

# THERMODYNAMICAL ASPECTS OF CLASSICAL LATTICE SYSTEMS <sup>1</sup>

C.-E. Pfister  
Ecole Polytechnique Fédérale de Lausanne  
Département de Mathématiques  
CH-1015 Lausanne, Switzerland

## CONTENTS

1. Introduction	2
2. Thermodynamics and statistical mechanics	4
2.1. Thermodynamics	6
2.2. Statistical mechanics	10
3. Asymptotically decoupled measures	20
3.1. Asymptotically decoupled measures, specific information gain	21
3.2. Local specification and specific information gain	32
3.3. Asymptotically I-null sequences	35
3.4. Large deviations of the empirical measure	37
3.5. Duality	42
4. Space of bounded total oscillation functions	44
4.1. Uniform boundary terms	45
4.2. The local specification $\Pi^f$	48
4.3. Change of reference measure in the large deviations principle	53
5. Conditional limit theorems	56
5.1. A concentration result	56
5.2. Existence of subgradients	59
5.3. Conditional limit theorems	61
5.4. Equivalence of ensembles	66
References	67

---

<sup>1</sup> Based on lectures delivered at IV-th Brazilian school of probability, Mambucaba, RJ, August 2000. Supported by Swiss National Science Foundation. These lectures have been slightly revised in December 2001. I thank in particular Alessandra Faggionato, Hans-Otto Georgii, Gonzalo Panizo, Daniel Ueltschi and Maria Eulalia Vares for comments, suggestions and corrections.

## 1. INTRODUCTION

In these lectures I consider the mathematical formalism of thermodynamics and statistical mechanics for lattice systems and show its tight connection with large deviations theory. The models I have in mind are the lattice models with compact state spaces, although the theory is developed in greater generality. The central notion studied here is the notion of an *asymptotically decoupled probability measure*. The main theme of these lectures is the *thermodynamical like aspects of the law of large numbers*. The main feature of the asymptotically decoupled probability measures is that the existence of important quantities, like the specific information gain, which have a thermodynamical interpretation in statistical mechanics, can be established using the standard and simple arguments developed for proving the existence of thermodynamical limits (see e.g. [YL]<sup>1</sup>). A large part of these lectures is original. However, the principal results and methods have been developed much earlier in a less general setting<sup>2</sup>. My presentation is strongly influenced by my long-standing collaboration with John Lewis and Wayne Sullivan in Dublin. It is based principally on [LP] and [LPS3]. The notion of asymptotically decoupled measures<sup>3</sup> has been introduced in lectures delivered in Grenoble in Spring 1999 on the same subject.

Section 2 serves as an introductory section. I consider a special model and present the main features of statistical mechanics of lattice systems. The interplay between mathematical physics and large deviations theory is displayed. No proof is given, but almost all results presented in that section are proved with complete proofs in much greater generality in sections 3 to 5. Section 2 is independent of the rest of the lectures.

In section 3 asymptotically decoupled measures are defined and examples are given. Large deviations properties of such measures  $\lambda$  are expressed through two functionals: the specific information gain  $h(\cdot|\lambda)$ , which is defined on the space of translation invariant probability measures  $\mathcal{M}_1^{+, \theta}$ , and  $p(\cdot|\lambda)$  (called the pressure), which is defined on the dual space of  $\mathcal{M}_1^{+, \theta}$ , the space of the quasilocal functions  $\mathcal{F}_{\text{qloc}}$ . These functionals are conjugate to each other. Convexity plays an important role in the thermodynamical formalism. The solutions of the equation  $h(\nu|\lambda) = 0$  and the related notion of asymptotically I-null sequences, extending the notion of asymptotically quasi-independence of Csizsár, are then studied. In subsection 3.4 a large deviations principle for the empirical measures in the multidimensional case is established.

In section 4 the space  $\mathcal{L}$  of bounded total oscillation functions is studied.  $\mathcal{L}$  is a dense subset of the quasilocal functions. Nice properties are true on  $\mathcal{L}$ . In particular

---

<sup>1</sup> The importance of the thermodynamical limit in statistical mechanics was recognized in the late thirties. See in particular the very interesting introduction of [KU].

<sup>2</sup> I did not write a review paper. The bibliography is very incomplete, and the historical remarks very rare. More information about earlier results can be found by consulting the references at the end of the lectures and in [EFS]. See [I] for a treatment of the thermodynamical formalism from the viewpoint of statistical mechanics for the classical and quantum cases, emphasizing convexity theory. [K] is written from the viewpoint of ergodic theory. See also [E], [G1], [Sim].

<sup>3</sup> After writing these notes, B. McGurk communicated to me reference [BD]. The mixing condition (S) of Bryc and Dembo is different from the asymptotically decoupled property. Previous results along these lines in chapter 6 of [DZ].

one can introduce local specifications and define what is called Gibbs measures in statistical mechanics. The key-point is that in  $\mathcal{L}$  one has a natural notion of boundary terms, which fails in the larger space  $\mathcal{F}_{\text{loc}}$ . Using a stronger version of an asymptotically decoupled probability measure one establishes two important results, theorems 4.1 and 4.2 in subsection 4.3. Theorem 4.2 describes how the large deviations principle is modified under the change of the reference measure.

In the last section, using the large deviations principle established in 3.4, I prove conditional limit theorems. This corresponds to the Gibbs conditioning principle. Here again there are many papers about that topic, both in probability theory and mathematical physics. The results of section 5 are very general, but they are “weak” in the following sense. One of the main aspects of the formalism, expressed in the physics language, is the coexistence of several equilibrium states for given thermodynamical parameters<sup>4</sup>. The conditional limit theorems presented here reflect this fact. The typical conclusion is “the set of cluster points of the sequence of conditioned measures is non-empty and each cluster point is a measure with well-defined thermodynamical parameters”. What seems to me important is the structure of these theorems. There is a compactness argument implying the existence of cluster points. This follows for example from the large deviations principle or more precisely from the upper bound of this principle and the compactness of the level-sets of the rate-function. The second part is the identification of the cluster points. In that part convexity theory plays the dominant role. Section 5 is written so that these two separate aspects are clearly displayed. Existence of the limit of the sequence of conditioned measures holds if there is no phase coexistence for these values of the thermodynamical parameters. The paper [Cs] of Csiszár “Sanov property, generalized I–projection and a conditional limit theorem” has been the main source of inspiration for [LPS3]. In short, in [LPS3] Csiszár’s paper is extended from the case of independent random variables to the case of dependent random fields. In these lectures I extend further the results of [LPS3]. When reading [MN] on the same subject I discovered the important reference [S], which is also inspired by [Cs], in which the results of Csiszár’s paper are extended to the case of dependent random variables. This paper is very close to [LPS3] in several aspects, although the scope is less general.

In these lectures the thermodynamical formalism of classical lattice models is exposed in details. However, statistical mechanics does not concern only these macroscopic aspects. In some sense, the boundary effects are the important effects in statistical mechanics, and they are never at the forefront in these lectures. They must be studied with different methods, often case by case.

Exposing the main concepts and the structure of the formalism leads me to take viewpoints, which are sometimes not those which one adopts when working on a specific lattice system. I also stress that lattice models are special; the results presented here are proved for these simple systems only.

---

<sup>4</sup>In most papers on the subject the cases considered correspond to a unique equilibrium state, i.e. there is a single probability measure corresponding to given thermodynamical parameters. The question of unicity of equilibrium state is a central question, but it cannot be studied in large deviations theory only, since boundary terms plays an essential role in the problem of coexistence of equilibrium states. It is however easy to prove that in the cases considered for example by Meda and Ney in [MN] there is always a single probability measure corresponding to given thermodynamical parameters (see e.g. [BLP]).

## 2. THERMODYNAMICS AND STATISTICAL MECHANICS

Classical lattice systems are random processes on  $\mathbb{Z}^d$  with values in a compact metric space. The terminology comes from physics, but such processes occur also in many different fields, like ergodic theory. The mathematical theory of equilibrium states of lattice systems gave rise to the thermodynamical formalism (see [Ru2]). In [Sin] Sinai introduced in ergodic theory concepts and methods inspired by this formalism. This is now a classical chapter of ergodic theory (see [B] and in particular [W]). There are also tight connections between statistical mechanics of lattice systems and Varadhan's treatment of large deviations<sup>5</sup>. In these lectures I consider this aspect of the formalism. In the present section I illustrate this fact by considering the relationship between thermodynamics and statistical mechanics for a particular case<sup>6</sup>. The main ideas are not new, but the perspective is different from that of Ruelle in [Ru2], chapters 1 to 4. Ruelle in his important paper [Ru1] provided a rigorous treatment of statistical thermodynamics, which was developed by Lanford in [L], who used these ideas to give a completely new proof of Cramér's theorem; this was the first step in an important development in the theory of large deviations. Statistical mechanics for classical particle systems is exposed in a systematic way in this formalism by Martin-Löf in [M-L]<sup>7</sup>.

There are two famous formulas<sup>8</sup> associated with the names of **Boltzmann** (1844-1906) and **Gibbs** (1839-1903),

$$S = k \ln W \quad \text{and} \quad e^{-\beta H}.$$

There is a third famous formula,  $e^{iHt}$ , which I shall not consider here, but there are deep connections between these three formulas.  $k \ln W$ ,  $e^{iHt}$ ,  $e^{-\beta H}$  define fundamental concepts, which are respectively, the entropy, the evolution operator and the equilibrium states. This paper is devoted to a detailed study of the relationships between  $S = k \ln W$  and  $e^{-\beta H}$ . Since I am emphasizing the structure of the formalism more importance is given to formula  $S = k \ln W$  than  $e^{-\beta H}$ .

Generally speaking a lattice system is an assembly of elementary subsystems, which are located at the nodes of a lattice. In the (two-dimensional) Ising model the subsystems are described by variables  $\sigma_x$ , called spin variables, which take only two values, and the lattice is  $\mathbb{Z}^2$ ;

$$\sigma_x = \pm 1, \quad x \in \mathbb{Z}^2 \quad \text{and} \quad \mathbb{Z}^2 := \{x = (x_1, x_2) : x_i = 0, \pm 1, \pm 2, \dots\}.$$

Macroscopic aspects of large systems is the subject of statistical mechanics. "Large systems" means that the elementary subsystems are indexed by the elements of sets

$$\Lambda(n) := \{x \in \mathbb{Z}^2 : -n \leq x_1 \leq n, -n \leq x_2 \leq n\},$$

---

<sup>5</sup>Basic references are [V], [DS] and [DZ].

<sup>6</sup>This choice corresponds to an Ising model, but all statements of section 2 hold for lattice systems with finite state spaces and finite range interactions, and even for a larger class of systems. See sections 3 and 4.

<sup>7</sup>See also [RZ] for a discussion of some basic questions in statistical mechanics in the setting of large deviations theory.

<sup>8</sup> $W$  stands for *Wahrscheinlichkeit*,  $H$  is the hamiltonian,  $\beta^{-1} = kT$ , where  $T$  is the absolute temperature and  $k$  is Planck constant, one of the fundamental constants of physics. In this paper these interpretations will not play a particular role.

whose cardinality  $|\Lambda(n)|$  is large. At first approximation such systems are described by infinite systems by taking the limit  $n \rightarrow \infty$  (thermodynamical limit). For my purpose I adopt the following viewpoint, which differs from the conventional one. Usually a lattice model is defined by giving a function, called hamiltonian, which already contains thermodynamical parameters (generalized chemical potentials). This function will be introduced later. I start with statistical thermodynamics, whose basic objects are macroscopic observables and macroscopic states. The macroscopic states are the values of a finite set of macroscopic variables. In this paper they are the values “at equilibrium” of these macroscopic variables. Statistical mechanics is then developed as a mathematical theory based on statistical thermodynamics. The macroscopic variables are sums of local variables, which are functions of the microscopic states of the system. In the example chosen, there are two real-valued macroscopic variables or equivalently a single  $\mathbb{R}^2$ -valued macroscopic variable. They are the macroscopic variables, whose values are needed to prescribe a thermodynamical state of an Ising model. The precise definitions are as follows. A configuration  $\omega$  of the lattice system is a mapping from  $\mathbb{Z}^2$  to  $\{-1, 1\}$ ,  $x \mapsto \omega(x)$ , whose value at  $x \in \mathbb{Z}^2$  specifies the state of the subsystem at  $x$ . The restriction of  $\omega$  to  $\Lambda \subset \mathbb{Z}^2$  is denoted by  $\omega_\Lambda$ . The space of all configurations is  $\Omega := \{-1, 1\}^{\mathbb{Z}^2}$ ; it is the space of **microscopic states**. I use two local variables to define the macroscopic variables:  $\sigma_x$ , such that  $\sigma_x(\omega) := \omega(x)$ , and  $e_x$ ,

$$e_x := -\frac{J}{2} \sum_{\substack{y \in \mathbb{Z}^2: \\ \|y-x\|=1}} \sigma_x \sigma_y \quad \text{where} \quad \|x-y\| := \sum_{i=1}^2 |y_i - x_i|.$$

$\mathbb{Z}^2$  acts as translation group. Its action on  $\Omega$  is defined by

$$(\theta_x \omega)_y := \omega_{x+y};$$

its action on a function  $g$  is  $(\theta_x g)(\omega) := g(\theta_x \omega)$ . Let  $\mathbf{f} : \Omega \mapsto \mathbb{R}^2$ ,

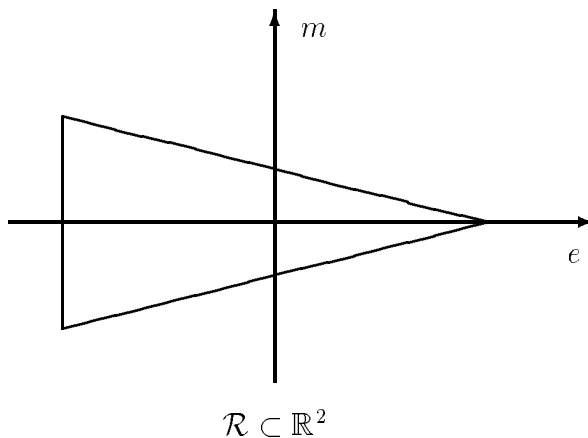
$$\mathbf{f}(\omega) := (e_0(\omega), \sigma_0(\omega)).$$

$\mathbf{f}$  is a **local observable**, whose values depend only on the states of the subsystems at  $(0,0)$ ,  $(0, \pm 1)$  and  $(\pm 1, 0)$ . Using the translations and  $\mathbf{f}$ ,  $\sum_{x \in \Lambda(n)} \theta_x \mathbf{f}$  defines a **macroscopic observable** with values in  $\mathbb{R}^2$ ; it is convenient to normalize it,

$$\mathbf{F}_n := \frac{1}{|\Lambda(n)|} \sum_{x \in \Lambda(n)} \theta_x \mathbf{f}.$$

The **macroscopic states** are the values of  $\mathbf{F}_n$  when  $n \rightarrow \infty$ , that is, a macroscopic state is an element of

$$\mathcal{R} := \{ \mathbf{x} = (e, m) \in \mathbb{R}^2 : |e/J| \leq 2, -1 + (e/J + 2)/4 \leq m \leq 1 - (e/J + 2)/4 \}.$$



**2.1. Thermodynamics.** At equilibrium the thermodynamical state of the Ising model is given by a macroscopic state  $(e, m) \in \mathcal{R}$  and the thermodynamical properties of the system are described by the thermodynamical entropy  $S$ , which is defined on  $\mathcal{R}$ <sup>9</sup>. Boltzmann's formula,  $S = k \ln W$ , relates the thermodynamical entropy  $S$  to the microscopic states of the system.  $W$  stands for *Wahrscheinlichkeit*, but the word has a different meaning as nowadays.  $W$  is the cardinality of a particular subset of microscopic configurations. If the system is confined in the subset  $\Lambda(n)$  and the macroscopic state is  $(e, m)$ , then  $W$  is the cardinality of

$$\{\omega_{\Lambda(n)} : \mathbf{F}_n(\omega) = (e, m)\},$$

the subset of all microscopic configurations compatible with  $(e, m)$ . It is an essential feature of statistical mechanics<sup>10</sup> that one basic problem is to “count” correctly microscopic configurations, that is, there is a natural way of measuring the size of subsets of  $\Omega$ . Here the natural measure is the counting measure.<sup>11</sup>

- If the system is not submitted to any physical influence or constraint, then all microscopic configurations  $\omega \in \Omega$  are equiprobable.

My viewpoint being a probabilistic one, I choose<sup>12</sup> as reference measure the uniform probability measure on  $\Omega$ , denoted by  $\lambda$ , instead of the counting measure. Hence  $(\Omega, \lambda)$  becomes a probability space,  $\mathbf{f}$  a random variable and  $\mathbf{F}_n$  a sum of i.i.d. random variables.

$S$  is an extensive quantity in the thermodynamical limit. In [Ru1] Ruelle gave a mathematical definition of  $s$ , the entropy per unit volume. For the definition of  $s(e, m)$  I follow [LP]. Let  $\mathbf{x} = (e, m)$  and

$$B_\varepsilon(\mathbf{x}) := \{\mathbf{y} \in \mathbb{R}^2 : \|\mathbf{y} - \mathbf{x}\| \leq \varepsilon\}$$

---

<sup>9</sup>It is not evident that any macroscopic state (in the interior of  $\mathcal{R}$ ) is an *equilibrium* macroscopic state under appropriate conditions. In our example this is the case, since the statistical thermodynamical entropy  $s(\mathbf{x})$  is well-defined (see (2.1)).

<sup>10</sup>Always in this paper equilibrium statistical mechanics.

<sup>11</sup>The choice of the reference measure is crucial. This is the choice made in equilibrium statistical mechanics. In other contexts one may choose another reference measure. The purpose of these lectures is to develop the theory for a large class of reference measures, which are the asymptotically decoupled probability measures.

<sup>12</sup>This choice is convenient here, but this is not always the case. The basic reference measure needs not be normalized. In statistical mechanics it is usually not normalized.

be the ball of center  $\mathbf{x}$  and radius  $\varepsilon$ . Using a standard subadditivity argument one shows that  $s(\mathbf{x})$  can be defined by

$$\begin{aligned} s(\mathbf{x}) &:= \min_{\varepsilon > 0} \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \text{Prob}_\lambda(\mathbf{F}_n(\omega) \in B_\varepsilon(\mathbf{x})) \\ &= \min_{\varepsilon > 0} \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \text{Prob}_\lambda(\mathbf{F}_n(\omega) \in B_\varepsilon(\mathbf{x})). \end{aligned} \quad (2.1)$$

The domain of  $s$  is  $\mathcal{R}$ :  $s(\mathbf{x}) > -\infty$  if and only if  $\mathbf{x} \in \mathcal{R}$ . Moreover,  $s(\mathbf{x}) \leq 0$  because  $\lambda$  is normalized; a variant of the subadditivity argument implies that it is concave<sup>13</sup>

$$s(a\mathbf{x} + (1-a)\mathbf{y}) \geq a s(\mathbf{x}) + (1-a) s(\mathbf{y}) \quad 0 < a < 1.$$

It is important to notice that  $s$ , which is defined on  $\mathcal{R}$ , is a functional of the *macroscopic observables*  $\mathbf{F}_n$  and the *reference measure*  $\lambda$ , although we do not write this dependence explicitly<sup>14</sup>. In statistical thermodynamics it is not the individual case, which is of interest; one tries to find statistical regularities based on the law of large numbers. Boltzmann fundamental contribution can be stated as follows.

- **Boltzmann's law:**  $s(\mathbf{x})$  is the thermodynamical entropy (per unit volume) of the lattice system, which is at equilibrium in the macroscopic state  $\mathbf{x}$ .

From the probability viewpoint, (2.1) shows that  $s$  is nothing else than the rate-function associated with the random variables  $\mathbf{F}_n$ <sup>15</sup>. All thermodynamical properties of the system, as well as large deviations properties of the sequence of random variables  $\mathbf{F}_n$ , are described by  $s$ . For example, let  $C \subset \mathbb{R}^2$  be an open convex subset; then

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \text{Prob}_\lambda(\mathbf{F}_n(\omega) \in C) = \max_{\mathbf{x} \in \text{cl} C} s(\mathbf{x}).$$

In words, the values  $\mathbf{y} = (e, m)$ , which are observed at the thermodynamical limit, when the macroscopic observables are constrained to have their values in  $C$ , are those in the closure of  $C$ , which maximize the thermodynamical entropy<sup>16</sup>,

$$\{\mathbf{y} \in \mathcal{R} \cap \text{cl} C : s(\mathbf{y}) = \max_{\mathbf{x} \in \text{cl} C} s(\mathbf{x})\}.$$

This is one version of the **maximum entropy principle**. Indeed, suppose that  $\mathbf{z} \in C$ ,  $\mathbf{z} \neq \mathbf{y}$  and  $s(\mathbf{z}) < s(\mathbf{y})$ ; if  $\varepsilon > 0$  is small enough, then  $s(\mathbf{y}) > \max_{\mathbf{x} \in B_\varepsilon(\mathbf{z})} s(\mathbf{x})$ . An elementary computation gives for large  $n$  the estimate

$$\text{Prob}_\lambda(\mathbf{F}_n \in B_\varepsilon(\mathbf{z}) \mid \mathbf{F}_n \in C) \asymp \exp(-|\Lambda(n)|[s(\mathbf{y}) - \max_{\mathbf{x} \in B_\varepsilon(\mathbf{z})} s(\mathbf{x})]).$$

<sup>13</sup>In general one does not have concavity of  $s$ , even in statistical mechanics. For example concavity of  $s$  is violated in mean-field models. Nevertheless in thermodynamics concavity is an essential property. If concavity is violated, then equilibrium thermodynamics does not apply. The results presented here do not directly apply to mean-field models. The main reason is that in these models the macroscopic states are values of non-linear functionals of macroscopic observables. This non-linearity leads in particular to the non-concavity of the entropy.

<sup>14</sup>The choice of the macroscopic observables, respectively macroscopic states, cannot be settled by mathematics only in the applications.

<sup>15</sup>See proposition 3.5 and theorem 3.3. Warning: the rate-function is often defined to be  $-s(\mathbf{x})$ .

<sup>16</sup>The maximum of  $s$  on  $\text{cl} C$  is not necessarily unique; non-unicity is related to phase coexistence. However, unicity of the maximum does not imply that there is no phase coexistence.

Convexity is a basic feature of thermodynamics, which has been emphasized and fully developed by Gibbs. Convexity gives an essential relationship between macroscopic states and thermodynamical parameters. The set of macroscopic states  $\mathcal{R}$  is a convex subset of the vector space  $E := \mathbb{R}^2$ . Let  $E'$  be the dual vector space, which is here again  $\mathbb{R}^2$ . Elements of  $E'$  are denoted by  $\mathbf{x}' \equiv (e', m')$ . The relation between  $E$  and  $E'$  is given by a pairing (a bilinear form), which is here the Euclidean scalar product,

$$E' \times E \rightarrow \mathbb{R} \quad \text{and} \quad (\mathbf{x}', \mathbf{x}) \mapsto \langle \mathbf{x}', \mathbf{x} \rangle \equiv e' \cdot e + m' \cdot m.$$

Let  $h : E \mapsto \mathbb{R}$ ; the conjugate function  $h^* : E' \mapsto \mathbb{R}$  is defined by (see e.g. [Ro])

$$h^*(\mathbf{x}') := \sup_{\mathbf{x} \in E} (\langle \mathbf{x}', \mathbf{x} \rangle - h(\mathbf{x})).$$

The function  $h^*$  is convex on  $E'$ ; if  $h$  is itself convex and lower semi-continuous, then the duality theorem states that

$$h^{**} = h \quad \text{i.e.} \quad h(\mathbf{x}) = \sup_{\mathbf{x}' \in E'} (\langle \mathbf{x}', \mathbf{x} \rangle - h^*(\mathbf{x}')).$$

Notice that for any  $\mathbf{x}'$  and  $\mathbf{x}$

$$h^*(\mathbf{x}') + h(\mathbf{x}) \geq \langle \mathbf{x}', \mathbf{x} \rangle. \quad (2.2)$$

Given  $h$ , a convex function, one defines for each  $\mathbf{x}$  a dual element  $\mathbf{x}'$ : two vectors  $\mathbf{x}'$  and  $\mathbf{x}$  are in **duality**<sup>17</sup>,  $\mathbf{x}' \leftrightarrow \mathbf{x}$ , if and only if

$$h^*(\mathbf{x}') + h(\mathbf{x}) = \langle \mathbf{x}', \mathbf{x} \rangle. \quad (2.3)$$

One may have two distinct points  $\mathbf{x}_1$  and  $\mathbf{x}_2$  such that  $\mathbf{x}' \leftrightarrow \mathbf{x}_1$  and  $\mathbf{x}' \leftrightarrow \mathbf{x}_2$ . This occurs when  $h$  is affine above the segment  $[\mathbf{x}_1, \mathbf{x}_2]$ ; such a situation is related to coexistence of several macroscopic states. The geometrical significance of (2.2) and (2.3) is simple and well-known. The affine function

$$\mathbf{y} \mapsto \langle \mathbf{x}', \mathbf{y} \rangle - h^*(\mathbf{x}') \quad (2.4)$$

is always below the graph of  $\mathbf{y} \mapsto h(\mathbf{y})$ , and its contact points with this graph are precisely the dual points of  $\mathbf{x}'$ . The graph of the affine function (2.4) is a tangent plane to the graph of  $h$ ;  $\mathbf{x}'$  is called a **subgradient** to  $h$  at  $\mathbf{x}$ . Stated differently,  $\mathbf{x}' \leftrightarrow \mathbf{x}$  if and only if

$$h(\mathbf{y}) \geq h(\mathbf{x}) + \langle \mathbf{x}', \mathbf{y} - \mathbf{x} \rangle \quad \forall \mathbf{y}.$$

If  $h$  is differentiable, then dual points  $(e', m') \leftrightarrow (e, m)$  are such that

$$\frac{\partial h(e, m)}{\partial e} = e' \quad \text{and} \quad \frac{\partial h(e, m)}{\partial m} = m'.$$

In the present case, since  $s$  is concave,  $h := -s$  is used for defining the dual relationship between elements of  $E$  and  $E'$ . ( $s$  is extended on  $\mathbb{R} \setminus \mathcal{R}$  by  $s(\mathbf{x}) := -\infty$ .) The next result is standard both in thermodynamics and in large deviations

---

<sup>17</sup>This notion is  $h$ -dependent, although we do not write explicitly this dependence in the notation.



theory<sup>18</sup>. It is at that point that formula  $e^{-\beta H}$  enters into the theory, when one determines the conjugate function of  $-s$  at  $\mathbf{x}'$ .  $-\beta H$  is given by<sup>19</sup>

$$\sum_{x \in \Lambda(n)} (e' \cdot e_x + m' \cdot \sigma_x) = \sum_{x \in \Lambda(n)} \frac{\beta J}{2} \sum_{\substack{y \in \mathbb{Z}^2: \\ \|x-y\|=1}} \sigma_x \sigma_y + \sum_{x \in \Lambda(n)} \beta h \sigma_x.$$

In this formula the usual notations of statistical mechanics are introduced. Hence,  $-e' \equiv \beta$ ,  $J$  is the **nearest neighbour coupling constant**,  $m' \equiv \beta h$ , and  $h$  is the **external magnetic field**. When  $\beta J > 0$  ( $e' < 0$ ) the model is ferromagnetic and when  $\beta J < 0$  it is antiferromagnetic<sup>20</sup>. The conjugate function  $(-s)^*$  is

$$\begin{aligned} p(\mathbf{x}') &:= \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \int_{\Omega} \exp \left( \sum_{x \in \Lambda(n)} (e' e_x(\omega) + m' \sigma_x(\omega)) \right) \lambda(d\omega) & (2.5) \\ &= \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \int_{\Omega} \exp \left( \sum_{x \in \Lambda(n)} \left( \sum_{\substack{y \in \mathbb{Z}^2: \\ \|x-y\|=1}} \frac{\beta J}{2} \sigma_x(\omega) \sigma_y(\omega) + \beta h \sigma_x(\omega) \right) \right) \lambda(d\omega) \\ &= \sup_{\mathbf{x}} (s(\mathbf{x}) + \langle \mathbf{x}', \mathbf{x} \rangle) := (-s)^*(\mathbf{x}'). \end{aligned}$$

An important property of  $p(\mathbf{x}')$  is that it is strictly convex. This has two consequences. First, the Boltzmann entropy is always differentiable. The dual point

---

<sup>18</sup> A general result in large deviations theory, Varadhan's theorem, describes how the large deviations properties are modified when one changes the reference measure  $\lambda$  (see [DS] 2.1.10 or [LP] theorem 6.3). Let  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$  be a continuous function and  $\nu_n$  be the probability measure

$$d\nu_n(\omega) := \exp [|\Lambda(n)|\varphi(\mathbf{F}_n(\omega)) + \psi_n(\omega) - |\Lambda(n)|Q_n] \lambda(d\omega),$$

where  $\psi_n(\omega)$  is a boundary term, which verifies

$$\lim_{n \rightarrow \infty} \sup_{\omega \in \Omega} \frac{1}{|\Lambda(n)|} |\psi_n(\omega)| = 0,$$

and  $Q_n$  is the normalization constant,

$$\exp (|\Lambda(n)| Q_n) := \int_{\Omega} \exp |\Lambda(n)| (\varphi(\mathbf{F}_n(\omega)) + \psi_n(\omega)) \lambda(d\omega).$$

If  $\mathbf{F}_n$  is defined on the probability space  $(\Omega, \nu_n)$ , then the new rate-function is

$$\begin{aligned} s_{\varphi}(\mathbf{x}) &:= \min_{\varepsilon > 0} \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \text{Prob}_{\nu_n}(\mathbf{F}_n(\omega) \in B_{\varepsilon}(\mathbf{x})) \\ &= \min_{\varepsilon > 0} \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \text{Prob}_{\nu_n}(\mathbf{F}_n(\omega) \in B_{\varepsilon}(\mathbf{x})) \\ &= s(\mathbf{x}) + \varphi(\mathbf{x}) - Q, \end{aligned}$$

with

$$\lim_{n \rightarrow \infty} Q_n \equiv Q = \sup_{\mathbf{x}} (s(\mathbf{x}) + \varphi(\mathbf{x})).$$

<sup>19</sup>  $-\beta H$  differs from the standard definition of the hamiltonian in  $\Lambda(n)$  by a boundary term.  $\beta$  is introduced here for making the link with the standard notations. See however next footnote!  $-\beta H$  is written later on  $\sum_{j \in \Lambda(n)} \theta_j f_{\mathbf{x}'}$ ; see (2.10).

<sup>20</sup> For spin systems like the Ising model it is better not to introduce the inverse temperature  $\beta$ , which can be positive and negative, and to use only the natural variables  $\mathbf{x}' = (e', m')$  of the dual space of  $\mathcal{R}$ , which are the thermodynamical parameters (generalized chemical potentials) related to macroscopic states via the duality relationship. It is also natural to consider the Ising ferromagnetic and antiferromagnetic models as a single lattice model.

$(e', m') \equiv \mathbf{x}'$  of  $(e, m) \equiv \mathbf{x}$  is given by

$$\frac{\partial s(e, m)}{\partial e} = -e' \quad \text{and} \quad \frac{\partial s(e, m)}{\partial m} = -m'. \quad (2.6)$$

Second, for any macroscopic state  $\mathbf{x} \in \text{int } \mathcal{R}$  there exists a unique dual point  $\mathbf{x}'$ . To summarize, the basic space in statistical thermodynamics is the space  $\mathcal{R}$  of macroscopic states. The thermodynamical entropy  $s$  is defined on  $\mathcal{R}$  by (2.1). There is a natural notion of dual points. To each macroscopic state  $\mathbf{x}$  there exists a unique dual point  $\mathbf{x}'$ , which defines the thermodynamical parameters for the state  $\mathbf{x}$ .

- **Equivalence of ensembles**<sup>21</sup> at the level of thermodynamics is the statement that  $s$  and  $p$  contain the same thermodynamical information, in the following precise sense:

$$p = (-s)^*, \quad p^* = -s \quad \text{and} \quad \mathbf{x}' \leftrightarrow \mathbf{x} \text{ iff } p(\mathbf{x}') - s(\mathbf{x}) = \langle \mathbf{x}', \mathbf{x} \rangle. \quad (2.7)$$

Instead of using  $\mathbf{x}$  and  $s$  for describing the thermodynamical properties of the system one can give an equivalent (dual) description using  $\mathbf{x}'$  and  $p \equiv (-s)^*$ . *Equivalence of ensembles holds even when there is phase coexistence, i.e. even if  $p$  is not differentiable and the set  $\{\mathbf{x} : \mathbf{x}' \leftrightarrow \mathbf{x}\}$  is not a singleton.* Equivalence of ensembles is a consequence of the concavity of  $s$ .

**2.2. Statistical mechanics.** In statistical mechanics the properties of a system are described by probability measures. Since I consider only homogeneous or translation invariant situations<sup>22</sup>,  $\mu \in \mathcal{M}_1^{+, \theta}$ , the space of translation invariant probability measures. A basic question is “what are the measures, which describe equilibrium situations?”. These measures are called “equilibrium states”. I do not define them as Gibbs states or through the variational principle (see (2.11)), which is a natural approach, but I illustrate the fact that the analysis of large deviations of the empirical measure leads to another natural approach, showing that, in some sense, statistical mechanics can be derived from Boltzmann’s law. Equilibrium states are defined with respect to a given macroscopic state  $\mathbf{x}$ . Then one *proves* that an equilibrium state verifies the variational principle and that it is a Gibbs state. This approach fits our viewpoint that the basic notions are those of macroscopic states and of thermodynamical entropy, and that *Boltzmann’s law is the central physical result, that is, the thermodynamical entropy is defined by (2.1)*. The rest is essentially mathematics.

There is a simple mathematical construction, the contraction principle, which leads to a fundamental formula for the entropy  $s$ , and which is the key-point for introducing the equilibrium states. The map  $\mathbf{F}_n$  is decomposed in a canonical way into  $T_{\Lambda(n)}$  and a linear map  $\Phi_{\mathbf{f}}$ .  $T_{\Lambda(n)}$  is the **empirical measure**; it is a map defined

---

<sup>21</sup> As already mentioned  $s$  defined by (2.1) needs not be concave. In such a case although  $p$  can be defined as the conjugate function of  $-s$ , one cannot have (2.7). But it may be true that for a given  $\mathbf{x}$  there exists a dual point  $\mathbf{x}'$  characterized by  $p(\mathbf{x}') - s(\mathbf{x}) = \langle \mathbf{x}', \mathbf{x} \rangle$ . We still say in that case that we have equivalence of ensembles (in a weaker sense) for the pair  $(\mathbf{x}', \mathbf{x})$ .

