\left([a_{\mathbb{L}_k}] \mid \pi^{-1}C_j^{(n)}\right)+\delta_n'}{\mu\left([a_{\mathbb{L}_k\setminus\Lambda}] \mid \pi^{-1}C_j^{(n)}\right)-\delta_n'}.\end{aligned} \tag{5.11}$$

Let us now evaluate the quotient

$$\frac{\mu\left([a_{\mathbb{L}_k}] \mid \pi^{-1}C_j^{(n)}\right)}{\mu\left([a_{\mathbb{L}_k\setminus\Lambda}] \mid \pi^{-1}C_j^{(n)}\right)}$$

for a cylindric even $C_j^{(n)} \subset \Sigma$. Since $C_j^{(n)}$ is a cylindric event, there exists $y' \in \Sigma$ and $W \Subset \mathbb{Z}^d$ such that $C_j^{(n)} = [y_W']$. Note that by the construction, we have $\mathbb{L}_n \subset W$, and $C_j^{(n)} = [y_W'] \subset [y_{\mathbb{L}_n}]$, hence, $y'_{\mathbb{L}_n} = y_{\mathbb{L}_n}$. One has

$$\begin{aligned}\frac{\mu\left([a_{\mathbb{L}_k}] \mid \pi^{-1}C_j^{(n)}\right)}{\mu\left([a_{\mathbb{L}_k\setminus\Lambda}] \mid \pi^{-1}C_j^{(n)}\right)} &= \frac{\mu\left([a_{\mathbb{L}_k}] \cap \pi^{-1}C_j^{(n)}\right)}{\mu\left([a_{\mathbb{L}_k\setminus\Lambda}] \cap \pi^{-1}C_j^{(n)}\right)} \\ &= \frac{\sum\limits_{\tilde{x}_{W\setminus\mathbb{L}_k}\in\pi^{-1}y'_{W\setminus\mathbb{L}_k}} \mu(a_\Lambda a_{\mathbb{L}_k\setminus\Lambda}\tilde{x}_{W\setminus\mathbb{L}_k})}{\sum\limits_{\tilde{x}_{W\setminus\mathbb{L}_k}\in\pi^{-1}y'_{W\setminus\mathbb{L}_k}} \sum\limits_{\bar{a}_\Lambda\in\pi^{-1}y_\Lambda} \mu(\bar{a}_\Lambda a_{\mathbb{L}_k\setminus\Lambda}\tilde{x}_{W\setminus\mathbb{L}_k})} \\ &\leq \sup_{\tilde{x}_{W\setminus\mathbb{L}_k}\in\pi^{-1}y'_{W\setminus\mathbb{L}_k}} \frac{\mu(a_\Lambda a_{\mathbb{L}_k\setminus\Lambda}\tilde{x}_{W\setminus\mathbb{L}_k})}{\sum\limits_{\bar{a}_\Lambda\in\pi^{-1}y_\Lambda} \mu(\bar{a}_\Lambda a_{\mathbb{L}_k\setminus\Lambda}\tilde{x}_{W\setminus\mathbb{L}_k})} \\ &= \sup_{\tilde{x}_{W\setminus\mathbb{L}_k}\in\pi^{-1}y'_{W\setminus\mathbb{L}_k}} \frac{\mu(a_\Lambda | a_{\mathbb{L}_k\setminus\Lambda}\tilde{x}_{W\setminus\mathbb{L}_k})}{\sum\limits_{\bar{a}_\Lambda\in\pi^{-1}y_\Lambda} \mu(\bar{a}_\Lambda | a_{\mathbb{L}_k\setminus\Lambda}\tilde{x}_{W\setminus\mathbb{L}_k})}.\end{aligned}$$

Similarly,

$$\frac{\mu\left([a_{\mathbb{L}_k}] \mid \pi^{-1}C_j^{(n)}\right)}{\mu\left([a_{\mathbb{L}_k\setminus\Lambda}] \mid \pi^{-1}C_j^{(n)}\right)} \geq \inf_{\tilde{x}_{W\setminus\mathbb{L}_k}\in\pi^{-1}y'_{W\setminus\mathbb{L}_k}} \frac{\mu(a_\Lambda | a_{\mathbb{L}_k\setminus\Lambda}\tilde{x}_{W\setminus\mathbb{L}_k})}{\sum\limits_{\bar{a}_\Lambda\in\pi^{-1}y_\Lambda} \mu(\bar{a}_\Lambda | a_{\mathbb{L}_k\setminus\Lambda}\tilde{x}_{W\setminus\mathbb{L}_k})}.$$

Gibbs measures are characterized by the uniform convergence of the finite dimensional conditional probabilities (c.f., Theorem 5.5). Thus, for any $\Lambda \subset \mathbb{L}_k \Subset \mathbb{Z}^d$

$$e_{\Lambda,k} = \sup_{a,\tilde{x},\bar{x},W} \left|\mu(a_\Lambda | a_{\mathbb{L}_k\setminus\Lambda} \tilde{x}_{W\setminus\mathbb{L}_k}) - \gamma_\Lambda^\Phi(a_\Lambda | a_{\mathbb{L}_k} \bar{x}_{\mathbb{L}_k{}^c})\right| \to 0$$

as $k \to \infty$. Therefore, one can conclude that

$$e_{\Lambda,k} := \sup_{a,\tilde{x},\bar{x},W} \left| \frac{\mu(a_\Lambda | a_{\mathbb{L}_k\setminus\Lambda} \tilde{x}_{W\setminus\mathbb{L}_k})}{\sum\limits_{\bar{a}_\Lambda \in \pi^{-1} y_\Lambda} \mu(\bar{a}_\Lambda | a_{\mathbb{L}_k\setminus\Lambda} \tilde{x}_{W\setminus\mathbb{L}_k})} - \frac{\gamma_\Lambda^\Phi(a_\Lambda | a_{\mathbb{L}_k\setminus\Lambda} \bar{x}_{W\setminus\mathbb{L}_k})}{\sum\limits_{\bar{a}_\Lambda \in \pi^{-1} y_\Lambda} \gamma_\Lambda^\Phi(\bar{a}_\Lambda | a_{\mathbb{L}_k\setminus\Lambda} \bar{x}_{W\setminus\mathbb{L}_k})} \right| \to 0, \text{ as } k \to \infty.$$

