Resenhas IME-USP 1993, Vol. 1, No. 2/3, 233 - 262.

Renormalization Transformations: Source of Examples and Problems in Probability and Statistics [1]

A. C. D. van Enter, R. Fernández[2]
and A. D. Sokal

Abstract: Renormalization transformations were introduced in statistical mechanics to study critical points. Their natural set-up, however, is within probability theory: they are maps between probability spaces, defined by suitable probability kernels. We review several interesting questions motivated by applications in physics as well as in other areas like image processing and speech recognition. Some of the questions refer to locality properties of the renormalized measures. In particular, it has been often assumed that the maps preserve quasilocality (= almost-Markovianness). We exhibit examples showing that this is not necessarily the case, and discuss the reasons for the loss of quasilocality. As a consequence, the renormalization procedure generates numerous examples of non-Gibbsian measures. Other questions pertain to the smoothness of the renormalization maps. We show that this is related to large-deviation properties. In particular, these maps provide examples of non-Gibbsian measures for which the relative entropy (information gain) density exists. A third category of questions corresponds to practical computational schemes for computing the parameters of renormalized measures. This is a largely unresolved issue of parameter estimation, for which we present some conjectures and partial results. We give a brief review of other manifestations of the important phenomenon of non-Gibbsianness, and we list some open probability-theoretic problems that prevent the extension of our analysis to products of non-compact spaces (unbounded spins).

Key words: Gibbs measure; non-Gibbsian measure; quasilocality; renormalization transformations.

Contents

[1] Invited talk at the V CLAPEM, São Paulo, June 28 – July 3, 1993.

[2] Speaker at the conference

1 Renormalization transformations

Renormalization transformations are a tool widely used by physicists. There is no precise general definition of what a "renormalization" is. Rather, this term refers to a circle of ideas, whose formalization is different (and not always clearly defined) for different fields of physics. In these notes we shall consider only the applications to classical statistical mechanics, where, as remarked long ago (the first published work was [5, 39, 21], but see for instance [55, 9, 40] for further references), the operations involved are most naturally formulated — and, we contend, most profitably studied — in the framework of probability theory.

Indeed, the renormalization transformations used in classical statistical mechanics are just maps between probability measures, defined by appropriate probability kernels. We recall that a ***probability kernel*** from a measure space (Ω = set of points, $\mathcal{F}$ = set of events) to another measure space $(\Omega', \mathcal{F}')$ is a map $T(\,\cdot\,|\,\cdot\,)\colon \mathcal{F}' \times \Omega \to [0,1]$ such that[3]:

[3] The way to order the arguments of a probability kernel changes with the application. Typical notations are: $T(\omega, A)$, $T(A|\omega)$, and $T_\omega(A)$. The first notation is used when T represents a kind of "transition probability" (as in the theory of Markov processes); the second notation is adapted to the interpretation of T as a conditional probability; and the third notation emphasizes that ω is a parameter ("boundary condition") indexing the probability measure on Ω'. Probability kernels defining renormalization transformations correspond to the first of these interpretations, but later in the exposition, we shall introduce kernels (associated to the notion of equilibrium) that

(a) For each fixed $\omega \in \Omega$, $T(\cdot\,|\omega)$ is a probability measure on $(\Omega', \mathcal{F}')$.

(b) For each fixed $A \in \mathcal{F}'$, $T(A|\,\cdot\,)$ is an $\mathcal{F}$-measurable function on Ω.

Each probability kernel T defines a map $\mathcal{T}$ between the corresponding spaces of probability measures:

$$\mathcal{T}: M_{+1}(\Omega, \mathcal{F}) \to M_{+1}(\Omega', \mathcal{F}') \tag{1}$$

with

$$(\mathcal{T}\mu)(A') = \int T(A'|\omega)\,\mu(d\omega) \tag{2}$$

for every $A' \in \mathcal{F}'$.

Renormalization transformations (in the statistical-mechanical setting) are particular cases of such maps, where typically $\Omega = \Omega' =$ a countably infinite Cartesian product. One can interpret $(\mathcal{T}\mu)(A')$ as some sort of average over all ω in the support of μ such that $T(A'|\omega) \neq 0$. Therefore, $\mathcal{T}\mu$ may be considered a "coarser" measure than μ, in which some information has been "averaged out" by the action of T. Alternatively, from the point of view of $\mathcal{T}\mu$, the ω's act as "hidden" degrees of freedom that together with μ determine the ("non-hidden") features of $\mathcal{T}\mu$. The kernel T should be chosen so that it removes the degrees of freedom which are uninteresting for the application at hand; succesive iterations of $\mathcal{T}$ should produce simpler and simpler measures that nevertheless contain all the interesting information.

The physical applications require that these renormalization maps enjoy some crucial properties — single-valuedness, continuity and smoothness, Gibbsianness — whose failure would force the standard paradigm into a substantial overhaul. These properties have been largely taken for granted, and little has been done in the way of rigorous proofs. As we shall see, a probability-theoretic analysis provides definite answers — unfortunately not always positive — for some of these questions.

It may be illuminating to point out that a renormalization transformation can be interpreted as a "degradation" of data. In fact, the above setting is completely analogous to the notion of noisy data used in image processing [23, 25, 11, 28, 20, 12], speech recognition [52] and other fields of applied probability theory (see e.g. the contribution by Prof. Basilis Gidas in this proceedings). In some cases, the noisy (= renormalized) measure is termed a *hidden Markov process* [52, 24, 46], because it arises from a Markov random field by "hiding" some degrees of freedom.

However, the approach, and the questions asked, vary according to whether the kernel T is interpreted as "noise" or as a "renormalization". In fact, the two points of view are complementary: In the first case, the interest lies in the *original* measure μ, and the effort is concentrated on estimating it starting from

---

correspond to the second and third interpretations. We shall stick to the second (conditional-probability-like) notation throughout this review.

the "incomplete" measure $T\mu$. In the second case, the interest is centered in the renormalized measure $T\mu$ itself, the original measure having unnecessary information. While in image processing the noise T is an unwanted and largely uncontrolled feature, in renormalization theory it is added on purpose in a controlled manner. When reconstructing an image one is interested in all features of the landscape; the noise T is an obstacle that should be removed by the inference scheme. In renormalization theory, instead, one wants to detect only the highest and/or broadest mountains, and the coarse-graining map T simplifies our task by "blurring out" smaller fluctuations.

For the sake of completeness, and given the nature of the present audience, we pause to present a brief introduction to the statistical mechanics of classical lattice systems, aimed at probabilists.

2 Elements of classical statistical mechanics (lattice systems)

We base the exposition in this section on the notion of *specification*. This constitutes an approach slightly more general than the usual one (based on *interactions*), and has the advantage that both Gibbsian and non-Gibbsian measures fit naturally within its framework. For a detailed presentation along the same lines, the reader can consult the book by Georgii [27]. A less exhaustive, review-style introduction can also be found in Section 2 of [57].

2.1 The space of configurations

Statistical mechanics deals with very large systems formed by many small constituents (for brevity called here *spins*). The basic ingredients of the mathematical formalism are:

Ω_0 = Space describing the possible configurations of a single constituent. Here, we choose to call it *single-spin space* (= single-constituent space = single-pixel space). It is equipped with a σ-field $\mathcal{F}_0$ and a probability measure μ_{Ω_0} (the *a priori single-spin measure*). Usually Ω_0 is equipped also with a topological structure compatible with $\mathcal{F}_0$ and μ_{Ω_0} (in the sense that open sets are measurable and of nonzero measure). In nearly all applications, Ω_0 is a complete separable metric space. When Ω_0 is a *compact* metric space, we say that we are working with a model of *bounded spins*.

$\mathcal{L}$ = Set labeling the different constituents (copies of Ω_0). For lattice systems, which model (for instance) atoms in a crystal, $\mathcal{L}$ is countable and often identifiable with $\mathbf{Z}^d$.

$\Omega = (\Omega_0)^{\mathcal{L}}$ = Space describing all possible configurations of the system, therefore called *configuration space*. It is equipped with the product topology and product σ-field, and with the *a priori* measure $\mu_0(d\omega) = \prod_{x\in\mathcal{L}} \mu_{\Omega_0}(d\omega_x)$.

Examples:

- The Ising model: Its single-spin space is $\Omega_0 = \{-1, 1\}$ with the discrete topology, and μ_{Ω_0} is the normalized counting measure.
- The N-vector model: $\Omega_0 = S^{N-1}$ = unit sphere in $\mathbf{R}^N$ with the usual topology, μ_{Ω_0} being the normalized Lebesgue measure.

2.2 The space of states

In classical statistical mechanics a *state* is a probability measure on the space of configurations; that is, it is a *random field* indexed by $\mathcal{L}$.