<sup>22</sup> The theory developed below is restricted to translation invariant measures although it is possible to consider periodic situations.

on  $\Omega$ , with values in the space  $\mathcal{M}_1^+$  of probability measures on  $\Omega$ <sup>23</sup>,

$$T_{\Lambda(n)}(\omega) := \frac{1}{|\Lambda(n)|} \sum_{x \in \Lambda(n)} \delta_{\theta_x \omega}.$$

$\mathbf{F}_n$  is decomposed into  $T_{\Lambda(n)}$  and  $\Phi_{\mathbf{f}} : \mathcal{M}_1^+ \rightarrow \mathbb{R}^2$ , which is  $\Phi_{\mathbf{f}}(\mu) := \langle \mathbf{f}, \mu \rangle$ .

$$\begin{array}{ccc} \Omega & \xrightarrow{T_{\Lambda(n)}} & \mathcal{M}_1^+ \\ & \searrow & \downarrow \Phi_{\mathbf{f}} \\ & & \mathbb{R}^2 \end{array} \quad \mathbf{F}_n = \Phi_{\mathbf{f}} \circ T_{\Lambda(n)}$$

$T_{\Lambda(n)}$  is a sum of local quantities whenever  $T_{\Lambda(n)}$  is tested against local functions. Then one proceeds as in subsection 2.1. One first introduces a topology on the space  $\mathcal{M}_1^+$ . Let  $\mathcal{F}_{\text{qloc}}$  be the space of functions, which are uniform limit of bounded local functions. An element of  $\mathcal{F}_{\text{qloc}}$  is a **quasilocal** function. A sequence of probability measures  $\mu_n$  converges to a probability measure  $\mu$  if and only if  $\lim_{n \rightarrow \infty} \langle g, \mu_n \rangle = \langle g, \mu \rangle$  for all quasilocal functions  $g$ . Let  $G$  be any open neighbourhood of  $\mu$ ; by arguments, which are almost the same as those used for proving formula (2.1), one obtains

$$\begin{aligned} \Sigma(\mu) &:= \inf_{G \ni \mu} \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \text{Prob}_{\lambda}(T_{\Lambda(n)}(\omega) \in G) \\ &= \inf_{G \ni \mu} \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \text{Prob}_{\lambda}(T_{\Lambda(n)}(\omega) \in G). \end{aligned} \quad (2.8)$$

The next theorem is fundamental.

**Theorem 2.1.** *1.  $\Sigma(\mu) = h_{Sh}(\mu) - \ln 2$  if  $\mu$  is translation invariant<sup>24</sup>,  $\Sigma(\mu) = -\infty$  otherwise.  $h_{Sh}(\mu)$  is the Shannon entropy<sup>25</sup> of  $\mu$ .*

*2. The Boltzmann entropy  $s$  associated with the macroscopic observables  $\mathbf{F}_n$  verifies for all macroscopic states  $\mathbf{x}$  the identity*

$$s(\mathbf{x}) = \sup \{ \Sigma(\mu) : \mu \in \mathcal{M}_1^+, \langle \mathbf{f}, \mu \rangle = \mathbf{x} \}. \quad (2.9)$$

*3. For all  $\mathbf{x}$  such that  $s(\mathbf{x}) > -\infty$  there exists  $\mu \in \mathcal{M}_1^+$  with  $s(\mathbf{x}) = \Sigma(\mu)$ .*

At the light of the previous theorem, definition 2.1 is natural. Suppose that the system is at equilibrium in the macroscopic state  $\mathbf{x}$ , with Boltzmann entropy  $s(\mathbf{x})$ .

<sup>23</sup>  $\delta_{\theta_x \omega}$  is the Dirac unit mass at  $\theta_x \omega$ .

<sup>24</sup> The factor  $\ln 2$  is a consequence of the normalization of  $\lambda$ . If the counting measure is used, then this factor does not occur.

<sup>25</sup>  $h_{Sh}(\mu) = -\lim_n \frac{1}{|\Lambda(n)|} \sum_{\omega_{\Lambda(n)}} \mu[\omega_{\Lambda(n)}] \log \mu[\omega_{\Lambda(n)}]$ , where  $[\omega_{\Lambda(n)}] := \{ \omega' : \omega'_{\Lambda(n)} = \omega_{\Lambda(n)} \}$ .

**Definition 2.1.** A translation invariant probability measure  $\mu$  is an equilibrium state corresponding to the macroscopic state  $\mathbf{x}$  if and only if

$$\Sigma(\mu) = s(\mathbf{x}) \quad \text{and} \quad \langle \mathbf{f}, \mu \rangle = \mathbf{x}.$$

The definition of equilibrium state is done directly at the thermodynamical limit since  $s(\mathbf{x})$  is defined in that limit. It is possible that  $\mu$  does not verify the law of large numbers. If  $\mu$  has a nontrivial ergodic decomposition, say  $\mu = \sum_{j=1}^k a_j \mu_j$  with  $a_j \geq 0$  and  $\sum_{j=1}^k a_j = 1$ , then  $\mathbf{x}$  can be decomposed as  $\mathbf{x} = \sum_j a_j \mathbf{x}_j$ ,  $\mathbf{x}_j := \langle \mathbf{f}, \mu_j \rangle$  and  $s(\mathbf{x}) = \sum_j a_j s(\mathbf{x}_j)$ . Indeed, using the concavity of  $s$ ,  $s(\mathbf{x}_j) \geq h_{Sh}(\mu_j) - \ln 2$  and then the fact that  $h_{Sh}$  is affine, one obtains

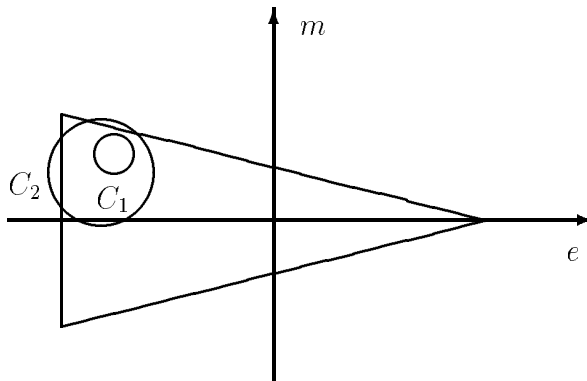
$$s(\mathbf{x}) = s\left(\sum_j a_j \mathbf{x}_j\right) \geq \sum_j a_j s(\mathbf{x}_j) \geq \sum_j a_j (h_{Sh}(\mu_j) - \ln 2) = h_{Sh}(\mu) - \ln 2 = s(\mathbf{x}).$$

Since Gibbs, affine portions in the graph of  $s$  are related to phase coexistence. See also the comments at the end of this section. For a nice introduction to the work of Gibbs see the introduction of Wightman for [I].

As a consequence of the previous theorem, if  $\mu$  is an equilibrium state for  $\mathbf{x}$  and  $\nu$  another translation invariant probability measure, such that  $\langle \mathbf{f}, \nu \rangle = \mathbf{x}$ , then

$$h_{Sh}(\mu) \geq h_{Sh}(\nu).$$

Equilibrium states have maximal Shannon entropy among all translation invariant probability measures verifying the constraint  $\langle \mathbf{f}, \nu \rangle = \mathbf{x}$ . This is another version of the **maximum entropy principle**<sup>26</sup>. Let  $C_1 \subset C_2$  be two closed convex subsets with nonempty interiors. Suppose that one first constrains the system so that  $\mathbf{x} \in C_1$ . The equilibrium value of the macroscopic observable is  $\mathbf{y}_1$ , which maximizes the Boltzmann entropy  $s$  over  $C_1$ ; let  $\mu_1$  be an equilibrium state for  $\mathbf{y}_1$ . Relaxing the constraint by imposing only that  $\mathbf{x} \in C_2$  leads to a new value  $\mathbf{y}_2$  of the macroscopic observable; let  $\mu_2$  be the new equilibrium state for  $\mathbf{y}_2$ . Then  $s(\mathbf{y}_2) \geq s(\mathbf{y}_1)$  and  $h_{Sh}(\mu_2) \geq h_{Sh}(\mu_1)$  in accordance with the second law of thermodynamics.



<sup>26</sup> “Maximum entropy principle” is a central theme in the works of Jaynes. An interesting selection of papers by Jaynes is [J]. The “maximum entropy principle” is used by Jaynes as a criterion for selecting the equilibrium states. For applying this principle one first defines the entropy of a measure (and justifies this choice). However we do not proceed in this way. Our starting point is the identification of the rate-function  $s$  with the thermodynamical entropy by Boltzmann’s law. (2.9) is a theorem and formula  $\Sigma(\mu) = h_{Sh}(\mu_j) - \ln 2$  is the result of a computation; it is not a definition of  $\Sigma(\mu)$  (see subsection 3.4).

The following construction gives equilibrium states. Let  $N_{\mathbf{x}}$  be the set of equilibrium states corresponding to the macroscopic state  $\mathbf{x} \in \mathcal{R}$ . Choose  $C_k \subset \mathbb{R}^2$ , a decreasing sequence of closed convex neighbourhoods of  $\mathbf{x}$ , such that

$$C_k \downarrow \{\mathbf{x}\} \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \text{Prob}_\lambda(\mathbf{F}_n \in C_n) = s(\mathbf{x}).$$

It is always possible to construct such sequences<sup>27</sup>. Then one defines conditioned measures  $\mu_n$  on  $\Omega$  by

$$\mu_n := \lambda[\cdot | \{\mathbf{F}_n \in C_n\}].$$

Notice that

$$\mu_n = \lambda[\cdot | \{T_{\Lambda(n)} \in \Phi_{\mathbf{f}}^{-1}(C_n)\}]$$

and that  $\Phi_{\mathbf{f}}^{-1}(\{\mathbf{x}\})$  is the fiber over  $\mathbf{x}$ ,  $\{\nu \in \mathcal{M} : \langle \mathbf{f}, \nu \rangle = \mathbf{x}\}$ , which contains the equilibrium states<sup>28</sup>. The empirical measures  $T_{\Lambda(n)}$ , considered as random variables on the probability spaces  $(\Omega, \mu_n)$ , verify a weak form of the law of large numbers. Let  $G \supset N_{\mathbf{x}}$  be an open neighbourhood of  $N_{\mathbf{x}}$ ; then

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \text{Prob}_{\mu_n}(T_{\Lambda(n)} \notin G) < 0.$$

Consequently the empirical measures are asymptotically concentrated on  $N_{\mathbf{x}}$ . This implies that the sequence of averaged measures

$$\tilde{\mu}_n := \frac{1}{|\Lambda(n)|} \sum_{x \in \Lambda(n)} \theta_x \mu_n,$$

has cluster points, and each cluster point is in  $N_{\mathbf{x}}$ . A cluster point  $\tilde{\mu}$  of  $\{\tilde{\mu}_n\}$  is called a **microcanonical state**.

As before one introduces a notion of duality. Let  $E := \mathcal{M}$  be the vector space of (signed) measures with the topology defined previously.  $\mathcal{M}_1^+$  is a convex subset of  $E$ . The dual space is  $E' := \mathcal{F}_{\text{qloc}}$ , the space of quasilocal functions; the pairing is defined on  $\mathcal{F}_{\text{qloc}} \times \mathcal{M} \rightarrow \mathbb{R}$  by

$$(g, \mu) \mapsto \langle g, \mu \rangle = \int_{\Omega} g(\omega) \mu(d\omega).$$

The conjugate function of  $-\Sigma(\mu)$  is determined by Varadhan's theorem ([DS] 2.1.10); it is an extension of  $p(\mathbf{x}')$  to all  $g \in \mathcal{F}_{\text{qloc}}$  (see (2.5))

$$\begin{aligned} p(g) &:= (-\Sigma)^*(g) = \sup \{ \Sigma(\nu) + \langle g, \nu \rangle : \nu \in \mathcal{M}_1^{+, \theta} \} \\ &= \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \int_{\Omega} \exp \left( \sum_{x \in \Lambda(n)} \theta_x g(\omega) \right) \lambda(d\omega). \end{aligned}$$

The notion of duality is defined using the convex function  $-\Sigma$ . The measure  $\nu$  is in **duality** with the quasilocal function  $g$ ,  $g \leftrightarrow \nu$ , if and only if

$$p(g) - \Sigma(\nu) = \langle g, \nu \rangle.$$

<sup>27</sup> See the stretching argument given after the definition of LD-regular sequences in section 5.

<sup>28</sup> Since the sequence  $\{C_k\}$  is decreasing and  $\Phi_{\mathbf{f}}$  continuous, the sequence  $\{\Phi_{\mathbf{f}}^{-1}(C_k)\}$  is also LD-regular with limit  $\Phi_{\mathbf{f}}^{-1}(\{\mathbf{x}\})$ . One can therefore apply theorem 5.1.

Since the domain of  $\Sigma(\nu)$  is contained in  $\mathcal{M}_1^{+, \theta}$ ,  $g \leftrightarrow \nu$  implies  $\nu \in \mathcal{M}_1^{+, \theta}$ . Let  $\mathbf{x}'$  be the dual point of  $\mathbf{x}$ ; one defines  $f_{\mathbf{x}'} \in \mathcal{F}_{\text{qloc}}$  by

$$f_{\mathbf{x}'}(\omega) := \langle \mathbf{x}', \mathbf{f}(\omega) \rangle = e' \cdot \epsilon_0(\omega) + m' \cdot \sigma_0(\omega), \quad (2.10)$$

so that  $p(f_{\mathbf{x}'}) = p(\mathbf{x}')$ .

- The duality statement  $f_{\mathbf{x}'} \leftrightarrow \mu$ , that is,

$$p(f_{\mathbf{x}'}) = \Sigma(\mu) + \langle f_{\mathbf{x}'}, \mu \rangle = \sup \{ \Sigma(\nu) + \langle f_{\mathbf{x}'}, \nu \rangle : \nu \in \mathcal{M}_1^{+, \theta} \}, \quad (2.11)$$

is the **variational principle** of statistical mechanics.

**Remark.**  $f_{\mathbf{x}'} \leftrightarrow \mu$  is equivalent to the statement that the affine functional  $g \mapsto \langle g, \mu \rangle + \Sigma(\mu)$  is a tangent plane to  $p$  at the quasilocal function  $f_{\mathbf{x}'}$ . Indeed, for any  $g \in \mathcal{F}_{\text{qloc}}$ ,

$$\begin{aligned} p(g) &\geq \Sigma(\mu) + \langle f_{\mathbf{x}'}, \mu \rangle + \langle g - f_{\mathbf{x}'}, \mu \rangle \\ &= p(f_{\mathbf{x}'}) + \langle g - f_{\mathbf{x}'}, \mu \rangle. \end{aligned}$$

**Theorem 2.2.** 1. Let  $\mathbf{x} \in \text{int } \mathcal{R}$  and  $\mathbf{x}'$  be the unique dual point of  $\mathbf{x}$ . Then *a* and *b* are equivalent.

- The translation invariant probability measure  $\mu$  is an equilibrium state for  $\mathbf{x}$ .
- $\mu$  and  $f_{\mathbf{x}'}$  are in duality<sup>29</sup> and  $\langle \mathbf{f}, \mu \rangle = \mathbf{x}$ .

2. Any ergodic equilibrium state<sup>30</sup>  $\nu$  for  $\mathbf{x}$  is of the form  $\nu = \lim_{n \rightarrow \infty} \nu_n$ , where

$$d\nu_n = \exp \left[ \sum_{x \in \Lambda(n)} \theta_x f_{\mathbf{x}'} + \psi_n - |\Lambda(n)| p_n(\mathbf{x}') \right] d\lambda,$$

$\mathbf{x}'$  is such that  $\mathbf{x}' \leftrightarrow \mathbf{x}$ ,  $\psi_n$  is a uniform boundary term<sup>31</sup>, and  $p(\mathbf{x}') = \lim_{n \rightarrow \infty} p_n(\mathbf{x}')$ . Conversely, if  $\lim_{n \rightarrow \infty} \nu_n = \nu$  is translation-invariant, then  $\nu$  is an equilibrium state for  $\mathbf{y} := \langle \mathbf{f}, \nu \rangle$ , and one has  $\mathbf{x}' \leftrightarrow \mathbf{y}$ .

3. Any equilibrium state for  $\mathbf{y}$  with  $\mathbf{x}' \leftrightarrow \mathbf{y}$  is a Gibbs state with respect to the local specification

$$\pi_\Lambda(\omega_\Lambda | \omega_{\Lambda^c}) = \lim_{n \rightarrow \infty} \frac{\exp \left( \sum_{x \in \Lambda(n)} \theta_x f_{\mathbf{x}'}(\omega_\Lambda, \omega_{\Lambda^c}) \right)}{\int \exp \left( \sum_{x \in \Lambda(n)} \theta_x f_{\mathbf{x}'}(\eta_\Lambda, \omega_{\Lambda^c}) \right) \lambda(d\eta_\Lambda)}.$$

Conversely, any translation invariant Gibbs measure  $\mu$  with respect to this local specification is an equilibrium state for  $\mathbf{y} := \langle \mathbf{f}, \mu \rangle$ , and one has  $\mathbf{x}' \leftrightarrow \mathbf{y}$ .

<sup>29</sup>  $\mu \leftrightarrow f_{\mathbf{x}'} \not\Rightarrow \langle \mathbf{f}, \mu \rangle = \mathbf{x}$ . If  $e' \ll -1$  (ferromagnetism and low temperatures) and  $m' = 0$  (zero magnetic field), one has  $(e', 0) \leftrightarrow (e, m)$  with  $|m| \leq m^*$ , where  $m^*$  is the spontaneous magnetization.

<sup>30</sup> For lattice models the set of equilibrium states for given thermodynamical parameters form a Choquet simplex. The extremal elements of this simplex are ergodic measures. In these lectures I shall not consider this aspect of the theory. See e.g. [Ru2] and [G1].

<sup>31</sup> Uniform boundary terms are defined in section 4. This means that

$$\lim_{n \rightarrow \infty} \sup_{\omega \in \Omega} \frac{1}{|\Lambda(n)|} |\psi_n(\omega)| = 0$$

and that there exists a function  $r : \mathbb{R}^+ \mapsto \mathbb{R}^+$  such that  $\lim_{x \rightarrow \infty} r(x) = 0$  and

$$\sup_{\omega, \eta} \{ |\psi_n(\omega) - \psi_n(\eta)| : \omega_i = \eta_i, i \notin \Lambda(m) \} \leq r(n - m) |\Lambda(m)| \quad \forall n > m.$$

4. If  $\mu$  is an equilibrium state for  $\mathbf{x}$  and  $\mathbf{x}' \leftrightarrow \mathbf{x}$ , then

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \text{Prob}_\mu \left( \left| \sum_{x \in \Lambda(n)} [\theta_x \mathbf{f}(\omega) - \langle \mathbf{f}, \mu \rangle] \right| \geq \varepsilon |\Lambda(n)| \right) \\ \leq \sup_{\mathbf{y} \notin B_\varepsilon(\mathbf{x})} (s(\mathbf{y}) + \langle \mathbf{x}', \mathbf{y} \rangle - p(\mathbf{x}')) . \end{aligned}$$

**Comments.** 1. An important consequence of theorem 2.2 is that the microcanonical states defined above, which are equilibrium states according to our definition, are also Gibbs states and are solutions of the variational principle. This fact is one form of the **equivalence of ensembles at the level of measures**. The key-hypothesis for this result is that  $\mathbf{x}'$  and  $\mathbf{x}$  are in duality. Stated differently, this is equivalent to the existence of a subgradient  $\mathbf{x}'$  to the Boltzmann entropy  $s$  at the macroscopic state  $\mathbf{x}$ , or to the validity of the equivalence of ensembles at the level of thermodynamics. In models like mean-field models, where the rate-function  $s$  may not be concave, the equivalence of ensembles at the level of thermodynamics may fail and one can show that the equivalence of ensembles at the level of measures also fails [LPS2]<sup>32</sup>.

- Equivalence of ensembles at the level of thermodynamics implies equivalence of ensembles at the level of measures.

It is also important to stress again that the equivalence of ensembles for models like those considered here *is valid when there is phase coexistence*.

2. If  $\mu$  is an equilibrium state for the macroscopic state  $\mathbf{x}$  and the Boltzmann entropy  $s$  is strictly concave at  $\mathbf{x}$ , then 4 gives a large deviations estimate for the macroscopic variables  $\mathbf{F}_n$  in a strong form, i.e. in the volume scale. Minlos in [M] insists rightly on the importance of such a result in statistical mechanics; he calls it “the principle of the representative character of the mean for the macroscopic observables  $\mathbf{F}_n$ ”. When such a result holds, then it is justified to identify the observed value of  $\mathbf{F}_n$ , for  $n$  large, with  $\langle \mathbf{f}, \mu \rangle$ . Notice that strict concavity of  $s$  at  $\mathbf{x}$  means that  $\mathbf{x}$  is the unique macroscopic state such that  $\mathbf{x}' \leftrightarrow \mathbf{x}$ .

More generally one has the following result. Assume that  $\mathbf{x}' \leftrightarrow \mathbf{x}$  and there is a unique  $\mathbf{x}'$  for all  $\mathbf{y}$  such that  $\mathbf{x}' \leftrightarrow \mathbf{y}$ . Let  $N^{\mathbf{x}'}$  be the set solutions  $\mu$  of the variational principle

$$p(\mathbf{x}') = \Sigma(\mu) + \langle f_{\mathbf{x}'}, \mu \rangle ,$$

and  $N_{\mathbf{x}}$  be the set of equilibrium states  $\mu$  for the macroscopic state  $\mathbf{x}$ . One has

$$N^{\mathbf{x}'} = \bigcup_{\substack{\mathbf{x} \in \mathcal{R} \\ \mathbf{x}' \leftrightarrow \mathbf{x}}} N_{\mathbf{x}} .$$

If  $\mu \in N^{\mathbf{x}'}$ , then the empirical measures  $T_{\Lambda(n)}$ , defined on the probability space  $(\Omega, \mathcal{F}, \mu)$ , obey a large deviations principle with rate-function (see theorem 4.2)

$$\Sigma(\nu) + \langle f_{\mathbf{x}'}, \nu \rangle - p(f_{\mathbf{x}'}) \quad \text{if } \nu \in \mathcal{M}_1^{+, \theta} .$$

(If  $\nu \notin \mathcal{M}_1^{+, \theta}$ , the rate-function is  $-\infty$ .) Let  $g$  be a local real-valued observable, and

$$G_n := \frac{1}{|\Lambda(n)|} \sum_{x \in \Lambda(n)} \theta_x g .$$

<sup>32</sup>See also the remark following the proof of lemma 5.1.

By the contraction principle, the random variables  $\{G_n\}$  obey a large deviations principle with rate-function

$$s_g(t) := \sup\{\Sigma(\nu) + \langle f_{\mathbf{x}'}, \nu \rangle - p(f_{\mathbf{x}'}) : \nu \in \mathcal{M}_1^{+, \theta} \text{ and } \langle g, \nu \rangle = t\}.$$

$s_g(t) \leq 0$ . One has  $s_g(t) = 0$  if and only if there exists  $\nu \in N^{\mathbf{x}'}$  such that  $\langle g, \nu \rangle = t$ , since  $\nu \in N^{\mathbf{x}'}$  if and only if  $\Sigma(\nu) + \langle f_{\mathbf{x}'}, \nu \rangle - p(f_{\mathbf{x}'}) = 0$ . Hence, if  $\langle g, \nu \rangle = \langle g, \nu' \rangle \forall \nu, \nu' \in N^{\mathbf{x}'}$ , for example if  $|N^{\mathbf{x}'}| = 1$ , then the principle of the representative character of the mean for the macroscopic observables  $G_n$  holds in the strong form:  $\forall \varepsilon > 0$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \text{Prob}_\mu(|G_n(\omega) - \langle g, \mu \rangle| \geq \varepsilon |\Lambda(n)|) \leq \sup_{t \notin B_\varepsilon(\langle g, \mu \rangle)} s_g(t) < 0.$$

3. Strict concavity of  $s$  does not imply uniqueness or ergodicity of the equilibrium state. For example, I could have chosen instead of  $\mathbf{F}_n$  the real-valued macroscopic observable

$$G_n := \frac{1}{|\Lambda(n)|} \sum_{x \in \Lambda(n)} e_x.$$

If one computes the entropy with respect to this macroscopic observable, then the entropy is always a strictly concave function<sup>33</sup>. The thermodynamical statistical entropy is defined with respect to a *given choice* of macroscopic observables like  $G_n$  or  $\mathbf{F}_n$ . So, if the choice is  $G_n$ , the corresponding entropy is strictly concave; nevertheless there is a phase transition with symmetry breaking if the dual parameter  $e' \ll -1$ . When this is the case the principle of the representative character of the mean does not hold in the strong form for the mean magnetization.

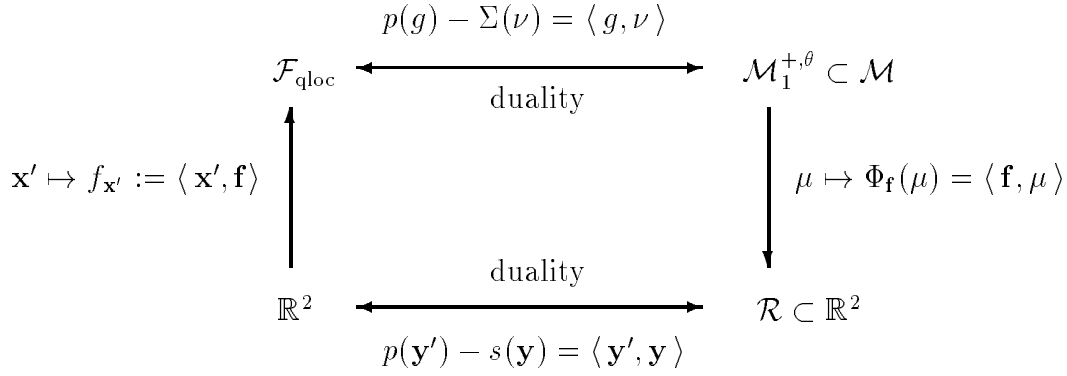
Similarly, if  $e'$  is positive (antiferromagnetic case), then the entropy  $s$  (defined by (2.1)) is strictly concave, because there exists a unique translation invariant equilibrium state [GH]. However, there is a phase transition for the antiferromagnet with different staggered mean magnetizations. Of course in such a case one can extend the whole formalism by defining macroscopic observables with respect to sublattices. Then an equilibrium state is a periodic probability measure, which maximizes the Shannon entropy. However, for a general lattice model one usually does not know in advance which macroscopic observables or which sublattices are relevant.

4. The next diagram summarizes the main thesis of this section.

---

<sup>33</sup>This result is probably true for the  $d$ -dimensional Ising model,  $d \geq 2$ . For the 2-dimensional Ising model, this follows from the fact that there is a unique translation invariant equilibrium state in the anti-ferromagnetic regime [GH], and that all translation equilibrium states are convex combinations of at most two translation invariant equilibrium states in the ferromagnetic regime [MM].





Starting with a macroscopic state  $\mathbf{x} \in \text{int } \mathcal{R}$  at the lower right corner one defines the equilibrium states  $\mu$  using the vertical right arrow:  $\mu$  has maximal Shannon entropy among the probability measures in  $\{\nu \in \mathcal{M}_1^{+, \theta} : \langle \mathbf{f}, \nu \rangle = \mathbf{x}\}$ . The key-point for this definition is the fact that

$$s(\mathbf{x}) = \sup\{\Sigma(\nu) : \nu \in \mathcal{M}_1^{+, \theta}, \langle \mathbf{f}, \nu \rangle = \mathbf{x}\}.$$

Using the duality at the thermodynamical level one constructs the dual point  $\mathbf{x}'$  to  $\mathbf{x}$ , which is uniquely defined by  $p(\mathbf{x}') - s(\mathbf{x}) = \langle \mathbf{x}', \mathbf{x} \rangle$  since  $p = (-s)^*$  is strictly convex.  $\mathbf{x}'$  determines the thermodynamical parameters for the macroscopic state  $\mathbf{x}$  and the one-site interaction-energy of the system is given by the quasilocal function  $f_{\mathbf{x}'}$ . Formally the hamiltonian of the model (including the thermodynamical parameters) is given by  $-\sum_j \theta_j f_{\mathbf{x}'}$ . One has  $p(f_{\mathbf{x}'}) = p(\mathbf{x}')$ . The dual points of  $f_{\mathbf{x}'}$  are the translation invariant probability measures, which are solutions of the variational principle  $p(f_{\mathbf{x}'}) - \Sigma(\nu) = \langle f_{\mathbf{x}'}, \nu \rangle$ . If  $f_{\mathbf{x}'} \leftrightarrow \nu$ , then  $\nu$  is an equilibrium state for the macroscopic state  $\mathbf{y} := \langle f, \nu \rangle$  and  $\mathbf{x}' \leftrightarrow \mathbf{y}$ . Indeed, suppose that  $\langle f, \nu \rangle = \mathbf{y}$ ; then  $\langle f_{\mathbf{x}'}, \nu \rangle = \langle \mathbf{x}', \mathbf{y} \rangle$  and

$$p(\mathbf{x}') - \langle \mathbf{x}', \mathbf{y} \rangle = \Sigma(\nu) \geq \Sigma(\mu) \quad \forall \mu \text{ such that } \langle f, \mu \rangle = \mathbf{y}.$$

Therefore  $\Sigma(\nu) = s(\mathbf{y})$  and  $\mathbf{x}' \leftrightarrow \mathbf{y}$ . Conversely, an equilibrium state  $\mu$  for the macroscopic state  $\mathbf{x}$ , where  $\mathbf{x}' \leftrightarrow \mathbf{x}$ , is a dual point to  $f_{\mathbf{x}'}$  since

$$s(\mathbf{x}) = p(\mathbf{x}') - \langle \mathbf{x}', \mathbf{x} \rangle = p(f_{\mathbf{x}'}) - \langle f_{\mathbf{x}'}, \mu \rangle = \Sigma(\mu).$$

On the other hand, when studying a particular model one usually starts with the upper left corner, i.e. with the hamiltonian or  $-\sum_j \theta_j f_{\mathbf{x}'}$ . Then one constructs the solutions of the variational principle as functions of the thermodynamical parameters  $\mathbf{x}'$ . In statistical physics the **phase diagram** of a model is given in the space of the thermodynamical parameters: for each value of  $\mathbf{x}'$  one gives the list of (pure) phases corresponding to the thermodynamical parameter  $\mathbf{x}'$ .

It is interesting to notice that, using only the fact that  $\Sigma(\nu)$  is affine on  $\mathcal{M}_1^{+, \theta}$ , and defining  $s(\mathbf{x}) := \sup\{\Sigma(\nu) : \nu \in \mathcal{M}_1^{+, \theta} \text{ and } \langle \mathbf{f}, \nu \rangle = \mathbf{x}\}$ , a similar diagram can be proven, starting with the upper right corner.  $s$  defined in this way is concave. In that generality there is no interpretation of  $s(\mathbf{x})$  as rate-function. As already mentioned my viewpoint in these lectures is different. The primary object is  $s(\mathbf{x})$ ;  $s(\mathbf{x})$  has a direct physical meaning: it is the Boltzmann entropy of the model, and therefore describes the thermodynamical properties of the model. The thermodynamical parameters  $\mathbf{x}'$  are defined by duality from  $s(\mathbf{x})$ .

5. The viewpoint adopted is a statistical one. For the thermodynamical formalism, which I exposed here, it is crucial that the macroscopic observables are sums of local quantities. The ergodic theorem for the dynamical system  $(\Omega; \mu; \theta_x, x \in \mathbb{Z}^2)$  ( $\mu \in \mathcal{M}_1^{+, \theta}$ ) is important in that context. One can prove that any (translation invariant) equilibrium states can be uniquely decomposed into ergodic probability measures, which are equilibrium states. Ergodic equilibrium states are interpreted as describing pure phases of the model (in contrast to mixed phases), because the law of large numbers is valid for any macroscopic observables  $\mathbf{F}_n$ .

### 6. Affine portions of the graph of the entropy.

The following hypothesis is made below: for all  $\mathbf{y}$ , there exists a unique  $\mathbf{x}'$  such that  $\mathbf{x}' \leftrightarrow \mathbf{y}$ . This is equivalent to the strict convexity of  $\mathbf{y}' \mapsto p(\mathbf{y}')$ . One has in this case

$$N^{\mathbf{x}'} = \bigcup_{\substack{\mathbf{x} \in \mathcal{R} \\ \mathbf{x}' \leftrightarrow \mathbf{x}}} N_{\mathbf{x}}.$$

For any  $\mu \in N^{\mathbf{x}'}$ , when the macroscopic observable  $\mathbf{F}_n$  is defined on the probability space  $(\Omega, \mathcal{F}, \mu)$ , the rate-function for  $\mathbf{F}_n$  is given by the formula

$$s_{\mathbf{x}'}(\mathbf{x}) := s(\mathbf{x}) + \langle \mathbf{x}', \mathbf{x} \rangle - p(\mathbf{x}').$$

Suppose that the graph of the entropy  $s : \mathbb{R}^k \rightarrow \mathbb{R}$  has an affine portion  $\mathcal{A} \subset \mathbb{R}^k \times \mathbb{R}$ . Let  $\mathbf{x} \mapsto p(\mathbf{x}') - \langle \mathbf{x}', \mathbf{x} \rangle$  be the supporting hyperplane containing  $\mathcal{A}$ , that is,

$$p(\mathbf{x}') - \langle \mathbf{x}', \mathbf{x} \rangle \geq s(\mathbf{x}) \quad \forall \mathbf{x},$$

and if  $(\mathbf{x}, s(\mathbf{x})) \in \mathcal{A}$ , then

$$p(\mathbf{x}') - \langle \mathbf{x}', \mathbf{x} \rangle = s(\mathbf{x}).$$

This implies that

$$\mathcal{A} = \{(\mathbf{x}, s(\mathbf{x})) : s_{\mathbf{x}'}(\mathbf{x}) = 0\}. \quad (2.12)$$

If  $\mathcal{A}$  has more than one point, then one has phase coexistence in the sense that  $N^{\mathbf{x}'}$  contains several elements. Even in this case the law of large numbers of the macroscopic observables  $\mathbf{F}_n$  defined on  $(\Omega, \mathcal{F}, \mu)$  may still be valid. However, in statistical mechanics one expects for models like the Ising model that phase coexistence is related to violation of the law of large numbers. The criterion for existence of phase transition of Berezin-Sinai [BS] is based on that property and on the fact that phase coexistence is often related to symmetry breakdown. See also [Do]. The original argument of Peierls [Pe] is precisely of that type. Peierls proved for the Ising model, that there exist  $\delta > 0$  and  $\varepsilon > 0$ , such that for  $n$  large enough,

$$\text{Prob} \left( \left| \frac{1}{|\Lambda(n)|} \sum_{i \in \Lambda(n)} \sigma_i \right| \geq \delta \right) \geq \varepsilon. \quad (2.13)$$

He proved (2.13) under the following conditions:  $m' = 0$  (zero magnetic field),  $e' \ll -1$  (ferromagnetism and low temperatures), and the boundary condition for the Gibbs measures in finite subsets  $\Lambda(n)$  is the free boundary condition. By symmetry, the expectation value of  $M_n := \frac{1}{|\Lambda(n)|} \sum_{i \in \Lambda(n)} \sigma_i$  with respect to these Gibbs measures is 0. This implies that the law of large numbers for the magnetization  $M_n$  is violated at the thermodynamical limit when the equilibrium state  $\mu$  is the limit of these Gibbs measures. There is breakdown of the symmetry of the model, since the measure  $\mu$

has a nontrivial ergodic decomposition into equilibrium states such that  $\mathbb{E}(\sigma_i) \neq 0$ . (2.13) and the symmetry of the model imply that  $s_{\mathbf{x}'}(e, m) = 0$  when  $|m| \leq \delta$ <sup>34</sup>, for those values of  $\mathbf{x}' = (e', 0)$  for which (2.13) is valid.