This allows us to conclude that for any cylindric event $C_j^{(n)}$ satisfying the conditions above, one has

$$\frac{\mu\left([a_{\mathbb{L}_k}] \mid \pi^{-1} C_j^{(n)}\right)}{\mu\left([a_{\mathbb{L}_k\setminus\Lambda}] \mid \pi^{-1} C_j^{(n)}\right)} - \frac{\gamma_\Lambda^\Phi(a_\Lambda | a_{\mathbb{L}_k\setminus\Lambda} \bar{x}_{W\setminus\mathbb{L}_k})}{\sum\limits_{\bar{a}_\Lambda \in \pi^{-1} y_\Lambda} \gamma_\Lambda^\Phi(\bar{a}_\Lambda | a_{\mathbb{L}_k\setminus\Lambda} \bar{x}_{W\setminus\mathbb{L}_k})}$$

is uniformly small (in the event $C_j^{(n)}$, $\bar{x}$, etc.). Finally, taking into account (5.11), and the fact that Gibbsian specifications are uniformly non-null, we can conclude that for all sufficiently large n, there exists $\delta_n'' \to 0$, such that

$$\begin{aligned}
\mu_n(a_\Lambda | a_{\mathbb{L}_k\setminus\Lambda}) &\le \max_j \frac{\mu\left([a_{\mathbb{L}_k}] \mid \pi^{-1} C_j^{(n)}\right) + \delta_n'}{\mu\left([a_{\mathbb{L}_k\setminus\Lambda}] \mid \pi^{-1} C_j^{(n)}\right) - \delta_n'} \\
&\le \sup_{\bar{x}} \frac{\gamma_\Lambda^\Phi(a_\Lambda | a_{\mathbb{L}_k\setminus\Lambda} \bar{x}_{W\setminus\mathbb{L}_k})}{\sum\limits_{\bar{a}_\Lambda \in \pi^{-1} y_\Lambda} \gamma_\Lambda^\Phi(\bar{a}_\Lambda | a_{\mathbb{L}_k\setminus\Lambda} \bar{x}_{W\setminus\mathbb{L}_k})} + \delta_n''.
\end{aligned}$$

Proceeding in completely similar fashion we can also conclude that for all sufficiently large n, there exists $\delta_n''' \to 0$, such that

$$\mu_n(a_\Lambda | a_{\mathbb{L}_k\setminus\Lambda}) \ge \inf_{\tilde{x}} \frac{\gamma_\Lambda^\Phi(a_\Lambda | a_{\mathbb{L}_k\setminus\Lambda} \tilde{x}_{W\setminus\Lambda_k})}{\sum\limits_{\bar{a}_\Lambda \in \pi^{-1} y_\Lambda} \gamma_\Lambda^\Phi(\bar{a}_\Lambda | a_{\mathbb{L}_k\setminus\Lambda} \tilde{x}_{W\setminus\mathbb{L}_k})} - \delta_n'''.$$

It follows that the limiting measure has conditional probabilities given by

$$\lambda(a_\Lambda | a_{\Lambda^c}) = \frac{\gamma_\Lambda^\Phi(a_\Lambda | a_{\Lambda^c})}{\sum\limits_{\bar{a}_\Lambda \in \pi^{-1} y_\Lambda} \gamma_\Lambda^\Phi(\bar{a}_\Lambda | a_{\Lambda^c})}$$

and is therefore a Gibbs measure. □

Now we are able to state two easy corollaries.

Corollary 5.22. *If $\nu = \mu \circ \pi^{-1}$ is a fuzzy Gibbs state and*

$$|\overline{\mathfrak{M}}_y| = 1$$

for all $y \in \Sigma$, then ν is Gibbs.

Clearly, $|\overline{\mathfrak{M}}_y| = 1$ is equivalent to all points in $y \in \Sigma$ being Tjur points, and hence we have a continuous measure disintegration $\{\mu_y\}_{y\in\Sigma}$, and thus ν is Gibbs by Theorem 5.10.

However, the sufficient conditions $|\overline{\mathfrak{M}}_y| = 1$ for all y is not easy to validate. Since $\overline{\mathfrak{M}}_y \subset \mathscr{G}_{\Omega_y}(\Phi)$ for all $y \in \Sigma$, we also have the following corollary.

Corollary 5.23. *If $\nu = \mu \circ \pi^{-1}$ is a fuzzy Gibbs state and*

$$|\mathscr{G}_{\Omega_y}(\Phi)| = 1$$

for all $y \in \Sigma$, then ν is Gibbs.