All concepts in statistical mechanics — as in the theory of random fields or processes — are defined via sequences of finite volumes ("windows"). This mimics how real systems are analyzed in practice. In particular, the (microscopic) *observables* — the measurable functions representing feasible measurements — are chosen to be (quasi)*local* functions, that is, functions that (almost) depend only on finitely many spins.

To formalize this, let us denote:

$\mathcal{F}$ = σ-algebra of measurable subsets (events) of Ω.

$\mathcal{F}_\Lambda$ = Sub-σ-algebra of $\mathcal{F}$ formed by the events that depend only on spins in $\Lambda \subset \mathcal{L}$.

$C_\Lambda(\Omega) = C(\Omega) \cap B(\Omega, \mathcal{F}_\Lambda)$ = Set of real-valued bounded continuous functions on Ω that depend only on spins in Λ [in the sense that $f(\omega) = f(\omega')$ whenever $\omega_i = \omega'_i$ for all $i \in \Lambda$].

$C_{\rm loc} = \bigcup_{\Lambda \text{ finite}} C_\Lambda(\Omega)$ = The set of *local* functions.

We can now introduce the (microscopic) *observables*. These are the *quasilocal functions*, that is the members of the set:

$$C_{\rm ql}(\Omega) \;=\; \overline{\bigcup_{\Lambda \text{ finite}} C_\Lambda(\Omega)}$$

where the closure is with respect to the supremum norm. Equivalently, an observable is a function that "depends weakly on distant spins" in the sense that

$$\lim_{\Lambda \uparrow \mathcal{L}} \sup_{\substack{\omega,\omega' \in \Omega \\ \omega_\Lambda = \omega'_\Lambda}} |f(\omega) - f(\omega')| \;=\; 0 \;. \tag{3}$$

In consistency with the experimental approach, two states are considered "distant" or "close" according to how the averages of observables differ. Mathematically, this corresponds to adopting the *weak quasilocal topology* for probability measures:

$$\mu_n \to \mu \quad \text{if} \quad \mu_n(f) \to \mu(f) \text{ for all } f \in C_{\mathrm{ql}}(\Omega) .$$

The space of probability measures, $M_{+1}(\Omega, \mathcal{F})$, with the weak quasilocal topology, is the *space of states* of classical statistical mechanics.[4]

Remarks:

- We emphasize that the convergence is required for each observable f, but *not* uniformly in f.

- Although the topology is defined in terms of *quasi*local functions, an $\epsilon/3$-argument shows that for probability measures it suffices to check convergence for $f \in C_{\mathrm{loc}}(\Omega)$.

- By its very definition, the weak quasilocal topology is insensitive to what happens over long distances: measures with very different long-range correlations can nevertheless be very "close" in the sense of this topology. This fact is at the root of some rather surprising properties of the space of states in general, and the space of Gibbs states in particular (see e.g. [27, Theorem 14.12], [37, Lemma IV.3.2], and [57, Section 2.6.7]).

- If Ω_0 is a compact metric space, then $C_{\mathrm{ql}}(\Omega)$ coincides with $C(\Omega)$, and the weak quasilocal topology coincides with the ordinary weak topology.

2.3 Finite-volume equilibrium distributions

Following the "finite-window" approach inherent to statistical mechanics, we first consider the notion of equilibrium for *finite* volumes.

The state of a finite volume Λ immersed in an exterior configuration ω_{Λ^c} is a probability measure. This measure changes with ω_{Λ^c}. Therefore, a finite volume in equilibrium with its exterior must be described by a *family* of probability measures on the interior, labeled by the exterior configurations. This corresponds to a probability kernel whose second "slot" has a measurable dependence on configurations on Λ^c, and whose first "slot" corresponds to measures on the events inside Λ.

It is technically simpler, however, to consider kernels defined over whole spaces, incorporating the mentioned dependences through suitable restrictions: Equilibrium in a volume Λ is described by a probability kernel $\pi_\Lambda : \mathcal{F} \times \Omega \to [0, 1]$ such that

[4] Warning: in the theory of random fields the space of configurations Ω is often called "state space". Here we call Ω the configuration space and reserve the word "state" for measures.

(a) For each $A \in \mathcal{F}$, the function $\pi_\Lambda(A|\cdot)$ is $\mathcal{F}_{\Lambda^c}$-measurable (i.e. it depends only on the exterior configuration).

(b) For each $B \in \mathcal{F}_{\Lambda^c}$, $\pi_\Lambda(B|\omega) = \chi_B(\omega)$ (i.e. it may act non-deterministically only on observables in the interior of Λ).

Moreover, these kernels for different volumes Λ should satisfy a natural compatibility condition: If a volume Λ' is in equilibrium with its exterior, then all its subvolumes should be in equilibrium with *their* exteriors. Mathematically this is expressed by the condition:

(c) If $\Lambda' \supset \Lambda$, then

$$\pi_{\Lambda'}(\,\cdot\,|\omega) \;=\; \int \pi_\Lambda(\,\cdot\,|\widetilde{\omega})\,\pi_{\Lambda'}(d\widetilde{\omega}|\omega) \tag{4}$$

for *all* configurations $\omega \in \Omega$.

A family $\Pi = \{\pi_\Lambda\}_{\Lambda\subset\mathcal{L},\,\Lambda\text{ finite}}$ satisfying (a), (b) and (c) is called a *specification*. Specifications are the *central* objects of classical statistical mechanics. Their actual expressions are determined by the physics of the problem. In fact, that is where all the physical input goes; once the specification is chosen, the game becomes an exercise in probability theory. This exercise is, however, usually non-trivial, and physical intuition can be very helpful!

We remark that conditions (a)–(c) are almost identical to the properties of regular conditional probabilities. The difference is that for specifications, properties (b) and (c) are required for *all* configurations ω, rather than for "almost all" as is the case for conditional probabilities. The reason is that in the present setting there is no underlying "unconditioned measure"; the notion of "almost all" is not defined [56].

2.4 Infinite-volume equilibrium

Infinite-volume equilibrium is defined extending the equilibrium condition (c) above: A probability measure μ on Ω is said to be *consistent with the specification* $\Pi = \{\pi_\Lambda\}$ if

$$\mu(\,\cdot\,) \;=\; \int \pi_\Lambda(\,\cdot\,|\omega)\,\mu(d\omega) \tag{5}$$

for all finite sets $\Lambda \subset \mathcal{L}$. Equivalently, each π_Λ coincides with a conditional probability of μ conditioned on the events outside Λ:

For each finite $\Lambda \subset \mathcal{L}$ and $A \in \mathcal{F}$, $E_\mu(\chi_A|\mathcal{F}_{\Lambda^c})(\,\cdot\,) = \pi_\Lambda(A|\,\cdot\,)$ μ-a.e.

We denote by $\mathcal{G}(\Pi)$ the set of all measures in $M_{+1}(\Omega,\mathcal{F})$ consistent with Π. Such measures are interpreted as describing possible states of a system whose equilibrium is described by the specification Π.[5]

[5] We are carefully *avoiding* calling such states "equilibrium states"; this term is usually reserved for states that are in addition *translation-invariant*.

The determination of the measures consistent with a given specification is the main task of classical statistical mechanics. In this regard, we can say that the basic problem of statistical mechanics is exactly the *inverse* of the classical probability problem of determining conditional probabilities. In the latter there is a measure given a priori, and one investigates the existence and properties of conditional probabilities. In statistical mechanics the conditional probabilities are given (by the physics of the problem, in the form of a specification), and the question is about the existence and properties of measures having the given conditional probabilities.

In the light of the interpretation of $\mathcal{G}(\Pi)$ as the set of states of a physical system, it is natural to ask the following questions:

- Is $\mathcal{G}(\Pi)$ non-empty? This may seem almost pedantic, but there do exist honest specifications for which one can prove that there are no consistent probability measures (e.g. Example (4.16) of [27] for compact spins; for non-compact spaces the examples are more widespread: massless Gausian models in dimension $d \leq 2$, solid-on-solid models in $d = 1$). Obviously, if $\mathcal{G}(\Pi)$ is empty we cannot proceed further.

- How large is $\mathcal{G}(\Pi)$? There exists a multitude of results giving conditions under which $\mathcal{G}(\Pi)$ consists of a single state (see e.g. [27, Chapters 8 and 9]). More interesting is the complementary case, namely when Π has *more than one* consistent measure. This corresponds to a physical system with several available thermodynamic states. This models what happens for example with water, which along certain curves in the temperature-pressure plane can equally well take the liquid form, the solid form, or a mixture of these; or what happens with magnets, which below a certain temperature can be magnetized in one or another direction. If the specification includes some adjustable parameters, a change in the cardinality of $\mathcal{G}(\Pi)$ under an infinitesimal variation of these parameters is called a *first-order phase transition*. We point out the obvious (in this presentation) fact that such phase transitions can only take place in infinite-volume systems. If the lattice $\mathcal{L}$ is *finite*, there is only one measure consistent with Π, namely $\pi_{\mathcal{L}}(\,\cdot\,|\omega)$ [which is independent of ω].