Suppose that a system is described by  $k$  macroscopic observables and that the macroscopic space  $\mathcal{R}$  is the subset  $\{\mathbf{x} \in \mathbb{R}^k : s(\mathbf{x}) > -\infty\}$ . We also assume that  $p(\mathbf{x}')$  is finite on  $\mathbb{R}^k$  and strictly convex. Let  $\mathcal{A}$  be an affine portion in the graph of the entropy  $s$ , and  $\mathbf{x}'$  be the unique point of  $\mathbb{R}^k$  such that  $\mathbf{x}' \leftrightarrow \mathbf{x}$  for all  $(\mathbf{x}, s(\mathbf{x})) \in \mathcal{A}$ . Set (see equation (2.12))

$$\Delta_{\mathbf{x}'} := \{\mathbf{x} \in \mathcal{R} : \mathbf{x}' \leftrightarrow \mathbf{x}\} = \{\mathbf{x} \in \mathcal{R} : s_{\mathbf{x}'}(\mathbf{x}) = 0\}.$$

If  $\mu \in N^{\mathbf{x}'}$  is an extremal element of the simplex  $N^{\mathbf{x}'}$ , then  $\mathbf{x} = \langle \mathbf{f}, \mu \rangle$  is an extremal element of the convex set  $\Delta_{\mathbf{x}'}$ . If we further assume that for each  $\mathbf{x} \in \Delta_{\mathbf{x}'}$  the set  $N_{\mathbf{x}}$  consists of a single probability measure  $\mu_{\mathbf{x}}$ , which means that  $\mu_{\mathbf{x}}$  is the unique  $\mu \in N^{\mathbf{x}'}$  with  $\langle \mathbf{f}, \mu \rangle = \mathbf{x}$ , then  $\mathbf{x}$  is an extremal element of  $\Delta_{\mathbf{x}'}$  if and only if  $\mu_{\mathbf{x}}$  is an extremal element of  $N^{\mathbf{x}'}$ . Similarly, one shows that any  $\mathbf{x} \in \Delta_{\mathbf{x}'}$  has a unique extremal decomposition. Hence  $\Delta_{\mathbf{x}'}$  is a simplex. When  $\mathcal{A}$ , or  $\Delta_{\mathbf{x}'}$ , is a  $m$ -dimensional simplex, there is coexistence of  $m + 1$  pure phases. The macroscopic states of the pure phases are the  $\mathbf{x}_i$ ,  $i = 1, \dots, m + 1$ , such that  $(\mathbf{x}_i, s(\mathbf{x}_i))$  are the extreme points of the simplex  $\mathcal{A}$ . The **Gibbs phase rule** is the statement that the family of  $m$ -dimensional simplices  $\Delta_{\mathbf{x}'}$  are parametrized by  $(k - m)$ -dimensional manifolds in the phase diagram of the system.

---

<sup>34</sup>To get this last result it is sufficient that in (2.13)  $\varepsilon = \varepsilon_n \equiv e^{-a_n}$ , with  $a_n \geq 0$  and  $\lim_{n \rightarrow \infty} a_n / |\Lambda(n)| = 0$ .

## 3. ASYMPTOTICALLY DECOUPLED MEASURES

The basic mathematical object in section 2 is  $s$ , which is identified with the thermodynamical entropy (per unit volume) by Boltzmann's law.  $s$  is a functional of the reference measure  $\lambda$  and of the macroscopic observables  $\mathbf{F}_n$ . The physical idea behind the existence of  $s$  is that the thermodynamical entropy  $S_\Lambda$  of the system confined in  $\Lambda$  is an extensive quantity. If  $\Lambda \subset \mathbb{Z}^d$  is a very large system and the system is decomposed into large subsystems confined in  $\Lambda_i$ , say

$$\Lambda = \bigcup_{i=1}^N \Lambda_i \quad , \quad i = 1, \dots, N,$$

then

$$S_\Lambda \simeq \sum_{i=1}^N S_{\Lambda_i}.$$

Taking the thermodynamical limit (infinite volume limit) in order to discard boundary effects,  $s$  can be defined by

$$s := \lim_{|\Lambda| \rightarrow \infty} \frac{S_\Lambda}{|\Lambda|}.$$

The main purpose of the lectures is to extend the whole formalism of section 2 to a large class of reference measures, called asymptotically decoupled measures, but keeping the same class of macroscopic observables. The principal results are established using arguments based directly on the above simple picture.

**Notation, basic setting.** The following setting is valid for the rest of the paper. For each  $x \in \mathbb{Z}^d$  let  $(\Omega_x, \mathcal{F}_x)$  be a copy of a given fixed standard Borel space. If  $x = (x_1, \dots, x_d) \in \mathbb{Z}^d$ , then  $|x| := \max_k |x_k|$ .  $|\Lambda|$  is the cardinality of a subset  $\Lambda \subset \mathbb{Z}^d$ ; the complement of  $\Lambda$  in  $\mathbb{Z}^d$  is written  $\Lambda^c := \mathbb{Z}^d \setminus \Lambda$ .  $(\Omega_\Lambda, \mathcal{F}_\Lambda)$  denotes the product space  $(\prod_{x \in \Lambda} \Omega_x, \prod_{x \in \Lambda} \mathcal{F}_x)$  and  $(\Omega, \mathcal{F})$  stands for the product space  $(\Omega_{\mathbb{Z}^d}, \mathcal{F}_{\mathbb{Z}^d})$ . An element of  $\Omega$  is denoted by  $\omega$  and an element of  $\Omega_\Lambda$  by  $\omega_\Lambda$ .  $\mathcal{F}_\Lambda$  also denotes the sub- $\sigma$ -algebra of  $\mathcal{F}$  generated by the projection  $p_\Lambda : \Omega \rightarrow \Omega_\Lambda$ ,  $\omega \mapsto p_\Lambda(\omega) := \omega_\Lambda$ ;  $f \in \mathcal{F}_\Lambda$  means that the real-valued function  $f$  is  $\mathcal{F}_\Lambda$ -measurable. Let  $x \in \mathbb{Z}^d$ ; the translation operator  $\theta_x$  acts on  $\mathbb{Z}^d$  by  $y \mapsto y + x$ . The action is lifted to  $\Omega$  by  $(\theta_x \omega)_y = \omega_{y+x}$ . The action of  $\theta_x$  on a function  $f$  is  $\theta_x f(\omega) = f(\theta_x \omega)$  and on a measure  $\nu$  it is defined by  $\int f d(\theta_x \nu) = \int (\theta_x f) d\nu$ . If  $B$  is a subset of a topological space, then  $\text{int } B$  is the interior of  $B$  and  $\text{cl } B$  is the closure.

**Definition 3.1.** A real valued function  $f$  on  $\Omega$  is called *local* if  $f$  is  $\mathcal{F}_\Lambda$ -measurable for some finite  $\Lambda$ . The symbol  $\mathcal{F}_{\text{loc}}$  denotes the space of all bounded local functions; the closure of  $\mathcal{F}_{\text{loc}}$  with respect to the sup-norm  $\|\cdot\|$  is denoted  $\mathcal{F}_{\text{qloc}}$ ;  $f \in \mathcal{F}_{\text{qloc}}$  is called *quasilocal*. The dual of the Banach space  $(\mathcal{F}_{\text{qloc}}, \|\cdot\|)$  is  $\mathcal{F}_{\text{qloc}}^*$ .

$\mathcal{M}$  is the vector space of all finite signed measures on  $(\Omega, \mathcal{F})$ ,  $\mathcal{M}_1^+$  the convex subset of probability measures and  $\mathcal{M}_1^{+, \theta}$  the convex subset of translation invariant probability measures. For any finite subset  $\Lambda \subset \mathbb{Z}^d$  there is an **average operator** acting on functions or measures,

$$\mathcal{A}_\Lambda := \frac{1}{|\Lambda|} \sum_{x \in \Lambda} \theta_x.$$

The topology on  $\mathcal{M}$  is the topology of the **bounded local convergence**<sup>35</sup> ( $\sigma(\mathcal{M}, \mathcal{F}_{\text{qloc}})$ -topology), that is, a sequence  $\{\nu_n\}$  of measures converges to the measure  $\nu$  if and only if

$$\lim_{n \rightarrow \infty} \int f d\nu_n = \int f d\nu \quad \forall f \in \mathcal{F}_{\text{qloc}}.$$

If  $\mathcal{F}_{\text{qloc}}$  is equipped with the  $\sigma(\mathcal{F}_{\text{qloc}}, \mathcal{M})$ -topology, then  $\mathcal{F}_{\text{qloc}}$  and  $\mathcal{M}$  form a dual pair of locally convex Hausdorff real topological vector spaces with pairing

$$\langle f, \nu \rangle = \int_{\Omega} f(\omega) \nu(d\omega).$$

The topology on  $\mathcal{M}_1^+$  or on  $\mathcal{M}_1^{+, \theta}$  is the induced topology. The topology on  $\mathcal{M}_1^+$  coincides with the  $\sigma(\mathcal{M}_1^+, \mathcal{F}_{\text{loc}})$ -topology, although the  $\sigma(\mathcal{M}, \mathcal{F}_{\text{loc}})$ -topology differs from the  $\sigma(\mathcal{M}, \mathcal{F}_{\text{qloc}})$ -topology on  $\mathcal{M}$ . Hence, a sequence  $\{\nu_n\}$  of probability measures converges to the probability measure  $\nu$  if and only if

$$\lim_{n \rightarrow \infty} \int f d\nu_n = \int f d\nu \quad \forall f \in \mathcal{F}_{\text{loc}}.$$

**3.1. Asymptotically decoupled measures, specific information gain.** The asymptotically decoupled measures are defined, and examples are given. Then a basic probability estimate is proved (lemma 3.1); lemmas 3.1 and 3.4 are the two main steps for proving a large deviations principle for these measures (theorem 3.3). An application of the method of proof of lemma 3.1 gives the existence of the specific information gain  $h(\nu|\mu)$  (proposition 3.2). Finally a variational formula for  $h(\nu|\mu)$  is proved in proposition 3.3.

Let  $n \in \mathbb{N}$  and

$$\Lambda(n) := \{x \in \mathbb{Z}^d : |x| \leq n\}.$$

If  $a \in \mathbb{Z}^d$ , then  $\Lambda(n) + a := \{y \in \mathbb{Z}^d : y = x + a, x \in \Lambda(n)\}$ . A scale is a diverging increasing sequence of positive numbers  $V_n$ . The volume scale corresponds to  $V_n := |\Lambda(n)|$ .

**Definition 3.2.** A positive measure  $\lambda$  on  $(\Omega, \mathcal{F})$  is asymptotically decoupled on the volume scale with parameters  $g$  and  $c$  if **a** and **b** hold.

**a.** There exist  $g : \mathbb{N} \mapsto \mathbb{N}$  and  $c : \mathbb{N} \mapsto [0, \infty)$ , such that

$$\lim_{n \rightarrow \infty} \frac{g(n)}{n} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{c(n)}{|\Lambda(n)|} = 0.$$

**b.**  $\forall a \in \mathbb{Z}^d, \forall n \in \mathbb{N}, \forall A \in \mathcal{F}_{\Lambda(n)+a}$ , and  $\forall B \in \mathcal{F}_{(\Lambda(n+g(n))+a)^c}$ ,

$$e^{-c(n)} \lambda[A] \lambda[B] \leq \lambda[A \cdot B] \tag{3.1}$$

$$\lambda[A \cdot B] \leq e^{c(n)} \lambda[A] \lambda[B]. \tag{3.2}$$

$\lambda$  is asymptotically decoupled on the volume scale from below if (3.1) holds only; it is asymptotically decoupled on the volume scale from above if (3.2) holds only.

**Comments.** 1.  $\lambda$  has the property of **weak dependence** defined in [LPS3] if and only if  $g(n) \equiv 0$ .

2. It is interesting, for example in ergodic theory, to consider specifically the case  $(\Omega_{\mathbb{N}}, \mathcal{F}_{\mathbb{N}}) = (\prod_{x \in \mathbb{N}} \Omega_x, \prod_{x \in \mathbb{N}} \mathcal{F}_x)$  and to modify the definition as follows (see [PS]).

<sup>35</sup> This convergence is called sometimes the  $\tau$ -convergence.

**Definition 3.3.** A positive measure  $\lambda$  on  $(\Omega_{\mathbb{N}}, \mathcal{F}_{\mathbb{N}})$  is asymptotically decoupled with parameters  $g$  and  $c$  if **a** and **b'** hold.

a. There exist  $g : \mathbb{N} \mapsto \mathbb{N}$  and  $c : \mathbb{N} \mapsto [0, \infty)$ , such that

$$\lim_{n \rightarrow \infty} \frac{g(n)}{n} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{c(n)}{n} = 0.$$

b'. Let  $I = [p, q]$ ,  $p \leq q \in \mathbb{N}$ ,  $n := q - p + 1$  and  $\mathcal{F}_{(q+g(n))^c} := \prod_{x > q+g(n)} \mathcal{F}_x$ . Then

$$e^{-c(n)} \lambda[A] \lambda[B] \leq \lambda[A \cdot B] \leq e^{c(n)} \lambda[A] \lambda[B].$$

for any  $A \in \mathcal{F}_I$  and any  $B \in \mathcal{F}_{(q+g(n))^c}$ .

3. Suppose that  $\lambda$  is asymptotically decoupled. Let  $A \in \mathcal{F}_{\Lambda(n)+a}$ ,  $B \in \mathcal{F}_{(\Lambda(n+g(n))+a)^c}$ . For each  $n \in \mathbb{N}$  and  $a \in \mathbb{Z}^d$  there exists a proper regular conditional probability kernel<sup>36</sup>  $\mathbb{K}_{\Lambda(n)+a}$ ,

$$\mathcal{F} \times \Omega \ni (F, \omega) \mapsto \mathbb{K}_{\Lambda(n)+a}^{\omega}(F) = \mathbb{E}_{\lambda}(F | \mathcal{F}_{(\Lambda(n+g(n))+a)^c})(\omega) \quad \lambda - \text{a.s.}$$

Then, for all  $B$ ,

$$e^{-c(n)} \lambda(A) \lambda(B) \leq \int_{\Omega} \mathbb{K}_{\Lambda(n+g(n))+a}^{\omega}(I_A) I_B(\omega) \lambda(d\omega) \leq e^{c(n)} \lambda(A) \lambda(B).$$

Therefore

$$e^{-c(n)} \lambda(A) \leq \mathbb{K}_{\Lambda(n+g(n))+a}^{\omega}(I_A) \leq e^{c(n)} \lambda(A) \quad \lambda - \text{a.s.} \quad (3.3)$$

Conversely, if (3.3) holds, then  $\lambda$  is asymptotically decoupled. It is useful to have a stronger property, namely (3.3) without restriction on  $\omega$ . This stronger version of an asymptotically decoupled measure is used in subsection 4.3.

**Definition 3.4.**  $\lambda \in \mathcal{M}_1^+$  is strongly asymptotically decoupled if it is asymptotically decoupled and (3.3) holds  $\forall a \in \mathbb{Z}^d$ ,  $\forall n \in \mathbb{N}$ ,  $\forall A \in \mathcal{F}_{\Lambda(n)+a}$  and  $\forall \omega \in \Omega$ .

4. Since only the volume scale is considered below, the words ‘‘on the volume scale’’ are skipped.

5. When  $\Omega_x$  is a finite set or a compact metric space, asymptotically decoupled measures form a relatively large class of interesting measures. There is a natural generalization of the notion of asymptotically decoupled probability measure, which is the following one.

**Definition 3.5.** A positive measure  $\lambda$  on  $(\Omega, \mathcal{F})$  is weakly asymptotically decoupled if for all  $\varepsilon > 0$  and  $\delta$ ,  $0 < \delta < 1$ , then **a** and **b** hold.

a. There exist  $g : \mathbb{N} \mapsto \mathbb{N}$  and  $c : \mathbb{N} \mapsto [0, \infty)$ , such that

$$\lim_{n \rightarrow \infty} \frac{|\Lambda(n)|}{|\Lambda(n+g(n))|} > 1 - \delta \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{c(n)}{|\Lambda(n)|} < \varepsilon.$$

b.  $\forall a \in \mathbb{Z}^d$ ,  $\forall n \in \mathbb{N}$ ,  $\forall A \in \mathcal{F}_{\Lambda(n)+a}$ , and  $\forall B \in \mathcal{F}_{(\Lambda(n+g(n))+a)^c}$ ,

$$e^{-c(n)} \lambda[A] \lambda[B] \leq \lambda[A \cdot B] \leq e^{c(n)} \lambda[A] \lambda[B].$$

<sup>36</sup>See subsection 3.2.

Propositions 3.1 and 3.2 are still true (essentially the same proofs) with this new definition. Lemma 3.1 also holds (with minor modifications). Hence theorem 3.3, which establishes a large deviations principle for the empirical measures, is still valid. However, the proof given below of proposition 3.3, which allows to identify the rate-function of the large deviations principle, does not hold.

**Examples.** 1. Product measures are asymptotically decoupled.

2. A Gibbs measure, solution of a local specification defined by an absolutely summable bounded potential, has the property of weak dependence (see [LPS3]), hence is asymptotically decoupled. This result is generalized in lemma 4.9.

3. Suppose that  $(\Omega_x, \mathcal{F}_x) \equiv (X, \mathcal{X})$  and let  $\pi$  be a Markov kernel defined on  $X \times \mathcal{X}$ , such that there exist a positive measure  $\nu$  on  $(X, \mathcal{X})$  and constants  $0 < a \leq b < \infty$ , so that

$$a \nu[A] \leq \pi(\omega, A) \leq b \nu[A] \quad \forall \omega \in X, \forall A \in \mathcal{X}.$$

A Markov chain on  $(\Omega_{\mathbb{N}}, \mathcal{F}_{\mathbb{N}})$  with transition kernel  $\pi$  has the property of weak dependence, hence it is asymptotically decoupled. This type of Markov chains is considered in [MN]; the Markov chains in [DV1] or in [S] are particular cases<sup>37</sup>. Indeed, one has

$$\frac{a}{b} \pi(\omega, A) \leq a \nu[A] \leq \pi(\omega', A) \quad \forall \omega, \omega' \in X, \forall A \in \mathcal{X}.$$

If  $\mu$  is the initial distribution of the chain,  $A \in \mathcal{F}_{[p,q]}$ ,  $B \in \mathcal{F}_{(q+1)^c}$ , then for any  $\eta_q$ ,

$$\begin{aligned} \nu[A \cdot B] &= \int \mu(d\omega_1) \pi(\omega_1, d\omega_2) \cdots \pi(\omega_{q-1}, d\omega_q) 1_A(\omega_p, \dots, \omega_q) \cdot \\ &\quad \int \pi(\omega_q, d\omega_{q+1}) \pi(\omega_{q+1}, d\omega_{q+2}) \cdots 1_B(\omega_{q+1}, \dots) \\ &\leq \frac{b}{a} \nu[A] \int \pi(\eta_q, d\omega_{q+1}) \pi(\omega_{q+1}, d\omega_{q+2}) \cdots 1_B(\omega_{q+1}, \dots). \end{aligned}$$

Integrating this inequality with respect to  $\int \mu(d\omega_1) \pi(\omega_1, d\omega_2) \cdots \pi(\omega_{q-1}, d\eta_q)$  gives the upper bound in  $\mathbf{b}'$  with  $g(n) \equiv 0$  and  $c(n) \equiv \text{const}$ .

4. Let  $\Omega = X^{\mathbb{Z}}$  and  $X$  be a finite set. Let  $\lambda$  be a stationary Markov chain with transition matrix  $M$ . Let  $[\omega_{i_1} \omega_{i_2} \cdots \omega_{i_k}]$  be the cylinder  $\{\omega' : \omega'_j = \omega_j, \forall j = i_1, \dots, i_k\}$ . Then

$$\lambda([\omega_1 \omega_2]) = \lambda([\omega_1]) M(\omega_1, \omega_2).$$

Suppose that  $\lambda$  is asymptotically decoupled. Let  $m := g(1) + 1$ . Then  $M^m$  has strictly positive entries, so the Markov chain is irreducible and aperiodic. Conversely, let  $\lambda$  be a stationary chain whose transition matrix  $M$  has the property that there exists an integer  $m > 0$  so that  $M^m$  has strictly positive entries. Then with

$$g(n) := m - 1 \quad \text{and} \quad c(n) := \sup_{\omega_1, \omega_2 \in S} \left| \ln \frac{M^m(\omega_1, \omega_2)}{\lambda([\omega_2])} \right|,$$

one can show that  $\lambda$  is asymptotically decoupled with these (constant) parameters.

5. Let  $\nu$  be a translation invariant, asymptotically decoupled probability measure on  $(\Omega, \mathcal{F})$  with  $\Omega = X^{\mathbb{Z}}$  and  $X$  a finite set (alphabet). For each  $n \in \mathbb{N}$ , let

$$\Omega_{\nu, n} := \{\eta_1 \eta_2 \cdots \eta_n \in \Omega_{[1, n]} : \nu[\{\omega : p_{[1, n]}(\omega) = \eta_1 \eta_2 \cdots \eta_n\}] > 0\}.$$

<sup>37</sup> The Markov chains in [DV2] are not asymptotically decoupled measures.

$\Omega_{\nu,n}$  is the set of all admissible words of length  $n$  of the shift space<sup>38</sup>

$$\Omega_{\nu} := \{ \omega \in \Omega : \forall \theta_x, \forall n \in \mathbb{N}, p_{[1,n]}(\theta_x \omega) \in \Omega_{\nu,n} \}.$$

$\Omega_{\nu}$  is a closed, translation invariant subset of  $\Omega$ ;  $\nu$  is naturally defined on  $\Omega_{\nu}$ ,  $\nu[\Omega_{\nu}] = 1$ . Since  $\nu$  is asymptotically decoupled the language of the shift  $\Omega_{\nu}$  has the following mixing property: if  $\eta_1 \eta_2 \cdots \eta_n$ ,  $\eta'_1 \eta'_2 \cdots \eta'_{m_1}$  and  $\eta''_1 \eta''_2 \cdots \eta''_{m_2}$  are three words of the language, with  $m_1$  and  $m_2$  arbitrary, then there exist words of the language  $\xi_1 \xi_2 \cdots \xi_k$  and  $\xi'_1 \xi'_2 \cdots \xi'_{k'}$  of lengths smaller than  $g(n)$ , such that

$$\eta''_1 \eta''_2 \cdots \eta''_{m_2} \xi'_1 \xi'_2 \cdots \xi'_{k'} \eta_1 \eta_2 \cdots \eta_n \xi_1 \xi_2 \cdots \xi_k \eta'_1 \eta'_2 \cdots \eta'_{m_1}$$

is an admissible word of length  $m_2 + k' + m + k + m_1$  of the shift.

6. In all examples above the measures are also strongly asymptotically decoupled.

**A basic probability estimate.** Lemma 3.1 gives a basic estimate for a nonnegative, translation invariant measure, which is asymptotically decoupled from below. The following objects are given.

- A positive, translation invariant measure  $\mu$  on  $(\Omega, \mathcal{F})$ , which is *asymptotically decoupled from below* with parameters  $g$  and  $c$ .
- A measurable space  $(E, \mathcal{B})$ , where  $E$  is a real locally convex topological vector space and  $\mathcal{B}$  its Borel  $\sigma$ -algebra.
- A nonnegative convex function  $\phi : E \rightarrow \mathbb{R}$  such that  $\inf\{\phi(x) : x \in E\} = 0$ .
- A measurable map  $Y : \Omega \rightarrow E$ .

For each  $\Lambda \subset \mathbb{Z}^d$ ,  $|\Lambda| < \infty$ , set

$$\widehat{T}_{\Lambda} := \mathcal{A}_{\Lambda} Y.$$

Lemma 3.1 gives a lower bound for the  $\mu$ -measure of the set

$$\mathcal{E}_{\Lambda, \phi}(a) := \{ \omega \in \Omega : \phi \circ \widehat{T}_{\Lambda}(\omega) < a \}.$$

**Lemma 3.1.** *The setting is as above; suppose that  $\exists r \in \mathbb{N}$  such that  $Y$  is  $\mathcal{F}_{\Lambda(r)}$ -measurable and*

$$b := \sup\{ \phi \circ \widehat{T}_{\Lambda}(\omega) : \omega \in \Omega \text{ and } \Lambda \subset \mathbb{Z}^d \} < \infty.$$

*Then  $\forall a, a', 0 < a' < a$ , and  $0 < \delta < 1$  there exist  $M(\delta, a', r)$  and  $N(m, \delta, a', r)$  such that for  $m \geq M$  and  $n \geq N$*

$$\frac{1}{|\Lambda(n)|} \ln \mu[\mathcal{E}_{\Lambda(n), \phi}(a)] \geq (1 - \delta) \frac{1}{|\Lambda(m)|} \ln \mu[\mathcal{E}_{\Lambda(m), \phi}(a')] - \frac{c(m)}{|\Lambda(m)|}.$$

**Proof.** Set

$$r' := \lceil g(m+r)/2 \rceil.$$

Let  $n > m + r + r'$ ; there exists a unique maximal  $k \in \mathbb{N}$  such that

$$2n + 1 = k[2(m + r + r') + 1] + j \quad 0 \leq j < 2(m + r + r') + 1.$$

One partitions the set  $\Lambda(n)$  in two steps. First one partitions it into  $k^d + 1$  disjoint subsets,  $k^d$  of them, denoted  $\Lambda'_q$ ,  $q = 1, \dots, k^d$ , are translates of  $\Lambda(m + r + r')$ , and the last one is

$$\Lambda'_{k^d+1} := \{ t \in \Lambda(n) : n - t_i < j, i = 1, \dots, d \}.$$

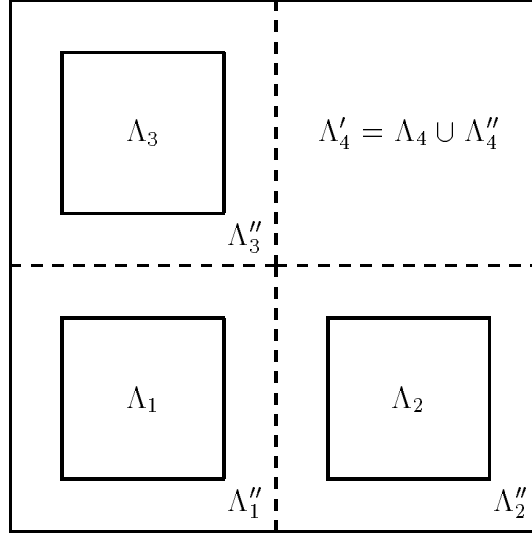
<sup>38</sup>See [LM] for an introduction to shift spaces.



Then each  $\Lambda'_q$  is further partitioned in two sets, one is a translate of  $\Lambda(m)$ , denoted by  $\Lambda_q$ , and the second one is a translate of  $\Lambda(m+r+r') \setminus \Lambda(m)$ , denoted by  $\Lambda''_q$ . Set

$$\Lambda_{k^d+1} := \left( \bigcup_{q=1}^{k^d} \Lambda''_q \right) \cup \Lambda'_{k^d+1}.$$

The final partition of  $\Lambda(n)$  is that given by  $\Lambda_q$ ,  $q = 1, \dots, k^d + 1$ .



Since  $\phi$  is convex

$$\phi \circ \widehat{T}_{\Lambda(n)} \leq \sum_{q=1}^{k^d} \frac{|\Lambda(m)|}{|\Lambda(n)|} \phi \circ \widehat{T}_{\Lambda_q} + \frac{|\Lambda_{k^d+1}|}{|\Lambda(n)|} \phi \circ \widehat{T}_{\Lambda_{k^d+1}} \leq \sum_{q=1}^{k^d} \frac{1}{k^d} \phi \circ \widehat{T}_{\Lambda_q} + \frac{|\Lambda_{k^d+1}|}{|\Lambda(n)|} \phi \circ \widehat{T}_{\Lambda_{k^d+1}}.$$

Since  $2n+1 \leq (k+1)(2(m+r+r')+1)$ ,

$$\begin{aligned} |\Lambda_{k^d+1}| &\leq dk^d 2(r+r')(2(m+r+r')+1)^{d-1} + d2(m+r+r')(2n+1)^{d-1} \\ &\leq dk^d |\Lambda(m)| \left( \frac{r+r'}{m+r+r'} + \frac{1}{k} \left(1 + \frac{1}{k}\right)^{d-1} \right). \end{aligned} \quad (3.4)$$

There exist  $M = M(\delta, a', r)$  and  $K = K(\delta, a', r)$  so that  $\forall m \geq M$  and  $\forall k \geq K$ ,

$$\frac{|\Lambda_{k^d+1}|}{|\Lambda(n)|} \leq \min \left\{ \delta, \frac{a-a'}{b} \right\}.$$

In particular,  $\forall m \geq M$  and  $\forall k \geq K$

$$1 \geq \frac{k^d |\Lambda(m)|}{|\Lambda(n)|} \geq 1 - \delta, \quad (3.5)$$

$$\frac{|\Lambda_{k^d+1}|}{|\Lambda(n)|} \|\phi \circ \widehat{T}_{\Lambda_{k^d+1}}\| \leq a - a', \quad (3.6)$$

and

$$\mathcal{E}_{\Lambda(n), \phi}(a) \supset \bigcap_{q=1}^{k^d} \mathcal{E}_{\Lambda_q, \phi}(a'). \quad (3.7)$$

For each  $q \leq k^d$  there exists  $a_q \in \mathbb{Z}^d$  so that  $\Lambda_q = \Lambda(m) + a_q$ ; the set  $\mathcal{E}_{\Lambda_q, \phi}(a')$  is  $\mathcal{F}_{\Lambda(m+r)+a_q}$ -measurable and  $\mathcal{E}_{\Lambda_p, \phi}(a')$ ,  $p \neq q$ , is  $\mathcal{F}_{(\Lambda(m+r)+g(m+r))+a_q}^c$ -measurable. Therefore, using the translation invariance of  $\mu$ ,

$$\begin{aligned} \frac{1}{|\Lambda(n)|} \ln \mu[\mathcal{E}_{\Lambda(n), \phi}(a)] &\geq \frac{1}{|\Lambda(n)|} \ln \mu \left[ \bigcap_{q=1}^{k^d} \mathcal{E}_{\Lambda_q, \phi}(a') \right] \\ &\geq (1 - \delta) \frac{1}{|\Lambda(m)|} \ln \mu[\mathcal{E}_{\Lambda(m), \phi}(a')] - \frac{c(m+r)}{|\Lambda(m+r)|}. \end{aligned}$$

□

Two canonical functionals  $p(f|\lambda)$  and  $h(\nu|\lambda)$ . As an application of the method of proof of lemma 3.1, two natural functionals are defined for a translation invariant probability measure  $\lambda$ , which is asymptotically decoupled. First, a convex functional  $p(f|\lambda)$  on  $\mathcal{F}_{\text{qloc}}$ ; second, the information gain  $h(\nu|\lambda)$  of  $\nu \in \mathcal{M}_1^{+, \theta}$  with respect to  $\lambda$ . Finally one proves the variational formula for  $h(\nu|\lambda)$ : on  $\mathcal{M}_1^{+, \theta}$   $h(\cdot|\lambda)$  is the conjugate of  $p(\cdot|\lambda)$ .

**Proposition 3.1.** *Let  $\lambda \in \mathcal{M}_1^{+, \theta}$  be asymptotically decoupled either from below or from above. Let  $f \in \mathcal{F}_{\text{qloc}}$  and set*

$$p_n(f|\lambda) := \frac{1}{|\Lambda(n)|} \ln \int_{\Omega} \exp \left[ \sum_{x \in \Lambda(n)} \theta_x f(\omega) \right] \lambda(d\omega).$$

Then

$$p(f|\lambda) := \lim_{n \rightarrow \infty} p_n(f|\lambda)$$

exists and defines a convex functional on the Banach space  $(\mathcal{F}_{\text{qloc}}, \|\cdot\|)$ ;

$$|p(f|\lambda) - p(g|\lambda)| \leq \|f - g\|.$$

$f \mapsto p(f|\lambda)$  is l.s.c. in the  $\sigma(\mathcal{F}_{\text{qloc}}, \mathcal{F}_{\text{qloc}}^*)$ -topology.

**Proof.** Let  $\lambda$  be asymptotically decoupled from below. Assume that  $f$  is  $\mathcal{F}_{\Lambda(r)}$ -measurable. One proceeds as in the proof of lemma 3.1; one introduces the same partition of  $\Lambda(n)$ ,  $\Lambda_q$ ,  $q = 1, \dots, k^d + 1$ . Given  $\varepsilon > 0$  and  $0 < \delta < 1$ , there exist  $M(\delta, \varepsilon, r)$  and  $N(m, \delta, \varepsilon, r)$  so that for all  $m \geq M$  and  $n \geq N$  (see (3.5) and (3.6))

$$\frac{|\Lambda_{k^d+1}|}{|\Lambda(n)|} \|\mathcal{A}_{\Lambda_{k^d+1}} f\| \leq \varepsilon \quad \text{and} \quad 1 \geq \frac{k^d |\Lambda(m)|}{|\Lambda(n)|} \geq 1 - \delta.$$

Translation invariance implies

$$p_n(f|\lambda) \geq \frac{k^d |\Lambda(m)|}{|\Lambda(n)|} p_m(f|\lambda) - \varepsilon - \frac{c(m+r)}{|\Lambda(m+r)|}. \quad (3.8)$$

From this follows

$$\liminf_{n \rightarrow \infty} p_n(f|\lambda) \geq \limsup_{m \rightarrow \infty} p_m(f|\lambda).$$

If  $\lambda$  is asymptotically decoupled from above, then, instead of (3.8), one has

$$p_n(f|\lambda) \leq \frac{k^d |\Lambda(m)|}{|\Lambda(n)|} p_m(f|\lambda) + \varepsilon + \frac{c(m+r)}{|\Lambda(m+r)|}.$$

Convexity of the functional follows from Hölder inequality,

$$\int_{\Omega} \exp \left[ \sum_{j \in \Lambda(n)} \theta_j (\alpha f(\omega) + (1 - \alpha)g(\omega)) \right] \lambda(d\omega) \leq \left( \int_{\Omega} \exp \left[ \sum_{j \in \Lambda(n)} \theta_j f(\omega) \right] \lambda(d\omega) \right)^{\alpha} \left( \int_{\Omega} \exp \left[ \sum_{j \in \Lambda(n)} \theta_j g(\omega) \right] \lambda(d\omega) \right)^{1-\alpha}.$$

Let  $0 \leq t \leq 1$ ; then

$$|p_n(f|\lambda) - p_n(g|\lambda)| = \left| \int_0^1 dt \frac{d}{dt} p_n(g + t(f - g)|\lambda) \right| \leq \|f - g\|.$$

Existence of  $p(f|\lambda)$  when  $f \in \mathcal{F}_{\text{qloc}}$  follows by continuity. Suppose that  $\lim_{k \rightarrow \infty} f_k = f$  in the  $\sigma(\mathcal{F}_{\text{qloc}}, \mathcal{F}_{\text{qloc}}^*)$ -topology, that is,  $\forall \mu \in \mathcal{F}_{\text{qloc}}^*$ ,  $\lim_{k \rightarrow \infty} \langle f_k, \mu \rangle = \langle f, \mu \rangle$ . Jensen's inequality implies

$$\begin{aligned} p_m(f_k|\lambda) &= \frac{1}{|\Lambda(m)|} \ln \int_{\Omega} \exp \left[ \sum_{x \in \Lambda(m)} \theta_x f(\omega) \right] \exp \left[ \sum_{x \in \Lambda(m)} \theta_x (f_k(\omega) - f(\omega)) \right] \lambda(d\omega) \\ &\geq p_m(f|\lambda) + \frac{1}{|\Lambda(m)|} \int_{\Omega} \left( \sum_{x \in \Lambda(m)} \theta_x (f_k(\omega) - f(\omega)) \right) \lambda_m^f(d\omega) \\ &= p_m(f|\lambda) + \langle f_k - f, \mu_m^f \rangle, \end{aligned}$$

where  $\lambda_m^f$  is the probability measure

$$\lambda_m^f(d\omega) := \exp \left[ \sum_{x \in \Lambda(m)} \theta_x f(\omega) - |\Lambda(m)| p_m(f|\lambda) \right] \lambda(d\omega)$$

and  $\mu_m^f$  the probability measure

$$\mu_m^f := \mathcal{A}_{\Lambda(m)} \lambda_m^f.$$

Since the unit ball in  $\mathcal{F}_{\text{qloc}}^*$  is compact in the  $\sigma(\mathcal{F}_{\text{qloc}}^*, \mathcal{F}_{\text{qloc}})$ -topology, there exists a convergent subsequence  $\mu_{m_i}^f$ ,  $\lim_{i \rightarrow \infty} \mu_{m_i}^f = \nu \in \mathcal{F}_{\text{qloc}}^*$ . Therefore

$$p(f_k|\lambda) \geq p(f|\lambda) + \langle f_k - f, \nu \rangle$$

and

$$\liminf_{k \rightarrow \infty} p(f_k|\lambda) \geq p(f|\lambda).$$

□

**Information gain.** Let  $\nu \in \mathcal{M}_1^+$  and  $\lambda \in \mathcal{M}_1^+$ . The information gain  $\mathcal{H}(\nu|\lambda)$  of  $\nu$  with respect to  $\lambda$  is

$$\begin{aligned} \mathcal{H}(\nu|\lambda) &:= \begin{cases} \int_{\Omega} \ln f(\omega) \nu(d\omega) & \text{if } \nu(d\omega) = f(\omega)\lambda(d\omega) \\ +\infty & \text{otherwise} \end{cases} \quad (3.9) \\ &= \sup_{g \in \mathcal{F}_b} \left( \int g d\nu - \ln \int e^g d\lambda \right). \end{aligned}$$

In (3.9)  $\mathcal{F}_b$  is the set of bounded,  $\mathcal{F}$ -measurable, real-valued functions and the convention  $0 \ln 0 := 0$  is used. If  $\mathcal{B}$  is a sub- $\sigma$ -algebra of  $\mathcal{F}$ , then  $\lambda|_{\mathcal{B}}$  is the restriction of  $\lambda$  to  $\mathcal{B}$ ; set

$$\mathcal{H}_{\mathcal{B}}(\nu|\lambda) := \mathcal{H}(\nu|_{\mathcal{B}}|\lambda|_{\mathcal{B}});$$

(3.9) holds with  $\mathcal{B}_b$ , the set of bounded,  $\mathcal{B}$ -measurable functions, instead of  $\mathcal{F}_b$ .  $\mathcal{H}(\nu|\lambda) \in [0, \infty]$  and  $\nu \mapsto \mathcal{H}(\nu|\lambda)$  is l.s.c. on  $\mathcal{M}_1^+$ .