Remark 5.24. Corollary 5.22 should be viewed as the proof of the ***easy*** part of the general hypothesis of van Enter-Fernández-Sokal on the necessary and sufficient conditions for preservation of the Gibbs property under renormalisation transformations. This hypothesis is formulated as follows [81, page 977]: the loss of Gibbsianity occurs when

(i) the internal spins have a phase transition; in our notation, for some $y \in \Sigma$

$$|\mathscr{G}_{\Omega_y}(\Phi)| > 1;$$

(ii) by varying boundary conditions in Σ, one can pick different phases in $\mathscr{G}_{\Omega_y}(\Phi)$.

We argue that Corollary 5.22 provides a way to represent the hypothesis in a compact way

$$\nu \text{ is Gibbs if and only if } |\overline{\mathfrak{M}}_y| = 1 \quad \text{for all } y \in \Sigma. \tag{5.12}$$

It is very easy to construct an example when $|\mathscr{G}_{\Omega_y}(\Phi)| > 1$, but $|\overline{\mathfrak{M}}_y| = 1$ for all y, and hence ν is Gibbs. Take $\Omega = \Omega_1 \times \Omega_2$, $\Sigma = \Omega_1$, π is the projection from Ω to Ω_1. Furthermore, let $\Phi_2 \in B(\Omega_2)$ be such that $|\mathscr{G}_{\Omega_2}(\Phi_2)| > 1$, and $\Phi_1 \in B(\Omega_1)$ is arbitrary. Then any Gibbs measure μ on Ω for $\Phi = \Phi_1 + \Phi_2$ has a form $\mu = \mu_1 \times \mu_2$, where $\mu_i \in \mathscr{G}_{\Omega_i}(\Phi_i)$, $i = 1, 2$. Clearly, $\nu = \mu \circ \pi^{-1} = \mu_1$ is Gibbs, while for any $y \in \Sigma = \Omega_1$, $\Omega_y = \Omega_2$, and hence, $\mathscr{G}_{\Omega_y}(\Phi) = \mathscr{G}_{\Omega_2}(\Phi_2)$ is not a singleton.

5.6 Conclusions and Outlook

We have shown that the existence of a continuous measure disintegration is a sufficient condition for preservation of Gibbsianity under renormalisation. Using the concepts of Tjur points we showed that this condition indeed covers the sufficiency of the van Enter-Fernández-Sokal hypothesis.

Our results can be used to simplify existing proofs of preservation of Gibbs property under renormalisation in some examples. For example, Haller & Kennedy [44] obtained their main result on the decimation of the two-dimensional Ising model (for $J < 1.36J_c$ the decimated Ising state is Gibbs) from the condition that the collection of measures μ_y is in the high-temperature phase uniformly in the image spin configuration , $y \in \mathcal{B}^{\mathbb{Z}^2}$, and hence $|\mathcal{G}_{\Omega_y}(\Phi)| = 1$ for all y.

More specifically, Haller and Kennedy derived the following general sufficient conditions for Gibbsianity of $\nu = \mu \circ \pi^{-1}$:

Theorem 5.25. *Suppose μ is a Gibbs measure on $\Omega = \mathcal{A}^{\mathbb{Z}^d}$ for interaction Φ, $\Sigma = \mathcal{B}^{\mathbb{Z}^d}$, and $\pi : \Omega \to \Sigma$ is a surjective single-block fuzzy map. Furthermore, suppose that there exist finite positive constants c and m such that for every finite set $V \Subset \mathbb{Z}^d$, every two sites $i, j \in V$, every boundary condition $\tilde{x} \in \Omega$, and all $y \in \Sigma$ one has*

$$|\mathbb{E}_{\mu_{y,V,\tilde{x}}}(x_i x_j) - \mathbb{E}_{\mu_{y,V,\tilde{x}}}(x_i)\mathbb{E}_{\mu_{y,V,\tilde{x}}}(x_j)| \leq c e^{-m\|i-j\|}, \tag{5.13}$$

where $\mu_{y,V,\tilde{x}}$ is the measure on $\mathcal{A}^V$ defined as

$$\mu_{y,V,\tilde{x}}(x_V) = \frac{\exp(-H_V(x_V \tilde{x}_{V^c}))}{\sum\limits_{\bar{x}_V \in \pi^{-1} y_V} \exp(-H_V(\bar{x}_V \tilde{x}_{V^c}))}.$$

Then $\nu = \mu \circ \pi^{-1}$ is Gibbs.

Haller and Kennedy point out that the condition (5.13) is similar to one of Dobrushin–Shlosman's many equivalent definitions of complete analyticity, with uniform (in $y \in \Sigma$) constants. The proof of Theorem 5.25 has two natural parts: first establishing uniqueness $|\mathcal{G}_{\Omega_y}(\Phi)| = 1$ for all y, and then establishing the Gibbsianity of $\nu = \mu \circ \pi^{-1}$. In view of obtained results, the second part is no longer required.

Similarly, the result of Häggström (Theorem 5.7), can be interpreted in the framework we proposed. Indeed, the sufficient condition $\beta < \beta_c(d, r_1)$ strongly suggests that among all fibres Ω_y, the 'first' fibre (with respect to the parameter β) where the hidden phase transition occurs is the fibre Ω_1. However, the proof of Theorem 5.7 in [43] uses a different method: namely, representation of the

Potts model using the random cluster measure. It would be very interesting to investigate whether our results could be helpful in providing an alternative proof of Theorem 5.7 and, possibly, closing the remaining gap.

In conclusion, we strongly believe that the van Enter-Fernández-Sokal hypothesis, confirmed in all known cases, is equivalent to requiring that all points are Tjur. However, the problem remains open. Moreover, it would be interesting to see whether identification of sets of non-Tjur points in some particular examples can shed light on the severity of the non-Gibbsianness of the renormalized measured.