While the answers to the preceding questions depend on the particularities of the specification, the following facts are valid for general specifications:

- $\mathcal{G}(\Pi)$ is a simplex.

- The extremal points of $\mathcal{G}(\Pi)$ have short-range correlations and trivial tail field.

We recall that the tail field is the σ-field of events that are unaffected by any alteration involving *finitely* many spins, that is, the events belonging to

$$\bigcap_{\substack{\Lambda \subset \mathcal{L} \\ \Lambda \text{ finite}}} \mathcal{F}_{\Lambda^c} .$$

This set can be interpreted as the set of *global events*, and the functions measurable with respect to it are called *observables at infinity* and can be considered the *global observables* of the system. Triviality on the tail field means that the measures of these events are either 0 or 1 or, equivalently, that observables at infinity are almost-everywhere constant.

Therefore, the measures that are extremal points of $\mathcal{G}(\Pi)$ are deterministic (i.e. do not exhibit fluctuations) when restricted to global observables, and have local fluctuations that become independent over long distances. These two properties characterize the macroscopic systems we observe in everyday life: their global characteristics (density, magnetization, etc.) do not appear to vary stochastically, and the local variations they may present are independent for distant points of the sample. Therefore, the extremal points of $\mathcal{G}(\Pi)$ are interpreted as the *macroscopic states* of the system described by the specification Π. *The determination and characterization of the extremal points of $\mathcal{G}(\Pi)$ is the central problem of statistical mechanics.*

2.5 Gibbsian specifications. Gibbs measures

At this point we should say something about the form taken by the specifications of the usual physical systems. It follows from the work of Boltzmann and Gibbs that systems in equilibrium at nonzero temperature should be described by *Gibbsian specifications*. Such specifications are in turn constructed from *interactions*. We now proceed to explain these two concepts.

An *interaction* is a family $\Phi = (\Phi_A)_{A \subset \mathcal{L},\, A \text{ finite}}$ of functions $\Phi_A \colon \Omega \to \mathbf{R}$, where each function Φ_A is interpreted as a contribution to the interaction energy of the spins in A. Therefore Φ_A is required to be $\mathcal{F}_A$-measurable (i.e. to depend only on the spins in the finite subset A).

As an example, the interaction for the Ising model can be written in the form:

$$\Phi_A(\omega) \;=\; \begin{cases} -h_x\omega_x & \text{if } A = \{x\} \\ -J_{xy}\omega_x\omega_y & \text{if } A = \{x,y\} \\ 0 & \text{otherwise} \end{cases} .$$

To define a Gibbsian specification one needs interactions that are *absolutely summable* in the sense that

$$\|\Phi\|_{\mathcal{B}^1} \;\stackrel{\text{def}}{=}\; \sup_{x\in\mathcal{L}} \sum_{\substack{\text{finite } A \,\subset\, \mathcal{L} \\ A \,\ni\, x}} \|\Phi_A\|_\infty < \infty \,. \tag{6}$$

(Roughly, the energy cost of changing one spin, leaving the rest fixed, must be finite.) This condition ensures that for each finite volume Λ one can construct the Hamiltonian

$$H^{\Phi}_{\Lambda}(\omega) \;=\; \sum_{\substack{\text{finite } A \,\subset\, \mathcal{L} \\ A \,\cap\, \Lambda \,\neq\, \emptyset}} \Phi_A(\omega) \,.$$

Note that this Hamiltonian includes not only the contributions coming from subvolumes of Λ (i.e. terms with $A \subset \Lambda$), but also contributions describing the interaction of Λ *with its exterior*.

The *Gibbsian specification* $\Pi^{\Phi} = \{\pi_{\Lambda}^{\Phi}\}$ for an absolutely summable interaction Φ is defined by the kernels (Boltzmann-Gibbs weights)

$$\pi_{\Lambda}^{\Phi}(A|\omega) = (\mathrm{Norm.})^{-1} \int \chi_A(\omega) \exp[-H_{\Lambda}^{\Phi}(\omega)] \prod_{x \in \Lambda} \mu_{\Omega_0}(d\omega_x) .$$

A *Gibbs measure* for an interaction Φ is a measure consistent with Π^{Φ}. A measure is *Gibbsian* if it is a Gibbs measure for some (absolutely summable) interaction.

Gibbs measures have been around us for over one hundred years. There is substantial experience with them; they are well-known, familiar objects. However, recent investigations in non-equilibrium statistical mechanics, as well as examples obtained from renormalization transformations (Section 5.2 below), have made clear the necessity to venture into the unknown territory of non-Gibbsian measures.

3 Gibbsianness and non-Gibbsianness

As a useful preliminary step, in this section we summarize some properties characteristic of Gibbsian measures. This leads naturally to a classification, by contraposition, of some possible types of non-Gibbsianness. For more details on the material of this section, the reader can consult reference [57].

3.1 Attributes of Gibbsian measures

Gibbs measures have properties that are highly desirable from the point of view of nonzero-temperature physics:

(G1) A measure can be a Gibbs measure for at most one Gibbsian specification. Among the consequences of this fact we mention:

- Specifications completely define a system; interactions defining the same specification must be considered physically equivalent.
- Since different multiples of an (nonzero!) interaction (=different temperatures) produce different specifications, a measure cannot be the Gibbs measure for the same system at two different temperatures.

(G2) A Gibbs measure μ is uniformly non-null (with respect to the *a priori* measure μ_0). This means that, for each finite $\Lambda \subset \mathcal{L}$:

- For each $A \in \mathcal{F}_{\Lambda}$,
$$\mu_0(A) > 0 \Longrightarrow \mu(A) > 0 .$$
(Basically, all open sets of configurations have nonzero measure.)

- Moreover, this property holds *uniformly* in the sense that there exist finite nonzero constants α_Λ, β_Λ such that

$$\alpha_\Lambda \mu_0(A) \;\leq\; \mu(A) \;\leq\; \beta_\Lambda \mu_0(A)$$

  for all $A \in \mathcal{F}_\Lambda$.

(G3) A Gibbs measure is *quasilocal* (*almost-Markovian*): It is consistent with a specification $\Pi = \{\pi_\Lambda\}$ which satisfies the equivalent conditions:

- The functions $\omega \mapsto \pi_\Lambda(f|\omega)$ are quasilocal for each Λ finite and each $f \in C_{\mathrm{ql}}$.
- The *direct* influence of far-away spins becomes negligible:

$$\lim_{\Lambda' \uparrow \mathcal{L}} \sup_{\substack{\omega_1, \omega_2 \in \Omega \\ (\omega_1)_{\Lambda'} = (\omega_2)_{\Lambda'}}} |\pi_\Lambda(f|\omega_1) - \pi_\Lambda(f|\omega_2)| \;=\; 0 \tag{7}$$

  for all $f \in C_{\mathrm{ql}}$.

[Here we have used the notation $\pi_\Lambda(f|\omega) = \int f(\omega')\,\pi(d\omega'|\omega)$.] The equivalence of the two conditions is a consequence of (3). From formula (7) we conclude that in Gibbsian measures, information from distant spins cannot be transmitted if the spins in the intermediate region are *fixed*. Alternatively, the only mechanism to transmit information from infinity is through the fluctuations of intermediate spins. (We remark that this transmission from infinity, via fluctuations, indeed takes place at a first-order phase transition; it does not contradict quasilocality.)

(G4) If there is a notion of *translation* — an action of $\mathbf{Z}^d$ on $\mathcal{L}$ by bijections — then the translation-invariant Gibbs measures have good large-deviation properties. More precisely:

- These measures satisfy a variational principle that constitutes a "thermodynamic description" of $\mathcal{G}_{\mathrm{inv}}(\Pi^\Phi)$.
- Their large-deviations probabilities are controlled by a well-defined rate function, called *relative entropy density* or *density of information gain* according to the sign adopted.

In the next subsection we discuss in more detail this last point, as it is very important for the "first fundamental theorem" on renormalization transformations to be presented in Section 5.1.