**Definition 3.6.** Let  $\lambda \in \mathcal{M}_1^{+, \theta}$ . The probability measure  $\nu \in \mathcal{M}_1^+$  has specific information gain  $h(\nu|\lambda)$  with respect to  $\lambda$  if

$$h(\nu|\lambda) := \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \mathcal{H}_{\mathcal{F}_{\Lambda(n)}}(\nu|\lambda) \quad \text{exists.}$$

To simplify the notation set  $\mathcal{H}_{\mathcal{F}_{\Lambda(n)}}(\nu|\lambda) := \mathcal{H}_n(\nu|\lambda)$ .

**Proposition 3.2.** Let  $\lambda \in \mathcal{M}_1^{+, \theta}$  be asymptotically decoupled from above.

- a. For any  $\nu \in \mathcal{M}_1^{+, \theta}$  the specific information gain of  $\nu$  with respect to  $\lambda$  exists.  $h(\cdot|\lambda)$  is a l.s.c. affine function on  $\mathcal{M}_1^{+, \theta}$ .
- b. The level-sets of  $h(\cdot|\lambda)$  in  $\mathcal{M}_1^{+, \theta}$  are compact.

**Proof.** Proof of a. The basic idea is the same as for the proof of lemma 3.1. The same notations are used. In particular  $\Lambda(n)$  is decomposed into  $\Lambda_q$ ,  $q = 1, \dots, k^d + 1$  (with  $r = 0$ ). Since

$$\mathcal{H}_{\mathcal{B}_1}(\nu|\lambda) \leq \mathcal{H}_{\mathcal{B}_2}(\nu|\lambda) \quad \text{if } \mathcal{B}_1 \subset \mathcal{B}_2,$$

one has

$$\mathcal{H}_n(\nu|\lambda) \geq \mathcal{H}_{\mathcal{D}_n}(\nu|\lambda) \quad \text{where } \mathcal{D}_n := \mathcal{F}_{\bigcup_{q=1}^{k^d} \Lambda_q}.$$

Since  $\lambda$  is asymptotically decoupled from above, replacing  $\lambda|_{\mathcal{D}_n}$  by the product measure  $\beta_n^m := \otimes_{q=1}^{k^d} \lambda|_{\mathcal{F}_{\Lambda_q}}$ , one obtains<sup>39</sup>

$$\mathcal{H}_{\mathcal{D}_n}(\nu|\lambda) \geq \mathcal{H}_{\mathcal{D}_n}(\nu|\beta_n^m) - k^d c(m) \geq \sum_{q=1}^{k^d} \mathcal{H}_{\mathcal{F}_{\Lambda_q}}(\nu|\lambda) - k^d c(m).$$

By translation invariance

$$\mathcal{H}_{\mathcal{D}_n}(\nu|\beta_n^m) \geq k^d \mathcal{H}_m(\nu|\lambda).$$

Let  $0 < \delta < 1$ ; there exists  $M$  so that for  $m \geq M$  (see (3.5)),

$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \mathcal{H}_n(\nu|\lambda) \geq \frac{\mathcal{H}_m(\nu|\lambda)}{|\Lambda(m)|} (1 - \delta) - \frac{c(m)}{|\Lambda(m)|}. \quad (3.10)$$

Taking  $\limsup_m$  and using the fact that  $\delta$  is arbitrary one obtains the existence of the specific information gain.

<sup>39</sup> Let  $(\Omega_1, \mathcal{F}_1)$  and  $(\Omega_2, \mathcal{F}_2)$  be measurable spaces and  $\Omega = \Omega_1 \times \Omega_2$  with  $\mathcal{F}$  the corresponding product  $\sigma$ -algebra. Let  $\rho$  and  $\lambda$  be probability measures on  $(\Omega, \mathcal{F})$  with  $\rho_1, \rho_2$  and  $\lambda_1, \lambda_2$  denoting the restrictions to  $\mathcal{F}_1, \mathcal{F}_2$  considered as sub- $\sigma$ -algebras of  $\mathcal{F}$ . Assume that  $\lambda = \lambda_1 \otimes \lambda_2$ . Then

$$\mathcal{H}(\rho|\lambda) = \mathcal{H}(\rho|\rho_1 \otimes \rho_2) + \mathcal{H}(\rho_1|\lambda_1) + \mathcal{H}(\rho_2|\lambda_2).$$

Let  $\nu_k$  be a convergent sequence to  $\nu$ . Since  $\mathcal{H}_m(\nu|\lambda)$  is l.s.c.  $\forall \delta, 0 < \delta < 1$ , and  $m$  sufficiently large,

$$\begin{aligned} \liminf_{k \rightarrow \infty} h(\nu_k|\lambda) &\geq \liminf_{k \rightarrow \infty} \frac{\mathcal{H}_m(\nu_k|\lambda)}{|\Lambda(m)|} (1 - \delta) - \frac{c(m)}{|\Lambda(m)|} \\ &\geq \frac{\mathcal{H}_m(\nu|\lambda)}{|\Lambda(m)|} (1 - \delta) - \frac{c(m)}{|\Lambda(m)|}. \end{aligned}$$

Hence  $\liminf_{k \rightarrow \infty} h(\nu_k|\lambda) \geq h(\nu|\lambda)$ . The affine character of  $h(\nu|\lambda)$  is a consequence of

$$\sum_1^m \alpha_j \mathcal{H}(\rho_j|\lambda) + \sum_1^m \alpha_j \ln \alpha_j \leq \mathcal{H}\left(\sum_1^m \alpha_j \rho_j|\lambda\right) \leq \sum_1^m \alpha_j \mathcal{H}(\rho_j|\lambda), \quad (3.11)$$

where  $\rho_j \in \mathcal{M}_1^+$  and  $\alpha_j > 0$  so that  $\sum_1^m \alpha_i = 1$ .

Proof of b. Let  $c < \infty$  and consider

$$K := \{\mu \in \mathcal{M}_1^{+, \theta} : h(\mu|\lambda) \leq c\}.$$

Since  $h$  is l.s.c., the set  $K$  is closed. A net in  $K$ ,  $\{\mu_\alpha\}_{\alpha \in D}$ , is locally equicontinuous if for each finite  $\Lambda \subset \mathbb{Z}^d$ , each sequence  $\{A_m\} \subset \mathcal{F}_\Lambda$ ,  $A_m \downarrow \emptyset$  as  $m \uparrow \infty$ , then

$$\lim_{m \rightarrow \infty} \limsup_{\alpha \in D} \mu_\alpha[A_m] = 0.$$

Each net in  $K$ ,  $\{\mu_\alpha\}_{\alpha \in D}$ , which is locally equicontinuous, has a cluster point in  $\mathcal{M}_1^+$  (see below, lemma 3.2 b). It is sufficient to check the local equicontinuity for  $\Lambda(m)$  with  $m$  sufficiently large. By (3.10) there exists  $m_0$  so that for  $m \geq m_0$  and  $\mu \in K$ ,

$$\mathcal{H}_m(\mu|\lambda) \leq |\Lambda(m)| 2c. \quad (3.12)$$

So let  $m \geq m_0$  and  $\{A_k\}_{k \geq 1}$ ,  $A_k \in \mathcal{F}_m$  and  $A_k \downarrow \emptyset$ . (3.12) implies that each  $\mu \in K$  is absolutely continuous with respect to  $\lambda$  on  $\mathcal{F}_m$ . Let  $f_m^\mu$  denote the Radon-Nikodym derivative of  $\mu|_{\mathcal{F}_m}$  with respect to  $\lambda|_{\mathcal{F}_m}$ . Given  $\varepsilon > 0$ , let  $\delta > 0$  so that

$$\varepsilon \ln \frac{\varepsilon}{\delta} \geq 2c|\Lambda(m)| + 1.$$

Let  $k$  be so large that  $\lambda[A_k] \leq \delta$ . Then, because<sup>40</sup>

$$f(\ln f)_- \leq (1 - f)_+ \leq 1,$$

$$\begin{aligned} \mu[A_k] &= \int_{\{f_m^\mu \leq \frac{\varepsilon}{\delta}\}} I_{A_k} f_m^\mu d\lambda + \int_{\{f_m^\mu > \frac{\varepsilon}{\delta}\}} I_{A_k} f_m^\mu d\lambda \\ &\leq \varepsilon + \left(\ln \frac{\varepsilon}{\delta}\right)^{-1} \int_{\{f_m^\mu > \frac{\varepsilon}{\delta}\}} f_m^\mu \ln f_m^\mu d\lambda \\ &\leq \varepsilon + \left(\ln \frac{\varepsilon}{\delta}\right)^{-1} (\mathcal{H}_m(\mu|\lambda) + 1) \\ &\leq 2\varepsilon. \end{aligned} \quad (3.13)$$

□

<sup>40</sup> For  $x \geq 0$ ,  $x \mapsto \psi(x) := x \ln x - x + 1$  is nonnegative and strictly convex. Hence,  $\psi(x) \geq 0$  implies  $x(\ln x)_- \leq (1 - x)_+ \leq 1$ .

The following two lemmas are very useful, in particular lemma 3.2 which is interesting independently of the context of these lectures. Lemma 3.2 is proved in [G1]; the hypothesis that  $(\Omega_x, \mathcal{F}_x)$  is a standard Borel space is used here.

**Lemma 3.2.** *Let  $\mu \in \mathcal{M}_1^+$  and  $\{\nu_\alpha\}_{\alpha \in D} \subset \mathcal{M}_1^+$  be a net such that for any  $m \in \mathbb{N}$*

$$\limsup_{\alpha \in D} \mathcal{H}_m(\nu_\alpha | \mu) \leq K_m < \infty .$$

- a.  $\{\nu_\alpha\}_{\alpha \in D}$  is locally equicontinuous.  
b. Any net, which is locally equicontinuous, has a cluster point.*

**Proof.** Part a of the lemma is proved as above. Let  $\{A_k\}_{k \geq 1} \subset \mathcal{F}_m$ ,  $A_k \downarrow \emptyset$ . There exists  $\alpha_m$  such that for all  $\alpha \geq \alpha_m$  (3.12) holds with  $K_m + 1$  instead of  $|\Lambda(m)|2c$ , that is

$$\mathcal{H}_m(\nu_\alpha | \mu) \leq K_m + 1 .$$

Given  $\varepsilon > 0$ , let  $\delta > 0$  be such that  $\varepsilon \ln \frac{\varepsilon}{\delta} \geq K_m + 2$ ; if  $k$  is so large that  $\mu[A_k] \leq \delta$ , then (3.13) holds: for all  $\alpha \geq \alpha_m$ ,  $\nu_\alpha[A_k] \leq 2\varepsilon$ . Hence

$$\lim_{k \rightarrow \infty} \limsup_{\alpha \in D} \nu_\alpha[A_k] = 0 .$$

For part b, see proposition 4.9 and corollary 4.10 in [G1].  $\square$

**Lemma 3.3.** *Let  $\lambda \in \mathcal{M}_1^{+\theta}$  be asymptotically decoupled from above with parameters  $g$  and  $c$ . Then, for any  $\mu \in \mathcal{M}_1^+$ ,*

$$\mathcal{H}_m(\mathcal{A}_{\Lambda(n)}\mu | \lambda) \leq \frac{(2m + 1 + \lceil g(m) \rceil)^d}{|\Lambda(n)|} \mathcal{H}_{n+m}(\mu | \lambda) + 2c(m) .$$

**Proof.** The proof is similar to the proof of the existence of the specific information gain. Since  $\lambda$  is translation invariant, for any  $\nu \in \mathcal{M}_1^+$  and any  $a \in \mathbb{Z}^d$ ,

$$\mathcal{H}_m(\theta_a \nu | \lambda) = \mathcal{H}_{\mathcal{F}_{\Lambda(m)+a}}(\nu | \lambda) .$$

One partitions  $\Lambda(n)$  into  $(2m + 1 + \lceil g(m) \rceil)^d$  subsets, denoted by  $\Lambda'_r(n)$ , such that

$$|\Lambda'_r(n)| \leq \left\lceil \frac{2n + 1}{(2m + 1 + \lceil g(m) \rceil)} \right\rceil^d \equiv k^d, \quad \inf_{i=1, \dots, d} \inf_{\substack{x \neq y \\ x, y \in \Lambda'_r(n)}} |x_i - y_i| \geq (2m + 1 + \lceil g(m) \rceil) .$$

For each  $r$ , let

$$\mathcal{F}_{n,r} := \bigcup_{j \in \Lambda'_r(n)} \mathcal{F}_{\Lambda(m)+j} \subset \mathcal{F}_{n+m} .$$

Since  $\lambda$  is asymptotically decoupled from above (see proof of proposition 3.2),

$$\mathcal{H}_{n+m}(\mu | \lambda) \geq \mathcal{H}_{\mathcal{F}_{n,r}}(\mu | \lambda) \geq \sum_{j \in \Lambda'_r(n)} \mathcal{H}_{\mathcal{F}_{\Lambda(m)+j}}(\mu | \lambda) - k^d c(m) .$$

Using (3.11)

$$\begin{aligned} \mathcal{H}_m(\mathcal{A}_{\Lambda(n)}\mu | \lambda) &\leq \frac{1}{|\Lambda(n)|} \sum_{x \in \Lambda(n)} \mathcal{H}_m(\theta_x \mu | \lambda) = \frac{1}{|\Lambda(n)|} \sum_{x \in \Lambda(n)} \mathcal{H}_{\mathcal{F}_{\Lambda(m)+x}}(\mu | \lambda) \\ &\leq \frac{(2m + 1 + \lceil g(m) \rceil)^d}{|\Lambda(n)|} \mathcal{H}_{n+m}(\mu | \lambda) + 2c(m) . \end{aligned}$$

□

Dual functional  $p^*$ . The dual functional  $p^*$  is defined on the dual space of  $\mathcal{F}_{\text{qloc}}$ , which is  $\mathcal{M}$ , as the conjugate functional of  $p$ , that is

$$p^*(\nu|\lambda) := \sup_{f \in \mathcal{F}_{\text{qloc}}} (\langle f, \nu \rangle - p(f|\lambda)).$$

**Proposition 3.3.** *Let  $\lambda \in \mathcal{M}_1^{+, \theta}$  be asymptotically decoupled from above. If  $\nu \in \mathcal{M} \setminus \mathcal{M}_1^{+, \theta}$ , then  $p^*(\nu|\lambda) = \infty$ . If  $\nu \in \mathcal{M}_1^{+, \theta}$ , then  $p^*(\nu|\lambda) = h(\nu|\lambda)$ .*

**Proof.** Suppose that  $\nu$  is not translation invariant. Then there exists  $f \in \mathcal{F}_{\text{loc}}$  and  $x \in \mathbb{Z}^d$ , so that  $\langle (f - \theta_x f), \nu \rangle \geq \varepsilon > 0$ . Set  $g := f - \theta_x f$ . Then

$$\limsup_{n \rightarrow \infty} \sup_{\omega} |\mathcal{A}_{\Lambda(n)} g| = 0.$$

Therefore, for any  $c > 0$ ,  $p(cg|\lambda) = 0$ , and

$$\sup_{c > 0} (\langle cg, \nu \rangle - p(cg|\lambda)) = \infty.$$

Similarly, if  $\langle 1, \nu \rangle \neq 1$  or  $\langle f, \nu \rangle \not\geq 0$  for all  $f \geq 0$ , then  $p^*(\nu|\lambda) = \infty$ .

Let  $\nu \in \mathcal{M}_1^{+, \theta}$ . By continuity of  $f \mapsto \langle f, \nu \rangle$  and of  $f \mapsto p(f|\lambda)$  in the  $\|\cdot\|$ -topology,

$$p^*(\nu|\lambda) = \sup_{f \in \mathcal{F}_{\text{qloc}}} (\langle f, \nu \rangle - p(f|\lambda)) = \sup_{f \in \mathcal{F}_{\text{loc}}} (\langle f, \nu \rangle - p(f|\lambda)).$$

Let  $f \in \mathcal{F}_{\Lambda(r)}$  be bounded; since  $\nu$  is translation invariant, (3.9) implies

$$\frac{1}{|\Lambda(n)|} \left[ \left\langle \sum_{j \in \Lambda(n)} \theta_j f, \nu \right\rangle - \ln \int_{\Omega} \exp \left( \sum_{j \in \Lambda(n)} \theta_j f(\omega) \right) \lambda(d\omega) \right] \leq \frac{1}{|\Lambda(n)|} \mathcal{H}_{n+r}(\nu|\lambda),$$

because  $\sum_{j \in \Lambda(n)} \theta_j f$  is  $\mathcal{F}_{n+r}$ -measurable. Taking  $n \rightarrow \infty$  one obtains  $\forall f \in \mathcal{F}_{\text{loc}}$ ,

$$\langle f, \nu \rangle - p(f|\lambda) \leq h(\nu|\lambda).$$

This proves  $p^*(\nu|\lambda) \leq h(\nu|\lambda)$ . To prove  $p^*(\nu|\lambda) \geq h(\nu|\lambda)$  one uses the setting of the proof of lemma 3.1. Let  $f \in \mathcal{F}_{\text{loc}}$ ; one partitions  $\Lambda(n)$ , as in the proof of the lemma 3.1, into  $\Lambda'_q$ ,  $q = 1, \dots, k^d + 1$ , with  $r = 0$  and  $r' = \lceil g(m)/2 \rceil$ .

$$\sum_{j \in \Lambda(n)} \theta_j f = \sum_{q=1}^{k^d} \sum_{j \in \Lambda'_q} \theta_j f + \sum_{j \in \Lambda'_{k^d+1}} \theta_j f. \quad (3.14)$$

For fixed  $m$ , (see last term of (3.4)),

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \left\| \sum_{j \in \Lambda'_{k^d+1}} \theta_j f \right\| = 0.$$

For the sake of clarity of the next argument we neglect this term. For any  $q = 1, \dots, k^d$  there exists  $a_q$  such that  $\Lambda'_q = \Lambda(m+r') + a_q$ ; therefore the double sum in (3.14) can be written

$$\sum_{q=1}^{k^d} \sum_{j \in \Lambda'_q} \theta_j f = \sum_{j \in \Lambda(m+r')} \sum_{q=1}^{k^d} \theta_{j+a_q} f.$$

Hölder inequality and translation invariance imply

$$\int_{\Omega} \exp \left( \sum_{j \in \Lambda(m+r')} \sum_{q=1}^{k^d} \theta_{j+a_q} f(\omega) \right) \lambda(d\omega) \leq \int_{\Omega} \exp \left( |\Lambda(m+r')| \sum_{q=1}^{k^d} \theta_{a_q} f(\omega) \right) \lambda(d\omega).$$

Let  $f$  be  $\mathcal{F}_m$ -measurable. Since  $\lambda \in \mathcal{M}_1^+$  is asymptotically decoupled from above

$$\int_{\Omega} \exp \left( |\Lambda(m+r')| \sum_{q=1}^{k^d} \theta_{a_q} f(\omega) \right) \lambda(d\omega) \leq \left( \int_{\Omega} \exp \left( |\Lambda(m+r')| f(\omega) \right) \lambda(d\omega) \right)^{k^d} e^{k^d c(m)}.$$

Taking into account the neglected term,

$$\begin{aligned} p^*(\nu|\lambda) &\geq \langle f, \nu \rangle - \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \int_{\Omega} \exp \left( \sum_{j \in \Lambda(n)} \theta_j f(\omega) \right) \lambda(d\omega) \\ &\geq \frac{\left( \langle |\Lambda(m+r')| f, \nu \rangle - \ln \int_{\Omega} \exp \left( |\Lambda(m+r')| f(\omega) \right) \lambda(d\omega) \right)}{|\Lambda(m+r')|} - \frac{c(m)}{|\Lambda(m)|}. \end{aligned}$$

Taking the supremum over all  $f \in \mathcal{F}_{\text{loc}} \cap \mathcal{F}_m$  one obtains

$$p^*(\nu|\lambda) \geq \frac{1}{|\Lambda(m+r')|} \mathcal{H}_m(\nu|\lambda) - \frac{c(m)}{|\Lambda(m)|}.$$

The result follows by taking  $m \rightarrow \infty$ .  $\square$

**3.2. Local specification and specific information gain.** Let  $\lambda \in \mathcal{M}_1^+$ . There exists a family of proper regular conditional probability kernels  $\mathbb{K} = \{\mathbb{K}_{\Lambda}\}$ , indexed by the finite subsets  $\Lambda$  of  $\mathbb{Z}^d$ , with the following properties (see theorem 3.2 of [So], as well as theorem 3.3 and the example p.544 following it).

1.  $\mathbb{K}_{\Lambda} : \mathcal{F} \times \Omega \rightarrow [0, 1]$  is a probability kernel and  $\forall F \in \mathcal{F}$ ,  $\omega \mapsto \mathbb{K}_{\Lambda}^{\omega}(F)$  is  $\mathcal{F}_{\Lambda^c}$ -measurable.
2. The kernels are **compatible**:  $\forall \Lambda_1 \subset \Lambda_2$   $\mathbb{K}_{\Lambda_2} = \mathbb{K}_{\Lambda_2} \mathbb{K}_{\Lambda_1}$ , that is

$$\mathbb{K}_{\Lambda_2}^{\omega}(F) = \int_{\Omega} \mathbb{K}_{\Lambda_1}^{\eta}(F) \mathbb{K}_{\Lambda_2}^{\omega}(d\eta) \quad \forall F \in \mathcal{F} \text{ and } \forall \omega \in \Omega.$$

3.  $\mathbb{K}_{\Lambda}$  is **proper**:  $\forall F_1 \in \mathcal{F}$ ,  $\forall F_2 \in \mathcal{F}_{\Lambda^c}$  and  $\forall \omega \in \Omega$ ,  $\mathbb{K}_{\Lambda}^{\omega}(F_1 \cdot F_2) = \mathbb{K}_{\Lambda}^{\omega}(F_1) I_{F_2}(\omega)$ , where  $I_F$  is the indicator function of  $F$ .
4.  $\mathbb{K}_{\Lambda}$  is a **regular conditional probability kernel**:  $\mathbb{K}_{\Lambda}^{\omega}(F) = \mathbb{E}_{\lambda}(F | \mathcal{F}_{\Lambda^c})(\omega)$   $\lambda$ -a.s..

**Definition 3.7.** 1. A local specification is a family of probability kernels  $\mathbb{K} = \{\mathbb{K}_{\Lambda}\}$  on  $\mathcal{F} \times \Omega$ , indexed by  $\Lambda \subset \mathbb{Z}^d$ ,  $|\Lambda| < \infty$ , and verifying 1, 2 and 3 above.

2. A local specification  $\mathbb{K}$  is **quasilocal** if for all  $\Lambda$  and all  $f \in \mathcal{F}_{\text{qloc}}$ , the function  $\mathbb{K}_{\Lambda}(f)$  is quasilocal, where

$$\omega \mapsto \mathbb{K}_{\Lambda}(f)(\omega) := \mathbb{K}_{\Lambda}^{\omega}(f) = \int_{\Omega} f(\eta) \mathbb{K}_{\Lambda}^{\omega}(d\eta).$$

3. A probability measure  $\mu$  is **compatible** with the local specification  $\mathbb{K}$  if and only if  $\mathbb{K}_{\Lambda}^{\omega}(F) = \mathbb{E}_{\mu}(F | \mathcal{F}_{\Lambda^c})(\omega)$ ,  $\mu$ -a.s.,  $\forall F \in \mathcal{F}$ .



**Remark.** A probability measure  $\nu$  is compatible with the local specification  $\mathbb{K}$  if and only if  $\forall \Lambda$  (see [G1] (1.24) p.17.)

$$\int_{\Omega} f(\omega) \nu(d\omega) = \int_{\Omega} \mathbb{K}_{\Lambda}^{\omega}(f) \nu(d\omega) \quad \forall f \in \mathcal{F}_{\text{loc}}.$$

For information gain, if  $\mathcal{H}(\nu|\lambda) = 0$ , then  $\nu = \lambda$ . For specific information gain the following theorem holds, see theorem 15.37 in [G1]. For the sake of completeness the proof of that important result is given.

**Theorem 3.1.** *Let  $\lambda \in \mathcal{M}_1^{+, \theta}$  be compatible with the local specification  $\mathbb{K}$ , which is assumed to be quasilocal. If  $\nu \in \mathcal{M}_1^{+, \theta}$  and  $h(\nu|\lambda) = 0$ , then  $\nu$  is also compatible with  $\mathbb{K}$ .*

**Proof.** The proof is taken from [G1]. It requires only that

$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \mathcal{H}_n(\nu|\lambda) = 0.$$

By assumption  $\mathcal{H}_n(\nu|\lambda) < \infty$  for all  $n \in \mathbb{N}$ ; therefore  $\mathcal{H}_{\mathcal{F}_{\Lambda}}(\nu|\lambda) < \infty$  for any finite subset  $\Lambda$ , so that there exists a  $\mathcal{F}_{\Lambda}$ -measurable function  $g_{\Lambda}$  such that  $\nu = g_{\Lambda} \lambda$  on  $\mathcal{F}_{\Lambda}$ . First, notice that it is sufficient to prove that, for any nonnegative bounded local function  $h$ , and for  $m$  large enough,

$$\int h(\omega) \nu(d\omega) = \int \mathbb{K}_{\Lambda(m)}^{\omega}(h) \nu(d\omega). \quad (3.15)$$

Indeed, let  $h \in \mathcal{F}_{\text{loc}}$ . Since the kernels  $\mathbb{K}_{\Lambda}$  are quasilocal, for any finite  $\Lambda$  and any  $\varepsilon > 0$  there exists a local function  $h_1$  such that  $\|h_1 - \mathbb{K}_{\Lambda}(h)\| \leq \varepsilon$ ; let  $\Lambda(m) \supset \Lambda$  such that (3.15) is true for  $h$  and  $h_1$ . Then,

$$\begin{aligned} \left| \int h(\omega) \nu(d\omega) - \int \mathbb{K}_{\Lambda}^{\omega}(h) \nu(d\omega) \right| &\leq \left| \int \mathbb{K}_{\Lambda(m)}^{\omega}(h) \nu(d\omega) - \int h_1(\omega) \nu(d\omega) \right| \\ &\quad + \left| \int h_1(\omega) \nu(d\omega) - \int \mathbb{K}_{\Lambda}^{\omega}(h) \nu(d\omega) \right| \\ &\leq \left| \int \mathbb{K}_{\Lambda(m)}^{\omega}(\mathbb{K}_{\Lambda}(h)) \nu(d\omega) - \int \mathbb{K}_{\Lambda(m)}^{\omega}(h_1) \nu(d\omega) \right| \\ &\quad + \varepsilon \\ &\leq \int \mathbb{K}_{\Lambda(m)}^{\omega} |\mathbb{K}_{\Lambda}(h) - h_1| \nu(d\omega) + \varepsilon \leq 2\varepsilon. \end{aligned}$$

Since  $\varepsilon$  is arbitrary

$$\int_{\Omega} h(\omega) \nu(d\omega) = \int_{\Omega} \mathbb{K}_{\Lambda}^{\omega}(h) \nu(d\omega)$$

for all  $h \in \mathcal{F}_{\text{loc}}$  and all finite  $\Lambda \subset \mathbb{Z}^d$ .

To prove (3.15) one first prove that it is sufficient to prove (3.16), and that (3.16) is implied by (3.17). Finally (3.17) is proved. Let  $h$  be a given nonnegative local function,  $\varepsilon > 0$  and  $m \in \mathbb{N}$ . Quasilocality of the kernels implies the existence of a  $\mathcal{F}_{\Lambda(n) \setminus \Lambda(m)}$ -measurable  $\tilde{h}$  such that  $\|\mathbb{K}_{\Lambda(m)}^{\omega}(h) - \tilde{h}\| \leq \varepsilon$ . Let  $\Delta$  be any finite subset,

such that  $h$  is  $\mathcal{F}_\Delta$ -measurable and  $\Delta \supset \Lambda(n)$ . Set  $\Delta_m := \Delta \setminus \Lambda(m)$ ,  $\nu \equiv g_{\Delta_m} \lambda$  on  $\mathcal{F}_{\Delta_m}$ , and  $\nu \equiv g_\Delta \lambda$  on  $\mathcal{F}_\Delta$ . Then, since  $\tilde{h}$  is  $\mathcal{F}_{\Delta_m}$ -measurable,

$$\begin{aligned} \int \tilde{h}(\omega) \nu(d\omega) &= \int \tilde{h}(\omega) g_{\Delta_m}(\omega) \lambda(d\omega) \\ &= \int \mathbb{K}_{\Lambda(m)}^\omega(h) g_{\Delta_m}(\omega) \lambda(d\omega) + \int (\tilde{h}(\omega) - \mathbb{K}_{\Lambda(m)}^\omega(h)) g_{\Delta_m}(\omega) \lambda(d\omega) \\ &= \int h(\omega) g_{\Delta_m}(\omega) \lambda(d\omega) + \int (\tilde{h}(\omega) - \mathbb{K}_{\Lambda(m)}^\omega(h)) g_{\Delta_m}(\omega) \lambda(d\omega). \end{aligned}$$

In the last equality one uses the fact that  $g_{\Delta_m}$  is  $\mathcal{F}_{\Delta_m}$ -measurable and  $\lambda$  is  $\mathbb{K}$ -compatible. Writing

$$\int \mathbb{K}_{\Lambda(m)}^\omega(h) \nu(d\omega) = \int \tilde{h}(\omega) \nu(d\omega) + \int (\mathbb{K}_{\Lambda(m)}^\omega(h) - \tilde{h}(\omega)) \nu(d\omega),$$

to prove (3.15), it is sufficient to prove that, given  $\varepsilon > 0$  and  $m$ , there exists  $\Delta$  so that

$$\int h(\omega) |g_\Delta(\omega) - g_{\Delta_m}(\omega)| \lambda(d\omega) \leq \varepsilon. \quad (3.16)$$

On the other hand an inequality of Csiszár gives

$$\mathcal{H}_{\mathcal{F}_\Delta}(\nu|\lambda) - \mathcal{H}_{\mathcal{F}_{\Delta_m}}(\nu|\lambda) \geq \frac{1}{2} \langle |g_\Delta - g_{\Delta_m}|, \lambda \rangle^2.$$

Therefore, to prove (3.15), it is sufficient to show that  $\forall \varepsilon > 0, \forall m, \forall n \geq m, \exists$  a finite subset  $\Delta \supset \Lambda(n)$  so that

$$\mathcal{H}_{\mathcal{F}_\Delta}(\nu|\lambda) - \mathcal{H}_{\mathcal{F}_{\Delta_m}}(\nu|\lambda) \leq \varepsilon. \quad (3.17)$$

So, let  $\varepsilon > 0$ ,  $m$  and  $n \geq m$  be given. Choose  $k$  large enough so that

$$\frac{\mathcal{H}_{\mathcal{F}_{\Lambda(kn)}}(\nu|\lambda)}{|\Lambda(kn)|} \leq \frac{\varepsilon}{|\Lambda(n)|}.$$

One partitions  $\Lambda(kn)$  into  $k^d$  disjoint hypercubes denoted by  $C_1, \dots, C_{k^d}$ , so that  $C_j \equiv \Lambda(n) + t_j$ .  $W_j := \bigcup_{i=1}^j C_i$  and  $W'_j := W_j \setminus (\Lambda(m) + t_j)$ . Then

$$\begin{aligned} \frac{1}{k^d} \sum_{i=1}^{k^d} \left( \mathcal{H}_{\mathcal{F}_{W_i}}(\nu|\lambda) - \mathcal{H}_{\mathcal{F}_{W'_i}}(\nu|\lambda) \right) &\leq \frac{1}{k^d} \left( \mathcal{H}_{\mathcal{F}_{W_1}}(\nu|\lambda) - \mathcal{H}_{\mathcal{F}_{W'_1}}(\nu|\lambda) \right) \\ &\quad + \frac{1}{k^d} \sum_{i=2}^{k^d} \left( \mathcal{H}_{\mathcal{F}_{W_i}}(\nu|\lambda) - \mathcal{H}_{\mathcal{F}_{W_{i-1}}}(\nu|\lambda) \right) \\ &\leq \frac{1}{k^d} \mathcal{H}_{\mathcal{F}_{\Lambda(kn)}}(\nu|\lambda) \\ &= \frac{|\Lambda(n)|}{|\Lambda(kn)|} \mathcal{H}_{\mathcal{F}_{\Lambda(kn)}}(\nu|\lambda) \leq \varepsilon. \end{aligned}$$

Since each term  $\mathcal{H}_{\mathcal{F}_{W_i}}(\nu|\lambda) - \mathcal{H}_{\mathcal{F}_{W'_i}}(\nu|\lambda) \geq 0$ , at least one of them is smaller than  $\varepsilon$ , say  $\mathcal{H}_{\mathcal{F}_{W_j}}(\nu|\lambda) - \mathcal{H}_{\mathcal{F}_{W'_j}}(\nu|\lambda) \leq \varepsilon$ . Then, using translation invariance, one chooses  $\Delta := W_j - t_j$ .  $\square$

**3.3. Asymptotically I-null sequences.** In [Cs] Csiszár introduced for a sequence of measures the notion of asymptotically quasi-independence with limiting measure  $Q$ . This notion is extended in [S] and in [LPS3]. The extension in [LPS3] of asymptotically quasi-independence to asymptotically I-null sequence is natural in statistical mechanics. Proposition 3.4 gives concentration properties of such sequences<sup>41</sup>. The main result is theorem 3.2, which is of the same kind as theorem 3.1.