Bibliography

[1] M. Aizenman, J. T. Chayes, L. Chayes, and C.M. Newman, *Discontinuity of the magnetization in one-dimensional* $1/|x-y|^2$ *Ising and Potts models*, J. Statist. Phys. **50** (1988), no. 1-2, 1–40.

[2] Sebastián Barbieri, Ricardo Gomez, Brian Marcus, Tom Meyerovitch, and Siamak Taati, *Gibbsian representations of continuous specifications: the theorems of Kozlov and Sullivan revisited*, ArXiv (2020).

[3] Leonard E. Baum and Ted Petrie, *Statistical Inference for Probabilistic Functions of Finite State Markov Chains*, Ann. Math. Statist. **37** (1966), no. 6, 1554–1563.

[4] Ron Begleiter, Ran El-Yaniv, and Golan Yona, *On prediction using variable order Markov models*, J. Artificial Intelligence Res. **22** (2004), 385–421 (electronic).

[5] Henry Berbee, *Chains with infinite connections: Uniqueness and Markov representation*, Probability Theory and Related Fields **76** (1987), no. 2, 243–253.

[6] Rodrigo Bissacot, Eric Endo, Aernout van Enter, and Arnaud Le Ny, *Entropic repulsion and lack of the* g*-measure property for Dyson models*, Communications in Mathematical Physics **363** (2018), 767–788.

[7] David Blackwell, *The entropy of functions of finite-state Markov chains*, Transactions of the first Prague conference on information theory, Statistical decision functions, random processes held at Liblice near Prague from November 28 to 30, 1956, 1957, pp. 13–20.

[8] Thomas Bogenschütz and Volker Mathias Gundlach, *Ruelle's transfer operator for random subshifts of finite type*, Ergodic Theory and Dynamical Systems **15** (1995), no. 3, 413–447.

[9] Thierry Bousch, *La condition de Walters*, Annales scientifiques de l'Ecole normale supérieure, 2001, pp. 287–311.

[10] Thierry Bousch and Oliver Jenkinson, *Cohomology classes of dynamically non-negative* C^k *functions*, Invent. Math. **148** (2002), no. 1, 207–217.

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

[12] Maury Bramson and Steven Kalikow, *Nonuniqueness in g-functions*, Israel J. Math. **84** (1993), no. 1-2, 153–160.

[13] Peter Bühlmann and Abraham J. Wyner, *Variable length Markov chains*, Ann. Statist. **27** (1999), no. 2, 480–513, DOI 10.1214/aos/1018031204.

[14] C. J. Burke and M. Rosenblatt, *A Markovian Function of a Markov Chain*, Ann. Math. Statist. **29** (1958), no. 4, 1112–1122.

[15] John L. Cardy, *Scaling and renormalization in statistical physics*, Cambridge lecture notes in physics, Cambridge University Press, Cambridge, 1996.

[16] J.-R. Chazottes and E. Ugalde, *Projection of Markov measures may be Gibbsian*, J. Statist. Phys. **111** (2003), no. 5-6, 1245–1272.

[17] Cleary J. and Witten I., *Data Compression Using Adaptive Coding and Partial String Matching*, IEEE Transactions on Communications **32** (1984), no. 4, 396-402, DOI 10.1109/TCOM.1984.1096090.

[18] Zaqueu Coelho and Anthony N. Quas, *Criteria for $\overline{d}$-continuity*, Trans. Amer. Math. Soc. **350** (1998), no. 8, 3257–3268.

[19] Manfred Denker and Mikhail Gordin, *Gibbs measures for fibred systems*, Adv. Math. **148** (1999), no. 2, 161–192.

[20] E. I. Dinaburg, *A correlation between topological entropy and metric entropy*, Dokl. Akad. Nauk SSSR **190** (1970), 19–22 (Russian).

[21] R. L. Dobrushin, *Description of a random field by means of conditional probabilities and conditions for its regularity*, Teor. Verojatnost. i Primenen **13** (1968), 201–229 (Russian, with English summary).

[22] ______, *Gibbsian random fields for lattice systems with pairwise interactions.*, Funkcional. Anal. i Priložen. **2** (1968), no. 4, 31–43 (Russian).

[23] ______, *Lecture given at the workshop "Probability and Physics"*, 1995. Renkum, Holland.

[24] R. L. Dobrushin and S. B. Shlosman, *Gibbsian description of 'non-Gibbsian' fields*, Russian Mathematical Surveys **52** (1997), no. 2, 285.

[25] E. B. Dynkin, *Markov processes*, Gosudarstv. Izdat. Fiz.-Mat. Lit., Moscow, 1963.

[26] Sergio Verdú and Tsachy Weissman, *The information lost in erasures*, IEEE Trans. Inform. Theory **54** (2008), no. 11, 5030–5058.

[27] Ai Hua Fan and Mark Pollicott, *Non-homogeneous equilibrium states and convergence speeds of averaging operators*, Math. Proc. Cambridge Philos. Soc. **129** (2000), no. 1, 99–115.

[28] M. Feder and N. Merhav, *Relations between entropy and error probability*, IEEE Transactions on Information Theory **40** (1994), no. 1, 259-266, DOI 10.1109/18.272494.

[29] Roberto Fernández and Grégory Maillard, *Chains with complete connections and one-dimensional Gibbs measures*, Electron. J. Probab. **9** (2004), no. 6, 145–176 (electronic).

[30] R. Fernández and G. Maillard, *Chains and specifications*, Markov Process. Related Fields **10** (2004), no. 3, 435–456.

