REPRESENTATION OF GIBBS FIELDS WITH SYNCHRONOUS RANDOM FIELDS.

LAURENT YOUNES.
GROUPE DIAM, CMLA (URA 1611)
ECOLE NORMALE SUPÉRIEURE DE CACHAN,
61 AVENUE DU PRÉSIDENT WILSON, 94 235 CACHAN CEDEX.

Abstract. We address the issue of the representation of Gibbs random fields over some configuration set by means of synchronous random fields, which lend themselves more efficiently to sampling on parallel devices. After describing the class of synchronous fields which is considered, we introduce a parametrization of synchronous fields by means of a potential. We give conditions under which it is one-to-one, and extend the results to the infinite lattice case. We also prove that every Gibbs field may be represented by such a synchronous field.
Keywords. Simulation of random fields. Parallel implementation.Gibbs distributions. Synchronous distributions.
AMS Classification. Primary: 60G60, 60K35, 65Y05. Secondary 60E25, 62F.

1. Introduction.

If S is a finite set of sites, and F a finite set of states, a random field on S with state space F is a probability distribution on the set of all configurations $\Omega_S = F^S$. A Gibbs field (or Gibbs distribution) is a random field π such that $\pi(x) > 0$ for all $x \in \Omega_S$. We denote by $\mathcal{P}^+ = \mathcal{P}^+(F, S)$ the set of all Gibbs fields on Ω_S. They generally are represented under the form

$$\pi(x) = \pi_U(x) = e^{-U(x)}/Z_U, \tag{1}$$

where U is the energy of π and Z_U is a normalizing constant. More precisely, being given an arbitrary element a in F, every Gibbs field can be written $\pi = \pi_U$ where U takes the form

$$U(x) = \sum_{C \subset S} u_C(x),$$

u_C being a function of x_s, $s \in C$, such that $u_C(x) = 0$ if $x_s = a$ for some $s \in C$ ([Ruelle 1978], [Georgii 1988]) (the potential is then said to be *normalized with respect to a*. One says that U is the energy associated to the *potential* $\mathbf{u} = (u_C,\ C \subset S)$. The Hammersley-Clifford theorem ([Besag 1974], [Geman 1991]) exhibits the relation between the potential and the structure of the conditional probabilities at one site given the states of all others.

When such models are employed in practical applications (eg. image analysis, neural networks, ...), it is most of the time necessary to have recourse to Monte Carlo sampling. This is due to the fact that, because of the inherent complexity of the processes which are modeled, one cannot make any direct computation of natural quantities of interest, such as probabilities of events or expectations of functions.

Unfortunately, this simulation step often dramatically slows down the methods in which it is required, and many attempts have been done to reduce the computation time they induce. A large number of papers are devoted to the determination of more efficient algorithms than the commonly used ones (see, for example, Swendsen and Wang 1987, Sokal 1989, Frigessi et al. 1990b, Besag and Green 1993, Smith and Roberts 1993,..., the list being not exhaustive). Other attempts aim at finding more efficient computer implementation of existing methods, and researchers have very early studied the possibility of using a parallel hardware ([Geman 1984], [Poggio 1985],...). The efficiency of a rigorous parallel implementation rapidly decreases as the complexity of the involved interactions increases, and the obtained improvement in efficiency appears, for this reason, as unsatisfactory. On the other hand, a non rigorous parallelization of the sampling algorithms must be done with care, and the study of the random fields which are simulated in that way requires the introduction of a formalism which is different from the Gibbsian one.

In [Younes 1993b], we have studied this synchronous formalism, mainly with a practical point of view, our purpose having been to construct models of random fields which can be sampled in parallel, and for which efficient algorithms may be devised in typical practical situations. The associated sampling algorithm requires building a generalized Markov chain of order $q \geq 1$ in Ω, for which the configuration $X^n \in \Omega$ at time n is obtained after synchronously and independently drawing the states $(X^n_s,\, s \in S)$ according to a transition probability which depends on $X^{n-q+1}, \dots, X^{n-1}$. If this Markov chain satisfied an additional condition which boils down to reversibility in the case $q+1 = 1$, we have shown in [Younes 1993b] they may be used in most of the practical situations in which the Gibbs representation is used, with the advantage of offering the possibility of being efficiently simulated on a parallel hardware. These fields have been called $(q+1)$-periodic synchronous random fields (see below for the reasons of this terminology).

The general parametric form of a periodic synchronous random field has been given in [Younes 1993b], reminiscent as the representation of Gibbs fields with a potential. However, when designing a parametric model in practice, it is very important to know whether this model is *identifiable*, that is whether the mapping which associates a probability distribution to a parameter is one-to-one. A sufficient condition for this fact in the case of te Gibbs representation is that the potential is normalized with respect to some $a \in F$. In this paper, we propose a special parametric form for synchronous fields, which is also associated to a potential, and provides identifiable models. This parametric form is also complete, in the sense that every Gibbs field may be represented in that way. Moreover, under some additional hypotheses on the potential, we will obtain estimates which will not depend on the size of the lattice S, and which will therefore yield results in the case of infinite integer lattices.

Representations of random fields, different from the Gibbsian one, already have been introduced in the litterature. Among them are the hidden Markov random fields ([Geman et al. 1993]), which are fields on $S = \mathbb{Z}^d$ which are images of some field, on $\mathbb{Z}^d$ also, with nearest neighbourg interactions, but with a large state space G than F. The unilateral approximation consists in representing a field as a Markov chain with respect to some ordering of the set S (cf. [Goutsias 1991]). Stochastic models introduced for neural networks, such as Boltzmann machines, also are alternative representations of Gibbs fields, a probability distribution on some configuration set Ω_S being represented as the marginal of a distribution on $\Omega_{S\cup H}$ (H being a set of "hidden" sites – or neurons) which is associated to a potential with only pairwise interactions (cf [Ackley et al. 1985], [Sussmann 1988], [Younes 1993a]).

This paper is organized as follows. In the next section, we give the definition of synchronous random fields, with the motivation which has led to it. In section 3, we introduce our parametrisation, which is based on a potential, and show that it approximates Gibbs distributions in a way which implies the exhaustivity of the class of synchronous fields. We shall also prove more precise results in the case when $S \subset \mathbb{Z}^d$ and the considered potential has bounded radius. Finally, in section 4, we see how these results may be extended to the case of infinite S ($S = \mathbb{Z}^d$).

2. Synchronous Random Fields.

In this paragraph, we summarize some necessary definitions and results concerning synchronous random fields. More practical motivations may be found in ([Younes 1993b]). We start with a formal definition, which do not refer to synchronous sampling, but has the advantage of being concise.

2.1. **Definition.** Let S be a finite set, F the state space and $\Omega = \Omega_S = F^S$. Let k be a positive integer. We denote by $\mathcal{D}_k(F, S)$ the set of all Gibbs distributions μ on Ω^k which satisfy the following conditions:

a. Invariance by circular permutation: for all $(x^1, \dots, x^k) \in \Omega^k$,
$$\mu(x^1, \dots, x^k) = \mu(x^k, x^1, \dots, x^{k-1}).$$
(We shall respect the following notational convention: superscripts are employed to index families of configurations, and subscripts to indicate the state of a given configuration at some site. So, x_s^l would refer to the value of $x^l \in \Omega$ at site $s \in S$: $x_s^l \in F$).

b. Conditional independance: the variables x_s^1, $s \in S$, are μ-conditionally independent given $x^2, \dots, x^k$.