**Definition 3.8.** Let  $(X, \mathcal{B})$  be a measurable space and  $\{\nu_n\}, \{\lambda_n\}$  be two sequences of probability measures on  $(X, \mathcal{B})$ . The sequence  $\{\nu_n\}$  is asymptotically I-null to the sequence  $\{\lambda_n\}$  on the scale  $V_n$  if

$$\lim_{n \rightarrow \infty} \frac{1}{V_n} \mathcal{H}(\nu_n | \lambda_n) = 0.$$

**Proposition 3.4.** Let  $(X, \mathcal{B})$  be a Hausdorff topological space and  $\mathcal{B}$  its Borel  $\sigma$ -algebra. Let  $\{\lambda_n\}$  be a sequence of probability measures on  $(X, \mathcal{B})$ , which is eventually concentrated on  $N$  at an exponential rate on the scale  $V_n$ . If  $\{\nu_n\}$  is a sequence of probability measures on  $(X, \mathcal{B})$ , which is asymptotically I-null to  $\{\lambda_n\}$  on the scale  $V_n$ , then  $\{\nu_n\}$  is eventually concentrated on  $N$ .

**Proof.** Let  $G$  be an open neighbourhood of  $N$ .

$$\begin{aligned} \mathcal{H}(\nu_n | \lambda_n) &\geq \nu_n[G] \ln \frac{\nu_n[G]}{\lambda_n[G]} + \nu_n[X \setminus G] \ln \frac{\nu_n[X \setminus G]}{\lambda_n[X \setminus G]} \\ &\geq -\ln 2 - \nu_n[X \setminus G] \ln \lambda_n[X \setminus G]. \end{aligned}$$

Since

$$\limsup_n \frac{1}{V_n} \ln \lambda_n[X \setminus G] < 0,$$

there exists  $\delta > 0$  such that, for all  $n$  sufficiently large,

$$\frac{1}{V_n} \ln \lambda_n[X \setminus G] < -\delta.$$

Thus

$$\frac{1}{V_n} \mathcal{H}(\nu_n | \lambda_n) \geq \frac{-\ln 2}{V_n} + \delta \cdot \nu_n[X \setminus G] \geq \frac{-\ln 2}{V_n};$$

but

$$\lim_n \frac{1}{V_n} \mathcal{H}(\nu_n | \lambda_n) = 0$$

by hypothesis, so that  $\lim_n \nu_n[X \setminus G] = 0$  and  $\lim_n \nu_n[G] = 1$ .  $\square$

There are natural sequences of probability measures, both in statistical mechanics and in large deviations theory (Varadhan's theorem), which are defined by an element of the dual space  $\mathcal{F}_{\text{qloc}}$ . Let  $f \in \mathcal{F}_{\text{qloc}}$ . Define a sequence of probability measures  $\nu_n$ , absolutely continuous with respect to  $\lambda$ , by

$$\frac{d\nu_n}{d\lambda}(\omega) := \exp \left[ \sum_{x \in \Lambda(n)} \theta_x f(\omega) + \psi_n(\omega) - |\Lambda(n)| p_n(f | \lambda) \right]. \quad (3.18)$$

<sup>41</sup> Concentration properties are studied later, in subsection 5.1. It is however natural to state this result here. For the definitions of ‘‘eventually concentrated’’ and ‘‘eventually concentrated at an exponential rate’’, see beginning of subsection 5.1.

The following convention in (3.18) is made: the boundary terms  $\psi_n$  are adjusted so that

$$\exp(|\Lambda(n)|p_n(f)) = \int \exp \left[ \sum_{x \in \Lambda(n)} \theta_x f(\omega) + \psi_n(\omega) \right] \lambda(d\omega).$$

This is always possible by adding to  $\psi_n$  a suitable constant  $c_n$  such that

$$\lim_{n \rightarrow \infty} \frac{c_n}{|\Lambda(n)|} = 0.$$

The basic criterion for the existence of cluster points is lemma 3.2. The next theorem gives another criterion, which is of interest in statistical mechanics.

**Theorem 3.2.** *Let  $\lambda$  be a translation invariant probability measure on  $(\Omega, \mathcal{F})$ , which is asymptotically decoupled, and  $f \in \mathcal{F}_{\text{qloc}}$ . Let  $\psi_n : \Omega \rightarrow \mathbb{R}$ , such that*

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \sup_{\omega \in \Omega} |\psi_n(\omega)| = 0.$$

Let  $\nu_n \in \mathcal{M}_1^+$  be defined by<sup>42</sup>

$$\frac{d\nu_n}{d\lambda}(\omega) := \exp \left[ \sum_{x \in \Lambda(n)} \theta_x f(\omega) + \psi_n(\omega) - |\Lambda(n)|p_n(f|\lambda) \right].$$

If  $\{\mu_n\}$  is asymptotically I-null to  $\{\nu_n\}$ , then the set of cluster points of the sequence  $\{\mathcal{A}_{\Lambda(n)}\mu_n\}_{n \geq 1}$  is non-empty, and  $h(\mu|\lambda) = \langle f, \mu \rangle - p(f|\lambda)$  for any cluster point  $\mu$  of  $\{\mathcal{A}_{\Lambda(n)}\mu_n\}_{n \geq 1}$ .

**Remark.** A simple application of theorem 3.2 is to choose  $\mu_n = \nu_n$ . Hence, the set of cluster points of the sequence  $\{\mathcal{A}_{\Lambda(n)}\nu_n\}$  is non-empty, and each cluster point  $\nu$  has the same specific information gain,  $h(\nu|\lambda) = \langle f, \nu \rangle - p(f|\lambda)$ . In the terminology of subsection 3.5,  $\nu$  and  $f$  are in duality.

**Proof.** For simplicity of notation one assumes that  $\psi_n(\omega) \equiv 0$ . By definition of the information gain

$$\begin{aligned} \frac{1}{|\Lambda(n)|} \mathcal{H}(\mu_n|\nu_n) &= \frac{1}{|\Lambda(n)|} \mathcal{H}(\mu_n|\lambda) - \int_{\Omega} \mathcal{A}_{\Lambda(n)} f(\omega) \mu_n(d\omega) + p_n(f|\lambda) \\ &= \frac{1}{|\Lambda(n)|} \mathcal{H}(\mu_n|\lambda) - \int_{\Omega} f(\omega) \mathcal{A}_{\Lambda(n)} \mu_n(d\omega) + p_n(f|\lambda). \end{aligned} \quad (3.19)$$

From lemma 3.3 and (3.19),

$$\begin{aligned} \limsup_n \frac{1}{|\Lambda(m + \lceil g(m) \rceil / 2)|} \mathcal{H}_m(\mathcal{A}_{\Lambda(n)} \mu_n|\lambda) &\leq \limsup_n \frac{1}{|\Lambda(n)|} \mathcal{H}_{n+m}(\mu_n|\lambda) + \frac{2c(m)}{|\Lambda(m)|} \\ &\leq \limsup_n \frac{1}{|\Lambda(n)|} \mathcal{H}(\mu_n|\lambda) + \frac{2c(m)}{|\Lambda(m)|} \\ &\leq \limsup_n \left( \frac{1}{|\Lambda(n)|} \mathcal{H}(\mu_n|\nu_n) + \int_{\Omega} f(\omega) \mathcal{A}_{\Lambda(n)} \mu_n(d\omega) - p_n(f|\lambda) \right) + \frac{2c(m)}{|\Lambda(m)|} \\ &\leq \limsup_n \int_{\Omega} f(\omega) \mathcal{A}_{\Lambda(n)} \mu_n(d\omega) - p(f|\lambda) + \frac{2c(m)}{|\Lambda(m)|} \\ &\leq \|f\| - p(f|\lambda) + \frac{2c(m)}{|\Lambda(m)|} < \infty. \end{aligned}$$

<sup>42</sup> Same convention as in (3.18).

By lemma 3.2 the set of cluster points of the sequence  $\{\mathcal{A}_{\Lambda(n)}\mu_n\}$  is non-empty.

Assume that  $\mu := \lim_{n \rightarrow \infty} \mathcal{A}_{\Lambda(n)}\mu_n$ . One computes

$$\lim_{m \rightarrow \infty} \frac{1}{|\Lambda(m)|} \mathcal{H}_m(\mu|\lambda).$$

Since the information gain is l.s.c.,

$$\mathcal{H}_m(\mu|\lambda) \leq \liminf_n \mathcal{H}_m(\mathcal{A}_{\Lambda(n)}\mu_n|\lambda).$$

Given  $m$ , by lemma 3.3,

$$\mathcal{H}_m(\mathcal{A}_{\Lambda(n)}\mu_n|\lambda) \leq \frac{|\Lambda(m + \lceil g(m) \rceil / 2)|}{|\Lambda(n)|} \mathcal{H}_{n+m}(\mu_n|\lambda) + 2c_m(\lambda).$$

(3.19) implies

$$\begin{aligned} \frac{1}{|\Lambda(n)|} \mathcal{H}_{n+m}(\mu_n|\lambda) &\leq \frac{1}{|\Lambda(n)|} \mathcal{H}(\mu_n|\lambda) \\ &= \frac{1}{|\Lambda(n)|} \mathcal{H}(\mu_n|\nu_n) + \left( \int_{\Omega} f(\omega) \mathcal{A}_{\Lambda(n)}\mu_n(d\omega) - p_n(f|\lambda) \right). \end{aligned}$$

Hence, since  $\lim_n \mathcal{A}_{\Lambda(n)}\mu_n = \mu$ , putting the three above inequalities together one obtains

$$\begin{aligned} \frac{1}{|\Lambda(m)|} \mathcal{H}_m(\mu|\lambda) &\leq \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(m)|} \mathcal{H}_m(\mathcal{A}_{\Lambda(n)}\mu_n|\lambda) \\ &\leq \frac{|\Lambda(m + \lceil g(m) \rceil / 2)|}{|\Lambda(m)|} \left( \int_{\Omega} f(\omega) \mu(d\omega) - p(f|\lambda) \right) + \frac{2c_m(\lambda)}{|\Lambda(m)|}. \end{aligned}$$

Therefore

$$h(\mu|\lambda) = \lim_{m \rightarrow \infty} \frac{1}{|\Lambda(m)|} \mathcal{H}_m(\mu|\lambda) \leq \langle f, \mu \rangle - p(f|\lambda).$$

On the other hand, for  $\mu \in \mathcal{M}_1^{+, \theta}$ ,  $\langle f, \mu \rangle - p(f|\lambda) \leq h(\mu|\lambda)$ .  $\square$

**3.4. Large deviations of the empirical measure.** In the whole section  $(\Omega, \mathcal{F}, \lambda)$  is a fixed probability space and  $\lambda$  is a translation invariant, asymptotically decoupled probability measure. Let  $Y : \Omega \rightarrow \mathcal{M}_1^+$ ,  $\omega \mapsto Y(\omega) := \delta_{\omega}$ . Let  $\Lambda$  be a finite subset of  $\mathbb{Z}^d$ . The **empirical measure**  $T_{\Lambda}$  is defined on  $\Omega$  by

$$T_{\Lambda}(\omega) := \frac{1}{|\Lambda|} \sum_{x \in \Lambda} \delta_{\theta_x \omega}.$$

The main result (theorem 3.3) is the existence of a large deviations principle<sup>43</sup>. The proof is done in several steps. The main steps are lemmas 3.1 and 3.4; the other steps are proved by general arguments. In the first step the empirical measures are considered as  $\mathcal{F}_{\text{qloc}}^*$ -valued random variables.  $\mathcal{F}_{\text{qloc}}^*$ , the dual of the Banach space  $(\mathcal{F}_{\text{qloc}}, \|\cdot\|)$ , is equipped with the  $\sigma(\mathcal{F}_{\text{qloc}}^*, \mathcal{F}_{\text{qloc}})$ -topology. One defines<sup>44</sup> a function  $s : \mathcal{F}_{\text{qloc}}^* \rightarrow [-\infty, 0]$ , which is the analogue of (2.1) in section 2.  $s$  is nontrivial only on the unit ball of  $\mathcal{F}_{\text{qloc}}^*$ , which is compact. Hence,  $s$  is the rate-function of a large deviations principle. The second step is the determination of the rate-function. In

<sup>43</sup>This is an extension of theorem 8.5 in [LPS3]; the proof is similar.

<sup>44</sup>To define  $s$  one needs only that  $\lambda$  is asymptotically decoupled from below.

the third step, using the natural embedding of  $\mathcal{M}_1^+$  into the unit ball of  $\mathcal{F}_{\text{qloc}}^*$ , one shows that  $s$  is also the rate-function of a large deviations principle for the empirical measures considered as  $\mathcal{M}_1^+$ -valued random variables.

1. **Definition of  $s$ .** Following [LP] one defines on  $\mathcal{F}_{\text{qloc}}^*$  an upper semicontinuous (u.s.c.) function  $s$ . Let  $\nu \in \mathcal{F}_{\text{qloc}}^*$  and  $\{G\}$  a base of open neighbourhoods of  $\nu$ . The claim is that one can define  $s(\nu|\lambda)$  as

$$\begin{aligned} s(\nu|\lambda) &:= \inf_G \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in G] \\ &= \inf_G \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in G]. \end{aligned} \quad (3.20)$$

First, if  $\nu$  is not translation invariant, or  $\langle 1, \nu \rangle \neq 1$ , or  $\nu$  is not positive, that is  $f \geq 0 \not\Rightarrow \langle f, \nu \rangle \geq 0$ , then  $s(\nu|\lambda) = -\infty$ . Indeed, suppose that  $\nu$  is not translation invariant. Then there exists  $x \in \mathbb{Z}^d$  and  $f \in \mathcal{F}_{\text{loc}}$ , so that  $\langle (f - \theta_x f), \nu \rangle = a > 0$ . Set  $g := f - \theta_x f$ . Then

$$\limsup_{n \rightarrow \infty} \sup_{\omega} |\mathcal{A}_{\Lambda(n)} g(\omega)| = \limsup_{n \rightarrow \infty} \sup_{\omega} |\langle g, T_{\Lambda(n)}(\omega) \rangle| = 0.$$

Therefore, if

$$G := \{\mu \in \mathcal{F}_{\text{qloc}}^* : |\langle g, \mu \rangle - \langle g, \nu \rangle| \leq a/2\},$$

then  $\{\omega : T_{\Lambda(n)}(\omega) \in G\} = \emptyset$  for  $n$  large enough, so that

$$s(\nu|\lambda) = \inf_G \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in G] = -\infty.$$

Similarly one proves that  $s(\nu|\lambda) > -\infty$  implies that  $\nu$  is positive and normalized. On the unit ball of  $\mathcal{F}_{\text{qloc}}^*$  the  $\sigma(\mathcal{F}_{\text{qloc}}^*, \mathcal{F}_{\text{qloc}})$ -topology coincides with the  $\sigma(\mathcal{F}_{\text{qloc}}^*, \mathcal{F}_{\text{loc}})$ -topology. Let  $\nu$  be an element of the unit ball and

$$U(\varepsilon; f_1, \dots, f_k) := \{\mu \in \mathcal{F}_{\text{qloc}}^* : \max_{i=1}^k |\langle f_i, \mu \rangle - \langle f_i, \nu \rangle| < \varepsilon\} \quad f_i \in \mathcal{F}_{\text{loc}}, \quad i = 1, \dots, k.$$

Open sets  $U(\varepsilon; f_1, \dots, f_k)$ ,  $f_i \in \mathcal{F}_{\text{loc}}$ ,  $k \in \mathbb{N}$  and  $\varepsilon > 0$ , form a base  $\{G\}$  of neighbourhoods of  $\nu$ . Let  $\phi_\nu$  be the nonnegative convex function

$$\phi_\nu(\mu) := \max_{i=1}^k |\langle f_i, \mu \rangle - \langle f_i, \nu \rangle|.$$

Lemma 3.1 implies the existence of a neighbourhood of  $\nu$ ,  $G' \subset G$ , such that

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in G] &\geq \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in G'] \\ &\geq \inf_{G'} \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in G']. \end{aligned}$$

Hence,

$$\begin{aligned} s(\nu|\lambda) &:= \inf_G \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in G] \\ &= \inf_G \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in G]. \end{aligned}$$

$s(\nu|\lambda)$  is nonpositive, u.s.c. and concave<sup>45</sup>.

**Proposition 3.5.** *Let  $\lambda$  be a translation invariant, asymptotically decoupled probability measure. Then, for each Borel set  $B$ ,*

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in B] &\geq \sup_{\mu \in \text{int } B} s(\mu|\lambda) \\ \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in B] &\leq \sup_{\mu \in \text{cl } B} s(\mu|\lambda). \end{aligned}$$

**Proof.** The lower bound is immediate. It follows from the second line in (3.20). Indeed, let  $G$  be open. By definition  $\forall \mu \in G$

$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in G] \geq s(\mu|\lambda).$$

The proof of the upper bound follows from the first line in (3.20). For any Borel set  $B$  define the increasing set-function

$$\bar{s}[B] := \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in B].$$

Let  $B_i, i = 1, \dots, n$ , be Borel sets. Then

$$\bar{s}\left[\bigcup_{1 \leq i \leq n} B_i\right] = \max_{1 \leq i \leq n} \bar{s}[B_i].$$

Indeed, e.g. for  $n = 2$ ,

$$\begin{aligned} \lambda[T_{\Lambda(n)} \in B_1] \vee \lambda[T_{\Lambda(n)} \in B_2] &\leq \lambda[T_{\Lambda(n)} \in B_1 \cup B_2] \\ &\leq 2\lambda[T_{\Lambda(n)} \in B_1] \vee \lambda[T_{\Lambda(n)} \in B_2]; \end{aligned}$$

moreover, for any pair of sequences  $\{a_n\}_{n \geq 1} \subset \overline{\mathbb{R}}$  and  $\{b_n\}_{n \geq 1} \subset \overline{\mathbb{R}}$ ,

$$\limsup_{n \rightarrow \infty} (a_n \vee b_n) = (\limsup_{n \rightarrow \infty} a_n) \vee (\limsup_{n \rightarrow \infty} b_n).$$

Let  $B$  be a closed Borel set and  $\mathcal{F}_{\text{qloc}}^{*,1}$  be the unit ball of  $\mathcal{F}_{\text{qloc}}^*$ . Then

$$\bar{s}[B] = \bar{s}[B \cap \mathcal{F}_{\text{qloc}}^{*,1}],$$

and  $K := B \cap \mathcal{F}_{\text{qloc}}^{*,1}$  is compact. If  $\bar{s}[K] = -\infty$ , then there is nothing to prove; the same is true if  $K = \emptyset$  and  $\sup_{\nu \in K} s(\nu|\lambda) = \infty$ . Assume that  $\sup_{\nu \in K} s(\nu|\lambda) < a \in \mathbb{R}$  and that  $K \neq \emptyset$ ; then  $s(\nu|\lambda) < a \forall \nu \in K$ . Hence for each  $\nu \in K$  there exists an open set  $G_\nu$  such that  $s(\nu|\lambda) \leq \bar{s}[G_\nu] < a$ . Since  $K$  is compact one can cover  $K$  by a finite number of these sets, say  $G_{\nu_1}, \dots, G_{\nu_n}$ . Set  $G := G_{\nu_1} \cup \dots \cup G_{\nu_n}$ . Then

$$\bar{s}[G] = \max_{1 \leq j \leq n} \bar{s}[G_{\nu_j}] < a.$$

<sup>45</sup> Upper semicontinuity follows directly from (3.20). Concavity follows from a slight modification of the proof of lemma 3.1. For details see [LPS3]. Notice that

$$\frac{1}{2}\phi_{\mu_1}(\nu) + \frac{1}{2}\phi_{\mu_2}(\nu) \geq \phi_{\frac{1}{2}\mu_1 + \frac{1}{2}\mu_2}(\nu).$$

Instead of (3.6) one defines  $N = N(M, a', \delta)$  so that for  $n \geq N$ ,

$$\mathcal{E}_{\Lambda(n), \phi_{\frac{1}{2}\mu_1 + \frac{1}{2}\mu_2}}(a) \supset \bigcap_{q, \text{ odd}} \mathcal{E}_{\Lambda_q, \phi_{\mu_1}}(a') \bigcap_{q, \text{ even}} \mathcal{E}_{\Lambda_q, \phi_{\mu_2}}(a').$$

Since  $K \subset G$ ,  $\bar{s}[K] < a$ . Taking the infimum over  $a$ , such that  $\sup_{\nu \in K} s(\nu|\lambda) < a$ , one obtains

$$\bar{s}[K] \leq \sup_{\nu \in K} s(\nu|\lambda).$$

□

2. **Determination of  $s$ .** Let  $f \in \mathcal{F}_{\text{qloc}}$ . The function  $\nu \mapsto \langle f, \nu \rangle$  on  $\mathcal{F}_{\text{qloc}}^*$  is continuous. Varadhan's theorem ([DS] 2.1.10) implies

$$\begin{aligned} p(f|\lambda) &= \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \int_{\Omega} \exp \left\{ \sum_{j \in \Lambda(n)} \theta_j f(\omega) \right\} \lambda(d\omega) \\ &= \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \int_{\Omega} \exp \{ |\Lambda(n)| \langle f, T_{\Lambda(n)}(\omega) \rangle \} \lambda(d\omega) \\ &= \sup_{\nu} \{ \langle f, \nu \rangle + s(\nu|\lambda) \}. \end{aligned}$$

By the duality theorem about conjugate convex functions [Br] one obtains

$$-s(\nu|\lambda) = \sup_{f \in \mathcal{F}_{\text{qloc}}} \{ \langle f, \nu \rangle - p(f|\lambda) \} = p^*(\nu|\lambda).$$

**Lemma 3.4.** *If  $s(\nu|\lambda) > -\infty$ , then  $\nu \in \mathcal{M}_1^{+, \theta}$ .*

**Proof.** Suppose that  $s(\nu|\lambda) \geq -a > -\infty$ . Let  $C$  be a closed convex neighbourhood of  $\nu$ . Then

$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in C] \geq -a.$$

Hence there exists  $n_{a,C}$  such for  $n \geq n_{a,C}$ ,

$$\frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in C] \geq -2a.$$

The basic identity

$$-\ln \lambda[T_{\Lambda(n)} \in C] = \mathcal{H}(\lambda[\cdot | T_{\Lambda(n)} \in C]|\lambda),$$

implies the upper bound

$$\begin{aligned} \frac{1}{|\Lambda(n)|} \mathcal{H}_n(\lambda[\cdot | T_{\Lambda(n)} \in C]|\lambda) &\leq \frac{1}{|\Lambda(n)|} \mathcal{H}(\lambda[\cdot | T_{\Lambda(n)} \in C]|\lambda) \\ &= -\frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in C] \leq 2a \quad \text{for } n \geq n_{a,C}. \end{aligned}$$

One constructs<sup>46</sup> a sequence  $C_n$  of closed convex neighbourhoods such that  $C_n \downarrow \{\nu\}$  and

$$\frac{1}{|\Lambda(n)|} \mathcal{H}_n(\lambda[\cdot | T_{\Lambda(n)} \in C_n]|\lambda) \leq 2a.$$

Then one constructs a sequence of translation invariant probability measures  $\{\mu_n\}$ . For any measure  $\mu \in \mathcal{M}$  one defines  $\mathcal{P}_n \mu$  by

$$\mathcal{P}_n \mu := \prod_{x \in \mathbb{Z}^d} \theta_{(2n+1)x}(\mu|_{\mathcal{F}_{\Lambda(n)}}).$$

<sup>46</sup> The argument is given in the remark following definition 5.2 in subsection 5.3.



Notice that  $\mu$  is a cluster point of  $\{\mu_k\}$  if and only if it is a cluster point of  $\{\mathcal{P}_k\mu_k\}$ . Also,  $\mu$  is a cluster point of  $\{\mathcal{A}_{\Lambda(k)}\mu_k\}$  if and only if it is a cluster point of  $\{\mathcal{A}_{\Lambda(k)}\mathcal{P}_k\mu_k\}$ . Define

$$\mu_n := \mathcal{A}_{\Lambda(n)}\mathcal{P}_n\lambda[\cdot | T_{\Lambda(n)} \in C_n].$$

$\mu_n \in \mathcal{M}_1^{+, \theta}$  and is ergodic with respect to  $\mathbb{Z}^d$ . Translation invariance and (3.11) imply

$$\frac{1}{|\Lambda(n)|} \mathcal{H}_n(\mathcal{A}_{\Lambda(n)}\lambda[\cdot | T_{\Lambda(n)} \in C_n]|\lambda) \leq \frac{1}{|\Lambda(n)|} \mathcal{H}(\lambda[\cdot | T_{\Lambda(n)} \in C_n]|\lambda) \leq 2a.$$

Given  $m$ , there exists  $\delta > 0$ , so that (see lemma 3.3 and (3.5))

$$\begin{aligned} \frac{1}{(1+\delta)|\Lambda(m)|} \mathcal{H}_m(\mathcal{A}_{\Lambda(n)}\lambda[\cdot | T_{\Lambda(n)} \in C_n]|\lambda) &\leq \frac{1}{|\Lambda(n)|} \mathcal{H}(\lambda[\cdot | T_{\Lambda(n)} \in C_n]|\lambda) + \frac{2c(m)}{|\Lambda(m)|} \\ &\leq 2a + \frac{2c(m)}{|\Lambda(m)|}. \end{aligned}$$

Hence

$$\limsup_{n \rightarrow \infty} \mathcal{H}_m(\mathcal{A}_{\Lambda(n)}\lambda[\cdot | T_{\Lambda(n)} \in C_n]|\lambda) \leq 2a(1+\delta)|\Lambda(m)| + (1+\delta)2c(m) \equiv K_m,$$

and one can apply lemma 3.2. Hence there exists a subsequence  $\{n_k\}_{k \geq 1}$  such that

$$\lim_{k \rightarrow \infty} \mu_{n_k} =: \mu \in \mathcal{M}_1^{+, \theta}.$$

Finally one shows that  $\mu = \nu$ . Indeed,  $\lim_{k \rightarrow \infty} \mu_{n_k} = \mu$  is equivalent to

$$\lim_{k \rightarrow \infty} \mathcal{A}_{\Lambda(n_k)}\lambda[\cdot | T_{\Lambda(n_k)} \in C_{n_k}] = \mu.$$

Let  $f \in \mathcal{F}_{\text{loc}}$ ; then

$$\langle f, \mathcal{A}_{\Lambda(n_k)}\lambda[\cdot | T_{\Lambda(n_k)} \in C_{n_k}] \rangle = \int_{\Omega} \langle f, T_{\Lambda(n_k)}(\omega) \rangle \lambda[d\omega | T_{\Lambda(n_k)} \in C_{n_k}],$$

so that, since  $C_{n_k}$  is a closed convex set,

$$\mathcal{A}_{\Lambda(n_k)}\lambda[\cdot | T_{\Lambda(n_k)} \in C_{n_k}] = \int_{\Omega} T_{\Lambda(n_k)}(\omega)[\cdot] \lambda[d\omega | T_{\Lambda(n_k)} \in C_{n_k}] \in C_{n_k}.$$

Since  $C_k \downarrow \{\nu\}$  it follows that  $\mu = \nu$ .  $\square$

**3. Main theorem.** Let  $T_{\Lambda(n)}$  be considered as a  $\mathcal{M}_1^+$ -valued random variable. There is a natural embedding of  $\mathcal{M}_1^+$  in  $\mathcal{F}_{\text{qloc}}^{*,1}$ , in particular, if  $B \subset \mathcal{M}_1^+$ , then  $\text{cl} B$ , the closure of  $B$  in  $\mathcal{M}_1^+$ , is equal to  $\overline{B} \cap \mathcal{M}_1^+$ , where  $\overline{B}$  is the closure of  $B$  in  $\mathcal{F}_{\text{qloc}}^{*,1}$ . Moreover, for all  $a \in \mathbb{R}$

$$\{\mu \in \mathcal{M}_1^+ : s(\mu|\lambda) \geq a\}$$

is compact in  $\mathcal{M}_1^{+, \theta}$ . The above results are summarized in the next theorem. For earlier results in the multidimensional case, see in particular [C], [FO], [O], [G2] and [LPS3].

**Theorem 3.3.** *Let  $\lambda$  be a translation invariant probability measure on  $(\Omega, \mathcal{F})$ , which is asymptotically decoupled. Then the empirical measures  $T_{\Lambda(n)}$  defined on the probability space  $(\Omega, \mathcal{F}, \lambda)$ , with  $\mathcal{M}_1^+$  equipped with the  $\sigma(\mathcal{M}_1^+, \mathcal{F}_{\text{qloc}})$ -topology, verify a*

large deviations principle on  $\mathcal{M}_1^+$ , with rate-function

$$s(\mu|\lambda) = \begin{cases} -h(\mu|\lambda) & \text{if } \mu \in \mathcal{M}_1^{+, \theta} \\ -\infty & \text{if } \mu \in \mathcal{M}_1^+ \setminus \mathcal{M}_1^{+, \theta}, \end{cases}$$

that is,  $s$  is u.s.c. with compact level-sets and for each Borel set  $B \subset \mathcal{M}_1^+$ ,

$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in B] \geq \sup_{\mu \in \text{int } B} s(\mu|\lambda)$$

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in B] \leq \sup_{\mu \in \text{cl } B} s(\mu|\lambda).$$

**Remark.** The rate-function coincides with  $-h(\mu|\lambda)$  only on  $\mathcal{M}_1^{+, \theta}$ . There are measures  $\mu$ , which are not translation-invariant with  $h(\mu|\lambda) < \infty$ . For example, periodic measures.

**3.5. Duality.** The convex, l.s.c., proper<sup>47</sup> function  $p(\cdot|\lambda)$  on  $\mathcal{F}_{\text{qloc}}$  allows to define the notion of dual points.

**Definition 3.9.** Let  $\lambda$  be a translation invariant probability measure, which is asymptotically decoupled. Let  $f \in \mathcal{F}_{\text{qloc}}$  and  $\nu \in \mathcal{M}$ . One says that  $f$  and  $\nu$  are dual points, or in duality, which is denoted by  $f \leftrightarrow \nu$ , if and only if

$$\langle f, \nu \rangle = p(f|\lambda) + p^*(\nu|\lambda).$$

**Remark.**  $p^*(\cdot|\lambda)$  is finite only on  $\mathcal{M}_1^{+, \theta}$  and  $f \leftrightarrow \nu$  is equivalent to

$$\nu \in \mathcal{M}_1^{+, \theta} \quad \text{and} \quad \langle f, \nu \rangle = p(f|\lambda) + h(\nu|\lambda). \quad (3.21)$$

Recall that for  $\nu \in \mathcal{M}_1^{+, \theta}$ ,  $\langle f, \nu \rangle \leq p(f|\lambda) + h(\nu|\lambda)$ . If  $f \leftrightarrow \nu$ , then in the physics terminology  $-f$  is the interaction energy per site, multiplied by  $\beta$  (inverse temperature) and  $\nu$  is an equilibrium state. (3.21) is the statement of the variational principle in the thermodynamical formalism.

Assume that  $f \leftrightarrow \nu$ . Then  $p(\cdot|\lambda)$  is subdifferentiable at  $f$  with subgradient  $\nu$ . Indeed,  $p(f|\lambda)$  is finite and for all  $g \in \mathcal{F}_{\text{qloc}}$

$$p(g + f|\lambda) \geq -h(\nu|\lambda) + \langle g + f, \nu \rangle = p(f|\lambda) + \langle g, \nu \rangle.$$

Conversely, if  $\nu$  is a subgradient of  $p$  at  $f$ , then  $f \leftrightarrow \nu$ . Indeed, for all  $g \in \mathcal{F}_{\text{qloc}}$

$$p(f|\lambda) - \langle f, \nu \rangle \leq p(f + g|\lambda) - \langle f + g, \nu \rangle,$$

which implies that

$$p(f|\lambda) - \langle f, \nu \rangle \leq -p^*(\nu|\lambda) = h(\nu|\lambda).$$

The set of all subgradients of  $p(\cdot|\lambda)$  at  $f$  is denoted by  $\partial p(f)$ . It is a closed convex subset of  $\mathcal{M}_1^{+, \theta}$  and the above remarks show that

$$\partial p(f) = \{\nu \in \mathcal{M}_1^{+, \theta} : f \leftrightarrow \nu\}.$$

Theorem 3.4 is a special case of a theorem of Varadhan. In section 4 theorem 4.2 is another (stronger) version of this theorem.

<sup>47</sup> I.e.  $p(f|\lambda) > -\infty, \forall f \in \mathcal{F}_{\text{qloc}}$ , and  $p(f|\lambda) < \infty$  for some  $f \in \mathcal{F}_{\text{qloc}}$ .

**Theorem 3.4.** *Let  $\lambda$  be a translation invariant probability measure on  $(\Omega, \mathcal{F})$ , which is asymptotically decoupled and  $f \in \mathcal{F}_{\text{qloc}}$ . Let  $\psi_n : \Omega \rightarrow \mathbb{R}$ , such that*

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \sup_{\omega \in \Omega} |\psi_n(\omega)| = 0.$$

Define<sup>48</sup>  $\nu_n \in \mathcal{M}_1^+$  by

$$\frac{d\nu_n}{d\lambda}(\omega) := \exp \left[ \sum_{x \in \Lambda(n)} \theta_x f(\omega) + \psi_n(\omega) - |\Lambda(n)| p_n(f|\lambda) \right].$$

Then the empirical measures  $T_{\Lambda(n)}$ , as random variables defined on the probability spaces  $(\Omega, \mathcal{F}, \nu_n)$ , verify a large deviations principle with rate-function  $s^f$ ,

$$\mu \mapsto s^f(\mu|\lambda) = s(\mu|\lambda) + \langle f, \mu \rangle - p(f|\lambda).$$

**Remark.** As noticed earlier theorem 3.2 also implies existence of solutions for the variational principle:  $\partial p(f)$  is non-empty and compact. One reaches here the same conclusion since  $\partial p(f) = \{\mu : s^f(\mu|\lambda) = 0\}$ . There is a dual set,  $\partial p^*(\nu)$ , and

$$\nu \in \partial p(f) \iff f \in \partial p^*(\nu).$$

The problem  $\partial p^*(\nu) \neq \emptyset$ , for a given  $\nu$ , is not considered here<sup>49</sup>.