[31] Roberto Fernández and Grégory Maillard, *Chains with complete connections: general theory, uniqueness, loss of memory and mixing properties*, J. Stat. Phys. **118** (2005), no. 3-4, 555–588.

[32] F. Fernández, A. Viola, and M. J. Weinberger, *Efficient Algorithms for Constructing Optimal Bi-directional Context Sets*, 2010 Data Compression Conference, 2010, pp. 179-188, DOI 10.1109/DCC.2010.23.

[33] Roberto Fernández, Sandro Gallo, and Grégory Maillard, *Regular g-measures are not always Gibbsian*, Electron. Commun. Probab. **16** (2011), 732–740.

[34] Harry Furstenberg, *PREDICTION THEORY*, ProQuest LLC, Ann Arbor, MI, 1958. Thesis (Ph.D.)–Princeton University.

[35] Christophe Gallesco and Daniel Y. Takahashi, *Mixing rates for potentials of non-summable variations*, arXiv (2020).

[36] A. Galves and E. Löcherbach, *Stochastic chains with memory of variable length.*, TICSP Series **38** (2008), 117-133.

[37] Hans-Otto Georgii, *Stochastische Felder und ihre Anwendung auf Interaktionssysteme*, Inst. für Angewandte Math., 1974.

[38] T. N. T. Goodman, *Relating topological entropy and measure entropy*, Bull. London Math. Soc. **3** (1971), 176–180.

[39] Robert B. Griffiths and Paul A. Pearce, *Position-Space Renormalization-Group Transformations: Some Proofs and Some Problems*, Phys. Rev. Lett. **41** (1978), 917–920.

[40] ______, *Mathematical properties of position-space renormalization-group transformations*, J. Statist. Phys. **20** (1979), no. 5, 499–545.

[41] Robert B. Griffiths, *Mathematical properties of renormalization-group transformations*, Phys. A **106** (1981), no. 1-2, 59–69 (English, with French summary). Statphys 14 (Proc. Fourteenth Internat. Conf. Thermodynamics and Statist. Mech., Univ. Alberta, Edmonton, Alta., 1980).

[42] Leonid Gurvits and James Ledoux, *Markov property for a function of a Markov chain: A linear algebra approach*, Linear Algebra and its Applications **404** (2005), 85 - 117.

[43] Olle Häggström, *Is the fuzzy Potts model Gibbsian?*, Ann. Inst. H. Poincaré Probab. Statist. **39** (2003), no. 5, 891–917.

[44] Karl Haller and Tom Kennedy, *Absence of renormalization group pathologies near the critical temperature. Two examples*, Journal of Statistical Physics **85** (1996), no. 5, 607–637.

[45] T. E. Harris, *On chains of infinite order*, Pacific J. Math. **5** (1955), 707–724.

[46] Paul Hulse, *An example of non-unique g-measures*, Ergodic Theory Dynam. Systems **26** (2006), no. 2, 439–445.

[47] R. B. Israel, *Banach algebras and Kadanoff transformations.*, Random Fields **II** (1981), 593–608.

[48] Anders Johansson and Anders Öberg, *Square summability of variations of g-functions and uniqueness of g-measure*, Math. res. Lett. **10** (2003), 587-601.

[49] Anders Johansson, Anders Öberg, and Mark Pollicott, *Phase transitions in long-range Ising models and an optimal condition for factors of g -measures*, Ergodic Theory and Dynamical Systems, 1–14.

[50] Steve Kalikow, *Random Markov processes and uniform martingales*, Israel J. Math. **71** (1990), no. 1, 33–54.

[51] Michael Keane, *Strongly mixing g-measures*, Invent. Math. **16** (1972), 309–324.

[52] John G. Kemeny and J. Laurie Snell, *Finite Markov chains*, The University Series in Undergraduate Mathematics, D. Van Nostrand Co., Inc., Princeton, N.J.-Toronto-London-New York, 1960.

[53] Thomas M. W. Kempton, *Factors of Gibbs measures for subshifts of finite type*, Bull. Lond. Math. Soc. **43** (2011), no. 4, 751–764.

[54] Bruce Kitchens and Selim Tuncel, *Finitary measures for subshifts of finite type and sofic systems*, Mem. Amer. Math. Soc. **58** (1985), no. 338, iv+68.

[55] O. K. Kozlov, *A Gibbs description of a system of random variables*, Problemy Peredači Informacii **10** (1974), no. 3, 94–103 (Russian).

[56] O. E. Lanford III and D. Ruelle, *Observables at infinity and states with short range correlations in statistical mechanics*, Comm. Math. Phys. **13** (1969), 194–215.

[57] Oscar E. Lanford, *Statistical Mechanics and Mathematical Problems* (A. Lenard, ed.), Springer Berlin Heidelberg, Berlin, Heidelberg, 1973.

[58] G. Langdon, *A note on the Ziv - Lempel model for compressing individual sequences (Corresp.)*, IEEE Transactions on Information Theory **29** (1983), no. 2, 284-287, DOI 10.1109/TIT.1983.1056645.

[59] François Ledrappier, *Principe variationnel et systèmes dynamiques symboliques*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **30** (1974), 185–202.

[60] Francois Ledrappier and Peter Walters, *A relativised variational principle for continuous transformations*, J. London Math. Soc. (2) **16** (1977), no. 3, 568–576.

[61] Jacob Ziv and Abraham Lempel, *Compression of individual sequences via variable-rate coding*, IEEE Trans. Inform. Theory **24** (1978), no. 5, 530–536.