We then define $\mathcal{S}_k(F, S)$ to be the set of all Gibbs fields on Ω which are a marginal distribution of some element of $\mathcal{D}_k(F, S)$ (we shall omit to indicate S and F when no confusion is possible). Elements of $\mathcal{S}_k$ are called *k-periodic synchronous distributions.*

2.2. **Interpretation in terms of sampling.** We now show the relation between this definition and synchronous Markov chains of order $k - 1$. Assume that μ is as above. Denote by $\mu^i(\,.\,|x^j,\, j \neq i)$, $i = 1, \dots, k$, the conditional distributions for μ of the component $x^i \in \Omega$ given x^j, $j \neq i$. By hypothesis, μ^i splits as a product over S, which we write :
$$\mu^i(x^i \,|\, x^j,\, j \neq i) = \prod_{s\in S} \mu_s^i(x_s^i \,|\, x^j,\, j \neq i)\,.$$

Then, consider the process $(X(n),\, n \geq 1)$ of configurations in Ω which is defined as follows. Its $k - 1$ first components $X(1), \dots, X(k - 1)$ are arbitrary, and for $n \geq k$, let $i = i(n)$ be the element of the class of n modulo k which is in $\{1, \dots, k\}$, so that $n = qk + i(n)$. Then, define the probability of $X(n) = x$ given the values of $X(p), p < n$ to be

$$\mu^{i(n)}(x \,|\, X(qk + 1), \dots, X(qk + i - 1), X((q - 1)k + i + 1), \dots, X(qk))\,. \tag{2}$$

Such a process is in fact a standard algorithm which is known to simulate μ (it is called the heat-bath, or Gibbs sampler in the litterature, see for example [Geman 1991] or [Sokal 1989]), that is, the joint distribution $X(qk+1),\ldots,X(qk+k)$ converges to μ when q tends to infinity. But, by circular permutation invariance, the probability in (2) may be written under he form (since $qk+i=n$)

$$P(X(n-k+1),\ldots,X(n-1)\,;\,x)$$

for a transition probability P from Ω^{k-1} to Ω which is independent on n, and, since it is the case for μ^i, this transition probability splits as a product of the kind (for any $(x^1,\ldots,x^{k-1})\in\Omega^{k-1}$, and any $x\in\Omega$):

$$P(x^1,\ldots,x^{k-1}\ ;\ x)=\prod_{s\in S}p_s(x^1,\ldots,x^{k-1}\,;\,x_s)\,, \tag{3}$$

the p_s being *local transition kernels* from Ω^{k-1} to F.

Thus, the process $X(n)$ may also be considered as a *homogeneous Markov chain of order* $k-1$, for which the transition from $X(p), p<n$ to $X(n)$ is done by *synchronously updating* all $X_s(n)$. Since the joint distribution of $X(qk+1),\ldots,X(qk+k)$ converges to μ, and all the marginal of μ are equal to π, the process $X(n)$ converges in distribution to π. Thus π may be simulated by a dynamic procedure which, at each time chunk, synchronously updates all the sites with probabilities which depend on the $k-1$ last outcomes of the process.

Conversely, let P be a positive transition probability of order $k-1$ which synchronously updates all the sites, and which is positive. Let μ be its k-step probability in stationary regime, ie. μ is the distribution of $X(n+1),\ldots,X(n+k)$ for the stationary Markov chain associated to P. Then, the distribution π of $X(n)$ is k-periodic as soon as μ satisfies property (a.) in definition (2.1). This alternative definition of k-periodicity clearly is equivalent to the first one (cf. [Younes 1993b]).

An interesting case is when $k=2$. (X^p) is then a standard Markov chain, and condition (a.) is equivalent to the fact that X^p is reversible. From this point of view, k-periodicity may be seen as a generalization of reversibility to $k-1$ order Markov chains.

At this point, the fact that a k-periodic synchronous distribution may efficiently be simulated on a parallel hardware should be clear. What should be less evident is why we needed to have recourse to Markov processes of order greater than 1 in our construction, and did not restrict to ordinary Markov chains. The reason is that, for ordinary Markov chains, reversibility is a very convenient property. When it is true, it allows to obtain an explicit description of the invariant probability π (since $\pi(x)/\pi(y)=P(y,x)/P(x,y)$), whereas, in the general case, π is only implicitely defined by $\pi.P=\pi$. Moreover, it seems quite difficult to develop feasible adaptations of standard practical algorithms which are employed under the Gibbsian formalism, when the models are synchronous and non-reversible, whereas such adaptations may be obtained in the reversible case. Thus, for practical reasons, using synchronous random fields of order 1 requires restricting to the class of reversible ones.

Unfortunately, as it is shown in ([Koslov and Vasilyev 1980]), reversible synchronous fields must satisfy some very constraining conditions. In fact, in this case, the 2-step distribution μ on Ω^2 (of which the synchronous field π is a marginal) must be of the kind

$$(4) \qquad \mu(x,y) = \exp[\sum_{st} h_{st}(x_s, y_t) + \sum_s h_s(x_s) + \sum_s h_s(y_s)]/Z\,.$$

with with $h_{st}(a,b) = h_{ts}(b,a)$ for all s and t in S and all a, b in F.

Since this class is to small to model the distributions which are needed in practice, the simpler way to enlarge it while remaining within a synchronous context was to consider Markov chains of higher order as we did. The condition of k-periodicity was enough to guarantee the feasibility of the algorithms in the applications, so that the practical problems linked to non-reversibility are successfully addressed..

The objective of the remaining is to study the representation of Gibbs fields by synchronous distributions in $\mathcal{S}_k(F,S)$. We shall exhibit an immersion of $\mathcal{S}_k(F,S)$ into $\mathcal{P}^+$, and show that this immersion is onto when $k = |S|$. In the particular case of local potential, we shall obtain uniform estimates with respect to $|S|$, which will be used for proving a local identifiability theorem in the case of infinite S.

3. Representation of Gibbs distributions by synchronous ones.

3.1. **A parametrization of synchronous distributions by a potential.** We now describe a framework under which a synchronous field may be defined with the help of a potential.

If $a \in F$, every Gibbs distribution on Ω may be uniquely written under the form $\pi_{U^a} = \exp(-U^a)/Z$ with

$$U^a(x) = \sum_{C \in S} u^a_C(x),$$

where u^a_C is a function such that

i- u^a_C only depends on $x_s,\ s \in C$.

ii- $u^a_C(x) = 0$ whenever $x_s = a$ for some $s \in C$.

A family of functions $\mathbf{u} = (u_C,\ C \subset S)$ which satisfies condition (i-) is called a potential. If (ii-) is also true for some a, one says that $\mathbf{u}$ is normalized with respect to a. When $u_C \equiv 0$ for $|C| > k$, one says that the range of $\mathbf{u}$ is bounded by k.

If a potential $\mathbf{u}$ is given, the associated energy is $U = \sum_C u_C$, and we shall also denote by $\pi_{\mathbf{u}}$ the Gibbs field π_U. We give a construction for approaching $\pi_{\mathbf{u}}$ by a synchronous distribution when $\mathbf{u}$ has range bounded by k. We assume that some ordering has been chosen on every subset $C \subset S$. When writing $C = \{c_1, \ldots, c_l\}$, we implicitely assume that the elements have been numbered consistently with it ($c_p < c_{p+1}$).