---

<sup>48</sup> With the convention for (3.18).

<sup>49</sup> One main problem discussed in [EFS] is to determine whether a given probability measure  $\nu$  is a Gibbs measure. In the perspective of these lectures this problem can be formulated as follows. Given  $\nu$ , does exist  $f \in \mathcal{L}$ , the space of bounded total oscillation functions, which is defined in the next section, such that  $f \in \partial p^*(\nu)$ . Theorem 4.1 gives sufficient conditions on the reference measure  $\lambda$ , so that the existence of  $f \in \mathcal{L}$  with  $f \in \partial p^*(\nu)$  implies that  $\nu$  is a Gibbs measure; more precisely, there exists a local specification defined by  $f$ , and  $\nu$  is compatible with that local specification (see next section). Subdifferentiability of  $p^*$  or  $h(\cdot|\lambda)$  is a subtle question. The answer depends of course in an essential manner of the choice of the dual space of  $\mathcal{M}$ . See in particular section 2.6.7 in [EFS].

## 4. SPACE OF BOUNDED TOTAL OSCILLATION FUNCTIONS

In the whole section  $\lambda$  is a fixed translation invariant probability measure, which is asymptotically decoupled with parameters  $g$  and  $c$ .

The spaces  $(\mathcal{F}_{\text{qloc}}, \mathcal{M})$  form a dual pair. However, one is primarily interested in the convex subset  $\mathcal{M}_1^+$ , whose induced topology coincides with the  $\sigma(\mathcal{M}_1^{+, \theta}, \mathcal{F}_{\text{loc}})$ -topology. Moreover, since  $|p(f|\lambda) - p(g|\lambda)| \leq \|f - g\|$ ,  $p$  is uniquely defined on  $\mathcal{F}_{\text{qloc}}$  by its values on  $\mathcal{F}_{\text{loc}}$ . Since the bilinear map  $\mathcal{F}_{\text{loc}} \times \mathcal{M} \rightarrow \mathbb{R}$ ,

$$\langle f, \mu \rangle = \int_{\Omega} f(\omega) \mu(d\omega),$$

is separating, the spaces  $(\mathcal{F}_{\text{loc}}, \mathcal{M})$  form also a dual pair, if they are equipped with the  $\sigma(\mathcal{F}_{\text{loc}}, \mathcal{M})$ -topology, resp.  $\sigma(\mathcal{M}, \mathcal{F}_{\text{loc}})$ -topology. The induced topology on  $\mathcal{M}_1^{+, \theta}$  is the same as before, the dual relations still hold,

$$-p(f|\lambda) = s^*(f|\lambda) \quad \text{and} \quad -s(\mu|\lambda) = p^*(\mu|\lambda).$$

In some respects,  $\mathcal{F}_{\text{qloc}}$  is too large and  $\mathcal{F}_{\text{loc}}$  is too small. There is a natural choice<sup>50</sup>,  $\mathcal{F}_{\text{loc}} \subset \mathcal{L} \subset \mathcal{F}_{\text{qloc}}$ , which is made in this section.

**Definition 4.1.** *Let  $f$  be a bounded function on  $\Omega$ . Let  $\Lambda \subset \mathbb{Z}^d$ ; the oscillation of  $f$  over  $\Lambda$  is*

$$\delta_{\Lambda}(f) := \sup\{|f(\omega) - f(\eta)| : \omega_j = \eta_j \forall j \notin \Lambda\}.$$

The total oscillation of  $f$  is

$$\Delta(f) := \sum_{j \in \mathbb{Z}^d} \delta_j(f).$$

The space of bounded total oscillation functions is

$$\mathcal{L} := \{f \in \mathcal{F}_{\text{qloc}} : \Delta(f) < \infty\}.$$

One of the nice features of the space  $\mathcal{L}$  is that one can perturb  $\lambda$ , as in (3.18),

$$\frac{d\nu_n}{d\lambda}(\omega) := \exp \left[ \sum_{x \in \Lambda(n)} \theta_x f(\omega) + \psi_n(\omega) - |\Lambda(n)| p_n(f|\lambda) \right],$$

and if  $\lim_{n \rightarrow \infty} \nu_n = \nu$  exists, then  $\nu$  is still asymptotically decoupled (lemma 4.3). In particular, if  $\nu \in \mathcal{M}_1^{+, \theta}$ , then the conclusion of theorem 3.4 are also true when the empirical measures  $T_{\Lambda(n)}$  are defined on the probability space  $(\Omega, \mathcal{F}, \nu)$ . One can also define a perturbed local specification  $\Pi^f$  to the local specification  $\mathbb{K}$  associated with  $\lambda$ . If  $\mathbb{K}$  is quasilocal, then the same is true for  $\Pi^f$ . The probability measures, which are compatible with  $\Pi^f$ , correspond to the Gibbs measures in statistical mechanics. All these developments are possible because in that space one has a good control of boundary terms (see below).

**Lemma 4.1.** *If  $A \subset B$ , then  $\delta_A(f) \leq \delta_B(f)$ . For all  $A$  and  $B$ ,  $\delta_{A \cup B}(f) \leq \delta_A(f) + \delta_B(f)$ .  $\Delta(f)$  is a semi-norm on  $\mathcal{L}$ ;  $\Delta(f) = 0$  if and only if  $f$  is constant. If  $f \in \mathcal{L}$ , then for any subset  $\Lambda \subset \mathbb{Z}^d$ ,  $\delta_{\Lambda}(f) \leq \sum_{j \in \Lambda} \delta_j(f)$ . For any  $j \in \mathbb{Z}^d$  and  $\Lambda$ ,*

$$\delta_{\Lambda+j}(\theta_j f) = \delta_{\Lambda}(f).$$

The space  $\mathcal{L}$  is dense in  $\mathcal{F}_{\text{qloc}}$ .

<sup>50</sup> The space  $\mathcal{L}$  is used (in a technical manner) in relation with Dobrushin uniqueness theorem in [L]. Its importance is that it is tightly connected to the space of absolutely summable potentials; on the other hand  $\mathcal{F}_{\text{qloc}}$  is connected to the ‘‘Big Banach space’’ of potentials. See [EFS].

**Proof.**  $\delta_A(f) \leq \delta_B(f)$  and  $\delta_{A \cup B}(f) \leq \delta_A(f) + \delta_B(f)$  are evident. The same is true for  $\delta_\Lambda(f) \leq \sum_{j \in \Lambda} \delta_j(f)$  when  $\Lambda$  is finite. Assume that the cardinality of  $\Lambda$  is infinite.  $\forall \varepsilon > 0$ , quasilocality implies the existence of a finite subset  $\Lambda_\varepsilon$  such that

$$\sup_{\substack{\omega, \eta: \\ \omega_{\Lambda_\varepsilon} = \eta_{\Lambda_\varepsilon}}} |f(\omega) - f(\eta)| \equiv \delta_{\Lambda_\varepsilon}(f) \leq \varepsilon.$$

Hence,

$$|f(\omega) - f(\eta)| \leq \delta_{\Lambda_\varepsilon \cap \Lambda}(f) + \sum_{j \in \Lambda_\varepsilon \cap \Lambda} \delta_j(f) \leq \varepsilon + \sum_{j \in \Lambda} \delta_j(f).$$

Since  $\varepsilon$  is arbitrary, this proves  $\delta_\Lambda(f) \leq \sum_{j \in \Lambda} \delta_j(f)$ . Let  $j \in \mathbb{Z}^d$  and  $\Lambda$ . Then

$$\begin{aligned} \delta_\Lambda(\theta_j f) &= \sup\{|\theta_j f(\omega) - \theta_j f(\eta)| : \omega_i = \eta_i \ \forall i \notin \Lambda\} \\ &= \sup\{|f(\theta_j \omega) - f(\theta_j \eta)| : (\theta_j^{-1} \theta_j \omega)_i = (\theta_j^{-1} \theta_j \eta)_i \ \forall i \notin \Lambda\} \\ &= \sup\{|f(\omega') - f(\eta')| : (\theta_j^{-1} \omega')_i = (\theta_j^{-1} \eta')_i \ \forall i \notin \Lambda\} \\ &= \sup\{|f(\omega') - f(\eta')| : \omega'_i = \eta'_i \ \forall i \notin \Lambda - j\} \\ &= \delta_{\Lambda - j}(f). \end{aligned}$$

□

**4.1. Uniform boundary terms.** The notion of boundary term,

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \sup_{\omega \in \Omega} |\psi_n(\omega)| = 0, \quad (4.1)$$

does not correspond to the notion which occurs in statistical mechanics. A stronger notion is therefore introduced below. Lemma 4.2 gives the basic estimate.

**Definition 4.2.** A sequence  $\psi_n : \Omega_n \rightarrow \mathbb{R}$  defines uniform boundary terms if

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \sup_{\omega \in \Omega} |\psi_n(\omega)| = 0$$

and if there exists a function  $r : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  such that  $\lim_{x \rightarrow \infty} r(x) = 0$  and for all  $n > m$

$$\delta_{\Lambda(m)}(\psi_n) \leq r(n - m) |\Lambda(m)|.$$

**Remark.** The definition of uniform boundary terms is formulated with respect to the sequence of subsets  $\Lambda(n)$  since only this sequence is considered in the paper.

Lemma 4.2 gives the basic method for decoupling. To formulate lemmas 4.2 and 4.3 the following notations are used. Let  $m \in \mathbb{N}$  and  $a \in \mathbb{Z}^d$ . Set

$$\begin{aligned} \Lambda_1(m) &:= \Lambda(m) \\ \Lambda_2(m) &:= \Lambda(m + \bar{g}(m)) \\ \Lambda_3(m) &:= \Lambda(m + \bar{g}(m) + \bar{g}(m + \bar{g}(m))) \\ \Lambda_4(m) &:= \Lambda(m + 2\bar{g}(m) + \bar{g}(m + \bar{g}(m))) \\ \Lambda_4(n, m, a) &:= \Lambda(n) \setminus (\Lambda_4(m) + a). \end{aligned}$$

**Lemma 4.2.** *Let  $f \in \mathcal{L}$ ,  $\psi_n$  be uniform boundary terms,  $\bar{g} : \mathbb{N} \rightarrow \mathbb{N}$ , increasing,  $\lim_n \bar{g}(n) = \infty$  and  $\lim_n \bar{g}(n)/n = 0$ . Let  $m \in \mathbb{N}$ ,  $a \in \mathbb{Z}^d$  and  $\xi \in \Omega$  be given; for any  $\omega \in \Omega$ , let  $\underline{\omega}$  and  $\bar{\omega}$  be defined by*

$$\underline{\omega}_k := \begin{cases} \omega_k & \text{if } k \notin \Lambda_3(m) + a \\ \xi_k & \text{if } k \in \Lambda_3(m) + a \end{cases} \quad \text{and} \quad \bar{\omega}_k := \begin{cases} \omega_k & \text{if } k \in \Lambda_2(m) + a \\ \xi_k & \text{if } k \notin \Lambda_2(m) + a \end{cases}.$$

Set

$$\alpha(m) := |\Lambda_3(m)| \sum_{i \notin \Lambda(\bar{g}(m))} \delta_i(f) + \delta_{\Lambda_3(m)+a}(\psi_n).$$

Then, for  $n$  large enough,

$$\left| \ln \int_{\Omega} \exp \left[ \sum_{i \in \Lambda_1(m)+a} \theta_i f(\bar{\omega}) + \sum_{i \in \Lambda_4(n,m,a)} \theta_i f(\underline{\omega}) + \psi_n(\underline{\omega}) \right] \lambda(d\omega) - |\Lambda(n)| p_n(f|\lambda) \right| \leq \|f\| \cdot |\Lambda_4(m) \setminus \Lambda_1(m)| + 2\alpha(m).$$

**Remark.** For fixed  $m$  and  $a$ ,  $\delta_{\Lambda_3(m)+a}(\psi_n) \rightarrow 0$  when  $n \rightarrow \infty$ .

**Proof.** Let  $j \in \Lambda_3(m) + a$  and  $k \in \Lambda_4(n, m, a)$ ; then  $k - j \notin \Lambda(\bar{g}(m))$ . Hence, by lemma 4.1

$$\sum_{i \in \Lambda_4(n,m,a)} \delta_i(\theta_j f) = \sum_{i \in \Lambda_4(n,m,a) - j} \delta_i(f) \leq \sum_{i \notin \Lambda(\bar{g}(m))} \delta_i(f).$$

Similarly, let  $j \in \Lambda_1 + a$ ; then  $\delta_{(\Lambda_2(m)+a)^c}(\theta_j f) \leq \sum_{i \notin \Lambda(\bar{g}(m))} \delta_i(f)$ . If  $n$  is large enough, then

$$\left| \sum_{i \in \Lambda_4(n,m,a)} \theta_i f(\omega) + \psi_n(\omega) - \sum_{i \in \Lambda_4(n,m,a)} \theta_i f(\underline{\omega}) + \psi_n(\underline{\omega}) \right| \leq \alpha(m) \quad (4.2)$$

and

$$\left| \sum_{i \in \Lambda_1+a} \theta_i f(\omega) - \sum_{i \in \Lambda_1+a} \theta_i f(\bar{\omega}) \right| \leq \alpha(m). \quad (4.3)$$

For  $n$  large enough,

$$\begin{aligned} |\Lambda(n)| p_n(f) &= \ln \int_{\Omega} \exp \left[ \sum_{j \in \Lambda(n)} \theta_j f(\omega) + \psi_n(\omega) \right] \lambda(d\omega) \\ &\geq \ln \int_{\Omega} \exp \left[ \sum_{j \in \Lambda_1+a} \theta_j f(\bar{\omega}) + \sum_{j \in \Lambda_4(n,m,a)} \theta_j f(\underline{\omega}) + \psi_n(\underline{\omega}) \right] \lambda(d\omega) \\ &\quad - |\Lambda_4(m) \setminus \Lambda_1(m)| \cdot \|f\| - 2\alpha(m). \end{aligned}$$

A similar upper bound holds.  $\square$

**Lemma 4.3.** *Let  $\lambda$  be asymptotically decoupled with parameters  $g$  and  $c$ ,  $f \in \mathcal{L}$  and  $\nu_n$  be the probability measure defined by<sup>51</sup>*

$$\frac{d\nu_n}{d\lambda}(\omega) := \exp \left[ \sum_{x \in \Lambda(n)} \theta_x f(\omega) + \psi_n(\omega) - |\Lambda(n)| p_n(f|\lambda) \right],$$

<sup>51</sup>Same convention as in (3.18).

where  $\psi_n$  are uniform boundary terms. Choose  $\bar{g}$  as in lemma 4.2 and so that  $\bar{g} \geq g$ . Suppose that  $\nu = \lim_{n \rightarrow \infty} \nu_n$  exists. Then  $\nu$  is asymptotically decoupled with parameters  $g_\nu(m) := 2\bar{g}(m) + \bar{g}(m + \bar{g}(m))$  and

$$c_\nu(m) \leq 4c(m + \bar{g}(m)) + 8|\Lambda_3(m)| \sum_{j \notin \Lambda(\bar{g}(m))} \delta_j(f) + 4|\Lambda_4(m) \setminus \Lambda_1(m)| \cdot \|f\|.$$

**Proof.**  $\lambda$  is asymptotically decoupled with parameters  $g$  and  $c$ , so that for any  $a \in \mathbb{Z}^d$ , bounded nonnegative functions  $G_1$  and  $G_2$ ,  $G_1 \in \mathcal{F}_{\Lambda_2(m)+a}$  and  $G_2 \in \mathcal{F}_{(\Lambda_3(m)+a)^c}$ ,

$$e^{-c(m+\bar{g}(m))} \int G_1 d\lambda \int G_2 d\lambda \leq \int G_1 G_2 d\lambda \leq e^{c(m+\bar{g}(m))} \int G_1 d\lambda \int G_2 d\lambda. \quad (4.4)$$

Assume that  $A \in \mathcal{F}_{\Lambda_1(m)+a}$  and  $B \in \mathcal{F}_{\Lambda_4(n,m,a)}$ . Set  $V(m) := \Lambda_4(m) \setminus \Lambda_1(m)$ . Let  $\varepsilon > 0$  and choose  $n$  large enough so that

$$\delta_{\Lambda_3(m)+a}(\psi_n) \leq \varepsilon.$$

Set  $\beta(m) := c(m + \bar{g}(m)) + 2\alpha_m + |V(m)| \cdot \|f\|$ ,

$$F_1(\omega) := \exp\left(\sum_{j \in \Lambda_1(m)+a} \theta_j f(\bar{\omega})\right),$$

and

$$F_2(\omega) := \exp\left(\sum_{j \in \Lambda_4(n,m,a)} \theta_j f(\underline{\omega}) + \psi_n(\underline{\omega})\right).$$

Then (4.2) and (4.3) imply

$$\begin{aligned} e^{-\beta(m)} e^{-|\Lambda(n)|p_n(f|\lambda)} \int_A F_1 d\lambda \int_B F_2 d\lambda &\leq \nu_n(A \cdot B) \\ &\leq e^{\beta(m)} e^{-|\Lambda(n)|p_n(f|\lambda)} \int_A F_1 d\lambda \int_B F_2 d\lambda. \end{aligned} \quad (4.5)$$

Lemma 4.2 and (4.4) imply

$$e^{-\beta(m)} \int_\Omega F_1 d\lambda \int_\Omega F_2 d\lambda \leq e^{|\Lambda(n)|p_n(f|\lambda)} \leq e^{\beta(m)} \int_\Omega F_1 d\lambda \int_\Omega F_2 d\lambda. \quad (4.6)$$

Multiplying and dividing (4.5) by  $\int_\Omega F_1 d\lambda \int_\Omega F_2 d\lambda$ , and using (4.6) and (4.4),

$$\begin{aligned} \nu_n(A \cdot B) &\leq e^{\beta(m)} \frac{\int_A F_1 d\lambda \int_\Omega F_2 d\lambda}{e^{|\Lambda(n)|p_n(f|\lambda)}} \frac{\int_\Omega F_1 d\lambda \int_B F_2 d\lambda}{\int_\Omega F_1 d\lambda \int_\Omega F_2 d\lambda} \\ &\leq e^{4\beta(m)} \nu_n(A) \nu_n(B). \end{aligned}$$

Taking now  $n \rightarrow \infty$  for given local events  $A$  and  $B$  one obtains, since  $\varepsilon$  is arbitrary,

$$\nu(A \cdot B) \leq e^{c_\nu(m)} \nu(A) \nu(B).$$

This result implies that the same inequality holds for arbitrary events  $A \in \mathcal{F}_{\Lambda(m)+a}$  and  $B \in \mathcal{F}_{(\Lambda(m)+g_\nu(m)+a)^c}$ . The lower bound is proved in the same manner.  $\square$

4.2. **The local specification  $\Pi^f$ .** Let  $f \in \mathcal{L}$ . Choose a fixed configuration<sup>52</sup>  $\omega^* \in \Omega$ . One defines two quasilocal functions on  $\Omega$ . However, this dependence is not written explicitly. For any finite subset  $\Lambda \subset \mathbb{Z}^d$ ,

$$H_\Lambda(\omega) := \lim_{n \rightarrow \infty} \sum_{t \in \Lambda(n)} (\theta_t f(\omega_\Lambda, \omega_{\Lambda^c}) - \theta_t f(\omega_\Lambda^*, \omega_{\Lambda^c})). \quad (4.7)$$

For any finite subsets  $\Lambda_2 \subset \mathbb{Z}^d$  and  $\Lambda_1 \subset \Lambda_2$ ,

$$W_{\Lambda_1, \Lambda_2}(\omega) := \lim_{n \rightarrow \infty} \sum_{t \in \Lambda(n)} (\theta_t f(\omega_{\Lambda_1}^*, \omega_{\Lambda_1^c}) - \theta_t f(\omega_{\Lambda_2}^*, \omega_{\Lambda_2^c})). \quad (4.8)$$

**Lemma 4.4.** *The limits (4.7) and (4.8) exist as uniform limits over  $\Omega$ .  $H_\Lambda$  and  $W_{\Lambda_1, \Lambda_2}$  are quasilocal and  $W_{\Lambda_1, \Lambda_2}$  is  $\mathcal{F}_{\Lambda_1^c}$ -measurable.*

$$\|H_\Lambda\| \leq \Delta(f) |\Lambda| \quad \text{and} \quad H_{\Lambda_2} = H_{\Lambda_1} + W_{\Lambda_1, \Lambda_2}, \quad \Lambda_1 \subset \Lambda_2.$$

**Proof.** For each  $t \in \mathbb{Z}^d$ ,  $\forall \omega \in \Omega$ ,

$$|\theta_t f(\omega_\Lambda, \omega_{\Lambda^c}) - \theta_t f(\omega_\Lambda^*, \omega_{\Lambda^c})| \leq \delta_\Lambda(\theta_t f) = \delta_{\Lambda-t}(f) \leq \sum_{s \in \Lambda} \delta_{s-t}(f).$$

Hence

$$\sum_{t \in \mathbb{Z}^d} |\theta_t f(\omega_\Lambda, \omega_{\Lambda^c}) - \theta_t f(\omega_\Lambda^*, \omega_{\Lambda^c})| \leq |\Lambda| \Delta(f).$$

The proof of the existence of the limit (4.8) is similar.  $\square$

**Definition 4.3.** *Let  $\mathbb{K}$  be the local specification associated with  $\lambda$  (defined at the beginning of subsection 3.2) and let  $f \in \mathcal{L}$ . Then, for all finite  $\Lambda \subset \mathbb{Z}^d$ , define  $\Pi_\Lambda^f : \mathcal{F} \times \Omega \rightarrow [0, 1]$  by*

$$\Pi_\Lambda^f(F | \omega) := \frac{\mathbb{K}_\Lambda^\omega(I_F \exp H_\Lambda)}{\mathbb{K}_\Lambda^\omega(\exp H_\Lambda)}.$$

$\Pi_\Lambda^f = \{\Pi_\Lambda^f\}$  is a local specification. Proof of  $\Pi_{\Lambda_2}^f = \Pi_{\Lambda_2}^f \Pi_{\Lambda_1}^f$ .

$$\begin{aligned} \mathbb{K}_{\Lambda_2}^\omega(I_F \exp H_{\Lambda_2}) &= \mathbb{K}_{\Lambda_2}^\omega(I_F \exp H_{\Lambda_1} \exp W_{\Lambda_1, \Lambda_2}) \\ &= \mathbb{K}_{\Lambda_2}^\omega(\mathbb{K}_{\Lambda_1}^\omega(I_F \exp H_{\Lambda_1}) \exp W_{\Lambda_1, \Lambda_2}) \\ &= \mathbb{K}_{\Lambda_2}^\omega(\mathbb{K}_{\Lambda_1}^\omega(\exp H_{\Lambda_1}) \Pi_{\Lambda_1}^f(I_F | \cdot) \exp W_{\Lambda_1, \Lambda_2}) \\ &= \mathbb{K}_{\Lambda_2}^\omega(\exp H_{\Lambda_1} \Pi_{\Lambda_1}^f(I_F | \cdot) \exp W_{\Lambda_1, \Lambda_2}) \\ &= \mathbb{K}_{\Lambda_2}^\omega(\exp H_{\Lambda_2} \Pi_{\Lambda_1}^f(I_F | \cdot)) \\ &= \mathbb{K}_{\Lambda_2}^\omega(\exp H_{\Lambda_2}) \Pi_{\Lambda_2}^f(\Pi_{\Lambda_1}^f(I_F | \cdot) | \omega). \end{aligned}$$

**Lemma 4.5.** *Let  $\lambda$  be asymptotically decoupled,  $f \in \mathcal{L}$  and  $\nu_n$  be the probability measure defined by<sup>53</sup>*

$$\frac{d\nu_n}{d\lambda}(\omega) := \exp \left[ \sum_{x \in \Lambda(n)} \theta_x f(\omega) + \psi_n(\omega) - |\Lambda(n)| p_n(f | \lambda) \right],$$

<sup>52</sup> This configuration is introduced for convenience. The local specification is independent of  $\omega^*$ . This configuration serves as reference configuration; by definition  $H_\Lambda(\omega^*) = 0$ .

<sup>53</sup> Same convention as in (3.18).



where  $\psi_n$  are uniform boundary terms. Suppose that  $\nu = \lim_{n \rightarrow \infty} \nu_n$  exists. If  $\forall g \in \mathcal{F}_{\text{qloc}}, \forall \Lambda \omega \mapsto \Pi_\Lambda^f(g | \omega)$  is quasilocal, then  $\nu$  is  $\Pi^f$ -compatible.

**Proof.** It is sufficient to prove

$$\mathbb{E}_\nu (gh) = \mathbb{E}_\nu (\Pi_\Lambda^f(g | \cdot) h)$$

for all positive local functions  $g$  and  $h$  which are respectively  $\mathcal{F}_\Lambda$ -measurable and  $\mathcal{F}_{\Lambda^c}$ -measurable. Since  $\omega \mapsto \Pi_\Lambda^f(g | \omega)$  is quasilocal,

$$\mathbb{E}_\nu (\Pi_\Lambda^f(g | \cdot) h) = \lim_{n \rightarrow \infty} \mathbb{E}_{\nu_n} (\Pi_\Lambda^f(g | \cdot) h).$$

One has

$$\begin{aligned} \mathbb{E}_{\nu_n} (\Pi_\Lambda^f(g | \cdot) h) &= \mathbb{E}_{\nu_n} \left( \frac{\mathbb{K}_\Lambda^\cdot(e^{H_\Lambda} g)}{\mathbb{K}_\Lambda^\cdot(e^{H_\Lambda})} h \right) \\ &= e^{-|\Lambda(n)|p_n(f)} \mathbb{E}_\lambda \left( \mathbb{K}_\Lambda^\cdot(e^{\sum_{i \in \Lambda(n)} \theta_i f + \psi_n}) \frac{\mathbb{K}_\Lambda^\cdot(e^{H_\Lambda} g)}{\mathbb{K}_\Lambda^\cdot(e^{H_\Lambda})} h \right). \end{aligned}$$

For all  $t \in \mathbb{Z}^d$ , define  $f_t^*$  by  $\omega \mapsto f_t^*(\omega) := (\theta_t f)(\omega_\Lambda^*, \omega_{\Lambda^c})$ ;  $f_t^*$  is  $\mathcal{F}_{\Lambda^c}$ -measurable. Similarly, define  $\psi^*$  by  $\omega \mapsto \psi_n^*(\omega) := \psi_n(\omega_\Lambda^*, \omega_{\Lambda^c})$ . Uniformly on  $\Omega$  (see proof of lemma 4.4),

$$\lim_{n \rightarrow \infty} \frac{\mathbb{K}_\Lambda^\cdot(e^{\sum_{i \in \Lambda(n)} (\theta_i f - f_t^*) + \psi_n - \psi_n^*})}{\mathbb{K}_\Lambda^\cdot(e^{H_\Lambda})} = 1. \quad (4.9)$$

Since  $\sum_{i \in \Lambda(n)} f_i^* + \psi_n^*$  is  $\mathcal{F}_{\Lambda^c}$ -measurable,

$$e^{\sum_{i \in \Lambda(n)} f_i^* + \psi_n^*} \mathbb{K}_\Lambda^\cdot(e^{H_\Lambda} g) = \mathbb{K}_\Lambda^\cdot(e^{\sum_{i \in \Lambda(n)} \theta_i f + \psi_n} e^{\psi_n^* - \psi_n + \sum_{t \notin \Lambda(n)} (\theta_t f - f_t^*)} g).$$

Uniformly on  $\Omega$  (see proof of lemma 4.4),

$$\lim_{n \rightarrow \infty} e^{\psi_n^* - \psi_n + \sum_{t \notin \Lambda(n)} (\theta_t f - f_t^*)} = 1. \quad (4.10)$$

Using (4.9), (4.10) and the  $\mathcal{F}_{\Lambda^c}$ -measurability of  $h$  one obtains

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}_{\nu_n} (\Pi_\Lambda^f(g | \cdot) h) &= \lim_{n \rightarrow \infty} \mathbb{E}_\lambda (\mathbb{K}_\Lambda^\cdot(e^{\sum_{i \in \Lambda(n)} \theta_i f + \psi_n - |\Lambda(n)|p_n(f)}) g) h) \\ &= \lim_{n \rightarrow \infty} \mathbb{E}_\lambda (e^{\sum_{i \in \Lambda(n)} \theta_i f + \psi_n - |\Lambda(n)|p_n(f)} g h) \\ &= \lim_{n \rightarrow \infty} \mathbb{E}_{\nu_n} (gh) = \mathbb{E}_\nu (gh). \end{aligned}$$

□

**Remark.** If for any finite  $\Lambda \subset \mathbb{Z}^d$  and any  $g \in \mathcal{F}_{\text{qloc}}$   $\omega \mapsto \mathbb{K}_\Lambda^\omega(g)$  is quasilocal, then the same is true for  $\Pi_\Lambda^f$  since  $H_\Lambda$  is quasilocal.

Lemmas 4.6 and 4.7 are two technical lemmas, which are needed for lemma 4.8, and in the next subsection for computing the information gain of a translation invariant measure with respect to a measure, which is  $\Pi^f$ -compatible.

**Definition 4.4.** Let  $A \subset \mathbb{Z}^d$  be a finite subset and  $\Lambda \subset \mathbb{Z}^d$ . The  $A$ -boundary of the set  $\Lambda$ ,  $\partial_A \Lambda$ , is the subset of  $\mathbb{Z}^d$

$$\partial_A \Lambda := \{j \in \Lambda : A + j \not\subset \Lambda\}.$$

The following properties are obvious from the definition,

$$\frac{|\partial_A \Lambda|}{|\Lambda|} \leq 1 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{|\partial_A \Lambda(n)|}{|\Lambda(n)|} = 0.$$

**Lemma 4.6.** *Let  $f \in \mathcal{L}$ . Then,  $\forall \varepsilon > 0$  there exists  $n_\varepsilon$  such that for any  $n \geq n_\varepsilon$ ,*

$$\sum_{t \notin \Lambda(n)} \delta_{\Lambda(n)}(\theta_t f) \leq \varepsilon |\Lambda(n)|.$$

**Proof.** Let  $\delta > 0$ ; then there exists  $n_\delta$  such that

$$\sum_{t \notin \Lambda(n_\delta)} \delta_t(f) \leq \delta.$$

One has

$$\sum_{t \notin \Lambda(n)} \delta_{\Lambda(n)}(\theta_t f) \leq \sum_{s \in \Lambda(n)} \sum_{t \notin \Lambda(n)} \delta_s(\theta_t f) \leq \sum_{s \in \Lambda(n)} \sum_{t \notin \Lambda(n)} \delta_{s-t}(f) \leq \sum_{s \in \Lambda(n)} \sum_{t \notin (\Lambda(n)-s)} \delta_t(f).$$

If  $(\Lambda(n) - s)^c \cap \Lambda(n_\delta) = \emptyset$ , then

$$\sum_{t \notin (\Lambda(n)-s)} \delta_t(f) \leq \sum_{t \notin \Lambda(n_\delta)} \delta_t(f) \leq \delta.$$

$(\Lambda(n) - s)^c \cap \Lambda(n_\delta) \neq \emptyset$  if and only if  $(\Lambda(n)^c - s) \cap \Lambda(n_\delta) \neq \emptyset$ . That is, there exist  $x \notin \Lambda(n)$  and  $y \in \Lambda(n_\delta)$  such that  $x - s = y$ . Hence

$$\{s \in \Lambda(n) : (\Lambda(n) - s)^c \cap \Lambda(n_\delta) \neq \emptyset\} = \{s \in \Lambda(n) : (s + \Lambda(n_\delta)) \not\subset \Lambda(n)\}.$$

In such a case one uses the trivial upper bound

$$\sum_{t \notin (\Lambda(n)-s)} \delta_t(f) \leq \Delta(f).$$

Therefore

$$\sum_{t \notin \Lambda(n)} \delta_{\Lambda(n)}(\theta_t f) \leq |\Lambda(n)| \left( \Delta(f) \frac{|\partial_{\Lambda(n_\delta)} \Lambda(n)|}{|\Lambda(n)|} + \delta \right).$$

Choose  $\delta := \varepsilon/2$  and  $n \geq n_\delta$  large enough, so that

$$\Delta(f) \frac{|\partial_{\Lambda(n_\delta)} \Lambda(n)|}{|\Lambda(n)|} \leq \varepsilon/2.$$

□

**Lemma 4.7.** *Let  $f \in \mathcal{L}$  and  $g : \mathbb{N} \rightarrow \mathbb{N}$  such that  $\lim_{n \rightarrow \infty} g(n)/n = 0$ . Then  $\forall \varepsilon > 0$  there exists  $n_\varepsilon$  such that for any  $n \geq n_\varepsilon$  there exists a bounded  $\mathcal{F}_{\Lambda(n)}$ -measurable function  $F_n$  verifying*

$$\|F_n - \sum_{t \in \Lambda(n+g(n))} \theta_t f\| \leq \varepsilon |\Lambda(n)|.$$

**Proof.** Let  $\delta > 0$ . There exists  $m_\delta$  such that  $\delta_{\Lambda(m_\delta)^c}(f) \leq \delta$ . Let  $\omega^*$  be a fixed element of  $\Omega$ , e.g. the one used in the definition of  $H_\Lambda$  (see (4.7)). Set

$$f_\delta(\omega) := f(\omega_{\Lambda(m_\delta)}, \omega_{\Lambda(m_\delta)^c}^*).$$

For any  $n$  such that  $n > m_\delta$  set

$$F_n := \sum_{t \in \Lambda(n-m_\delta)} \theta_t f_\delta.$$

Then

$$\begin{aligned} \|F_n - \sum_{t \in \Lambda(n+g(n))} \theta_t f\| &\leq \sum_{t \in \Lambda(n-m_\delta)} \|\theta_t f_\delta - \theta_t f\| + \sum_{t \in \Lambda(n+g(n)) \setminus \Lambda(n-m_\delta)} \|\theta_t f\| \\ &\leq \delta |\Lambda(n)| + \|f\| |\Lambda(n+g(n)) \setminus \Lambda(n-m_\delta)|. \end{aligned}$$

□

**Lemma 4.8.** *Let  $f \in \mathcal{L}$ . Let  $\lambda$  be strongly asymptotically decoupled. Then*

$$\frac{e^{H_{\Lambda(n)}(\omega)}}{\mathbb{K}_{\Lambda(n)}^\omega(e^{H_{\Lambda(n)}})} = \exp\left(\sum_{t \in \Lambda(n)} \theta_t f(\omega) + \psi_n(\omega) - |\Lambda(n)| p_n(f)\right),$$

where  $\psi_n$  are uniform boundary terms<sup>54</sup>.