3.2 Density of information gain

Given two probability measures μ and ν on a measure space (Ω, Σ), the *information gain* (or *Kullback-Leibler information*, or minus the *relative entropy*) of μ relative to ν is defined as

$$I(\mu|\nu) \;=\; \begin{cases} \displaystyle\int \left(\log \tfrac{d\mu}{d\nu}\right) d\mu \;=\; \int \left(\tfrac{d\mu}{d\nu} \log \tfrac{d\mu}{d\nu}\right) d\nu & \text{if } \mu \ll \nu \\ +\infty & \text{otherwise} \end{cases} \tag{8}$$

This object determines the probability of a *large deviation*. Roughly, if $X_1, \ldots, X_n$ are independent samples drawn from a probability distribution ν, then

$$\mathrm{Prob}_\nu(X_1, \ldots, X_n \text{ is typical for } \mu) \sim e^{-nI(\mu|\nu)} . \tag{9}$$

Two important properties of the information gain are:

(I1) $I(\mu|\nu) = 0$ if and only if $\mu = \nu$. This, together with the preceding formula, suggests the interpretation of $I(\mu|\nu)$ as a measure of "how different" μ and ν are. [Note, however, that $I(\mu|\nu) \neq I(\nu|\mu)$ in general.]

(I2) If $\mathcal{F}' \subset \mathcal{F}$, $I(\mu_{\mathcal{F}'}|\nu_{\mathcal{F}'}) \leq I(\mu_{\mathcal{F}}|\nu_{\mathcal{F}})$. That is, "coarser" measurements are less revealing.

The notion of information gain is, however, useless in the context of infinite-volume statistical-mechanical systems, because nearly all the measures of interest turn out to be mutually singular. Therefore, by the second line in (8), the relative information gain is always $+\infty$. However, the reason for this infinity is well known: When two translation-invariant states are restricted to a finite "window" Λ, the relative information gain is typically proportional (asymptotically) to the volume of Λ. The right object in the statistical-mechanical setting is, therefore, the *density of information gain*, in which this volume divergence is divided out by defining

$$i(\mu|\nu) = \lim_{\Lambda \nearrow \mathcal{L}} \frac{1}{|\Lambda|} I(\mu_\Lambda|\nu_\Lambda) , \tag{10}$$

where $\mu_\Lambda = \mu|_{\mathcal{F}_\Lambda}$, and the sequence of sets Λ is suitably chosen (e.g. growing concentric cubes).

The limit (10) is known to exist if ν is a Gibbs measure and both μ and ν are translation-invariant. For such measures we have, instead of (9):

$$\mathrm{Prob}_\nu(\omega_\Lambda \text{ is typical for } \mu) \sim e^{-|\Lambda|\, i(\mu|\nu)} ,$$

where ω_Λ represents a configuration of spins inside the finite volume Λ.

An important difference between information gain and its density is that for the latter the analogue of (I1) is *false*. In fact, if $\nu \in \mathcal{G}_{\mathrm{inv}}(\Pi^\Phi)$,

$$i(\mu|\nu) = 0 \iff \mu \in \mathcal{G}_{\mathrm{inv}}(\Pi^\Phi) , \tag{11}$$

where we have denoted by $\mathcal{G}_{\mathrm{inv}}(\Pi^\Phi)$ the set of *translation-invariant* Gibbs states consistent with Π^Φ. Indeed, for different Gibbs states consistent with the same specification, the large deviations decay subexponentially in the volume, typically as an exponential of the perimeter (or the surface area) of Λ.

3.3 Attributes of non-Gibbsian measures

Examples of non-Gibbsian measures have previously been found and studied in:

- Zero-temperature statistical mechanics (see the review by Dobrushin and Shlosman [16] and Appendix B of [57]).

- Non-equilibrium phenomena, especially stationary states of stochastic dynamical processes [45, 54, 49].

This is not surprising, because there is no reason to expect Gibbsian measures outside the setting for which they were tailored, namely *nonzero-temperature equilibrium* statistical mechanics. What is tantalizing is the fact that even in this setting non-Gibbsian measures may also show up. The first such example was found by Lebowitz and Maes [43], in their study of the entropic repulsion of a surface by a wall. Later, it was discovered that the application of renormalization techniques also leads to non-Gibbsianness in some situations where a non-compact symmetry group is broken [17], or in the presence of a first-order phase transition [38, 58, 59, 60, 61, 57]. We review the last type of examples in Section 5.2 below.

Let us present a tentative classification of the known examples of non-Gibbsian measures — based, for lack of better understanding, on which of the Gibbsian properties (G1)–(G4) are violated.

~~(G1)~~ A convex combination of measures that are Gibbsian for different specifications (e.g. for different temperatures) cannot be Gibbsian. If it were, it would be consistent with all the different specifications at the same time, against (G1).

~~(G2)~~ Measures that fail to be uniformly non-null are non-Gibbsian. They come in two flavors:

- Measures that are not even non-null: they give zero measure to some open set. This is a mild case of non-Gibbsianness. Often Gibbsianness can be restored by introducing hard-core (excluded-volume) interactions to account for the forbidden configurations [50, 2, 53].

- Measures that are non-null but non-uniformly so. This behavior is usually associated to some unbounded interaction; it is the typical situation for long-range models with unbounded spins. Sometimes Gibbsianness can be restored by excluding "by hand" a well-chosen set of "catastrophic" configurations [44, 10].

~~(G3)~~ Measures that do not have any system of quasilocal conditional probabilities cannot be Gibbsian. This is associated to the presence of some "hidden" variables that transmit information from arbitrarily far away even when the "non-hidden" spins are fixed. This is the type of non-Gibbsianness that appears through renormalization transformations, as will be discussed below.

~~(G4)~~ Measures that do not satisfy the large-deviation estimates proven for Gibbs measures cannot be Gibbsian. There are two types of examples:

- Measures whose large-deviation probability is too large. Among these we find the stationary measures of some dynamic processes (voter model [45], Martinelli-Scoppola cluster dynamics [49]), and the sign field of an (an)harmonic crystal [17, 57]. In these cases it is proven that $i(\delta_+|\mu) = 0$, hence, as δ_+ is not Gibbsian, by (11) neither is μ.

- Measures whose large-deviation probability is too small. One example is the restriction of a 2-dimensional Ising state (below the critical temperature) to a coordinate axis, first studied by Schonmann [54]. Schonmann proves that two such states, arising from the + and − phases of the Ising model at the same temperature, cannot be compatible with different specifications; but the density of information gain between them is strictly positive. By (11), if there is a common specification, it cannot be Gibbsian. The same conclusions are valid for restrictions of d-dimensional low-temperature Ising states to $(d-1)$-dimensional coordinate hyperplanes [18].

Besides the examples and results collected here, very little is known about non-Gibbsian measures. We shall comment more about this state of ignorance in Section 6 below.

4 Position-space renormalization transformations. The questions

After this short encounter with the key notions of statistical mechanics, we return to the main subject of this work: the renormalization transformations. For the rest of this paper we shall consider only translation-invariant measures.

4.1 Definition and examples

As discussed in Section 1, a renormalization transformation, in the framework of classical lattice statistical mechanics, is defined by a probability kernel T from a measure space $(\Omega = \Omega_0^{\mathcal{L}}, \mathcal{F})$, called the *original* or *object system*, to another measure space $(\Omega' = (\Omega_0')^{\mathcal{L}'}, \mathcal{F}')$ called the *image* or *renormalized system*. This kernel defines a transformation $\mathcal{T}: M_{+1}(\Omega, \mathcal{F}) \to M_{+1}(\Omega', \mathcal{F}')$ in the form

$$(\mathcal{T}\mu)(\,.\,) \;=\; \int T(\,.\,|\omega)\,\mu(d\omega)\,.$$

Such a transformation is called *strictly local* if there exists a $K < \infty$ (compression factor) such that

For each $A \in \mathcal{F}'_{\Lambda'}$, the function $T(A|\cdot)$ is $\mathcal{F}_\Lambda$-measurable for a suitable (van Hove) sequence of volumes Λ, Λ' with $\limsup \dfrac{|\Lambda|}{|\Lambda'|} \le K$.

Informally, the renormalized configurations in Λ' are determined by a volume of original spins not exceeding $K|\Lambda'|$.

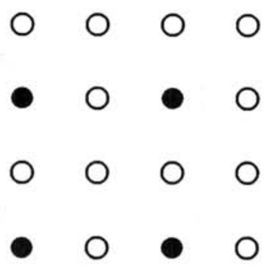

Figure 1: Decimation with block of size 2×2. Black circles represent renormalized spins; the spins corresponding to white circles are averaged over.

Often T has the structure of a product:

$$T(d\omega'|\omega) \;=\; \prod_{x\in\mathcal{L}'} \tilde{T}(d\omega'_x|\omega_{B_x})\;;$$

that is, each renormalized spin ω'_x is determined only by original spins in a finite set B_x. In this case ω'_x is interpreted as the *block spin* associated to the "block" B_x.

In addition, we demand that $\mathcal{T}$ carry translation-invariant measures into translation-invariant measures.