If $\mathbf{u}$ have range smaller than k, we define an energy on Ω^k, denoted $\tilde{U}^k$ as follows. First, consider $C \subset S$ with $|C| \leq k$: $C = \{c_1, \ldots, c_l\}$, $l \leq k$. For $X = (x^1, \ldots, x^k) \in \Omega^k$ set

$$(5) \qquad \tilde{u}^k_C(X) = \frac{1}{k} \sum_{p=0}^{k-1} u_C(x^p_{c_1}, \ldots, x^{p+l-1}_{c_l}),$$

where superscripts must be understood modulo k in the set $\{1, \ldots, k\}$. (We shall use this convention throughout this paper without making any new reference to it). We now define, for $X \in \Omega^k$,

$$(6) \qquad \tilde{U}^k(X) = \sum_{C \subset S} \tilde{u}^k_C(X).$$

For $\lambda, \lambda' \in F$ we put $d(\lambda, \lambda') = 1$ if $\lambda \neq \lambda'$ and 0 if not. For $x, y \in \Omega$, we denote by $d_S(x, y)$, or by $d(x, y)$ when no confusion is possible, the quantity $\sum_{s \in S} d(x_s, y_s)$. Then, for $X = (x^1, \ldots, x^k) \in \Omega^k$, we set

$$V(X) = \sum_{p=1}^{k} d_S(x^p, x^{p+1}). \tag{7}$$

Finally, with the notations above, we define, for every positive number, α, the distribution $\mu^\alpha_{k,\mathbf{u}}$ on Ω^k by

$$\mu^\alpha_{k,\mathbf{u}}(X) = \frac{1}{Z^\alpha_{k,U}} \exp\left[-\alpha V(X) - \tilde{U}^k(X)\right]. \tag{8}$$

We leave to the reader to check that

Proposition 1. *For $\alpha > 0$ and all potential $\mathbf{u}$ with range at most k, $\mu^\alpha_{k,\mathbf{u}}$ is in $\mathcal{D}_k(F, S)$.*

We denote by $\nu^\alpha_{k,\mathbf{u}}$ the element of $\mathcal{S}(F, S)$ which is associated to $\mu^\alpha_{k,\mathbf{u}}$, ie. the first marginal of $\mu^\alpha_{k,\mathbf{u}}$. Everytime it will not induce confusion, we shall drop some of the indices α, k or $\mathbf{u}$ to simplify the notations. Since $\tilde{U}^k(x, \ldots, x) = U(x)$, and the term $-\alpha V(X)$ in (8) penalizes differences between configurations x^k, $\nu^\alpha_{k,\mathbf{u}}$ converges to $\pi_{\mathbf{u}}$ when $\alpha \to \infty$ (we prove a more precise statement in theorem 1). One obtains in this way that $\mathcal{S}_n$ is dense in $\mathcal{P}^+$.

For a potential $\mathbf{u} = (u_C)$, we let $|\mathbf{u}|$ be the smallest number M such that, for all $x, y \in \Omega$, for all $C \subset S$,

$$|u_C(x) - u_C(y)| \leq M \sum_{s \in C} d(x_s, y_s).$$

We also let $N_s(\mathbf{u})$ be number of sets C such that $u_C \not\equiv 0$ and $s \in C$ and $N(\mathbf{u})$ the maximum of $(N_s(u))$.

Our first results estimate the proximity between $\pi_{\mathbf{u}}$ and $\nu^\alpha_{\mathbf{u}}$. We have,

Theorem 1. *Assume that Λ is an energy function on Ω and $\tilde{\Lambda}$ an energy function on Ω^k satisfying, for $X = (x^1, \ldots, x^k) \in \Omega^k$*

$$|\tilde{\Lambda}(X) - \Lambda(x^1)| \leq \Delta . V(X).$$

Let $\pi = \pi_\Lambda$ be the Gibbs distribution with energy Λ, and ν be the first marginal of the Gibbs distribution on Ω^k with energy $\tilde{\Lambda} + \alpha V$.

Then, one has

$$|\log \frac{\nu(x)}{\pi(x)}| \leq |S| g(k, \Delta, \alpha), \tag{9}$$

with

$$g(k, \Delta, \alpha) \leq 2k(|F| - 1) e^{\Delta} e^{-\alpha} \tag{10}$$

for $\alpha \geq \log k + \Delta + \log |F| + 1$.

We have the corollary

Corollary 1. *Let $\mathbf{u}$ be as above. Then*

$$\left| \log \frac{\nu^\alpha_{k,\mathbf{u}}(x)}{\pi_{\mathbf{u}}(x)} \right| \leq |S| g(k, |u| N(u), \alpha)$$

This result trivially implies that $\nu_{k,\mathbf{u}}^\alpha$ converges to $\pi_\mathbf{u}$ when α tends to infinity. For fixed α and n going to infinity, this also provides an estimate of the specific Kullback information between these probabilities. Its proof is trivial from theorem 1 (which will be proved in the next section) and the lemma

Lemma 1. *Let $X = (X^1, \dots, X^k)$ in Ω^k. We have*

$$|\tilde{U}(X) - U(x^1)| \le N(\mathbf{u})|\mathbf{u}|V(X) \tag{11}$$

To prove lemma 1, let $x = x^1$. We have $\tilde{U}(X) - U(x) = \sum_C \tilde{u}_C(X) - u_C(x)$. Now, let $C = \{c_1, \dots, c_l\}$, with $l \le k$. We have

$$\begin{aligned}
|\tilde{u}_C(X) - u_C(x)| &\le \frac{1}{k}\sum_{p=0}^{k-1} |u_C(x_{c_1}^p, \dots, x_{c_l}^{p+l-1}) - u_C(x^1)| \\
&\le \frac{1}{k}\sum_{p=0}^{k-1} |\mathbf{u}| \sum_{q=1}^{l} d(x_{c_q}^{p+q-1}, x_{c_q}^1) \\
&= |\mathbf{u}| \sum_{s\in C} \frac{1}{k}\sum_{p=1}^{k} d(x_s^p, x_s^1)
\end{aligned}$$

Hence

$$\begin{aligned}
|\tilde{U}(X) - U(x^1)| &\le |\mathbf{u}| \sum_{C}\sum_{s\in C} \frac{1}{k}\sum_{p=1}^{k} d(x_s^p, x_s^1) \\
&= |\mathbf{u}|\frac{1}{k}\sum_{p=1}^{k}\sum_{s\in S} N_s(\mathbf{u}) d(x_s^p, x_s^1) \\
&\le N(\mathbf{u})|\mathbf{u}|\frac{1}{k}\sum_{p=1}^{k} d_S(x^1, x^p) \\
&\le N(\mathbf{u})|\mathbf{u}|\frac{1}{k}\sum_{p=1}^{k} V(X) \\
&= N(\mathbf{u})|\mathbf{u}|V(X)
\end{aligned}$$

□

Also the following estimate (proved in the next paragraph, together with theorem 1) will be useful. Let $\partial\Omega$ be the subset of Ω^k containing all elements $(x, \dots, x)$ with $x \in \Omega$.

Lemma 2. *One has, with the hypotheses of theorem 1, and denoting by μ the Gibbs distribution on Ω^k with energy $\tilde{U}$*

$$\mu_\mathbf{u}^\alpha(\partial\Omega \mid x_1 = x) \ge 1 - 2nk(|F| - 1)e^{\Delta}e^{-\alpha}. \tag{12}$$

for $\alpha \ge \log n + \log k + \Delta + \log|F| + 1$.

As a consequence, we have

Proposition 2. *Let μ be like in theorem 1. Let $C \subset S$ and f be a function defined on Ω^k, depending only on $x_s^1, \dots, x_s^k$ with $s \in C$, such that $|f(x)| \le 1$. One has, for $x \in \Omega$,*

$$|E(f \mid x^1 = x) - f(x, \dots, x)| \le k|C|(|F| - 1)e^{-\alpha+\Delta}\,. \tag{13}$$

where the expectation is with respect to μ.

Proof of proposition 2. One has

$$E(f \mid x^1 = x) = E\left[\, E(f \mid x^1 = x \text{ and } x^l_t, l = 2, \dots, k,\ s \in S \setminus C) \mid x^1 = x\right],$$

so that it suffices to estimate

$$E(f \mid x^1 = x \text{ and } x^l_t, l = 2, \dots, k,\ t \in S \setminus C) - f(x, \dots, x), \tag{14}$$

for given values of $x^l_t,\ l = 2, \dots, k,\ t \in S \setminus C$.