**Proof.** For all  $t \in \mathbb{Z}^d$  define  $f_t^*$  by  $\omega \mapsto f_t^*(\omega) := (\theta_t f)(\omega_{\Lambda(n)}^*, \omega_{\Lambda(n)^c})$ . Set

$$\Phi_n(\omega) := \sum_{t \in \Lambda(n)} (\theta_t f(\omega) - f_t^*(\omega)).$$

Since  $\forall t \in \mathbb{Z}^d$   $f_t^*$  is  $\mathcal{F}_{\Lambda(n)^c}$ -measurable,

$$\frac{e^{H_{\Lambda(n)}(\omega)}}{\mathbb{K}_{\Lambda(n)}^\omega(e^{H_{\Lambda(n)}})} = \frac{e^{\sum_{t \in \Lambda(n)} \theta_t f(\omega) + \Phi_n(\omega)}}{\mathbb{K}_{\Lambda(n)}^\omega(e^{\sum_{t \in \Lambda(n)} \theta_t f + \Phi_n})}.$$

$\Phi_n$  are uniform boundary terms. Indeed, lemma 4.6 gives (4.1). Let  $n > m$ ; then

$$\delta_{\Lambda(m)}(\Phi_n) \leq \sum_{t \notin \Lambda(n)} \delta_{\Lambda(m)}(\theta_t f).$$

Set

$$r(k) := \sum_{j \notin \Lambda(k)} \delta_j(f).$$

Since  $f \in \mathcal{L}$ ,  $\lim_{k \rightarrow \infty} r(k) = 0$ . Moreover, if  $t \notin \Lambda(n)$  and  $s \in \Lambda(m)$ , then  $s + t \notin \Lambda(n-m)$ . Hence

$$\sum_{t \notin \Lambda(n)} \delta_{\Lambda(m)}(\theta_t f) \leq |\Lambda(m)| \sum_{t \notin \Lambda(n-m)} \delta_t(f) = |\Lambda(m)| r(n-m).$$

This proves that  $\Phi_n$  are uniform boundary terms. Consider now

$$\exp(\Phi'_n(\omega)) := \frac{\mathbb{K}_{\Lambda(n)}^\omega(e^{\sum_{t \in \Lambda(n)} f + \Phi_n})}{\int e^{\sum_{t \in \Lambda(n)} \theta_t f(\eta)} \lambda(d\eta)}.$$

Set  $\psi_n := \Phi_n - \Phi'_n$ .  $\Phi'_n$  are uniform boundary terms. Indeed, since  $\Phi'_n$  is  $\mathcal{F}_{\Lambda(n)^c}$ -measurable one must only verify (4.1). Let  $\varepsilon > 0$ . Applying lemma 4.7 (in a slightly different version), there exists a function  $F'_n$ ,  $\mathcal{F}_{\Lambda(n-g(n))}$ -measurable, such that for  $n$  large enough

$$\|F'_n - \sum_{t \in \Lambda(n)} \theta_t f\| \leq \varepsilon |\Lambda(n)|.$$

<sup>54</sup> Here the convention of (3.18) does not hold!

Therefore,

$$e^{-\varepsilon|\Lambda(n)|} \int e^{F'_n(\eta)} \lambda(d\eta) \leq \int e^{\sum_{t \in \Lambda(n)} \theta_t f(\eta)} \lambda(d\eta) \leq e^{\varepsilon|\Lambda(n)|} \int e^{F'_n(\eta)} \lambda(d\eta).$$

By lemma 4.6, for  $n$  large enough,

$$\|\Phi_n\| \leq \varepsilon |\Lambda(n)|.$$

Since  $\lambda$  is strongly asymptotically decoupled

$$\mathbb{K}_{\Lambda(n)}^\omega(e^{\sum_{t \in \Lambda(n)} \theta_t f + \Phi_n}) \leq e^{2\varepsilon|\Lambda(n)|} \mathbb{K}_{\Lambda(n)}^\omega(e^{F'_n} | \omega) \leq e^{2\varepsilon|\Lambda(n)|} e^{c(n)} \int e^{F'_n(\eta)} \lambda(d\eta).$$

A similar lower bound holds. This proves the lemma.  $\square$

Assuming that  $\lambda$  is strongly asymptotically decoupled one can prove that any probability measure, which is  $\Pi^f$ -compatible, is also strongly asymptotically decoupled. The proof of this result is similar to the proof of the lemma 4.3.

**Lemma 4.9.** *Let  $f \in \mathcal{L}$ . Let  $\lambda$  be strongly asymptotically decoupled. Then any probability measure  $\mu$ , which is  $\Pi^f$ -compatible, is strongly asymptotically decoupled.*

**Proof.**  $\lambda$  is strongly asymptotically decoupled with parameters  $g$  and  $c$ . Let  $\bar{g} : \mathbb{N} \rightarrow \mathbb{N}$  be increasing,  $\lim_{n \rightarrow \infty} \bar{g}(n) = \infty$ ,  $\lim_{n \rightarrow \infty} \bar{g}(n)/n = 0$  and  $g(n) \leq \bar{g}(n)$  for all  $n \in \mathbb{N}$ . From (3.3) (used twice), for any  $\omega$  and  $\eta$ , and any  $A \in \mathcal{F}_{\Lambda(m)}$ ,

$$e^{-2c(m)} \mathbb{K}_{\Lambda(n+g(m))+a}^\eta(I_A) \leq \mathbb{K}_{\Lambda(m+g(m))+a}^\omega(I_A) \leq e^{2c(n)} \mathbb{K}_{\Lambda(m+g(m))+a}^\eta(I_A).$$

One proves the result for  $a = 0$ ,  $n \in \mathbb{N}$  and  $A \in \mathcal{F}_{\Lambda(n)}$ . Let  $\bar{n} := n + \bar{g}(n)$ . Define

$$\begin{aligned} \Lambda_1 &:= \Lambda(n) \\ \Lambda_2 &:= \Lambda(\bar{n}) \\ \Lambda_3 &:= \Lambda(\bar{n} + g(\bar{n})). \end{aligned}$$

Let  $\xi \in \Omega$  be given. Define for any  $\omega$

$$\bar{\omega}_k := \begin{cases} \omega_k & \text{if } k \in \Lambda_2 \\ \xi_k & \text{otherwise.} \end{cases}$$

For any  $t \in \mathbb{Z}^d$ ,  $\omega \mapsto f_t(\omega) := (\theta_t f)(\bar{\omega})$  is  $\mathcal{F}_{\Lambda(\bar{n})}$ -measurable. Hence  $F_1$  defined by

$$\omega \mapsto F_1(\omega) := \sum_{t \in \Lambda_1} f_t(\omega)$$

is  $\mathcal{F}_{\Lambda(\bar{n})}$ -measurable. Since  $\lambda$  is strongly asymptotically decoupled  $\Pi_{\Lambda_3}^f$  has the representation of lemma 4.8. In particular there exists  $a(n) \geq 0$ ,  $\lim_{n \rightarrow \infty} a(n) = 0$ , such that

$$\|\psi_{\bar{n}+g(\bar{n})}\| \leq a(n) |\Lambda(n)|.$$

Similarly, there exists  $b(n) \geq 0$ ,  $\lim_{n \rightarrow \infty} b(n) = 0$ , such that

$$\|f\| |\Lambda_3 \setminus \Lambda_1| \leq b(n) |\Lambda(n)|.$$

Furthermore,

$$|F_1(\omega) - \sum_{t \in \Lambda_1} \theta_t f(\omega)| \leq \delta_{(\Lambda(\bar{g}(n)))^c}(f) |\Lambda(n)|.$$

Hence

$$\left| \sum_{t \in \Lambda_3} \theta_t f(\omega) + \psi_{\bar{n}+g(\bar{n})}(\omega) - F_1(\omega) \right| \leq (\delta_{(\Lambda(\bar{g}(n)))^c}(f) + a(n) + b(n)) |\Lambda(n)| \equiv d(n) |\Lambda(n)|.$$

For any  $\omega$  and  $\eta$

$$\begin{aligned} \Pi_{\Lambda_3}^f(A|\omega) &\leq e^{-|\Lambda_3|p_{\bar{n}+g(\bar{n})}(f)} e^{d(n)|\Lambda(n)|} \mathbb{K}_{\Lambda_3}^\omega(I_A e^{F_1}) \\ &\leq e^{-|\Lambda_3|p_{\bar{n}+g(\bar{n})}(f)} e^{d(n)|\Lambda(n)|+2c(\bar{n})} \mathbb{K}_{\Lambda_3}^\eta(I_A e^{F_1}) \\ &\leq e^{2d(n)|\Lambda(n)|+2c(\bar{n})} \Pi_{\Lambda_3}^f(A|\eta). \end{aligned}$$

Integrating with respect to  $\mu(d\omega)$  one obtains

$$\mu(A) \leq e^{2d(n)|\Lambda(n)|+2c(\bar{n})} \Pi_{\Lambda_3}^f(A|\eta).$$

Similarly one shows that

$$\mu(A) \geq e^{2d(n)|\Lambda(n)|+2c(\bar{n})} \Pi_{\Lambda_3}^f(A|\eta).$$

Hence  $\mu$  is strongly asymptotically decoupled with parameters

$$g_f(n) := \bar{g}(n) + g(\bar{n}) \quad \text{and} \quad c_f(n) := 2d(n)|\Lambda(n)| + 2c(\bar{n}).$$

□

**4.3. Change of reference measure in the large deviations principle.** This subsection is devoted to the computation of the specific information gain of a translation invariant probability measure with respect to a probability measure, which is  $\Pi^f$ -compatible. It is required that  $f \in \mathcal{L}$  and that the basic reference measure  $\lambda$  is strongly asymptotically decoupled. The results are summarized in theorem 4.1. Then, as a corollary, one obtains theorem 4.2, which can be considered as an extension of Varadhan's theorem about the change of reference measure in the large deviations principle.

**Lemma 4.10.** *Let  $f \in \mathcal{L}$ . Let  $\mu$  be a  $\Pi^f$ -compatible probability measure and  $\lambda$  be strongly asymptotically decoupled. Then  $\forall \varepsilon > 0$  there exists  $n_\varepsilon$  such that for any  $n \geq n_\varepsilon$  and any  $A \in \mathcal{F}_{\Lambda(n)}$ ,*

$$e^{-2c_n(\lambda)} e^{-\varepsilon|\Lambda(n)|} \frac{\langle I_A e^{F_n}, \lambda \rangle}{\langle e^{F_n}, \lambda \rangle} \leq \langle I_A, \mu \rangle \leq e^{2c_n(\lambda)} e^{\varepsilon|\Lambda(n)|} \frac{\langle I_A e^{F_n}, \lambda \rangle}{\langle e^{F_n}, \lambda \rangle}. \quad (4.11)$$

$F_n$  is a bounded  $\mathcal{F}_{\Lambda(n)}$ -measurable function<sup>55</sup>. The restrictions of  $\mu$  and  $\lambda$  to  $\mathcal{F}_{\Lambda(n)}$  are mutually absolutely continuous.

**Proof.** Since  $\mu$  is  $\Pi^f$ -compatible one has

$$\langle I_A, \mu \rangle = \int_{\Omega} \Pi_{\Lambda(n+g(n))}^f(I_A|\omega) \mu(d\omega).$$

To prove (4.11) one proves (4.11) with  $\mu$  replaced by  $\Pi_{\Lambda(n+g(n))}^f(\cdot|\omega)$ , uniformly in  $\omega$ .

Let  $\varepsilon > 0$ .  $H_\Lambda$  is defined by (4.7). Therefore, by lemma 4.6, if  $n$  is large enough, then

$$\|H_{\Lambda(n+g(n))} - \sum_{t \in \Lambda(n+g(n))} (\theta_t f - f_t^*)\| \leq \frac{\varepsilon}{4} |\Lambda(n)|,$$

<sup>55</sup> See lemma 4.7

where here  $\omega \mapsto f_t^*(\omega) := (\theta_t f)(\omega_{\Lambda(n+g(n))}^*, \omega_{\Lambda(n+g(n))})$ .  $\forall t$   $f_t^*$  is  $\mathcal{F}_{\Lambda(n+g(n))}$ -measurable.

If in the definition of  $\Pi_{\Lambda(n+g(n))}^f$  one replaces  $H_{\Lambda(n+g(n))}$  by

$$\sum_{t \in \Lambda(n+g(n))} (\theta_t f - f_t^*),$$

then one can cancel the function  $\sum_{t \in \Lambda(n+g(n))} f_t^*$  in the denominator and numerator. Applying lemma 4.7 with  $\varepsilon/4$  one obtains, provided that  $n$  is large enough,

$$e^{-\varepsilon|\Lambda(n)|} \frac{\mathbb{K}_{\Lambda(n+g(n))}^\omega(I_A e^{F_n})}{\mathbb{K}_{\Lambda(n+g(n))}^\omega(e^{F_n})} \leq \Pi_{\Lambda(n+g(n))}^f(I_A | \omega) \leq e^{\varepsilon|\Lambda(n)|} \frac{\mathbb{K}_{\Lambda(n+g(n))}^\omega(I_A e^{F_n})}{\mathbb{K}_{\Lambda(n+g(n))}^\omega(e^{F_n})}.$$

Since  $\lambda$  is strongly asymptotically decoupled

$$e^{-c(n)} \langle I_A e^{F_n}, \lambda \rangle \leq \mathbb{K}_{\Lambda(n+g(n))}^\omega(I_A e^{F_n}) \leq e^{c(n)} \langle I_A e^{F_n}, \lambda \rangle$$

and

$$e^{-c_n(\lambda)} \langle e^{F_n}, \lambda \rangle \leq \mathbb{K}_{\Lambda(n+g(n))}^\omega(e^{F_n}) \leq e^{c_n(\lambda)} \langle e^{F_n}, \lambda \rangle.$$

Putting these results together one obtains (4.11) with  $\Pi_{\Lambda(n+g(n))}^f(\cdot | \omega)$  instead of  $\mu$ . Integrating with respect to  $\mu$  gives the desired result.  $\square$

**Lemma 4.11.** *Let  $f \in \mathcal{L}$ . Let  $\mu$  be a  $\Pi^f$ -compatible probability measure and  $\lambda$  be strongly asymptotically decoupled. Then for any translation invariant probability measure  $\nu$*

$$h(\nu | \mu) = \lim_{m \rightarrow \infty} \frac{1}{|\Lambda(m)|} \mathcal{H}_m(\nu | \mu) = h(\nu | \lambda) - \langle f, \nu \rangle + p(f).$$

**Proof.** Let  $\varepsilon > 0$  and suppose that  $n$  is large enough so that lemmas 4.7 and 4.10 apply. If  $\forall n$  the restriction of  $\nu$  on  $\mathcal{F}_{\Lambda(n)}$  is absolutely continuous with respect to  $\mu$ , then by lemma 4.10 the same is true for  $\lambda$  in place of  $\mu$ . Let us consider this case, since otherwise the result is evident. Let  $A \in \mathcal{F}_{\Lambda(n)}$ . Then

$$\mu(A) := \int_{\Omega} I_A(\omega) g_n(\omega) \lambda(d\omega)$$

with  $g_n$   $\mathcal{F}_{\Lambda(n)}$ -measurable, and by lemma 4.10

$$e^{-2c_n(\lambda)} e^{-\varepsilon|\Lambda(n)|} \mu_n(A) \leq \mu(A) \leq e^{2c_n(\lambda)} e^{\varepsilon|\Lambda(n)|} \mu_n(A),$$

where

$$\mu_n(A) := \frac{\langle I_A e^{F_n}, \lambda \rangle}{\langle e^{F_n}, \lambda \rangle} \equiv \int_{\Omega} I_A(\omega) k_n(\omega) \lambda(d\omega)$$

and  $k_n$   $\mathcal{F}_{\Lambda(n)}$ -measurable. Therefore,  $\lambda$ -a.s.,

$$e^{-2c_n(\lambda)} e^{-\varepsilon|\Lambda(n)|} k_n(\omega) \leq g_n(\omega) \leq e^{2c_n(\lambda)} e^{\varepsilon|\Lambda(n)|} k_n(\omega).$$

One has

$$\begin{aligned} \mathcal{H}_n(\nu | \lambda) - \langle \ln k_n, \nu \rangle - 2c_n(\lambda) - \varepsilon|\Lambda(n)| &\leq \mathcal{H}_n(\nu | \mu) \leq \\ \mathcal{H}_n(\nu | \lambda) - \langle \ln k_n, \nu \rangle + 2c_n(\lambda) + \varepsilon|\Lambda(n)| &. \end{aligned}$$

By lemma 4.7 and translation invariance of  $\nu$

$$\begin{aligned} - \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \langle \ln k_n, \nu \rangle &= \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \langle e^{F_n}, \lambda \rangle - \frac{1}{|\Lambda(n)|} \int_{\Omega} F_n(\omega) \nu(d\omega) \\ &= p(f) - \langle f, \nu \rangle. \end{aligned}$$

Since  $\varepsilon > 0$  is arbitrary the lemma is proved.  $\square$

The above results about the specific information gain are summarized in theorem 4.1.

**Theorem 4.1.** *Let  $f \in \mathcal{L}$  and  $\lambda$  be strongly asymptotically decoupled and translation invariant. Assume that  $\mu \in \mathcal{M}_1^+$  is  $\Pi^f$ -compatible and  $\nu \in \mathcal{M}_1^{+, \theta}$ .*

1.  $h(\nu|\mu)$  exists for any  $\nu \in \mathcal{M}_1^{+, \theta}$  and  $h(\nu|\mu) = h(\nu|\lambda) - \langle f, \nu \rangle + p(f)$ .
2. If  $\nu = \mu$ , then  $h(\mu|\lambda) = \langle f, \mu \rangle - p(f)$ , that is,  $\mu$  is a dual point to  $f$ .
3. If  $\nu$  is  $\Pi^f$ -compatible, then  $h(\nu|\mu) = 0$ .
4. If the local specification<sup>56</sup>  $\mathbb{K}$  is quasilocal, then  $h(\nu|\mu) = 0$  implies that  $\nu$  is  $\Pi^f$ -compatible.

**Proof.** 1. follows from lemma 4.11. If  $\nu = \mu$ , then  $h(\nu|\mu) = 0$ , hence 2. follows from 1. If  $\nu$  is  $\Pi^f$ -compatible, then  $h(\nu|\lambda) = \langle f, \nu \rangle - p(f)$ , hence  $h(\nu|\mu) = 0$ . If  $\mathbb{K}$  is quasilocal, then by the remark after lemma 4.5,  $\Pi^f$  is also quasilocal. Then 4. follows from theorem 3.1 applied to  $\Pi^f$  and  $\mu$ .  $\square$

The next theorem is about the change of the reference measure in the large deviations principle.

**Theorem 4.2.** *Let  $f \in \mathcal{L}$  and  $\lambda$  be strongly asymptotically decoupled and translation invariant. Assume that the local specification  $\mathbb{K}$  is quasilocal. If  $\mu$  is a dual point to  $f$ , then the empirical measures defined on the probability space  $(\Omega, \mathcal{F}, \mu)$  verify a large deviations principle with rate-function*

$$s^f(\nu|\lambda) := \begin{cases} -h(\nu|\lambda) + \langle f, \nu \rangle - p(f) & \text{if } \nu \in \mathcal{M}_1^{+, \theta} \\ -\infty & \text{if } \nu \in \mathcal{M}_1^+ \setminus \mathcal{M}_1^{+, \theta}. \end{cases}$$

**Proof.** By hypothesis,  $\mu \in \mathcal{M}_1^{+, \theta}$  and  $h(\mu|\lambda) = \langle f, \mu \rangle - p(f)$ . Let  $\nu$  be  $\Pi^f$ -compatible. Then

$$h(\mu|\nu) = h(\mu|\lambda) - \langle f, \mu \rangle + p(f) = 0.$$

Hence  $\mu$  is  $\Pi^f$ -compatible. But, lemma 4.9 implies that  $\mu$  is asymptotically decoupled, so that theorem 3.3 is true with  $\mu$  instead of  $\lambda$ . The rate-function is computed using again theorem 4.1.  $\square$

---

<sup>56</sup> Defined at the beginning of subsection 3.2.

## 5. CONDITIONAL LIMIT THEOREMS

Conditional limit theorems are obtained as consequences of two general results: proposition 5.1 and corollary 5.1, which follow from the upper bound of the large deviations principle, and proposition 5.2, which is a general result of convex analysis. Proposition 5.1 is a concentration result for the laws of the empirical measures, which is formulated in terms of a variational problem for the rate-function  $s$ . Proposition 5.2 is a statement about the subdifferentiability of  $s$  at the solutions of this variational problem. Existence of the solutions of this variational problem is a consequence of the large deviations principle. It is useful to treat these two results in some generality. The conditional limit theorems are proved and analyzed in subsection 5.3.

**5.1. A concentration result.** The following setting is assumed in the whole subsection 5.1.  $Y_n$  is a sequence of random variables defined on probability spaces  $(\Omega_n, \mathcal{F}_n, \lambda_n)$  with values in the measurable space  $(X, \mathcal{B})$ .  $X$  is a Hausdorff topological space and  $\mathcal{B}$  a  $\sigma$ -algebra on  $X$ , such that each point of  $X$  has a local base of measurable neighbourhoods of  $\mathcal{B}$ . The law of  $Y_n$  is denoted by  $\mathbb{M}_n := \lambda_n \circ Y_n^{-1}$ .

**Definition 5.1.** *The sequence  $Y_n$ , resp.  $\mathbb{M}_n$ , is eventually concentrated on  $A \subset X$  if and only if for each measurable neighbourhood  $G$  of  $A$*

$$\lim_{n \rightarrow \infty} \mathbb{M}_n[X \setminus G] = 0.$$

*The sequence is eventually concentrated on  $A$  at an exponential rate on the scale  $V_n$  if and only if for each measurable neighbourhood  $G$  of  $A$*

$$\limsup_{n \rightarrow \infty} \frac{1}{V_n} \ln \mathbb{M}_n[X \setminus G] < 0.$$

Proposition 5.1 is an improved version of theorems 2.2 and 3.1 in [LPS3] (see [LPS4]).

**Proposition 5.1.** *The setting above is assumed. Let  $t : X \rightarrow [-\infty, 0]$  be a u.s.c. function with compact level-sets, such that*

$$\inf_{\substack{G \ni x \\ \text{open}}} \limsup_{n \rightarrow \infty} \frac{1}{V_n} \ln \lambda_n[Y_n \in G] \leq t(x);$$

for all  $B \in \mathcal{B}$

$$\limsup_{n \rightarrow \infty} \frac{1}{V_n} \ln \lambda_n[Y_n \in B] \leq \sup_{x \in \text{cl } B} t(x).$$

Let  $K_n \in \mathcal{F}_n$  such that

$$\liminf_{n \rightarrow \infty} \frac{1}{V_n} \ln \lambda_n[K_n] \geq a > -\infty. \quad (5.1)$$

Let  $C_n$  be the image of  $K_n$  under the map  $Y_n$  and  $C := \bigcap_n \text{cl} \left( \bigcup_{k \geq n} C_k \right)$ .

If  $\nu_n := \lambda_n[\cdot | K_n]$ , then (for  $n$  large enough) for any  $B \in \mathcal{B}$

$$\limsup_{n \rightarrow \infty} \frac{1}{V_n} \ln \nu_n[Y_n \in B] \leq \sup_{x \in \text{cl } B \cap C} t(x) - a. \quad (5.2)$$



The set  $N_C(a) := \{x \in C : t(x) \geq a\}$  is non-empty and compact. The sequence of probability measures  $\nu_n$  is eventually concentrated on  $N_C(a)$  at an exponential rate on the scale  $V_n$ .

**Comments.** 1. Let  $X$  be a Hausdorff topological space and  $C_n \subset X$ ,  $n \geq 1$ . The **K-lower limit** of  $\{C_n\}$ ,  $K - \liminf_{n \rightarrow \infty} C_n$ , is the set of all points  $x \in X$  with the following property: for every open neighbourhood  $U$  of  $x$  there exists  $k \in \mathbb{N}$  such that  $U \cap C_n \neq \emptyset$  for every  $n \geq k$ . The **K-upper limit** of  $\{C_n\}$ ,  $K - \limsup_{n \rightarrow \infty} C_n$ , is the set of all points  $x \in X$  with the following property: for every open neighbourhood  $U$  of  $x$  and for every  $k \in \mathbb{N}$  there exists  $n \geq k$  such that  $U \cap C_n \neq \emptyset$ . If there exists  $C \subset X$  such that  $C = K - \limsup_{n \rightarrow \infty} C_n = K - \liminf_{n \rightarrow \infty} C_n$ , then the sequence  $C_n$  **converges to  $C$  in the sense of Kuratowski**. Convergence in the sense of Kuratowski is related to the  $\Gamma$ -convergence of functions.  $C_n$  converges to  $C$  in the sense of Kuratowski if and only if the sequence of indicator functions  $I_{C_n}$   $\Gamma$ -converges to the indicator function  $I_C$  (see p.43 [D]). The set  $C$  in proposition 5.1 is equal to the **K-upper limit** of  $\{C_n\}$ .

2. The real number  $a$  in proposition 5.1 is bounded by

$$a \leq \inf_m \limsup_{n \rightarrow \infty} \frac{1}{V_n} \ln \lambda_n [Y_n \in \text{cl} \left( \bigcup_{k \geq m} C_k \right)] \leq \sup_{x \in C} t(x)$$

since  $t$  is u.s.c. and has compact level-sets (see proof of theorem 5.2). This also shows that hypothesis (5.1) implies that  $C \neq \emptyset$ .

**Proof.** Set for  $B \in \mathcal{B}$

$$\bar{t}_C[B] := \limsup_{n \rightarrow \infty} \frac{1}{V_n} \ln \nu_n [Y_n \in B].$$

For any  $B_1$  and  $B_2$

$$\bar{t}_C[B_1 \cup B_2] = \bar{t}_C[B_1] \vee \bar{t}_C[B_2]. \quad (5.3)$$

Let  $x \in X$  and  $G$  a measurable neighbourhood of  $x$ . One first proves that

$$\inf_{G \ni x} \bar{t}_C[G] \leq \begin{cases} t(x) - a & \text{if } x \in C \\ -\infty & \text{otherwise.} \end{cases} \quad (5.4)$$

One has

$$\begin{aligned} \bar{t}_C[G] &\leq \limsup_{n \rightarrow \infty} \frac{1}{V_n} \ln \lambda_n [\{Y_n \in G\} \cap K_n] - a \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{V_n} \ln \lambda_n [Y_n \in G] - a. \end{aligned}$$

Therefore

$$\inf_{G \ni x} \bar{t}_C[G] \leq t(x) - a.$$

Suppose that  $x \notin C$ . Either  $t(x) = -\infty$  and (5.4) is true, or  $t(x) \geq b > -\infty$ . Since  $x \notin C$ , there exists  $n$  such that  $x \notin \text{cl}(\cup_{k \geq n} C_k)$ . Hence there exists an open set  $O \ni x$ , which is disjoint from  $\text{cl}(\cup_{k \geq n} C_k)$  and a measurable neighbourhood  $G$ ,  $O \supset G \ni x$ . Therefore

$$\bar{t}_C[G] \leq \limsup_{n \rightarrow \infty} \frac{1}{V_n} \ln \lambda_n [Y_n \in G \cap K_n] - a = -\infty. \quad (5.5)$$

This proves (5.4). Proof of (5.2).

$$\bar{t}_C[B] \leq \limsup_{n \rightarrow \infty} \frac{1}{V_n} \ln \lambda_n[Y_n \in B] - a \leq \sup_{x \in \text{cl} B} t(x) - a. \quad (5.6)$$

If  $\bar{t}_C[B] = -\infty$ , then (5.2) is true. Assume that  $\bar{t}_C[B] > -\infty$ . (5.6) implies that

$$-\infty < a + \bar{t}_C[B] \leq \sup_{x \in \text{cl} B} t(x).$$

Since  $t$  has compact level-sets it reaches its maximum on closed sets. Hence

$$\text{cl} B \cap \{x : t(x) \geq a + \bar{t}_C[B]\} \neq \emptyset.$$

Let  $K := \{x : t(x) \geq a + \bar{t}_C[B]\}$  and let  $G$  be a measurable neighbourhood of the level-set  $K$ . One has  $\text{cl}(X \setminus G) \cap K = \emptyset$ . Thus

$$\begin{aligned} \bar{t}_C[X \setminus G] &\leq \limsup_{n \rightarrow \infty} \frac{1}{V_n} \ln \lambda_n[Y_n \in X \setminus G] - a \\ &\leq \sup_{x \in \text{cl}(X \setminus G)} t(x) - a \\ &< a + \bar{t}_C[B] - a = \bar{t}_C[B]. \end{aligned}$$

Hence

$$\begin{aligned} \bar{t}_C[B] &\leq \bar{t}_C[\text{cl} B \cap G] \vee \bar{t}_C[\text{cl} B \cap (X \setminus G)] \\ &\leq \bar{t}_C[\text{cl} B \cap G] \vee \bar{t}_C[X \setminus G] \\ &\leq \bar{t}_C[\text{cl} B \cap G]. \end{aligned} \quad (5.7)$$

Given  $\varepsilon > 0$  and  $x \in X$ , let  $G_x$  be a measurable neighbourhood of  $x$  such that (see (5.5))

$$\bar{t}_C[G_x] \leq \begin{cases} t(x) - a + \varepsilon & \text{if } x \in C \\ -\infty & \text{if } x \notin C. \end{cases} \quad (5.8)$$

One constructs a measurable neighbourhood of  $K$  as follows: since  $K \cap \text{cl} B$  is compact one can find  $x_i \in K \cap \text{cl} B$ ,  $i = 1, \dots, k$ , and measurable neighbourhoods  $G_{x_i}$  of  $x_i$ , which cover  $K \cap \text{cl} B$ . Let  $U$  be any measurable neighbourhood of  $K$ ; then

$$G := (U \setminus \text{cl} B) \cup (U \cap [G_{x_1} \cup \dots \cup G_{x_k}])$$

is a measurable neighbourhood of  $K$  with the property that  $\text{cl} B \cap G \subset G_{x_1} \cup \dots \cup G_{x_k}$ . Given  $\varepsilon > 0$ , it follows from (5.7) and (5.3) that

$$\bar{t}_C[\text{cl} B] \leq \bar{t}_C[\text{cl} B \cap G] \leq \bar{t}_C[\cup_{x_i} G_{x_i}] \leq \sup_{x_i} \bar{t}_C[G_{x_i}]. \quad (5.9)$$

If  $\bar{t}_C[B] > -\infty$  holds, then inequalities (5.9) and (5.8) imply that  $\text{cl} B \cap C \neq \emptyset$ . Since  $\bar{t}_C[X] = 0$  one concludes in particular that  $C \neq \emptyset$ . Therefore

$$\bar{t}_C[\text{cl} B] \leq \sup_{x \in \text{cl} B \cap C} t(x) - a + \varepsilon \quad \forall \varepsilon > 0.$$

This proves (5.2). By (5.2)

$$0 = \bar{t}_C[X] \leq \sup_{x \in C} t(x) - a;$$

hence the set

$$N_C(a) = \{x \in C : t(x) \geq a\}$$

is non-empty and compact. Let  $G$  be a measurable neighbourhood of the compact set  $N_C(a)$ ; since  $\text{cl}(X \setminus G) \cap N_C(a) = \emptyset$  one has

$$\bar{t}_C[X \setminus G] \leq \sup_{x \in \text{cl}(X \setminus G)} t(x) - a = t(x^*) - a < 0$$

because the supremum is attained on  $\text{cl}(X \setminus G)$  at some point  $x^*$ .  $\square$

**Corollary 5.1.** *If in theorem 5.1 we replace hypothesis (5.1) by*

$$\lim_{n \rightarrow \infty} \frac{1}{V_n} \ln \lambda_n[K_n] = \sup_{x \in C} t(x) \equiv a > -\infty, \quad (5.10)$$

*then the same conclusions hold with*

$$N_C(a) := \{x \in C : t(x) = \sup_{x' \in C} t(x')\}.$$

**5.2. Existence of subgradients.** In this subsection the following assumptions are made.  $X$  is a closed subset of a locally Hausdorff real topological vector space  $E$ .  $E'$  is the topological dual of  $E$ ; on  $E$ , resp.  $E'$ , one chooses the  $\sigma(E, E')$ -topology, resp.  $\sigma(E', E)$ -topology. The topology of  $X$  coincides with the induced topology. The u.s.c. function  $t$  is extended to  $\widehat{E}$  by setting  $t(x) := -\infty$  for all  $x \in \widehat{E} \setminus X$ . The concave envelope of  $t$  is denoted by  $\widehat{t}$ .

The conclusion of corollary 5.1 is that  $N_C(a)$  is a non-empty subset of  $X$ . The solutions of the variational problem  $\sup_{x' \in C} t(x')$  are the elements of  $N_C(a)$ . The main result of proposition 5.2 is the proof of the subdifferentiability of  $t$  on  $N_C(a)$ , when  $C$  is a closed convex subset and  $t$  coincides with  $\widehat{t}$  on  $C$ . It is convenient to set

$$F_1(x) := -\widehat{t}(x),$$

and

$$F_2(x) := \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{if } x \notin C. \end{cases}$$

Both functions  $F_1$  and  $F_2$  are proper, l.s.c. and convex. Under the above assumptions

$$N_C(a) = \{x \in E : \inf_{y \in E} (F_1 + F_2)(y) = (F_1 + F_2)(x)\}.$$

The points of  $N_C(a)$  are minima of  $F := F_1 + F_2$ . Hence, if  $x \in N_C(a)$ , then  $0 \in \partial F(x)$ . One always has

$$\partial F(x) \supset \partial F_1(x) + \partial F_2(x).$$

The point is to show that

$$\partial F(x) = \partial F_1(x) + \partial F_2(x). \quad (5.11)$$

**Proposition 5.2.** *Under the above setting, let  $F_1$  be a proper l.s.c. convex function and  $C$  a non-empty closed convex subset. Assume that*

$$N_C := \{x \in C : \inf_{y \in C} F_1(y) = F_1(x)\} \neq \emptyset \quad \text{and} \quad \inf_{y \in C} F_1(y) \in \mathbb{R}.$$

*If either  $\text{int } C \neq \emptyset$  or  $F_1$  is continuous at some point of  $C$ , then*

1. *there exists  $x' \in E'$ , which is bounded below on  $C$ ;*
2.  *$x' \in \partial F_1(x)$  for all  $x \in N_C$ ;*
3.  *$-x' \in \partial F_2(x)$  for all  $x \in N_C$ ;*

4.  $N_C \subset \partial F_1^*(x')$ , where  $F_1^*$  is the conjugate function of  $F_1$  at  $x'$ .

**Remark.** The subgradients of  $F_2$  have a simple geometrical interpretation: if  $0 \neq x' \in \partial F_2(x)$ , then (see (5.13))

$$0 \geq \langle x', z - x \rangle, \quad \forall z \in C;$$

thus  $x'$  is the exterior normal to the closed half-plane  $\{z \in E : \langle x', z \rangle \leq \langle x', x \rangle\}$  containing  $C$ . Conversely, if  $x \in C$  and the closed half-plane  $\{z \in E : \langle x', z \rangle \leq \langle x', x \rangle\}$  contains  $C$ , then  $x' \in \partial F_2(x)$ .

**Proof.** The main point is the proof of (5.11). Since either  $C$  has an interior point and  $\inf_{y \in C} F_1(y)$  is finite or  $F_1$  is continuous at some point of  $C$ , there exists  $z \in \text{dom } F_2 \cap \text{dom } F_1$ , where  $F_1$  or  $F_2$  is continuous, say  $F_1$ . One can apply proposition 5.6 [ET] to get (5.11). For completeness the proof of that result is given. This is a consequence of Hahn-Banach theorem.