[62] József Lőrinczi, Christian Maes, and Koen Vande Velde, *Transformations of Gibbs measures*, Probab. Theory Related Fields **112** (1998), no. 1, 121–147, DOI 10.1007/s004400050185. MR1646444

[63] Christian Maes and Koen Vande Velde, *The fuzzy Potts model*, J. Phys. A **28** (1995), no. 15, 4261–4270.

[64] Brian Marcus, Karl Petersen, and Susan Williams, *Transmission rates and factors of Markov chains*, Conference in modern analysis and probability (New Haven, Conn., 1982), Contemp. Math., vol. 26, Amer. Math. Soc., Providence, RI, 1984, pp. 279–293.

[65] A. Moffat and R. Neal, *Arithmetic Coding Revisited*, Proceedings of the Conference on Data Compression, 1995, pp. 202–.

[66] S. Dachian and B. S. Nahapetian, *Description of specifications by means of probability distributions in small volumes under condition of very weak positivity*, J. Statist. Phys. **117** (2004), no. 1-2, 281–300.

[67] J. von Neumann, *Zur Operatorenmethode in der klassischen Mechanik*, Ann. of Math. (2) **33** (1932), no. 3, 587–642.

[68] Mordechai Nisenson, Ido Yariv, Ran El-Yaniv, and Ron Meir, *Towards Behaviometric Security Systems: Learning to Identify a Typist* (Nada Lavrač, Dragan Gamberger, Ljupčo Todorovski, and Hendrik Blockeel, eds.), Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[69] O. Onicescu and G. Mihoc, *Sur les chaines statistiques*, C. R. Acad. Sci. Paris **200** (1935), 511-512.

[70] Lars Onsager, *Crystal statistics. I. A two-dimensional model with an order-disorder transition*, Phys. Rev. (2) **65** (1944), 117–149.

[71] E. Ordentlich, M. J. Weinberger, and T. Weissman, *Efficient pruning of bi-directional context trees with applications to universal denoising and compression*, 2004.

[72] E. Ordentlich, M. J. Weinberger, and T. Weissman, *Multi-directional context sets with applications to universal denoising and compression*, Proceedings. International Symposium on Information Theory, 2005. ISIT 2005., 2005, pp. 1270-1274, DOI 10.1109/ISIT.2005.1523546.

[73] E. Ordentlich, M. J. Weinberger, and C. Chang, *On Multi-Directional Context Sets*, IEEE Transactions on Information Theory **57** (2011), no. 10, 6827-6836, DOI 10.1109/TIT.2011.2165818.

[74] Marion Rachel Palmer, William Parry, and Peter Walters, *Large sets of endomorphisms and of g-measures*, The structure of attractors in dynamical systems (Proc. Conf., North Dakota State Univ., Fargo, N.D., 1977), Lecture Notes in Math., vol. 668, Springer, Berlin, 1978, pp. 191–210.

[75] Jorma Rissanen, *A universal data compression system*, IEEE Trans. Inform. Theory **29** (1983), no. 5, 656–664.

[76] Dana Ron, Yoram Singer, and Naftali Tishby, *The power of amnesia: Learning probabilistic automata with variable memory length*, Machine Learning **25** (1996), no. 2, 117–149, DOI 10.1007/BF00114008.

[77] David Ruelle, *Thermodynamic formalism*, Encyclopedia of Mathematics and its Applications, vol. 5, Addison-Wesley Publishing Co., Reading, Mass., 1978. The mathematical structures of classical equilibrium statistical mechanics; With a foreword by Giovanni Gallavotti and Gian-Carlo Rota.

[78] Douglas Lind and Brian Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, Cambridge, 1995.

[79] Brian Marcus, *Symbolic dynamics and connections to coding theory, automata theory and system theory*, Different aspects of coding theory (San Francisco, CA, 1995), Proc. Sympos. Appl. Math., vol. 50, Amer. Math. Soc., Providence, RI, 1995, pp. 95–108.

[80] Ja. G. Sinaĭ, *Gibbs measures in ergodic theory*, Uspehi Mat. Nauk **27** (1972), no. 4(166), 21–64 (Russian).

[81] Aernout C. D. van Enter, Roberto Fernández, and Alan D. Sokal, *Regularity properties and pathologies of position-space renormalization-group transformations: scope and limitations of Gibbsian theory*, J. Statist. Phys. **72** (1993), no. 5-6, 879–1167.

[82] Wolfgang Stadje, *The evolution of aggregated Markov chains*, Statistics & Probability Letters **74** (2005), no. 4, 303 - 311.

[83] Örjan Stenflo, *Uniqueness in g-measures*, Nonlinearity **16** (2002), 403.

[84] W. G. Sullivan, *Potentials for almost Markovian random fields*, Comm. Math. Phys. **33** (1973), 61–74.

[85] Tue Tjur, *Conditional probability distributions*, Institute of Mathematical Statistics, University of Copenhagen, Copenhagen, 1974. Lecture Notes, No. 2.

[86] ______, *A Constructive Definition of Conditional Distributions*, Preprint, Institute of Mathematical Statistics, University of Copenhagen, 1975.

[87] Evgeny Verbitskiy, *On factors of g-measures*, Indag. Math. (N.S.) **22** (2011), no. 3-4, 315–329.

[88] ———, *Thermodynamics of the binary symmetric channel*, Pacific Journal of Mathematics for Industry **8** (2016), no. 1, 2, DOI 10.1186/s40736-015-0021-5.

[89] P. Volf, *Weighting Techniques in Data Compression Theory and Algorithms*, Technische Universiteit Eindhoven, 2002. Ph.D. thesis.