Let μ_C be the distribution on Ω^k_C, equal to the conditional distribution for μ given $x^l_t,\ l = 1, \dots, k,\ t \in S \setminus C$. It is the Gibbs field in $\mathcal{P}^+(F, C)$ with energy

$$\tilde{\Lambda}_C(y^1_C, \dots, y^k_C) = \tilde{\Lambda}(y^1, \dots, y^k),$$

configurations y^l being extended outside C by $y^l_t = x^l_t,\ t \notin C$. Define $\Lambda_C(y_C) = \tilde{\Lambda}_C(y_C, \dots, y_C)$. Then Λ and $\tilde{\Lambda}$ satisfy the hypotheses of theorem 1 with the same value of Δ. The expectation in equation (14) is the conditional expectation for μ_C given that $x^1_C = x_C$. Thus, denoting by E_C the expectation with respect to μ_C,

$$\begin{aligned} E_C(f \mid x^1_C = x_C) - f(x_C, \dots, x_C) &= E_C(f\ ;\ X_C \notin \delta\Omega_C \mid x^1_C = x_C) \\ &\leq 1 - \mu(\delta\Omega_C \mid x^1_C = x_C) \\ &\leq k|C|(|F| - 1)e^{-\alpha+\Delta}\,. \end{aligned}$$

□

Remark: Our purpose is to study the parametrization $\mathbf{u} \to \nu^\alpha_{k,\mathbf{u}}$ for fixed α. We do not aim at approaching Gibbs fields by synchronous ones, but to directly use synchronous modeling in practice. However, for large enough α, we show that this parametrization shares some properties of the Gibbsian parametrization, in particular the fact of being one-to-one.

3.2. **Proof of theorem 1 and lemma 2.** We have

$$\frac{\nu(x)}{\pi(x)} = \frac{\sum_{X\ ;\ x^1 = x} e^{-\alpha V(X) - \tilde{U}(X) + U(x)}}{\left(\sum_X e^{-\alpha V(X) - \tilde{U}(X)} / \sum_{x^1} e^{-U(x^1)}\right)}$$

which yields

$$\frac{\nu(x)}{\pi(x)} \leq \frac{\sum_{X\ ;\ x^1 = x} e^{-(\alpha-\Delta)V(X)}}{\left(\sum_X e^{-(\alpha+\Delta)V(X) - U(x^1)} / \sum_{x^1} e^{-U(x^1)}\right)}\,. \tag{15}$$

But

$$\sum_X e^{-(\alpha+\Delta)V(X) - U(x^1)} = \sum_y e^{-U(y)} \sum_{X\ ;\ x^1 = y} e^{-(\alpha+\Delta)V(X)}\,;$$

Since, as one can easily get convinced,

$$\sum_{X\ ;\ x^1 = y} e^{-(\alpha+\Delta)V(X)}$$

does not depend on the configuration $y \in \Omega$, (15) simplifies in

$$\frac{\nu(x)}{\pi(x)} \leq \frac{\sum_{X\ ;\ x^1 = x} e^{-(\alpha-\Delta)V(X)}}{\sum_{X\ ;\ x^1 = x} e^{-(\alpha+\Delta)V(X)}}\,.$$

Theorem 1 is a consequence of the lemma

Lemma 3. *For $x \in \Omega$ and $\beta \in \mathbb{R}$, one has, letting $f = |F|$,*

$$\sum_{X \in \Omega^k \ ; \ x^1 = x} e^{-\beta V(X)} = \left[\frac{B_k(\beta)}{f}\right]^n \tag{16}$$

where

$$B_k(\beta) = (f-1)\left(1 - e^{-\beta}\right)^k + \left(1 + (f-1)e^{-\beta}\right)^k .$$

and $V(x^1, \dots, x^k) = \sum_{s \in S} \sum_{p=1}^k d(x_s^p, x_s^{p+1})$.

Indeed, using (16) we see that we can set

$$g(k, \Delta, \alpha) = \log \frac{B_k(\alpha - \Delta)}{B_k(\alpha + \Delta)}. \tag{17}$$

The upper bound given in (10) relies on the following elementary lemma

Lemma 4. *If $0 \le x \le \log 2/k$, then $(1+x)^k \le 1 + 2kx$.*

(of which we leave the easy proof to the reader). Using the facts that $B_k(\beta)$ is smaller than $f(1+(f-1)e^{-\beta})^k$, that the denominator in (17) is always larger than f (see below), and that $\log(1+x) \le x$, we get the estimate (10).

Let's describe how lemma 3 may be obtained. We have

$$\begin{aligned}
A &= \sum_{X \ ; \ x^1 = x} e^{-\beta V(X)} \\
&= \sum_{X, \ x^1 = x} \exp\left(-\beta \sum_{p=1}^k \sum_{l=1}^n d(x_l^p, x_l^{p+1})\right) \\
&= \prod_{l=1}^n \left[\sum_{x_l^2 \in F, \dots, x_l^k \in F} \exp\left(-\beta \sum_{p=1}^k d(x_l^p, x_l^{p+1})\right)\right] \\
&= \left[\sum_{c^2, \dots, c^k \in F} \exp\left(-\beta \sum_{p=1}^k d(c^p, c^{p+1})\right)\right]^n .
\end{aligned}$$

where in the last term c^1 is fixed as an arbitrary element of F, the result being independent of its value; A is therefore equal to

$$\left[\frac{1}{f} \sum_{c^1, \dots, c^n \in F} \exp\left(-\beta \sum_{p=1}^k d(c^p, c^{p+1})\right)\right]^n ,$$

where $f = |F|$. We now check that

$$\begin{aligned}
B_k(\beta) &:= \textstyle\sum_{c^1, \dots, c^k \in F} \exp\left(-\beta \sum_{p=1}^k d(c^p, c^{p+1})\right) \\
&= (f-1)\left(1 - e^{-\beta}\right)^k + \left(1 + (f-1)e^{-\beta}\right)^k
\end{aligned} \tag{18}$$

Recall that $d(c^i, c^j) = 1 - \delta_{c^j}(c^i)$ and that by convention c^{k+1} is equal to c^1. To evaluate B, we order the terms by the number of indices p such that $c^p \ne c^{p+1}$. If this number is q, we have $\binom{k}{q}$ possibilities for choosing them. This yields $q+1$ regions in the set $\{1, \dots, q+1\}$. Denote by a_q the number of ways for coloring these regions, ie. the number of $(q+1)$-uples $(\gamma^1, \ldots, \gamma^{q+1})$ in F such that $\gamma^p \neq \gamma^{p+1}$ and $\gamma^{q+1} = \gamma^1$. We have

$$B_k(\beta) = \sum_{q=0}^{k} \binom{k}{q} a_q e^{-q\beta} . \tag{19}$$

To compute a_q, let b_q denote the number of ways of choosing $\gamma^1, \ldots, \gamma^{q+1}$ in F such that $\gamma^p \neq \gamma^{p+1}$, but without the constraint $\gamma^{q+1} = \gamma^1$. For this, we can choose any γ^1 in F, and then any γ^i in $F \setminus \{\gamma^{i-1}\}$. Thus, $b_q = f(f-1)^q$. For a_p, assume we have chosen the values of $\gamma^1, \ldots, \gamma^{q-1}$. If $\gamma^{q-1} = \gamma^1$, it remains $f-1$ possibilities for fixing γ^q, and $f-2$ if $\gamma^{q-1} \neq \gamma^1$. Thus, one has the identity :

$$a_q = (f-1)a_{q-2} + (f-2)(b_{q-2} - a_{q-2}) = a_{q-2} + (f-2)f(f-1)^{q-2}.$$

Using $a_0 = f$, and $a_1 = 0$, we get $a_q = (f-1)^q + (-1)^q(f-1)$. Equation (19) then gives formula (18). Note that we always have $B_k(\beta) > a_0 = f$, and since $a_q \leq (f-1)[1 + (f-1)^{q-1}]$, we have

$$B_k(\beta) \leq (f-1)(1+e^{-\beta})^k + (1+(f-1)e^{-\beta})^k \leq f(1+(f-1)e^{-\beta})^k .$$