Examples:

The following are among the most popular renormalization transformations used for lattice systems (see e.g. [7]):

- *Decimation transformations*: $\mathcal{L} = \mathcal{L}' = \mathbb{Z}^d$, $\Omega_0 = \Omega'_0$, $B_x = b\times b$ blocks, and

$$\tilde{T}(\omega_{B_x}|d\omega'_x) = \delta(\omega'_x - \omega_{bx})\, d\omega'_x\;.$$

  These transformations are deterministic: A particular spin of the block is chosen as the renormalized spin, ignoring ("integrating over") the configurations of the other spins (Figure 1).

- *Majority-rule transformations* (Ising model): $\mathcal{L} = \mathcal{L}' = \mathbb{Z}^d$, $\Omega_0 = \Omega'_0 = \{-1,1\}$, $B_x = b$-translate of a fixed block B_0 (usually of size small enough so that no two blocks overlap), and

$$\tilde{T}(\omega_{B_x}|d\omega'_x) =$$
$$\left[\chi\Bigl(\sum_{y\in B_x}\omega_y > 0\Bigr)\,\delta(\omega'_x - 1) + \chi\Bigl(\sum_{y\in B_x}\omega_y < 0\Bigr)\,\delta(\omega'_x + 1)\right.$$

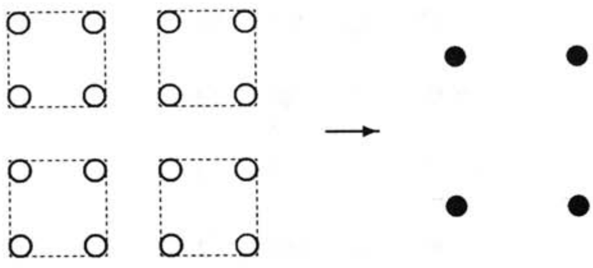

Figure 2: General renormalization transformations with block-size 2×2.

$$+ \chi\left(\sum_{y\in B_x} \omega_y = 0\right)\frac{1}{2}\Big(\delta(\omega'_x - 1) + \delta(\omega'_x + 1)\Big)\Bigg]\, d\omega'_x \, .$$

In words, the renormalized spin takes the value of the majority of the spins of the block; in case of ties a coin is flipped. This last part of the prescription makes the transformation (slightly) stochastic (Figure 2).

- *Kadanoff transformations* (Ising model): Same setting as in previous example, and

$$\check{T}(\omega_{B_x}|d\omega'_x) = \frac{\exp\left(p\,\omega'_x \sum_{y\in B_x}\omega_y\right)}{2\cosh\left(p\sum_{y\in B_x}\omega_y\right)}\left[\frac{1}{2}\Big(\delta(\omega'_x - 1) + \delta(\omega'_x + 1)\Big)\right] d\omega'_x \, ,$$

for some fixed $p > 0$. These are fully stochastic transformations, which converge to the majority-rule transformation in the limit $p \to \infty$. These transformations have been used also in image processing (p related to the probability of "data corruption" at each spin site) [23, 25, 11, 28, 20] and speech recognition [52]. In these contexts, the renormalized measures are called ***hidden Markov models*** [52, 24, 46].

- *Block averaging*: $\mathcal{L} = \mathcal{L}' = \mathbb{Z}^d$, $B_x = b$-translate of a fixed block B_0 and

$$\check{T}(\omega_{B_x}|d\omega'_x) = \delta\left(\omega'_x - c\sum_{y\in B_x}\omega_y\right) d\omega'_x \, ,$$

where c is a rescaling factor. If $c = |B_0|^{-1}$, the renormalized spin is indeed the block-average of the original ones. We observe that $\Omega'_0 \neq \Omega_0$; for instance if $\Omega_0 = \{-1, 1\}$ and $c = 1$, then $\Omega'_0 = \{-|B_0|, -|B_0| + 2, \ldots, |B_0| - 2, |B_0|\}$, This is an example of a *linear* transformation, and as such it is of independent interest because of its connections with central-limit theorems [36, 4, 22, 13, 8].

Non-examples:

The following important transformations are *not* strictly local:

- Quasi-local transformations (e.g. momentum-space transformations [62, 3]):

$$\check{T} = \delta \left(\omega'_x - \sum_{y \in \mathcal{L}} F(bx - y)\omega_y \right) d\omega'_x$$

  for some length rescaling factor $b > 1$ and some non-local kernel F decreasing at infinity. These transformations are widely used in the study of systems with unbounded spins.

- Restriction of a d-dimensional model to a $(d-1)$-dimensional coordinate hyperplane. The compression factor K is infinite. These transformations are relevant to the study of stationary states of cellular automata [29, 54].

4.2 Renormalization transformations at the level of interactions

To study the properties of a statistical-mechanical system, the transformations $\mathcal{T}$ are applied iteratively to a starting Gibbs measure μ. In this way, we "integrate out" the "fine" details of μ, and concentrate attention on the important "long-distance" features.

In fact, this type of transformation was designed initially (by Kadanoff, Wilson and others) to study the so-called *critical points*, which correspond to situations where the fluctuations exhibit long-range correlations. At these points, the behavior has been observed to be *universal* in the sense that it is the same for different models, depending only on some very gross features like the dimension, type of symmetry, and whether the interaction is finite-range. Moreover, this universal behavior can be fully characterized by a reduced number of parameters — the so-called critical exponents — which relate the divergence of the correlation length and other observables to the deviation from criticality of the temperature and other parameters *of the interaction* (fields or couplings). Therefore, from the point of view of physics, it is of more direct interest to see how *the interaction* changes upon renormalization.

Thus, physicists usually call a renormalization transformation the map $\mathcal{R}$ induced at the level of interactions, according to the diagram

$$\begin{array}{ccc} \mu & \xrightarrow{\mathcal{T}} & \mu' \\ \uparrow & & \downarrow \\ \Phi & \xrightarrow{\mathcal{R}} & \Phi' \end{array}$$

This approach, however, relies on the assumption that *the renormalized measure is Gibbsian* — that is, that there *exists* some summable interaction Φ' such that $\mu' \in \mathcal{G}(\Pi^{\Phi'})$. This assumption turns out to be *false*, in certain situations, for the transformations of the preceding examples.

4.3 The questions

In addition to the hypothesis of Gibbsianness, renormalization physicists-style involves the assumption that $\mathcal{R}$ is a *single-valued* and *smooth* map. This is a delicate issue in the presence of a *first-order phase transition*, i.e. when $|\mathcal{G}(\Pi^{\Phi})| > 1$:

Indeed, in this case each of the measures in $\mathcal{G}(\Pi^{\Phi})$ gives rise to a *different* renormalized measure, and it is not obvious that they are all consistent with the *same* renormalized interaction Φ'. Hence, assuming Gibbsianness of the renormalized measures, we have two possible situations:

$$\begin{array}{ccc} \mu & \overset{\mathcal{T}}{\substack{\longrightarrow \\ \longrightarrow \\ \longrightarrow}} & \mu' \\ \uparrow\uparrow\uparrow & & \downarrow \\ \Phi & \xrightarrow{\mathcal{R}} & \Phi' \end{array} \qquad \text{or} \qquad \begin{array}{ccc} \mu & \overset{\mathcal{T}}{\substack{\longrightarrow \\ \longrightarrow \\ \longrightarrow}} & \mu' \\ \uparrow\uparrow\uparrow & & \downarrow\downarrow\downarrow \\ \Phi & \overset{\mathcal{R}}{\substack{\longrightarrow \\ \longrightarrow \\ \longrightarrow}} & \Phi' \end{array} \qquad ?$$

If the alternative on the right occurs, then the map $\mathcal{R}$ is *multivalued* at the first-order phase transition; and moreover, it is presumably *discontinuous* with respect to small changes in Φ that select one or the other of the original Gibbs measures μ. These "pathologies" would be absent in the alternative on the left — the so-called *standard scenario*.

Physicists have long assumed that the left alternative is the correct one (that is why it is called the standard scenario!) [51, 41, 19]; however, a series of numerical results [6, 42, 14, 30] suggested that discontinuities and multivaluedness did occur in certain cases.

In addition, some analytic work of Griffiths and Pearce [32, 33, 31] pointed out the almost unavoidable occurrence of "peculiarities" for the most common renormalization transformations. The nature of these "peculiarities" was later clarified by Israel [38] — at least for the 2×2 decimation transformation of the Ising model — who showed that the issue was non-Gibbsianness rather than the lack of single-valuedness or smoothness.

To summarize, renormalization theory faced two big questions:

- (*Raised by physicists.*) Is the standard scenario valid?—Is the renormalization transformation $\mathcal{R}$ always *single-valued* and *continuous*?

- (*Raised by mathematical physicists.*) Is the renormalized measure always Gibbsian?—Is the map $\mathcal{R}$ always *defined*?

5 The answers

The answers to the above questions are: a qualified "yes" for the first one, and "no" for the second one. In fact:

- The map $\mathcal{R}$ is single-valued and continuous (in an appropriate topology) *whenever it is defined*.