[90] Peter Walters, *Ruelle's Operator Theorem and g-Measures*, Transactions of the American Mathematical Society **214** (1975), 375–387.

[91] Peter Walters, *Invariant measures and equilibrium states for some mappings which expand distances*, Trans. Amer. Math. Soc. **236** (1978), 121–153.

[92] ———, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York-Berlin, 1982.

[93] ———, *Relative pressure, relative equilibrium states, compensation functions and many-to-one codes between subshifts*, Trans. Amer. Math. Soc. **296** (1986), no. 1, 1–31.

[94] ———, *Erratum: "A necessary and sufficient condition for a two-sided continuous function to be cohomologous to a one-sided continuous function"*, Dyn. Syst. **18** (2003), no. 3, 271–278.

[95] ———, *Regularity conditions and Bernoulli properties of equilibrium states and g-measures*, J. London Math. Soc. (2) **71** (2005), no. 2, 379–396.

[96] ———, *A natural space of functions for the Ruelle operator theorem*, Ergodic Theory Dynam. Systems **27** (2007), no. 4, 1323–1348.

[97] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdu, and M. J. Weinberger, *Universal discrete denoising: known channel*, IEEE Transactions on Information Theory **51** (2005), no. 1, 5-28.

[98] F. M. J. Willems, Y. M. Shtarkov, and T. J. Tjalkens, *The context-tree weighting method: basic properties*, IEEE Transactions on Information Theory **41** (1995), no. 3, 653-664, DOI 10.1109/18.382012.

[99] Jisang Yoo, *On Factor Maps that Send Markov Measures to Gibbs Measures*, Journal of Statistical Physics **141** (2010), no. 6, 1055–1070.

[100] J. Yu and S. Verdu, *Schemes for Bidirectional Modeling of Discrete Stationary Sources*, IEEE Transactions on Information Theory **52** (2006), no. 11, 4789-4807.

[101] Jiming Yu, *Bidirectional modeling of finite -alphabet sources*, 2007 (English).

[102] Michiko Yuri, *Zeta functions for certain non-hyperbolic systems and topological Markov approximations*, Ergodic Theory Dynam. Systems **18** (1998), no. 6, 1589–1612.

Samenvatting

Dit proefschrift presenteert resultaten over kansmaten gedefinieerd op roostersystemen, waarbij aan ieder punt op een rooster $\mathbb{L}$ een element uit een eindige verzameling $\mathscr{A}$ wordt toegekend. De verzameling van toestanden is in dit geval gegeven door $\mathscr{A}^{\mathbb{L}}$. Discrete tijd stochastische processen zijn een bekend voorbeeld, waarbij $\mathbb{L} = \mathbb{Z}$ of $\mathbb{L} = \mathbb{Z}_+ = \{0, 1, 2, ...\}$. Meer algemene roosters worden gebruikt in modellering binnen de informatietheorie, dynamische systemen en statistische mechanica.

Een bekend voorbeeld is de Markovketen, waarbij op een gegeven tijdstip de kansverdeling voor het volgende tijdstip is bepaald door de huidige toestand, onafhankelijk van het verleden. Een veralgemenisering hiervan is een proces waarbij de voorwaardelijke kansen voor de huidige toestand, gegeven het verleden, omschreven kunnen worden door een functie die continu is in de producttopologie. Dit betekent dat de invloed van het verre verleden verdwijnt. In dynamische systemen worden kansmaten op $\mathscr{A}^{\mathbb{L}}$ met dergelijke eigenschappen bestudeerd onder de naam g-maten.

Vergelijkbare kansmaten worden op $\mathscr{A}^{\mathbb{Z}^d}$, $d \geq 1$, gebruikt voor de modelering van thermodynamische systemen onder de naam Gibbsmaten. Hierbij representeert het rooster gewoonlijk niet tijd maar plaats. In dit geval zijn de voorwaardelijke kansen voor een positie in het rooster, gegeven de configuratie op alle andere roosterposities, continu.

Het eerste hoofdstuk is een introductie en samenvatting van de resultaten.

In het tweede hoofdstuk vergelijken we g-maten met Gibbsmaten in één dimensie. Een recent resultaat laat zien dat er g-maten zijn die niet ook een Gibbsmaat zijn. Met andere woorden, een maat met eenzijdig continue voorwaardelijke kansen heeft niet noodzakelijk tweezijdig continue voorwaardelijke kansen. Het voornaamste resultaat uit het tweede hoofdstuk is een voldoende en noodzakelijke voorwaarde op de eenzijdige voorwaardelijke kansen van een g-maat zodat deze g-maat ook een Gibbsmaat is. Daarbij bestuderen we ook de gevolgen van dit criterium. We bespreken onder andere een aantal bekende uniciteitscriteria voor g-maten waaruit volgt dat de maat een Gibbsmaat is. Aan de andere kant laten we zien dat uniciteit niet noodzakelijk is voor een g-maat om een Gibbsmaat te zijn. In dit hoofdstuk behandelen we tevens het vergelijkbare probleem of een g-maat een g-maat blijft onder het inverteren van tijd. Met andere woorden, zijn de voorwaardelijke kansen in de omgekeerde richting ook continu? Hier lossen we het open probleem op of een g—maat met een potentiaal $\log g$ in de Bowen klasse een g-maat blijft onder het inverten van tijd.