□

We now prove lemma 2. We have

$$\mu_U^\alpha(\Omega \setminus \delta\Omega \mid x_1 = x) = \frac{\sum_{Y \in \Omega^n,\ y^1 = x} e^{-\tilde{\Lambda}(Y) - \alpha V(Y)} - e^{-\Lambda(x)}}{\sum_{Y \in \Omega^n,\ y^1 = x} e^{-\tilde{\Lambda}(Y) - \alpha V(Y)}},$$

which is smaller than

$$\frac{\left\{\frac{2}{|F|}(1+e^{-\alpha+\Delta})^k + (1-\frac{2}{|F|})\left(1+(|F|-1)e^{-\alpha+\Delta}\right)^k\right\}^n - 1}{\left\{\frac{2}{|F|}(1+e^{-\alpha-\Delta})^k + (1-\frac{2}{|F|})\left(1+(|F|-1)e^{-\alpha-\Delta}\right)^k\right\}^n},$$

which yields equation (12) after an application of lemma 4. □

3.3. **Representation by elements of $\mathcal{S}_k$: general potentials.** In this section, we study the functions $\mathbf{u} \to \nu_{\mathbf{u}}^\alpha$ for potentials $\mathbf{u}$ of range smaller than k. More precisely, let $a \in F$ and let $\mathbf{a} \in \Omega$ be the configuration with state a at every site. Denote by $\mathcal{R}_k$ the set of potentials, normalized with respect to a, which have range bounded by k. We identify $\mathcal{R}_k$ with the vector space

$$\prod_{|C| \leq k} \mathbb{R}^{(|F|-1)^{|C|}},$$

and represent its elements by $\mathbf{u} = (u_{C,x_C} = u_C(x_C)\ C \subset S,\ |C| \leq k,\ x_C \in (F \setminus \{a\})^C)$. If $\mathbf{u} \in \mathcal{R}_n$, its energy $U = U_{\mathbf{u}}$ has been defined by

$$U(y) = \sum_C u_C(y_C) = \sum_{C, x_C} u_{C,x_C} \delta_{x_C}(y_c).$$

For $\mathbf{u} \in \mathcal{R}_k$, we set $\|\mathbf{u}\| = \max_{C,x_C} u_C(x_C)$, and always use the associated operator norm for linear mapping between $\mathcal{R}_k$ and $\mathcal{R}_{k'}$.

Note that we have also defined another norm for $\mathbf{u}$, namely $|\mathbf{u}|$ which was defined as the smallest number M such that, for all $C \subset S$, $x, y \in \Omega$,

$$|u_C(x) - u_C(y)| < M \sum_{s \in C} d(x_s, y_s) .$$

For potential $\mathbf{u}$ in $\mathcal{R}_k$, we have

$$\|\mathbf{u}\|/k \leq |\mathbf{u}| \leq 2\|\mathbf{u}\|\,.$$

For a configuration $x \in \Omega$, denote by x_C^a the configuration with state x_s at site s for $s \in C$, and with state a at s for $s \notin C$.

To $\mathbf{u} \in \mathcal{R}_k$ and $\alpha > 0$, we have associated a synchronous random field $\nu_{\mathbf{u}}^\alpha \in \mathcal{S}(F,S)$. To describe the relation btween $\mathbf{u}$ and $\nu_{\mathbf{u}}^\alpha$, we consider the mapping

$$\begin{aligned} \psi_k^\alpha : \quad \mathcal{R}_k &\longrightarrow \mathcal{R}_k \\ \mathbf{u} \quad &\to (-\log \frac{\nu_{\mathbf{u}}^\alpha(x_C^a)}{\nu_{\mathbf{u}}^\alpha(\mathbf{a})},\ C \subset S,\ |C| \leq k,\ x_s \in F \setminus \{a\}, s \in C) \end{aligned}$$

Let k be fixed and consider the family ψ^α of endomorphisms of $\mathcal{R}_k$. The first remark is that, when $\alpha \to \infty$, $\psi^\alpha(\mathbf{u})$ converges to a limit $\psi^\infty(\mathbf{u})$, because $\nu_{\mathbf{u}}^\alpha$ converges to $\pi_{\mathbf{u}}$. The expression of ψ^∞, which comes from a straightforward computation, is given by the next proposition.

Proposition 3. *The limit of ψ^α when $\alpha \to \infty$ is a linear, invertible, mapping on $\mathcal{R}_k$, given by*

$$\psi^\infty(\mathbf{u}) = \left(\sum_{B \subset C} u_{B,x_B}, C \subset S,\ |C| \leq k,\ x_s \in F \setminus \{a\}, s \in C\right)\,. \tag{20}$$

One has

$$(\psi^\infty)^{-1}(\mathbf{u}) = \left(\sum_{C \subset B} (-1)^{|B-C|} u_{C,x_C}, B \subset S,\ |B| \leq k,\ x_s \in F \setminus \{a\}, s \in B\right)\,. \tag{21}$$

For $\mathbf{u} \in \mathcal{R}_k$, denote by $d_{\mathbf{u}}\psi^\alpha$ and $d_{\mathbf{u}}^2\psi^\alpha$ the first and second derivative of ψ^α with respect to $\mathbf{u}$. We first study the behaviour of these derivatives for large α. The following theorem is proved in he next section :

Theorem 2. *Let $n = |S|$. For all $M > 0$, there exists a number $\alpha_{kn}(M)$, such that for all $\mathbf{u} \in \mathcal{R}_k$, with $|\mathbf{u}| \leq M$, for all $\alpha > \alpha_{kn}(M)$, the differential $d_{\mathbf{u}}\psi$ is invertible, and*

$$2^{k-2} \leq \|(d_{\mathbf{u}}\psi)^{-1}\| \leq 2^{k+1}. \tag{22}$$

One may take

$$\alpha_{kn} = \binom{n-1}{k-1} M + (k+3)\log 2 + \log(k^2\, dim\mathcal{R}_k)\,. \tag{23}$$

Moreover, the norm of the second derivative of ψ is always bounded by $2(dim\mathcal{R}_k)^2$.

As a consequence, we obtain the following fact; denote by $\mathcal{O}_k(M)$ the set of all potentials $\mathbf{u}$ in $\mathcal{R}_k$ such that $|\mathbf{u}| < M$. Moreover, denote by $\mathcal{B}_k(\mathbf{u}, r)$ the open ball (for the norm $\|\mathbf{u}\|$ in $\mathcal{R}_k$) with center $\mathbf{u}$ and radius r.

Theorem 3. *There exists two positive numbers, r_{kn} and ρ_{kn}, depending on k and $n = |S|$, such that, for all $M > 0$, for all $\alpha > \alpha_{kn}(M)$, for all $\mathbf{u}$ such that $\mathcal{B}_k(\mathbf{u}, r_{kn}) \subset \mathcal{O}_k(M)$, ψ^α is a diffeomorphism from some open set $\mathcal{V} \subset \mathcal{B}_k(\mathbf{u}, r_{kn})$ onto its image, which contains the open ball $\mathcal{B}_k(\psi^\alpha(\mathbf{u}), \rho_{kn})$.*

Furthermore, there exists $\overline{\alpha}_{kn}(M) \geq \alpha_{kn}(M)$ such that, for all $\alpha > \overline{\alpha}_{kn}(M)$, ψ^α restricted to $\mathcal{O}_k(M)$ is one to one.