- Many examples show that the map $\mathcal{R}$ can *fail to be defined* at or in the vicinity of a first-order phase transition.

We present in this section a brief discussion of these issues. For further details the reader should consult [57]. (Other summaries, oriented towards an audience of physicists, have appeared in [58, 59, 60, 61].)

5.1 Validity of the standard scenario

The proof that the standard scenario is valid, whenever the renormalized measure is Gibbsian, is contained in two "fundamental theorems". The first of them deals with the issue of multivaluedness.

> **First fundamental theorem** (single-valuedness). *If μ and ν are Gibbs measures for the same interaction, then either $\mathcal{T}\mu$ and $\mathcal{T}\nu$ are both non-Gibbsian, or else there exists a unique (modulo physical equivalence) interaction Φ' for which both $\mathcal{T}\mu$ and $\mathcal{T}\nu$ are Gibbs measures.*

SKETCH OF THE PROOF. The proof is based on the following two facts:

- One can determine whether two measures (at least one of which is Gibbsian) are Gibbsian for the *same* interaction by inspecting the density of information gain between them: according to (11), this quantity is zero if the two measures are both Gibbsian for the *same* interaction, and is nonzero otherwise.

- The information gain does not increase when the measures are restricted to a "coarser" σ-algebra.

With these observations, the proof basically reduces to the following line:

$$\mu, \nu \in \mathcal{G}_{\rm inv}(\Pi^{\Phi}) \implies i(\mu|\nu) = 0 \implies i(\mathcal{T}\mu|\mathcal{T}\nu) = 0 \, . \tag{12}$$

(In the last implication the finiteness of the compression factor is needed.)

From the rightmost identity in (12) we conclude that there are only two possibilities:

- One of $\mathcal{T}\mu$, $\mathcal{T}\nu$ is Gibbsian. In this case, by (11), the other one is also Gibbsian, and for the same interaction. (The uniqueness modulo physical equivalence of this interaction is a well-known result due to Griffiths and Ruelle [34].)
- Neither $\mathcal{T}\mu$ nor $\mathcal{T}\nu$ is Gibbsian. ■

We remark that whenever $\mathcal{T}\mu$ and $\mathcal{T}\nu$ are non-Gibbsian, the last implication of (12) provides, as a by-product, examples of non-Gibbsian measures for which the density of information gain exists and is zero. On the other hand, Schonmann's restriction [54] provides examples of non-Gibbsian measures for which the density of information gain exists and is strictly positive.

Regarding continuity of the renormalization map at the level of interactions, we have:

Second fundamental theorem (continuity). *The map $\mathcal{R}$ is continuous in the $\mathcal{B}^0$-norm (in fact, Lipschitz continuous) on the domain where it is defined.*

SKETCH OF THE PROOF. The $\mathcal{B}^0$-norm is defined as

$$||\Phi||_{\mathcal{B}^0} \stackrel{\text{def}}{=} \sum_{\substack{\text{finite } A \subset \mathcal{L} \\ A \ni 0}} \frac{||\Phi_A||_\infty}{|A|} .$$

It defines a larger interaction space than the $\mathcal{B}^1$-norm defined in (6) (it allows interactions that are more strongly multi-body), and it is the natural norm for measuring *bulk* energy contributions. In particular one has the very important *identity*

$$\left\| \log \frac{\pi_\Lambda^{\Phi_1}}{\pi_\Lambda^{\Phi_2}} \right\|_{\infty/\text{const.}} = |\Lambda| \, ||\Phi_1 - \Phi_2||_{\mathcal{B}^0/\text{p.e.}} + o(|\Lambda|) , \tag{13}$$

where "p.e." (short for "physical equivalence") indicates that one must identify functions differing by constants or by functions having zero mean with respect to all translation-invariant measures. This identity proves the uniqueness result of Griffiths and Ruelle. On the other hand, since probability kernels act like averages,

$$\left\| \log \frac{\pi_\Lambda^{\Phi'_1}}{\pi_\Lambda^{\Phi'_2}} \right\|_{\infty/\text{Const.}} \leq \left\| \log \frac{\pi_\Lambda^{\Phi_1}}{\pi_\Lambda^{\Phi_2}} \right\|_{\infty/\text{Const.}} . \tag{14}$$

Combining (13) and (14), and using the finiteness of the compression factor K, we get

$$\|\Phi_1' - \Phi_2'\|_{\mathcal{B}^0/\mathrm{p.e.}} \;\leq\; K\, \|\Phi_1 - \Phi_2\|_{\mathcal{B}^0/\mathrm{p.e.}} \,,$$

which is the precise statement of the second fundamental theorem. ∎

5.2 Non-Gibbsianness of the renormalized measures

We turn now to the issue of non-Gibbsian renormalized measures, that is to situations in which the map $\mathcal{R}$ — renormalization at the level of interactions — is not defined. As mentioned in Section 4.3, Israel [38] exhibited the first example of this phenomenon. Generalizing his idea, we have been able to show that all the real-space renormalization transformations listed above — decimation, Kadanoff, (some) majority-rule, and (even-block) block-averaging — yield instances of non-Gibbsianness. That is, when these transformations are applied to the Ising model we have that:

> *For $J_{xy} = J$ large enough (low temperature) and $h_x = h$ small enough,* one *renormalization step leads to a non-Gibbsian renormalized measure.*

More precisely, this statement holds for all the above transformations except majority rule, for $d \geq 3$. For $d = 2$ the result holds for all the examples, but (we have proven it so far) only for $h = 0$. For block averaging we have a much stronger result: non-Gibbsianness occurs for arbitrary h (and low temperature), for all $d \geq 2$.

This non-Gibbsianness is caused by the mechanism labeled ~~(G3)~~ in Section 3.3. The ingredients of the argument are:

- Once the renormalized spins are fixed, the original spins form a constrained system. Nevertheless, at least for some special renormalized-spin configurations, the constraints allow for a considerable amount of fluctuations.

- These remanent fluctuations of the original spins act as "hidden" variables that transmit information from infinity.

- For this transmission to take place, the *constrained* original system must undergo a first-order phase transition. Usually (but not always) this means that the (unconstrained) original system must be at low temperature and small field. However, the original system *need not* be sitting *on* a first-order phase-transition surface; it need only be "near enough" to one so that the *constrained* original system can be placed onto a first-order phase-transition surface by a suitable choice of renormalized-spin configuration.

We remark that in image-processing language this shows that, in the limit of infinite window, the noisy image may fail to be Gibbsian. This may lead to surprises when using inference schemes designed with Gibbs measures in mind. At this stage, however, it is not clear what are the surprises, if any, in store. Analogously, using speech-recognition nomenclature, the example of the Kadanoff transformation shows that hidden Markov models can be very far from Markovian: in fact they may even fail to be Gibbsian.

6 More questions and open problems

By now, we hope to have convinced the reader of the ubiquity of non-Gibbsian measures. This fact, along with the present lack of systematic study, makes the subject of non-Gibbsianness a particularly appealing field of investigation. In order to suggest directions for further research, we close this exposition with a list of questions and open problems that we consider especially important.

6.1 Practical consequences of non-Gibbsianness?

We have seen that non-Gibbsianness can be unexpectedly present in very concrete applied problems, such as the processing of noisy images or speech, or as a result of widely used computational procedures, like renormalization transformations. Therefore, there is a rather pressing need to investigate the practical consequences of this phenomenon. In this regard, it is natural to formulate the following question:

> **How does one detect, numerically, that one is working with a non-Gibbsian measure?**

In other words, which are the possible "surprises" mentioned above that could hit a Gibbsianness-oriented statistician (or statistical physicist) when his/her target turns out not to be a Gibbs measure? This is probably a vast and difficult question, but we would like to contribute with some meditations.

The question fits into a more general parameter-estimation problem: Suppose one has a measure μ and one makes a numerical experiment under the wrong assumption that it belongs to a certain family $\mathcal{G}$. If the experiment is well done, it will pick up the measure $\mu^{\mathcal{G}} \in \mathcal{G}$ "closest" to μ. The key issue here is how this "closest" measure is defined, or put in another way, what "closeness" means in this context.