In het derde hoofdstuk bestuderen we toepassingen van de relaties tussen voorwaardelijke kansen in g-maten en Gibbsmaten. Een groot aantal algoritmes is in staat om de voorwaardelijke kansen voor de volgende toestand van een proces te schatten, gegeven de vorige toestanden. Dit kan worden gezien als het schatten van de voorwaardelijke kansen van een g-maat. Deze algoritmes kunnen ook de voorwaardelijke kansen van een Gibbsmaat schatten via de relaties die zijn gebruikt in het tweede hoofdstuk. Een toepassing hiervan in informatietheorie is dat een compressiealgoritme gebruikt kan worden als een algoritme dat ruis van een discreet signaal verwijdert. In het derde hoofdstuk vergelijken we zes verschillende algoritmes, die normaal gebruikt worden voor compressie of voorspelling, wanneer ze worden toegepast voor het schatten van tweezijdige kansen. Hierbij beoordelen we de algoritmes gebruikmakend van een metriek op de ruimte van kansmaten en ook via de geschiktheid voor het verwijderen van ruis. De data waarop we dit bestuderen is grotendeels kunstmatig, maar de algoritmes worden ook toegepast op engelse tekst.

In het vierde hoofdstuk bestuderen we verborgen Markovmaten. Stel een Markovproces is niet direct te observeren omdat verschillende symbolen niet onderscheidbaar zijn. Het resulterende proces is dan in het algemeen niet Markov. Sterker nog, de voorwaardelijke kansen kunnen een willekeurig sterke afhankelijkheid hebben van een willekeurig ver verleden. Een belangrijke vraag is onder welke voorwaarden deze afhankelijkheid nog regulier is. In dit hoofdstuk bedoelen we daarmee dat de kansmaat een g-maat is. Dit vraagstuk, voor verscheidene definities van regulier, is een bekend probleem in de studie van Markovmaten. In dit hoofdstuk benaderen we het probleem met een techniek die eerder is toegepast voor factoren van g-maten. Deze techniek, het aantonen van het bestaan van een continue maat disintegratie, wordt toegepast om regulariteit van factoren van g-maten te onderzoeken. De Markovpotentialen die we bestuderen in dit hoofdstuk zijn minder complex, maar in plaats daarvan kunnen er verboden transities zijn. We tonen aan dat deze voorwaarde strikt meer algemeen is dan de bekende voorwaarden voor dit vraagstuk. Niettemin is de voorwaarde niet een noodzakelijke voorwaarde voor het behoud van regulariteit onder een factorafbeelding. Hierbij maken we gebruik van de resultaten van Tue Tjur op het gebied van maatdisintegraties.

In het vijfde en laatste hoofdstuk bestuderen we met dezelfde techniek de regulariteit van factoren van Gibbsmaten, ook wel gerenormaliseerde Gibbsmaten genoemd. In deze context zijn er geen verboden configuraties, maar generaliseren we naar een hogere dimensie van het rooster en naar Gibbspotentialen. De definitie van regulier is dat de factormaat een Gibbsmaat is. We tonen aan dat het bestaan van een continue disintegratie van de kansmaat inderdaad een voldoende voorwaarde is voor regularteit van de factormaat. Vervolgens tonen we aan dat het bestaan van een continue maatdisintegratie volgt uit bestaande resulaten door van Enter, Fernández en Sokal. De vraag of deze voorwaarde noodzakelijk is voor een factor van een Gibbsmaat om regulier te zijn, blijft open.

Acknowledgements

The projects in this thesis were made possible with the support of many wonderful people that I would like to thank in this section.

First, I would like to thank the individual with the largest contribution to my PhD: my supervisor Prof. Dr. Evgeny Verbitskiy. The effort you put into our joint projects was impressive and the contribution to the mathematical ideas invaluable.

Second I would like to thank Prof. Dr. Roberto Fernandez, who, as one of my master thesis supervisors recommended this position. And who, as co-author, by his expertise, kindness and passion for the topic, was a source of many inspiring conversations.

The probability group in Leiden was an important context for learning more about mathematics, especially on topics outside of what is covered in this thesis. My thanks go to my second promotor Prof. Dr. Frank den Hollander for bringing this group together and also for his feedback on my thesis and accompanying propositions.

I would also like to thank Prof. Dr. Aernout van Enter for providing extensive feedback on my manuscript.

Many at the Mathematical Institute contributed indirectly. From among the staff I would like to thank Charlene Kalle for many supportive discussions. From among the PhD students, there are too many to name, since so many of you have been such amazing friends.

Finally I would like to thank the people closest to me. My parents, Johan en Yvonne, for the loving and supportive upbringing, and my siblings Leonie en Mariska for always believing in me. And finally, Liza, who is always there for me and makes my life so much brighter.

Curriculum vitae

Steven Berghout was born in Krimpen aan den IJssel on Octber 18, 1987. He attended high school in Rotterdam, starting in 2000, and obtained his diploma in 2005. That same year he started with the Bachelor programmes Physics and Astronomy and Mathematics at Utrecht University. He finished this programme with a bachelor thesis on the topic of Big Bang nucleosynthesis.

Subsequently, he followed and completed the master programmes in Theoretical Physics and Mathemetical Sciences at Utrecht University. These programmes were completed with a master thesis titled "The Einstein-Smoluchowski Equation in the One Dimensional Exclusion Process" in 2014. In the mean time, from 2007 onwards, he acted as a teaching assistant in several courses in both the Physics and the Mathematics departments.

After a brief intermezzo working in IT, he started his PhD at Leiden University with Prof. Dr. Evgeny Verbitskiy as his daily supervisor and Prof. Dr. Frank den Hollander as his second promotor. Currently he works as a design engineer in the metrology department of ASML.