We therefore obtain a result stating that, at least for bounded potential and for large enough α, our parametrisation is one-to-one. At this level of generality, α still depends on the cardinality n of the set S, which may seem is unsatisfying, given the fact that Gibbs field models usuually defined from finite range potentials are formally not dependent on S. We will see later how this can be addressed in the case of regular lattices, and local potentials. The present result a direct consequence of the inverse mapping theorem, of which we take the following standard version :

> *If φ is a function (defined on an open subset of a Banach space $\mathcal{X}$, into a Banach space $\mathcal{Y}$), with $\varphi(0) = 0$, $d_0\varphi = I$, and if $\delta > 0$ is such that, φ is defined on the open ball $\mathcal{B}(0, \delta)$, and*

$$\max(|x|, |x'|) < \delta \Rightarrow |x - \varphi(x) - x' + \varphi(x')| < c|x - x'|\,, \tag{24}$$

> *with $c < 1$ then, there is an open neighbourhood of 0, $\mathcal{V}$, in $\mathcal{X}$ such that φ is a diffeomorphism from $\mathcal{V}$ onto the open ball $\mathcal{B}(0, \delta(1-c))$ on $\mathcal{Y}$.*

Applying this theorem to

$$\varphi(\,.\,) = (d_{\mathbf{u}_0}\psi)^{-1}[\psi(\,.\, + \mathbf{u}_0) - \psi(\mathbf{u}_0)]\,,$$

and using the fact that $\|d^2_{\mathbf{u}}\psi\| < C$ in $\mathcal{R}_k$, with $C = 2(\dim\mathcal{R}_k)^2$, we see that the inequality (24) is true for φ with $c = 2C\delta\|(d_{\mathbf{u}_0}\psi)^{-1}\|$. Taking $\delta = (4C\|(d_{\mathbf{u}_0}\psi)^{-1}\|)^{-1}$, which is smaller than $r_{kn} = 2^{-k}/C$ for $\alpha > \alpha_{kn}(M)$, we obtain the fact that there exists some open set included in $\mathcal{B}_k(\mathbf{u}_0, r_{kn})$ which is diffeomorphic, by ψ^α, to the set

$$\psi(\mathbf{u}_0) + (d_{\mathbf{u}_0}\psi).\mathcal{B}_k(0, \delta/2)\,.$$

Since $\delta/2 > 2^{-k-4}/C$ and $\|(d_{\mathbf{u}_0}\psi)^{-1}\| \leq 2^{k+1}$, we see that this set contains the ball $\mathcal{B}_k(\psi(\mathbf{u}_0), \rho_{kn})$ with

$$\rho_{kn} = 2^{-2k-5}/C\,. \tag{25}$$

To prove the second claim of proposition 3, assume that, for all $\alpha' > \alpha_{kn}(M)$, there exists an $\alpha > \alpha'$ and two potentials $\mathbf{u}$ and $\mathbf{u}'$ in $\mathcal{R}_k$, with $\max(|\mathbf{u}|, |\mathbf{u}'|) \leq M$ and $\psi^\alpha(\mathbf{u}) = \psi^\alpha(\mathbf{u}')$. One can then construct a sequence $\alpha_p \to \infty$ and two sequences $\mathbf{u}_p$ and $\mathbf{u}'_p$, which may be assumed to converge (to $\mathbf{u}_\infty$ and $\mathbf{u}'_\infty$), such that

$$\psi^{\alpha_p}(\mathbf{u}_p) = \psi^{\alpha_p}(\mathbf{u}'_p)\,.$$

Since ψ^∞ is one-to-one, we must have $\mathbf{u}_\infty = \mathbf{u}'_\infty$, and we obtain the fact that there exists no neighbourhood of $\mathbf{u}^\infty$ on which ψ^α is a diffeomorphism for all large enough α, which is a contradiction to the preceding result.

Thus, there exists an large enough $\overline{\alpha}(M)$ such that, for all $\alpha > \overline{\alpha}(M)$, ψ^α is a diffeomorphism.

□

We also have this interesting corollary, which is valid for fixed $\mathbf{u}$:

Corollary 2. *Let $a \in F$. If $\mathbf{u} \in \mathcal{R}_k$ is given, then ψ^α_k is locally invertible at $\mathbf{u}$, excepted for a finite number of α.*

Proof: The function $\gamma(\alpha, \mathbf{u})$ which associates to $(\alpha, \mathbf{u}) \in \mathbb{R} \times \mathcal{R}_k$ the determinant of $d_U \psi^\alpha$ is analytic (it is a rational function of the variables e^α and $e^{u_{C,x_C}}$). The fact that, for fixed $\mathbf{u}$ this function cannot vanish if $\alpha > \alpha_{k,n}(\mathbf{u})$, implies that the set of α such that $\gamma(\alpha, U) = 0$ is finite. □

And, as a second corollary, the representation theorem

Theorem 4. *If S is a finite set of sites, of cardinality n, and F is a finite state space, then*

$$\mathcal{S}_n(F, S) = \mathcal{P}^+(F, S).$$

(every Gibbs field is a n-reversible synchronous random field).

Proof: It suffices to show that

$$\bigcup_\alpha \psi_n^\alpha(\mathcal{R}_n) = \mathcal{R}_n,$$

since this means that for every Gibbs distribution π on Ω, there exists an n-periodic synchronous random field ν such that, for all $x \in (F \setminus \{a\})^S$ and all $B \subset S$,

$$\log \frac{\nu(x_B^a)}{\nu(\mathbf{a})} = \log \frac{\pi(x_B^a)}{\pi(\mathbf{a})},$$

which is exactly

$$\log \frac{\nu(x)}{\nu(\mathbf{a})} = \log \frac{\pi(x)}{\pi(\mathbf{a})},$$

for all $x \in \Omega$, which implies $\pi = \nu$.

Thus, let $\mathbf{v}_0 \in \mathcal{R}_n$, and set $\mathbf{u}_0 = (\psi_n^\infty)^{-1}(\mathbf{v}_0)$. We know that $\psi_n^\alpha(\mathbf{u}_0)$ converges to $\mathbf{v}_0$, so let α_0 be such that, for all $\alpha > \alpha_0$,

$$\|\psi_n^\alpha(\mathbf{u}) - \mathbf{u}_0\| < \rho_{nn}/2\,,$$

where ρ_{nn} is the number given in theorem 3 for $k = n$. Let M be large enough so that the ball $\mathcal{B}_n(\mathbf{u}_0, r_{nn})$ is included in $\mathcal{O}_n(M)$. Then, for all $\alpha > \alpha_{nn}(M)$, we know that the ball $\mathcal{B}_k(\psi_n^\alpha(\mathbf{u}_0), \rho_{nn})$ lies in $\psi_n^\alpha(\mathcal{R}_n)$, which implies that $\mathbf{v}_0 = \psi_n^\alpha(\mathbf{u})$ for some $\mathbf{u}$.

3.4. **Proof of theorem 2.** We first study the differential of ψ in $\mathbf{u}$, $d_\mathbf{u}\psi_k^\alpha$. Denote by $\tilde{K}_{C,y_C}$ the function, defined on Ω^k by

$$\tilde{K}_{C,y_C}(x^1, \ldots, x^k) = \frac{1}{k} \sum_{p=0}^{k-1} \prod_{q=1}^{l} \delta_{y_{c_q}}(x_{c_q}^{k+q}),$$

where $C = \{c_1, \ldots, c_l\}$, so that $\tilde{U}$ given in equation (6) may be written

$$\tilde{U}(X) = \sum_{C \subset S} \sum_{y_C} u_{C,y_C} \tilde{K}_{C,y_C}(X).$$

We have (here and in the sequel, $E_\mathbf{u}^\alpha$ refers to expectation with respect to $\mu_\mathbf{u}^\alpha$, the Gibbs field on Ω^k with energy $\tilde{U} + \alpha V$):

Proposition 4.

$$\frac{d}{du_{C,y_C}} \log \frac{\nu_\mathbf{u}^\alpha(x_B^a)}{\nu_\mathbf{u}^\alpha(\mathbf{a})} = E_U^\alpha[\tilde{K}_{C,y_C} \mid x^1 = \mathbf{a}] - E_U^\alpha[\tilde{K}_{C,y_C} \mid x^1 = x_B^a]. \tag{26}$$

(27)
$$\frac{d^2}{du_{C,y_C}du_{C',z_{C'}}}\log\frac{\nu_{\mathbf{u}}^\alpha(x_B^a)}{\nu_{\mathbf{u}}^\alpha(\mathbf{a})} = cov_{\mathbf{u}}^\alpha(\tilde{K}_{C,y_C},\tilde{K}_{C',z_{C'}} \mid x^1=\mathbf{a}) - cov_{\mathbf{u}}^\alpha(\tilde{K}_{C,y_C},\tilde{K}_{C',z_{C'}} \mid x^1 = x_B^a)\ .$$