One reasonable answer could be: The "closest" measure should in practice be determined via some optimal estimation method, for instance maximum likelihood. Now, in the ideal limit of an infinite random sample, the maximum-likelihood estimate converges to the minimizer of the information gain $I(\mu|\,\cdot\,)$ (note the order of arguments!) [35]. Therefore, this information gain could provide a possible measure of "closeness". For the theory of random fields or processes, however, such an approach does not work because in the infinite-volume limit the information gain diverges. In view of the discussion of Section 3.2, it is natural then to resort to the *density* of information gain. We are therefore led to the following reasonable postulate:

> *The measure(s) $\mu^{\mathcal{G}}$ closest to μ is (are) defined to be the one(s) which minimize $i(\mu|\cdot)$.*

Now, in a realistic experiment, such a minimizer is determined via successive approximations. In our case, where $\mathcal{G}$ is the set of Gibbsian measures, each approximation involves a successively finer determination of the putative renormalized interaction. Hence in our situation the last question can be transcribed in the form:

> **Consider an increasing sequence of subspaces $V_1 \subset V_2 \subset \cdots$ whose union is dense in $\mathcal{B}^1$, and let Φ_n be the interaction of the measure in $\bigcup_{\Phi \in V_n} \mathcal{G}(\Pi^{\Phi_n})$ closest to μ in the above sense. What are the properties of the sequence $\{\Phi_n\}$?**

In this regard, we offer two conjectures:

- If $\mu \in \mathcal{G}(\Pi^{\Phi})$, then $\Phi_n \to \Phi$ in $\mathcal{B}^1$. A partial result is proven in [57]: There is indeed a converging sequence $\{\widetilde{\Phi}_n\} \to \Phi$ of *almost* minimizers, but it is not known whether the *exact* minimizers Φ_n always converge.
- If μ is not Gibbsian, then $||\Phi_n||_{\mathcal{B}^1} \to \infty$. A preliminary analysis done in [57] shows that there is another possibility to contend with, namely that the sequence $\{\Phi_n\}$ does not converge at all (due either to oscillations or to a mean-field-type dependence of the couplings).

At this point it may be useful to remember the numerical discontinuities and multivaluedness apparently detected in the case of renormalization transformations [6, 42, 14, 30]. If the renormalized measures are indeed Gibbsian, such phenomena are ruled out by the "fundamental theorems" presented in Section 5.1; the *apparent* discontinuity must be an artifact of the truncation to a small subspace V_n, and it ought to disappear as $n \to \infty$. On the other hand, these discontinuities could be a manifestation of non-Gibbsianness. In [1] a toy example is presented where a relation between discontinuities/multivaluedness and non-Gibbsianness is explicitly exhibited, although for non-Gibbsianness due to lack of non-nullness (mechanism ~~(G2)~~ of Section 3.3) rather than due to lack of quasilocality. This example, however, does not really belong to a probabilistic setting because it involves complex interactions.

6.2 "Degrees" of non-Gibbsianness?

A question, not unrelated to the ones of the previous subsection, is whether one can establish some sort of hierarchy of non-Gibbsianness. A suggestive analogy (due to Joel Lebowitz) is provided by the irrational numbers, which can be classified according to the rate at which they are approximated by rationals (Diophantine approximation). If a similar classification were possible for non-Gibbsianness, it would certainly be more useful than the one proposed in Section 3.3.

Such program has the intrinsic difficulty that there is no unique way to estimate "rate of approximation" in measure spaces. Perhaps a more practical approach would be to characterize non-Gibbsianness according to the severity of the concrete, e.g. numerical, manifestations. In this regard, the following questions could serve as guidelines:

- When can Gibbsianness be restored by removing "by hand" some small set of "pathological" configurations? Some aspects to consider:
  - The set of "pathological" configuration is quite likely a tail event, and therefore it has measure either one or zero for the measures of interest (*extremal* Gibbs measures and their images under renormalization). These two possibilities could be interpreted respectively as signaling a "large" or a "small" set.
  - In general, it is foreseeable that one can find *some* set of configurations whose removal yields a Gibbsian measure. The question is whether this can be accomplished by removing only a *small* set of configurations. Current investigations on Schonmann's example could be illustrative in this regard [48, 47, 18].
- When is non-Gibbsianness so weak that no realistic numerical experiment will detect it? One such case would of course happen if the set of "pathological" configurations has measure zero and its removal restores Gibbsianness.
- Under what conditions there is still some sort of "thermodynamic description" (density of information gain, variational principle) also for non-Gibbsian measures? When such a description exists, one could perhaps extend to non-Gibbsian measures many of the concepts — and perhaps also the intuitions — developed for Gibbsian measures. See [45] for some pioneer work in this direction.

6.3 Pervasiveness of non-Gibbsianness?

Given the growing fauna of non-Gibbsian measures, one may certainly wonder whether there is some convincing way to estimate how "large" the set of such measures is. The question is perhaps a little vague and typically academic; the two partial answers we have are of the same nature:

- The set of Gibbsian measures is dense in $M_{+1}(\Omega, \mathcal{F})$ in the weak topology [57]. This is a weak result, because, as commented above, the weak topology is insensitive to long-range-order properties.

- The set of Gibbsian measures is a set of first Baire category (countable union of nowhere dense sets) in the space $M_{+1}(\Omega, \mathcal{F})$. In this sense, Gibbsian measures are "exceptional". This has recently been proven by Israel (unpublished).

6.4 Extension to unbounded spins

Finally, we mention as an open problem the extension of all the considerations of the present work to the case of unbounded spins (Ω_0 non-compact). This is indeed an extremely interesting topic, but it faces numerous difficulties; among them:

- Unless the interaction is strictly finite-range,
  - One faces the problem of defining-out "catastrophic" configurations that lead to meaningless Boltzmann weights (i.e. divergence of the sum defining H_Λ^Φ). This is at present a painful model-dependent process.
  - The Gibbs measures are not quasilocal.
- It is not clear what is the "largest reasonable" space of interactions.
- The theory of large deviations is still in the making [15, 26].

Acknowledgments

R.F. thanks the organizers of the V CLAPEM for the invitation to present these results, and to the Instituto de Matemática e Estatística of Universidade de São Paulo for the hospitality during the conference. The research of the first author (A.C.D.v.E.) has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences (KNAW). The research of the second author (R.F.) was supported by the Fonds National Suisse. The research of the third author (A.D.S.) was supported in part by U.S. National Science Foundation grant DMS-9200719 and by U.S. Department of Energy contract DE-FG02-90ER40581.

References

[1] M. Asorey, J. G. Esteve, R. Fernández, and J. Salas. Rigorous analysis of renormalization group pathologies in the 4-state clock model. *Nucl. Phys.*, B392:593–618, 1993.

[2] M. B. Averintsev. Gibbs description of random fields whose conditional probabilities may vanish. *Probl. Inform. Transmission*, 11:326–334, 1975.

[3] M. J. Bissett. A perturbation-theoretic derivation of Wilson's 'incomplete integration' theory. *J. Phys. C*, 6:3061–3070, 1973.

[4] P. M. Bleher and P. Major. Critical phenomena and universal exponents in statistical physics. On Dyson's Hierarchical model. *Ann. Prob.*, 15:431–477, 1987.

[5] P. M. Bleher and Ja. G. Sinai. Investigation of the critical point in models of the type of Dyson's hierachical models. *Commun. Math. Phys.*, pages 23–42, 1973.

[6] H. W. J. Blöte and R. H. Swendsen. First-order phase transitions and the three-state Potts model. *Phys. Rev. Lett.*, 43:799–802, 1979.

[7] T. W. Burkhardt and J. M. J. van Leeuwen, editors. *Real-Space Renormalization*. Springer-Verlag, Berlin–Heidelberg–New York, 1982.

[8] E. A. Carlen and A. Soffer. Entropy production by block spin summation and central limit theorems. *Commun. Math. Phys.*, 140:339–371, 1992.

[9] M. Cassandro and G. Jona-Lasinio. Critical point behaviour and probability theory. *Adv. Phys.*, 27:913–941, 1978.

[10] M. Cassandro, E. Olivieri, A. Pellegrinotti, and E. Presutti. Existence and uniqueness of DLR measures for unbounded spin systems. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 41:313–334, 1978.

[11] F. S. Cohen and D. B. Cooper. Simple parallel hierarchical and relaxation algorithms for segmenting noncausal Markovian random fields. *IEEE Trans. Pattern Anal. Machine Intell.*, 9:195–219, 1987.

[12] F. Comets and B. Gidas. Parameter estimation for Gibbs distributions from partially observed data. *Ann. Appl. Prob.*, 2:142–170, 1992.

[13] J. de Coninck and C. M. Newman. The magnetization-energy scaling limit in high dimension. *J. Stat. Phys.*, 59:1451–1467, 1990.

[14] K. Decker, A. Hasenfratz, and P. Hasenfratz. Singular renormalization group transformations and first order phase transitions (II). Monte Carlo renormalization group results. *Nucl. Phys.*, B295 [FS21]:21–35, 1988.

[15] J.-D. Deuschel, D. W. Stroock, and H. Zessin. Microcanonical distributions for lattice gases. *Commun. Math. Phys.*, 139:83–101, 1991.

[16] R. L. Dobrushin and S. B. Shlosman. The problem of translation invariance of Gibbs states at low temperatures. *Sov. Sci. Rev. C Math. Phys.*, 5:53–196, 1985.