Let  $x^*$  be any subgradient in  $\partial(F_1 + F_2)(x)$ ; this means that  $(F_1 + F_2)(x)$  is finite, hence  $F_1(x)$  and  $F_2(x)$  are finite, and for all  $y \in E$

$$F_1(y) + F_2(y) \geq F_1(x) + F_2(x) + \langle x^*, y - x \rangle. \quad (5.12)$$

Let

$$y \mapsto g(y) := F_1(y) - F_1(x) - \langle x^*, y - x \rangle;$$

the function  $g$  is a l.s.c. convex function on  $E$  which is continuous at  $z$ . Define the convex sets  $\mathcal{C}_1$  and  $\mathcal{C}_2$  in  $E \times \mathbb{R}$ :

$$\mathcal{C}_1 := \{(y, a) : g(y) \leq a\}$$

and

$$\mathcal{C}_2 := \{(y, a) : a \leq F_2(x) - F_2(y)\}.$$

Relation (5.12) implies that  $\mathcal{C}_1$  and  $\mathcal{C}_2$  have only boundary-points in common; since  $\mathcal{C}_1$  is the epigraph of  $g$  and  $g$  is continuous at  $z$ , the set  $\mathcal{C}_1$  has a non-empty interior. One can separate  $\mathcal{C}_2$  and  $\text{int } \mathcal{C}_1$  by a closed hyperplane. The hyperplane cannot be vertical; indeed, if the hyperplane were vertical, then one could separate  $\text{dom } F_1$  and  $\text{dom } F_2$ ; this is impossible since there exists  $z \in \text{dom } F_1 \cap \text{dom } F_2$  which is a continuity point of  $F_1$ . Consequently, the separating hyperplane is of the form

$$y \mapsto \langle x'', y \rangle + \alpha, \quad x'' \in E', \quad \alpha \in \mathbb{R},$$

and for all  $y$

$$F_2(x) - F_2(y) \leq \langle x'', y \rangle + \alpha \leq F_1(y) - F_1(x) - \langle x^*, y - x \rangle.$$

Putting  $y = x$  one has  $\alpha = -\langle x'', x \rangle$ ; hence

$$F_2(y) \geq F_2(x) + \langle -x'', y - x \rangle, \quad \forall y$$

and

$$F_1(y) \geq F_1(x) + \langle x^* + x'', y - x \rangle, \quad \forall y.$$

Therefore  $x^*$  can be decomposed into  $x^* = (x^* + x'') + (-x'')$  with  $x^* + x'' \in \partial F_1$  and  $-x'' \in \partial F_2$ . This proves (5.11).

(5.11) implies the existence of  $x' \in E'$  such that  $x' \in \partial F_1(x)$  and  $-x' \in \partial F_2(x)$ . In particular

$$F_2(u) \geq \langle -x', u - x \rangle, \quad \forall u \in E,$$

which is equivalent to

$$\langle x', u \rangle \geq \langle x', x \rangle, \quad \forall u \in C. \quad (5.13)$$

The functional  $x'$  is therefore bounded below on  $C$  and 1. is proved. Let  $x$  and  $y$  be distinct elements of  $N_C$ ; since  $F_1(x) = F_1(y)$  and  $\langle x', y - x \rangle \geq 0$  by (5.13), for all  $z$  one has

$$\begin{aligned} F_1(z) &\geq F_1(x) + \langle x', z - x \rangle \\ &= F_1(y) + \langle x', z - y \rangle + \langle x', y - x \rangle \\ &\geq F_1(y) + \langle x', z - y \rangle. \end{aligned}$$

Therefore  $x' \in \partial F_1(y) \forall y \in N_C$ , and  $N_C \subset \partial F_1^*(x')$ .  $\square$

**5.3. Conditional limit theorems.** The main notion, which is introduced here, is the notion of LD-regular sequences.

**Definition 5.2.** *Let  $Y_n$  be a sequence of random variables, with values in the topological Hausdorff space  $X$ , and which verify a large deviations principle with rate-function  $s$  on the scale  $V_n$ . A sequence  $\{C_n\}$ ,  $C_n \subset X$ , is LD-regular with limit  $C \subset X$  if*

$$C = \bigcap_{n \in \mathbb{N}} \text{cl} \left( \bigcup_{k \geq n} C_k \right)$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{V_n} \ln \text{Prob} [Y_n \in C_n] = \sup_{x \in C} s(x) > -\infty.$$

Let the setting be as in section 3 with  $X = \mathcal{M}_1^+$ ,  $E = \mathcal{M}$  and  $E' = \mathcal{F}_{\text{loc}}$ .  $Y_n = T_{\Lambda(n)}$  is defined on  $(\Omega, \mathcal{F}, \lambda)$  with  $\lambda \in \mathcal{M}_1^{+, \theta}$  an asymptotically decoupled probability measure. The scale  $V_n := |\Lambda(n)|$ .

The main hypothesis (5.10) of corollary 5.1 is closely related to Sanov's property of Csiszár [Cs], which in the present context has the following statement. A set  $D \subset \mathcal{M}_1^+$  has **Sanov's property** if

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda [T_{\Lambda(n)} \in D] = \sup_{\nu \in D} s(\nu | \lambda) > -\infty.$$

If  $D \subset \mathcal{M}_1^+$  is convex and there exists  $\nu \in \text{int } D$  such that  $s(\nu | \lambda) > -\infty$ , then  $D$  has Sanov's property; (5.10) is valid<sup>57</sup> with  $K_n \equiv T_{\Lambda(n)}^{-1} D$  since in that case

$$\sup_{\nu \in \text{int } D} s(\nu | \lambda) = \sup_{\nu \in \text{cl } D} s(\nu | \lambda).$$

In [LPS3] and [LPS4] Sanov's property is replaced by the notion of LD-regular sequence for the purpose of proving conditional limit theorems.

**Remark.** An interesting special case of LD-regular sequences is when  $C = \{\nu\}$  is a singleton. In the present setting this means that

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda [T_{\Lambda(n)} \in C_n] = s(\nu | \lambda) > -\infty.$$

<sup>57</sup>This follows from the large deviations principle and because  $s$  is concave. See lemma 6.1 in [LPS3].

Choosing suitably the sets  $C_n$ , it is always possible to find such LD-regular sequence, whenever  $s(\nu|\lambda) > -\infty$ . Indeed, let  $G_n$ ,  $n \geq 1$ , be a decreasing sequence of open sets, such that  $\text{cl}G \downarrow \{\nu\}$ . An LD-regular sequence  $\{C_n\}$  is constructed by stretching the sequence  $\{G_n\}$ . Since  $G_m$  is open the lower bound of the large deviations principle gives  $\forall m$

$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in G_m] \geq \sup_{\mu \in G_m} s(\mu|\lambda) \geq s(\nu|\lambda).$$

Therefore there exists an increasing diverging sequence  $\{N_k\}$  starting with  $N_0 = 1$ , such that, if

$$C_n := G_k \quad \text{for } n \text{ such that } N_{k-1} \leq n < N_k,$$

then

$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in C_n] \geq s(\nu|\lambda).$$

On the other hand, since  $\text{cl}G \downarrow \{\nu\}$ , one has (see proof of theorem 5.2)

$$\limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in C_n] \leq s(\nu|\lambda).$$

Theorem 5.1 is a general result about conditional limit theorems. It is stated for sequences  $\{C_n\}$ , which are slightly more general than LD-regular.

**Theorem 5.1.** *Let  $\lambda$  be a translation invariant, asymptotically decoupled probability measure on  $(\Omega, \mathcal{F})$ . Let  $K_n \subset \Omega$ ,  $C_n := \{T_{\Lambda(n)}(\omega) \in \mathcal{M}_1^+ : \omega \in K_n\}$  and  $\mu_n := \lambda[\cdot | K_n]$ . Assume that*

$$C = \bigcap_{n \in \mathbb{N}} \text{cl} \left( \bigcup_{k \geq n} C_k \right)$$

and

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[K_n] = \sup_{\mu \in C} s(\mu|\lambda) > -\infty.$$

*Then the set of cluster points of  $\{\mathcal{A}_{\Lambda(n)}\mu_n\}$  is non-empty and any cluster point is in the closed convex hull  $\text{cl}(\text{conv } N_C)$  of the non-empty compact set*

$$N_C := \left\{ \mu \in C : s(\mu|\lambda) = \sup_{\nu \in C} s(\nu|\lambda) \right\}.$$

**Proof.** The proof of the existence of cluster points is the same as the proof of lemma 3.4. By corollary 5.1 the empirical measures  $T_{\Lambda(n)}$  defined on the probability spaces  $(\Omega, \mathcal{F}, \nu_n)$  are concentrated at exponential rate on the non-empty compact set  $N_C$ . Assume that  $\nu$  is a cluster point of the sequence  $\nu_n := \mathcal{A}_{\Lambda(n)}\mu_n$  and that  $\nu \notin \text{cl}(\text{conv } N_C)$ . One can separate strictly the convex compact set  $\{\nu\}$  and the disjoint set  $\text{cl}(\text{conv } N_C)$ . This implies the existence of a local bounded function  $f$  and  $\delta > 0$ , such that for all  $\alpha \in \text{cl}(\text{conv } N_C)$

$$\langle f, \alpha \rangle \geq \delta \quad \text{and} \quad \langle f, \nu \rangle \leq 0. \quad (5.14)$$

Let  $G := \{\alpha : \langle f, \alpha \rangle > \delta/2\}$ ;  $G \supset N_C$  and there exist  $\kappa > 0$  and  $n_\kappa$  such that

$$\mu_n[T_{\Lambda(n)} \notin G] \leq e^{-\kappa n} \quad \forall n \geq n_\kappa.$$

Since

$$\langle f, \nu_n \rangle = \int_{\Omega} \langle T_{\Lambda(n)}(\omega), f \rangle \mu_n(d\omega),$$

$\lim_{n \rightarrow \infty} \langle f, \nu_n \rangle \geq \delta/2$ , which contradicts (5.14).  $\square$

**Remark.** Since the whole analysis is based on the empirical measure only statements about the convergence of averaged measures  $\mathcal{A}_{\Lambda(n)}\lambda[\cdot|K_n]$  are available. If  $N_C = \{\mu\}$  is a singleton, then

$$\lim_{n \rightarrow \infty} \mathcal{A}_{\Lambda(n)}\lambda[\cdot|K_n] = \mu.$$

The limiting measure  $\mu$  is characterized by a maximum entropy property:  $s(\mu|\lambda) > s(\nu|\lambda) \forall \nu \in C, \nu \neq \mu$ .

The next conditional limit theorem is a variant of theorem 5.1. It is a consequence of corollary 5.1 and proposition 5.2. It is stated in a dual manner.

**Theorem 5.2.** *Let  $\lambda$  be a translation invariant, asymptotically decoupled probability measure on  $(\Omega, \mathcal{F})$ . Let  $\{C_n\} \subset \mathcal{M}_1^+$ , such that*

$$C = \bigcap_{n \in \mathbb{N}} \text{cl} \left( \bigcup_{k \geq n} C_k \right).$$

*Assume that  $C$  is convex and that there exists  $\nu \in \text{int } C$  with  $s(\nu|\lambda) > -\infty$ . Define  $\mu_n := \lambda[\cdot|T_{\Lambda(n)} \in C_n]$ . Then the set of cluster points of  $\{\mathcal{A}_{\Lambda(n)}\mu_n\}$  is non-empty, and there exists  $f \in \mathcal{F}_{\text{qloc}}$  such that any cluster point of  $\{\mathcal{A}_{\Lambda(n)}\mu_n\}$  is in  $\partial p(f|\lambda)$ .*

**Proof.** By hypothesis  $C$  is convex,  $\text{int } C \neq \emptyset$  and  $\exists \nu \in \text{int } C$  with  $s(\nu|\lambda) > -\infty$ . Therefore (see comment after definition 5.2)

$$\sup_{\nu \in \text{int } C} s(\nu|\lambda) = \sup_{\nu \in \text{cl } C} s(\nu|\lambda) > -\infty.$$

Let  $D_n := \text{cl} \left( \bigcup_{k \geq n} C_k \right)$ ; since  $s$  is u.s.c. and has compact level-sets  $\exists \nu_n \in D_n$  such that  $s(\nu_n|\lambda) = \sup_{\nu \in D_n} s(\nu|\lambda)$ . Let  $\nu^*$  be a cluster point of  $\{\nu_n\}$ .  $D_n$  is closed and  $\nu_m \in D_n \forall m \geq n$ , hence  $\nu^* \in D_n \forall n$ , which implies that  $\nu^* \in \bigcap_n D_n = C$ . Therefore

$$\begin{aligned} \sup_{\nu \in \text{int } C} s(\nu|\lambda) &\leq \liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in C_n] \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in C_n] \\ &\leq \inf_m s(\nu_m|\lambda) \leq s(\nu^*|\lambda) \leq \sup_{\nu \in \text{cl } C} s(\nu|\lambda). \end{aligned}$$

Hence the sequence  $\{C_n\}$  is LD-regular and theorem 5.1 applies. By proposition 5.2 there exists  $f \in \mathcal{F}_{\text{qloc}}$ , such that  $N_C$  is a subset of the convex set  $\partial p(f|\lambda)$ .  $\square$

The third conditional theorem is formulated in the spirit of statistical mechanics. It makes use of the notion of asymptotically I-null sequences. When  $\{\mu_n\}$  is a sequence of conditioned measures, it is possible to estimate  $\mathcal{H}(\mu_n|\nu_n)$  with  $\nu_n$  defined by (3.18) in terms of thermodynamical functions. The following setting is assumed. Let  $\mathbf{f} : \Omega \rightarrow \mathbb{R}^k$  be a quasilocal  $\mathbb{R}^d$ -valued function and define

$$\mathbf{F}_n := \frac{1}{|\Lambda(n)|} \sum_{x \in \Lambda(n)} \theta_x \mathbf{f}.$$

The random variables  $\mathbf{F}_n$  defined on the probability space  $(\Omega, \mathcal{F}, \lambda)$  verify a large deviations principle, whose rate-function is denoted by  $s(\mathbf{x}|\mathbf{f})$ ,  $\mathbf{x} \in \mathbb{R}^k$ .  $s(\cdot|\mathbf{f})$  is concave and its domain is

$$\text{dom } s := \{\mathbf{x} \in \mathbb{R}^k : s(\mathbf{x}|\mathbf{f}) > -\infty\}.$$

Let  $\{C_n\}$  be LD-regular with limit  $C$  in  $\mathbb{R}^k$ . Define  $K_n := \{\mathbf{F}_n \in C_n\}$  and

$$\mu_n^C := \lambda[\cdot | \mathbf{F}_n \in C_n].$$

The dual space of  $E := \mathbb{R}^k$  is  $E' = \mathbb{R}^k$  and the pairing is the Euclidean scalar product. Elements of  $E'$  are denoted by  $\mathbf{x}'$ . For each  $\mathbf{x}'$  define the quasilocal function

$$f_{\mathbf{x}'}(\omega) := \langle \mathbf{x}', \mathbf{f}(\omega) \rangle$$

and the probability measures  $\nu_n^{\mathbf{x}'}$  by (3.18) with  $f = f_{\mathbf{x}'}$ , that is <sup>58</sup>

$$\begin{aligned} \frac{d\nu_n^{\mathbf{x}'}}{d\lambda}(\omega) &:= \exp \left[ \sum_{j \in \Lambda(n)} \theta_j f_{\mathbf{x}'}(\omega) - |\Lambda(n)| p_n(f_{\mathbf{x}'}) \right] \\ &= \exp \left[ |\Lambda(n)| \langle \mathbf{x}', \mathbf{F}_n(\omega) \rangle - |\Lambda(n)| p_n(f_{\mathbf{x}'}) \right]. \end{aligned}$$

One compares the sequences  $\{\mu_n^C\}$  and  $\{\nu_n^{\mathbf{x}'}\}$ .

**Lemma 5.1.** *The setting is as above. Assume that the laws of the random variables  $\mathbf{F}_n$  defined on the probability spaces  $(\Omega, \mathcal{F}, \mu_n^C)$  are eventually concentrated on the non-empty compact subset  $N \subset \mathbb{R}^k$ . Then*

$$\limsup_n \frac{1}{|\Lambda(n)|} \mathcal{H}(\mu_n^C | \nu_n^{\mathbf{x}'}) \leq \sup_{\mathbf{x} \in N} (p(f_{\mathbf{x}'}) - \langle \mathbf{x}', \mathbf{x} \rangle - \sup_{\mathbf{y} \in C} s(\mathbf{y}|\mathbf{f})).$$

**Proof.** Since the sequence  $\{C_n\}$  is LD-regular with limit  $C$ ,

$$a := \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[K_n] = \sup_{\mathbf{y} \in C} s(\mathbf{y}|\mathbf{f}) > -\infty.$$

By definition

$$\frac{1}{|\Lambda(n)|} \mathcal{H}(\mu_n^C | \nu_n^{\mathbf{x}'}) = - \int_{K_n} \langle f_{\mathbf{x}'}, T_{\Lambda(n)}(\omega) \rangle \lambda[d\omega | K_n] + p_n(f_{\mathbf{x}'}) - \frac{1}{|\Lambda(n)|} \ln \lambda[K_n].$$

Therefore

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \mathcal{H}(\mu_n^C | \nu_n^{\mathbf{x}'}) &\leq - \liminf_{n \rightarrow \infty} \int_{K_n} \langle \mathbf{x}', \mathbf{F}_n(\omega) \rangle \lambda[d\omega | K_n] \\ &\quad + p(f_{\mathbf{x}'}) - a. \end{aligned}$$

One needs the following auxiliary result (5.15). Let  $\mathbb{M}_n$  be the law of  $\mathbf{F}_n$  and  $g : \mathbb{R}^k \rightarrow \mathbb{R}$  be l.s.c., bounded below on  $B \subset \mathbb{R}^k$ , such that  $\lim_{n \rightarrow \infty} \mathbb{M}_n[B] = 1$ . Then

$$\inf_{\mathbf{x} \in N} g(\mathbf{x}) \leq \liminf_{n \rightarrow \infty} \int_B g(\mathbf{x}) \mathbb{M}_n(d\mathbf{x}). \quad (5.15)$$

Indeed, for each  $n \geq 1$  and any open set  $G \supset N$

$$\int_B g(\mathbf{x}) \mathbb{M}_n(d\mathbf{x}) \geq \left( \inf_{\mathbf{x} \in G} g(\mathbf{x}) \right) \mathbb{M}_n[B \cap G] + \left( \inf_{\mathbf{x} \in B} g(\mathbf{x}) \right) \mathbb{M}_n[B \setminus G].$$

<sup>58</sup> Same convention as in (3.18) for the boundary terms  $\psi_n$ , if those terms are included.



Since  $G \supset N$ ,  $\lim_{n \rightarrow \infty} \mathbb{M}_n[G] = 1$ . Since  $\lim_{n \rightarrow \infty} \mathbb{M}_n[B] = 1$ , it follows that given  $\varepsilon > 0$ , for  $n$  sufficiently large,

$$\mathbb{M}_n[B \cap G] \geq 1 - \varepsilon \quad \text{and} \quad \mathbb{M}_n[B \setminus G] \leq \varepsilon.$$

Thus for any open neighbourhood  $G$  of  $N$

$$\liminf_{n \rightarrow \infty} \int_B g(\mathbf{x}) \mathbb{M}_n(d\mathbf{x}) \geq \inf_{\mathbf{x} \in G} g(\mathbf{x}).$$

The result follows since  $g$  is l.s.c.. Applying (5.15) with  $B$  a bounded subset of  $\mathbb{R}^k$ , such that  $f_{\mathbf{x}'}(\omega) \in B \forall \omega$ , one obtains

$$\liminf_{n \rightarrow \infty} \int_{K_n} \langle \mathbf{x}', \mathbf{F}_n(\omega) \rangle \lambda[d\omega|K_n] = \liminf_{n \rightarrow \infty} \int_B \langle \mathbf{x}', \mathbf{y} \rangle \mathbb{M}_n(d\mathbf{y}) \geq \inf_{\mathbf{x} \in N} \langle \mathbf{x}', \mathbf{x} \rangle.$$

Thus

$$\begin{aligned} \limsup_n \frac{1}{|\Lambda(n)|} \mathcal{H}(\mu_n^C | \nu_n^{\mathbf{x}'}) &\leq - \inf_{\mathbf{x} \in N} \langle \mathbf{x}', \mathbf{x} \rangle + p(f_{\mathbf{x}'} | \lambda) - a \\ &\leq \sup_{\mathbf{x} \in N} (p(f_{\mathbf{x}'} | \lambda) - \langle \mathbf{x}', \mathbf{x} \rangle - a). \end{aligned}$$

□

**Remark.** Lemma 5.1 is the first part of lemma 5.1 in [LPS3]. Lemma 5.1 in [LPS3] has an interesting second part dealing with the case of a rate-function  $s$ , which is not concave as it is the case in mean-field models for example. To be concrete, let  $\varphi : \mathbb{R}^k \rightarrow \mathbb{R}^q$  be a continuous function and  $E = E' = \mathbb{R}^q$ . Let  $Y_n := \varphi \circ \mathbf{F}_n$ . The random variables  $Y_n$  defined on  $(\Omega, \mathcal{F}, \lambda)$  verify a large deviations principle, whose rate-function is given by the contraction principle,

$$s(\mathbf{x}) := \sup\{s(\mathbf{y}|\mathbf{f}) : \varphi(\mathbf{y}) = \mathbf{x}\}.$$

In that case define  $\nu_n^{\mathbf{x}'}$  by

$$\frac{d\nu_n^{\mathbf{x}'}}{d\lambda}(\omega) := \exp [|\Lambda(n)| \langle \mathbf{x}', Y_n(\omega) \rangle - |\Lambda(n)| p_n(\mathbf{x}')],$$

where

$$p_n(\mathbf{x}') := \lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \int_{\Omega} \exp [|\Lambda(n)| \langle \mathbf{x}', Y_n(\omega) \rangle] \lambda(d\omega).$$

Let  $\{C_n\}$  be as above and  $K_n := \{Y_n \in C_n\}$ . Let  $\widehat{s}$  be the concave envelope of  $s$ . Suppose that  $N$  is a non-empty compact concentration set for the random variables  $Y_n$  defined on the probability spaces  $(\Omega, \mathcal{F}, \mu_n^C)$ . Then<sup>59</sup>

$$\liminf_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \mathcal{H}(\mu_n^C | \nu_n^{\mathbf{x}'}) \geq \inf_{\mathbf{x} \in N} (\widehat{s}(\mathbf{x}) - s(\mathbf{x})).$$

Thus, if  $\inf_{\mathbf{x} \in N} (\widehat{s}(\mathbf{x}) - s(\mathbf{x})) > 0$ , then it is impossible that the sequence  $\{\mu_n^C\}$  is asymptotically  $I$ -null to the sequence  $\{\nu_n^{\mathbf{x}'}\}$ .

**Theorem 5.3.** *Let  $\lambda$  be a translation invariant, asymptotically decoupled probability measure on  $(\Omega, \mathcal{F})$ . Let  $\mathbf{f} : \Omega \rightarrow \mathbb{R}^k$  be a quasilocal function and  $s(\cdot | \mathbf{f})$  the rate-function of the large deviations principle of the random variables  $\mathbf{F}_n = \mathcal{A}_{\Lambda(n)} \mathbf{f}$  defined on the probability space  $(\Omega, \mathcal{F}, \lambda)$ . Let  $\{C_n\}$  be a LD-regular sequence in  $\mathbb{R}^k$  with*

<sup>59</sup> The lower bound is proved in lemma 5.1 in [LPS3] without any particular assumption on the form of  $Y_n$ .

limiting set  $C$ , which is assumed to be convex. Assume that  $s(\cdot|\mathbf{f})$  is continuous at some  $\mathbf{x} \in C$  or that  $\text{int } C \neq \emptyset$ . Then the set of cluster points of  $\{\mathcal{A}_{\Lambda(n)}\mu_n^C\}$  is non-empty. There exists  $\mathbf{x}'$ , which is a subgradient of  $s(\cdot|\mathbf{f})$ , such that any cluster point of  $\{\mathcal{A}_{\Lambda(n)}\mu_n^C\}$  is in  $\partial p(f_{\mathbf{x}'})$  with  $f_{\mathbf{x}'} := \langle \mathbf{x}', \mathbf{f} \rangle$ .

**Proof.** Corollary 5.1 implies that

$$N_C = \{\mathbf{x} \in \mathbb{R}^k : \sup_{\mathbf{y} \in C} s(\mathbf{y}|\mathbf{f}) = s(\mathbf{x}|\mathbf{f})\}$$

is a concentration set for the random variables  $\mathbf{F}_n$  defined on the probability spaces  $(\Omega, \mathcal{F}, \mu_n^C)$ . Proposition 5.2 implies that there exists a subgradient  $\mathbf{x}' \in \mathbb{R}^k$ , such that  $\mathbf{x}' \leftrightarrow \mathbf{x}$  for all  $\mathbf{x} \in N_C$ ,

$$p(f_{\mathbf{x}'}) - s(\mathbf{x}|\mathbf{f}) = \langle \mathbf{x}', \mathbf{x} \rangle \quad \forall \mathbf{x} \in N_C.$$

Applying lemma 5.1

$$\limsup_n \frac{1}{|\Lambda(n)|} \mathcal{H}(\mu_n^C | \nu_n^{\mathbf{x}'}) \leq \sup_{\mathbf{x} \in N_C} (p(f_{\mathbf{x}'}) - \langle \mathbf{x}', \mathbf{x} \rangle - \sup_{\mathbf{y} \in C} s(\mathbf{y}|\mathbf{f})) = 0.$$

The theorem follows from theorem 3.2.  $\square$

**5.4. Equivalence of ensembles.** In the previous setting, suppose that  $\mathbf{x} \in \text{dom } s$  and that  $s(\mathbf{x}|\mathbf{f})$  is continuous at  $\mathbf{x}$ . Suppose further that there exists a unique  $\mathbf{x}'$  such that  $\mathbf{x}' \leftrightarrow \mathbf{x}$  (and that the same is true for any  $\mathbf{y}$  such that  $\mathbf{x}' \leftrightarrow \mathbf{y}$ ). One can construct a decreasing sequence of closed balls  $B_{\varepsilon_n}(\mathbf{x})$ , with  $\varepsilon_n \downarrow 0$ , such that

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[\mathbf{F}_n \in B_{\varepsilon_n}] = s(\mathbf{x}|\mathbf{f}). \quad (5.16)$$

Theorem 5.1 implies that the sequence of conditioned measures<sup>60</sup>

$$\mathcal{A}_{\Lambda(n)}\lambda[\cdot | \mathbf{F}_n \in B_{\varepsilon_n}] = \mathcal{A}_{\Lambda(n)}\lambda[\cdot | T_{\Lambda(n)} \in \Phi_{\mathbf{f}}^{-1} B_{\varepsilon_n}],$$

which are interpreted in statistical mechanics as microcanonical states, has cluster points and each cluster point is in  $\partial p(f_{\mathbf{x}'})$ . Notice that (5.16) means simply that asymptotically the entropy of the microcanonical states, as defined in statistical mechanics, is equal to the Boltzmann entropy of the macroscopic state  $\mathbf{x}$ .

Conversely, given  $\nu \in p(f_{\mathbf{x}'})$ , one can construct a decreasing sequence  $\{C_n\}$  of closed neighbourhoods of  $\nu$  such that  $C_n \downarrow \{\nu\}$  and

$$\lim_{n \rightarrow \infty} \frac{1}{|\Lambda(n)|} \ln \lambda[T_{\Lambda(n)} \in C_n] = s(\mathbf{x}|\mathbf{f}).$$

Theorem 5.3 implies that

$$\lim_{n \rightarrow \infty} \mathcal{A}_{\Lambda(n)}\lambda[\cdot | T_{\Lambda(n)} \in C_n] = \nu.$$

<sup>60</sup>  $\Phi_{\mathbf{f}}$  is defined as in section 2 by  $\Phi_{\mathbf{f}}(\mu) = \langle \mathbf{f}, \mu \rangle$ .

## REFERENCES

- [B] R. Bowen, Equilibrium states and ergodic theory of Anosov diffeomorphisms, Lecture Notes in Mathematics 470, Springer, Berlin (1975).
- [BS] F.A. Berezin, Y. Sinai, Existence of a phase transition for the lattice gas with interparticle attraction, *Transaction Moscow Math. Soc.* **17** 219-235 (1967).
- [Br] A. Brønsted, Conjugate convex functions in topological vector spaces *Mat. Fys. Medd. Dan. Vid. Selk.* **34**, no.2, 3-27 (1964).
- [BLP] J. Bricmont, J.L. Lebowitz, C.-E. Pfister, On the Equivalence of Boundary Conditions, *J. Stat. Phys.* **21**, 573-582 (1979).
- [BD] W. Bryc, A Dembo, Large deviations and strong mixing *Ann. Inst. H. Poincaré* **32**, 549-569 (1996).
- [C] F. Comets, Grandes déviations pour des champs de Gibbs sur  $\mathbb{Z}^d$ , *C. R. Acad. Sci. Série I*, **303**, 511-514 (1986).
- [Cs] I. Csiszár, Sanov property, generalized I-projection and a conditional limit theorem *Ann. Prob.* **12**, 768-793 (1984).
- [D] G. Dal Maso, An introduction to  $\Gamma$ -convergence, Birkhäuser, Basel (1993)
- [DS] J.-D. Deuschel, D.W. Stroock, Large Deviations, Academic Press, New-York (1989).
- [Do] R.L. Dobrushin, Existence of phase transitions in models of a lattice gas, Proc. Berkeley Symposium on Prob. Th. and Statistics, vol. III, 73-87 (1965).
- [DV1] M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time - I, *Commun. Pure Appl. Math.*, **28**, 1-47 (1975).
- [DV2] M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time - III, *Commun. Pure Appl. Math.*, **29**, 389-461 (1976).
- [DZ] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, Jones and Bartlett Publishers, Boston (1993).
- [E] R.S. Ellis, Entropy, Large Deviations, and Statistical Mechanics, Springer, Berlin (1985).
- [EFS] A.C.D. van Enter, R. Fernández, A.D. Sokal, Regularity properties and pathologies of position-space renormalization-group transformation: scope and limitations of Gibbsian theory, *J. Stat. Phys.* **72**, 879-1167 (1993).
- [ET] I. Ekeland, R. Temam, Analyse convexe et problèmes variationnels, Dunod, Gauthier-Villars, Paris-Bruxelles-Montréal (1974).
- [FO] H. Föllmer, S. Orey, Large deviations for the empirical field of a Gibbs measure, *Ann. Prob.* **16**, 961-977 (1988).
- [G1] H.O. Georgii, Gibbs measures and phase transitions, Walter de Gruyter, Berlin, New-York (1988).
- [G2] H.O. Georgii, Large deviations and maximum entropy principle for interacting random fields on  $\mathbb{Z}^d$ , *Ann. Prob.* **21**, 1845-1875 (1993).
- [GH] H.O. Georgii, Y. Higuchi, Percolation and the number of phases in the 2D Ising model, *J. Math. Phys.* **41**, 1153-1169 (2000).
- [I] R.B. Israel, Convexity in the theory of lattice gases, Princeton University Press, Princeton (1979).
- [J] E.T. Jaynes, Papers on probability, statistics and statistical physics. R.D. Rosenkrantz, ed., D.Reidel Publishing Co., Dordrecht, Boston, London (1983).
- [K] G. Keller, Equilibrium states in ergodic theory, Cambridge University Press, Cambridge (1998).
- [KU] B. Kahn, G.E. Uhlenbeck, On the Theory of Condensation *Physica* **5**, 399-416 (1938).
- [L] O.E. Lanford, Entropy and equilibrium states in classical mechanics, in Statistical Mechanics and Mathematical Problems. A. Lenard, ed., Lecture Notes in Physics 20, Springer, Berlin (1973).
- [LM] D. Lind, B. Markus, Symbolic Dynamics and Coding, Cambridge University Press, Cambridge (1995).
- [LP] J.T. Lewis, C.-E. Pfister, Thermodynamic probability theory, *Russian Math. Surveys* **50**, 279-317 (1995).
- [LPS1] J.T. Lewis, C.-E. Pfister and W.G. Sullivan, Large Deviations and the Thermodynamic Formalism: a new proof of the equivalence of ensembles, in On Three Levels, M. Fannes, C. Maes, A. Verbeure eds., Plenum Press 183-193, (1994).

- [LPS2] J.T. Lewis, C.-E. Pfister and W.G. Sullivan, The Equivalence of Ensembles for Lattice Systems: Some Examples and a Counterexample, *J. Stat. Phys.* **77**, 397-419 (1994).
- [LPS3] J.T. Lewis, C.-E. Pfister and W.G. Sullivan, Entropy, Concentration of Probability and Conditional Limit Theorems, *Markov Processes and Related Fields*, **1** 319–386 (1995).
- [LPS4] J.T. Lewis, C.-E. Pfister and W.G. Sullivan, Generic points for stationary measures via large deviation theory *Markov Processes and Related Fields* **5**, 235–267 (1999).
- [M-L] A. Martin-Löf, Statistical Mechanics and the Foundations of Thermodynamics. Lecture Notes in Physics 101, Springer, Berlin (1979)
- [MN] A. Meda, P. Ney, The Gibbs conditioning principle for Markov chains, in Perplexing Problems in Probability, M. Bramson, R. Durrett eds., Progress in Probability **44**, 385-398, Birkhäuser, Basel (1999).
- [MM] A. Messenger, S. Miracle-Sole, Equilibrium states of the two-dimensional Ising model in the two-phase region *Commun. Math. Phys.* **40**, 187-196 (1975).
- [M] R.A. Minlos, Lectures on Statistical Physics, *Russian Math. Surveys* **23**, 137-196 (1968)
- [O] S. Olla, Large deviation for Gibbs random fields, *Prob. Th. Rel. Fields* **77**, 343-357 (1988).
- [Pe] R. Peierls, On Ising's model of ferromagnetism, *Proc. Camb. Phil. Soc.* **32**, 477-481 (1936).
- [PS] C.-E. Pfister, W.G. Sullivan, Billingsley dimension, preprint (2000).
- [Ro] R.T. Rockafellar, Convex Analysis. Princeton University Press, Princeton (1970)
- [RZ] S. Roelly, H. Zessin, The Equivalence of Equilibrium Principles in Statistical Mechanics and some Applications to Large Particle Systems, *Expositiones Mathematicae* **11**, 385-405 (1993)
- [Ru1] D. Ruelle, Correlation functionals, *J. Math. Phys.* **6**, 201-220 (1965).
- [Ru2] D. Ruelle, Thermodynamic Formalism: the Mathematical Structures of Classical Equilibrium Statistical Mechanics. Addison-Wesley, Reading, Mass. (1978)
- [S] C. Schroeder,  $I$ -projection and conditional limit theorems for discrete parameter Markov processes, *Ann. Prob.* **21**, 721-758 (1993).
- [Sim] B. Simon, The Statistical Mechanics of Lattice Gases, Princeton Univ. Press, Princeton (1993).
- [Sin] Y. Sinai, Gibbs measures in ergodic theory, *Russian Math. Surveys* **27**, 21-69 (1972).
- [So] A.D. Sokal, Existence of compatible families of proper regular conditional probabilities, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **56**, 537-548 (1981).
- [V] S.R.S. Varadhan, Large Deviations and Applications. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania (1984).
- [W] P. Walters, An Introduction to Ergodic Theory, Springer, Berlin (1981).
- [YL] C.N. Yang, T.D. Lee, Statistical Theory of equations of State and Phase Transitions. I. Theory of Condensation, *Phys. Rev.* **87**, 404-409 (1952).