Proof: These identities are applications of the following well-known result: let $P_\theta(\omega_1,\omega_2)$ be a probability distribution over the finite set $\Omega_1\times\Omega_2$, and let Q_θ be its marginal over Ω_1. Assume that P_θ is twice differentiable with respect to the parameter $\theta\in\mathbb{R}^d$. Then, one has :

$$\frac{d}{d\theta}\log Q_\theta(\omega) = E_P(\frac{d}{d\theta}\log Q_\theta \mid \omega_1=\omega).$$

$$\frac{d^2}{d\theta^2}\log Q_\theta(\omega) = \mathrm{var}_P(\frac{d}{d\theta}\log P_\theta \mid \omega_1=\omega) + E_P(\frac{d^2}{d\theta^2}\log P_\theta \mid \omega_1=\omega).$$

We apply this result with $P_\theta=\mu_{\mathbf{u}}$ and $Q_\theta=\nu_{\mathbf{u}}$. The computation of the derivatives of $\log\mu_{\mathbf{u}}$ with respect to $\mathbf{u}$ is easy, since $\mu_{\mathbf{u}}$ is an exponential family of probability measures: one gets

$$\frac{d}{du_{C,y_c}}\log\mu(x^1,\ldots,x^n) = E[\tilde{K}_{C,y_C}(x^1,\ldots,x^n)] - \tilde{K}_{C,y_C}(x^1,\ldots,x^n),$$

and

$$\frac{d^2}{du_{C,y_C}du_{C',z_{C'}}}\log\mu(x^1,\ldots,x^n) = cov(\tilde{K}_{C,y_C},\tilde{K}_{C',z_{C'}}).$$

Taking conditional expectations yields the expressions in proposition 4. □

Assume that $|\mathbf{u}|\le M$. We have defined $N(\mathbf{u})$ to be the maximum, over all $s\in S$, of the number of sets $C\subset S$ such that $s\in C$ and $u_C\neq 0$. This number is smaller than $\binom{n-1}{k-1}$ for $\mathbf{u}\in\mathcal{R}_k$. Set $\Delta = M\binom{n-1}{k-1}$, which is therefore larger than $N(\mathbf{u})|\mathbf{u}|$. In this proof, we shall always assume that $\alpha > 2\log k + |F| + \Delta + 1$.

According to lemma 2 and proposition 4, $-\frac{d}{du_{C,y_C}}\log\frac{\nu_{\mathbf{u}}^\alpha(x_B^a)}{\nu_{\mathbf{u}}^\alpha(\mathbf{a})}$ is, for large enough α, close to

$$\tilde{K}_{C,y_C}(x_B^a,\ldots,x_B^a) = \prod_{s\in C}\delta_{y_s}[(x_B^a)_s],$$

which is zero $C\not\subset B$, and $\prod_{s\in C}\delta_{y_s}(x_s)$ otherwise (note that, by definition of $\mathcal{R}_k$, $y_s\neq a$). This is in fact the coefficient at the (B,x_B) line and (C,y_C) column of the mapping ψ^∞ considered as a matrix.

Now, write

$$d_{\mathbf{u}}\psi = \psi^\infty + R,$$

which defines R. If $\|R(\psi^\infty)^{-1}\|\le 1$, the system $\lambda' = (\psi^\infty + R)\lambda$, for $\lambda'\in\mathcal{R}_k$ and $\lambda\in\mathcal{R}_k$, yields

$$\lambda = (\psi^\infty)^{-1}\lambda' - (\psi^\infty)^{-1}R\lambda = (\psi^\infty)^{-1}(\lambda' - [R(\psi^\infty)^{-1}]\lambda' + \cdots + (-1)^q[R(\psi^\infty)^{-1}]^q\lambda' + \cdots)$$

which uniquely defines λ. Moreover, we have

$$\|\lambda\| \le \frac{\|(\psi^\infty)^{-1}\|}{1-\|R(\psi^\infty)^{-1}\|}.$$

If $\lambda' = \psi^\infty(\lambda)$ equation (21) implies that $|\lambda_{C,x_C}| \le 2^{|C|}\|\lambda'\|$, which yields $\|(\psi^\infty)^{-1}\|\le 2^k$.

Moreover, fixing a C with $|C| = k$, and letting $\lambda'_{B,x_B} = 1$ is $B \subset C$ and $|C| - |B|$ is even, and 0 otherwise, we see that the norm of $(\psi^\infty)^{-1}$ is larger than the number of subsets of C which a cardinality of the same parity as $|C|$, which is 2^{k-1}. Since a lower bound to $\|(\psi^\infty + R)^{-1}\|$ is

$$\|(\psi^\infty)^{-1}\| \left(\frac{1 - 2\|R(\psi^\infty)^{-1}\|}{1 - \|R(\psi^\infty)^{-1}\|} \right) ,$$

the inequality

$$\|R(\psi^\infty)^{-1}\| < 1/4$$

is more than enough for (22) to be true. Therefore, we shall choose α such that $\|R\| \leq 2^{-k-2}$.

To estimate $\|R\|$, we use proposition 2, since every component of the matrix R is the sum of two terms which have the form estimated in equation (14), with $|C| \leq k$. We see that, for $\lambda \in \mathcal{R}_k$, $\|\lambda\| = 1$,

$$\|R\lambda\| \leq \max_{B, y_B} \sum_{C, x_C} R_{C,x_C}(B, y_B)\lambda_{C,x_C} \leq 2k^2 e^{-\alpha+\Delta} \mathrm{dim}\mathcal{R}_k \ , \tag{28}$$

so that the first part of theorem 2 comes if we take

$$\alpha > \alpha_{kn} = \log(2^{k+2} 2k^2 n e^{\Delta} \mathrm{dim}\mathcal{R}_k) \, . \tag{29}$$

We now estimate the second derivative of ψ^α. Denote, for short, by $A_{(C,y_C),(C',z_{C'})}(B, x_B)$ the expression of the the second derivative of the corresponding coefficient of ψ given in equation (27). The norm of the second derivative of ψ is the supremum, for $\lambda, \overline{\lambda} \in \mathcal{R}_k$, $\|\lambda\| = \|\overline{\lambda}\| = 1$, of the norm of the vector

$$\sum_{(C,y_C),(C',z_{C'})} \lambda_{C,y_C} \overline{\lambda}_{C',z_{C'}} A_{(C,y_C),(C',z_{C'})}(B, x_B)$$

in $\mathcal{R}_k$, that is

$$\max_{B, x_B} \sum_{(C,y_C),(C',z_{C'})} \lambda_{C,y_C} \overline{\lambda}_{C',z_{C'}} A_{(C,y_C),(C',z_{C'})}(B, x_B) \, .$$

It is smaller than

$$(\mathrm{dim}\mathcal{R}_k)^2 \max A_{(C,y_C),(C',z_{C'})}(B, x_B) \leq 2(\mathrm{dim}\mathcal{R}_k)^2 \, .$$

□

3.5. **Representation by elements of $\mathcal{S}_k$: case of local potentials on the integer lattice.** In equation (23), which gives the value of $\alpha_{kn}(\mathbf{u})$, two terms depend on n, namely

$$\mathrm{dim}\mathcal{R}_k = \sum_{l=1}^{k} \binom{n}{l} (|F| - 1)^l,$$

and $\binom{n-1}{k-1}$, which is the upper bound of $N(\mathbf{u})$ for $\mathbf{u} \in \mathcal{R}_k$. We can get rid of this dependancy by specifying additional constraints to the potential.