[17] T. C. Dorlas and A. C. D. van Enter. Non-Gibbsian limit for large-block majority-spin transformations. *J. Stat. Phys.*, 55:171–181, 1989.

[18] R. Fernández and C. E. Pfister. Work in progress.

[19] M. E. Fisher and A. N. Berker. Scaling for first-order phase transitions in thermodynamic and finite systems. *Phys. Rev.*, B26:2507–2513, 1982.

[20] A. Frigessi and M. Piccioni. Parameter estimation for two-dimensional Ising fields corrupted by noise. *Stoch. Proc. Appl.*, 34:297–311, 1990.

[21] G. Gallavotti and G. Jona-Lasinio. Limit theorems for multidimensional Markov processes. *Commun. Math. Phys.*, pages 301–307, 1975.

[22] K. Gawędzki and A. Kupiainen. A rigorous block spin approach to massless lattice theories. *Commun. Math. Phys.*, 77:31–64, 1980.

[23] D. Geman. Random fields and inverse problems in imaging. In *Ecole d'Eté de Probabilités de Saint-Flour XVIII – 1988*, pages 116–193. Springer-Verlag (Lecture Notes in Mathematics #1427), Berlin–Heidelberg–New York, 1990.

[24] S. Geman. Hidden Markov models for image analysis. Lecture at Istituto per le Applicazioni del Calcolo, Roma (July 1990).

[25] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6:721–741, 1984.

[26] H.-O. Georgii. Large deviations and maximum entropy principle for interacting random fields on $\mathbb{Z}^d$. To appear in *Ann. Prob.*

[27] H.-O. Georgii. *Gibbs Measures and Phase Transitions.* Walter de Gruyter (de Gruyter Studies in Mathematics, Vol. 9), Berlin–New York, 1988.

[28] B. Gidas. A renormalization group approach to image processing problems. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11:164–180, 1989.

[29] S. Goldstein, R. Kuik, J. L. Lebowitz, and C. Maes. From PCA's to equilibrium systems and back. *Commun. Math. Phys.*, 125:71–79, 1989.

[30] A. Gonzalez-Arroyo, M. Okawa, and Y. Shimizu. Monte Carlo renormalization-group study of the four-dimensional Z_2 gauge theory. *Phys. Rev. Lett.*, 60:487–490, 1988.

[31] R. B. Griffiths. Mathematical properties of renormalization-group transformations. *Physica*, 106A:59–69, 1981.

[32] R. B. Griffiths and P. A. Pearce. Position-space renormalization-group transformations: Some proofs and some problems. *Phys. Rev. Lett.*, 41:917–920, 1978.

[33] R. B. Griffiths and P. A. Pearce. Mathematical properties of position-space renormalization-group transformations. *J. Stat. Phys.*, 20:499–545, 1979.

[34] R. B. Griffiths and D. Ruelle. Strict convexity ("continuity") of the pressure in lattice systems. *Commun. Math. Phys.*, 23:169–175, 1971.

[35] P. J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In L. M. LeCam and J. Neyman, editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. I*, pages 221–233. University of California Press, Berkeley–Los Angeles, 1967.

[36] D. Iagolnitzer and B. Souillard. Random fields and limit theorems. In J. Fritz, J. L. Lebowitz, and D. Szász, editors, *Random Fields (Esztergom, 1979), Vol. II*, pages 573–591. North-Holland, Amsterdam, 1981.

[37] R. B. Israel. *Convexity in the Theory of Lattice Gases.* Princeton Univ. Press, Princeton, N.J., 1979.

[38] R. B. Israel. Banach algebras and Kadanoff transformations. In J. Fritz, J. L. Lebowitz, and D. Szász, editors, *Random Fields (Esztergom, 1979), Vol. II*, pages 593–608. North-Holland, Amsterdam, 1981.

[39] G. Jona-Lasinio. The renormalization group: A probabilistic view. *Nuovo Cim.*, pages 99–119, 1975.

[40] G. Jona-Lasinio. Large fluctuations of random fields and renormalization group: Some perspectives. In J. Fröhlich, editor, *Scaling and Self-similarity in Physics*, Boston–Basel–Stuttgart, 1983. Birkhäuser.

[41] W. Klein, D. J. Wallace, and R. K. P. Zia. Essential singularities at first-order phase transitions. *Phys. Rev. Lett.*, 37:639–642, 1976.

[42] C. B. Lang. Renormalization study of compact $U(1)$ lattice gauge theory. *Nucl. Phys.*, B280 [FS18]:255–275, 1987.

[43] J. L. Lebowitz and C. Maes. The effect of an external field on an interface, entropic repulsion. *J. Stat. Phys.*, 46:39–49, 1987.

[44] J. L. Lebowitz and E. Presutti. Statistical mechanics of unbounded spin systems. *Commun. Math. Phys.*, 50:195–218, 1976.

[45] J. L. Lebowitz and R. H. Schonmann. Pseudo-free energies and large deviations for non-Gibbsian FKG measures. *Prob. Th. Rel. Fields*, 77:49–64, 1988.

[46] B. G. Leroux. Maximum-likelihood estimation for hidden Markov models. *Stoch. Proc. Appl.*, 40:127–143, 1992.

[47] J. Lőrinczi. Some results on the projected two-dimensional Ising model. In preparation for the Proceedings of the NATO Advanced Workshop 'On Three Levels', July 19–24, 1993, Leuven, Belgium.

[48] C. Maes and K. van de Velde. Defining relative energies for the projected Ising measure. *Helv. Phys. Acta*, 65:1055–1068, 1992.

[49] F. Martinelli and E. Scoppola. A simple stochastic cluster dynamics: rigorous results. *J. Phys. A*, 24:3135–3157, 1991.

[50] J. Moussouris. Gibbs and Markov random systems with constraints. *J. Stat. Phys.*, 10:11–33, 1974.

[51] B. Nienhuis and M. Nauenberg. First-order phase transitions in renormalization-group theory. *Phys. Rev. Lett.*, 35:477–479, 1975.

[52] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77:257–286, 1989.

[53] D. Ruelle. *Thermodynamic Formalism*. Addison-Wesley, Reading, Massachusetts, 1978.

[54] R. H. Schonmann. Projections of Gibbs measures may be non-Gibbsian. *Commun. Math. Phys.*, 124:1–7, 1989.

[55] Ya. G. Sinai. Self-similar probability distributions. *Theor. Prob. and its Appl.*, 21:64–80, 1976.

[56] A. D. Sokal. Existence of compatible families of proper regular conditional probabilities. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 56:537–548, 1981.

[57] A. C. D. van Enter, R. Fernández, and A. D. Sokal. Regularity properties and pathologies of position-space renormalization-group transformations: Scope and limitation of Gibbsian theory. To appear in *J. Stat. Phys.*, September 1993.

[58] A. C. D. van Enter, R. Fernández, and A. D. Sokal. Regularity properties and pathologies of position-space renormalization-group transformations. *Nucl. Phys. B (Proc. Suppl.)*, 20:48–52, 1991.

[59] A. C. D. van Enter, R. Fernández, and A. D. Sokal. Renormalization transformations in the vicinity of first-order phase transitions: What can and cannot go wrong. *Phys. Rev. Lett.*, 66:3253–3256, 1991.

[60] A. C. D. van Enter, R. Fernández, and A. D. Sokal. Non-Gibbsian states. To be published in the proceedings of the 1992 Prague Workshop on Phase Transitions. R. Kotecký, editor. World Scientific, Singapore, 1993.

[61] A. C. D. van Enter, R. Fernández, and A. D. Sokal. Non-Gibbsian states for renormalization-group transformations and beyond. In N. Boccara, E. Goles, S. Martinez, and P. Picco, editors, *Proceedings of the NATO Advanced Studies Institute workshop on Cellular Automata and Cooperative Systems [Les Houches 1992]*, pages 141–152, Dordrecht, 1993. Kluwer (NATO ASI Series, Series C: Mathematical and Physical Sciences, Volume 396).

[62] K. G. Wilson and J. Kogut. The renormalization group and the ϵ-expansion. *Phys. Reports*, 12C:75–200, 1974.

Aernout C. D. van Enter
Institute for Theoretical Physics
Rijksuniversiteit Groningen
P.O. Box 800
9747 AG Groningen
THE NETHERLANDS
e-mail aenter@th.rug.nl

Roberto Fernández
Institut de Physique Théorique
Ecole Polytechnique Fédérale de Lausanne
PHB – Ecublens
CH–1015 Lausanne
SWITZERLAND
e-mail fernandez@eldp.epfl.ch

Alan D. Sokal
Department of Physics
New York University
4 Washington Place
New York, NY 10003
USA
e-mail sokal@acf4.nyu.edu