We consider the case when S is a subset of $\mathbb{Z}^d$, or S is a d-dimensional torus $S = \prod_{i=1}^d \mathbb{Z}/n_i\mathbb{Z}$. For $s \in S$, we let $|s| = \max_{i=1}^d |s_i|$ (taking the representation of s_i of smaller modulus in the case of the torus. We say that a potential $\mathbf{u}$ on S has radius h if $u_{C,x_C} = 0$ for all C with diameter larger than h. Denote by $\mathcal{H}_h$ the set of potentials with radius h. Setting $k = (2h)^d$, we have $\mathcal{H}_h \subset \mathcal{R}_k$. When $\mathbf{u} \in \mathcal{H}_h$, we have $N(\mathbf{u}) \leq 2^k$ which is independent of n. We let $k = (2h)^d$ in the following.

We may define a mapping ψ_h^α, from on $\mathcal{H}_h$ into $\mathcal{H}_h$, by associating to $\mathbf{u}$ the collection of the $-\log \nu_k^\alpha(x_B^a)/\nu_k^\alpha(\mathbf{a})$ for $\mathrm{diam}(B) \le h$. We can carry over all details of the proof of theorem 2 with $\mathcal{H}_h$ instead of $\mathcal{R}_k$, but the estimate of $\|R\|$ in equation (28) can be significantly improved. In the present case, one has (cf. prop 2)

Proposition 5. *Let $\mathbf{u} \in \mathcal{H}_h$, $\alpha > 0$, and $\mu = \mu_{k,\mathbf{u}}^\alpha$, with $k = (2h)^d$. Let $B_1, B_2 \subset S$ and f be a function defined on Ω^k, depending only on $x_s^1, \dots, x_s^k$ with $s \in B_1$, such that $0 \le f(x) \le 1$. Let $x, y \in \Omega$ be such that $x_s = y_s$ for all $s \in S \setminus B_2$. Then*

$$|E(f \mid x^1 = x) - E(f \,|\, x^1 = y)| \le k|B_1|\gamma_*^{\frac{d(B_1,B_2)}{h}-2}. \tag{30}$$

where the expectation is with respect to μ, and $\gamma_* = 2k^2 e^{-\alpha + N(u)|u|}$.

Moreover, if $\gamma_* < 1$, and g is another function depending on $x_s^1, \dots, x_s^k$ with $s \in B_2$, such that $0 \le g(x) \le 1$, one has

$$|cov(f, g \,|\, x^1 = x)| \le |B_1||B_2|\gamma_*^{\frac{d(B_1,B_2)}{h}-2}. \tag{31}$$

We prove this proposition at the end of this section. Fix an $M > 0$, and consider $\mathbf{u}$ such that $|\mathbf{u}| < M$. Applying proposition 5 to R, together with the estimate of proposition 2, which still holds, we get that, if $|C| \le k$

$$|R_{C,x_C}(B, y_B)| \le 2k^2 \min\left(e^{-\alpha+\Delta}, \gamma_*^{\frac{d(B,C)}{h}-2}\right),$$

in which we may take for Δ an upper bound for $N(\mathbf{u})|\mathbf{u}|$ for $\mathbf{u} \in \mathcal{H}_h$ and $|\mathbf{u}| \le M$, that is $\Delta = \Delta(h, M) = 2^k M$ (with $k = (2h)^d$).

We must estimate $A = \sum_{C,x_C} |R_{C,x_C}(B, y_B)|$. Pick some site $s_0 \in B$. Since only at most $|F|^k$ configurations x_C are concerned for each C, and each site s is contained in at most 2^k sets of diameter less than h, there exists a constant $K(d, h, |F|)$ such that

$$A \le K. \sum_{s \in S} \min(e^{-\alpha+\Delta}, \gamma_*^{\frac{d(s,s_0)}{h}-4}).$$

(the letter K stands for a generic constant, of which we shall not trace the value).

We have, for $p > 4h$,

$$\sum_{s, d(s,s_0) > p} \gamma_*^{\frac{d(s,s_0)}{h}-4} < \sum_{q > p} \sum_{s, d(s,s_0) = q} \gamma_*^{q/h-4} < \sum_{q > p-4h} (2q+8h)^d \gamma_*^{q/h} < K(d)(p)^d \gamma_*^{p/h-4}.$$

This implies, taking $p = 5h$, $A \le K(d, |F|, h)[(5h)^d e^{-\alpha+\Delta} + \gamma_*^{1/h}]$. Since $\gamma_* = C(h)e^{-\alpha+\Delta}$, there exists α_h^0 independent on $|S|$, such that $\|R\| < 2^{-k-2}$ for $\alpha > \alpha_h^0$. The rest of the proof of theorem 2 remaining unchanged, we have proved the first part of

Theorem 5. *Assume that $S \subset \mathbb{Z}^d$, or $S = \prod_{i=1}^d \mathbf{Z}/n_i\mathbf{Z}$. For all $M > 0$, there exists $\alpha_h(M) > 0$, independent on $|S|$, such that, for all $\alpha > \alpha_h(M)$, for all $\mathbf{u} \in \mathcal{H}_h$ with $|\mathbf{u}| < M$, $d_\mathbf{u}\psi^\alpha$ is invertible and*

$$2^{k-2} \le \|(d_\mathbf{u}\psi^\alpha)^{-1}\| \le 2^{k+1},$$

with $k = (2h)^d$.

Moreover, there exists a constant C_h, independent on $|S|$, such that, for $\alpha > \alpha_h(M)$,

$$\|d_\mathbf{u}^2\psi^\alpha\| \le C_h,$$

This theorem therefore states that the model is identifiable for α larger than a lower bound which does not depends on the size of S.

Proof of theorem 5:

It remains to prove the second part, that is, to estimate the second derivative of ψ with a bound which does not depend on $|S|$. Denote, as in paragraph 3.4, by $A_{(C,y_C),(C',z_{C'})}(B,x_B)$ the partial second derivative of ψ with respect to u_{C,y_C} and $u_{C',z_{C'}}$. According to proposition 4, it is given by

(32) $$cov^{\alpha}_{\mathbf{u}}(\tilde{K}_{C,y_C},\tilde{K}_{C',z_{C'}} \mid x^1=\mathbf{a}) - cov^{\alpha}_{\mathbf{u}}(\tilde{K}_{C,y_C},\tilde{K}_{C',z_{C'}} \mid x^1=x^a_B)\ .$$

By proposition 5, this quantity may be bounded in two ways. First, each covariance term is smaller than a quantity of the kind

$$K(h,|F|)\gamma_*^{\frac{d(C,C')}{h}-2}\ .$$

Moreover, since (32) also involves differences between conditional expectations given $\mathbf{a}$ or x^a_B, it is also smaller than

$$K(h,|F|)\gamma_*^{\frac{\min(d(C,B),d(C',B))}{h}-2}\ .$$

The norm of the second differential $d^2_{\mathbf{u}}\psi$ is given by

$$\max_{B,x_B}\sum_{(C,y_C),(C',z_{C'})}|A_{(C,y_C),(C',z_{C'})}(B,x_B)|,$$

and is therefore smaller than

$$K(h,|F|)\sum_{(C,y_C),(C',z_{C'})}\min\left(1,\gamma_*^{\frac{d(C,C')}{h}-2},\gamma_*^{\frac{\min(d(C,B),d(C',B))}{h}-2}\right).$$

Using the fact that, for each C, there is at most $|F|^k$ configurations $x_C\in\Omega_C$, that each site s may be element of at most $2^{(2h)^d}$ sets C, the above sum is smaller than

$$K(h,|F|)\sum_{s,t\in S}\min\left(1,\gamma_*^{\frac{|s-t|}{h}-2},\gamma_*^{\frac{\min(|s-s_0|,|t-s_0|)}{h}-2}\right),$$

where s_0 is any fixed element of B (in all the preceding estimates, $K(h,|F|)$ means a function of only h and $|F|$, but for which the expression may vary from line to line).

Let $\eta=\gamma_*^{\frac{1}{h}}$, and denote by G the sum above. We have, noting that, for each $s\in S$, there are at most $2^d(2p+1)^{d-1}$ sites t such that $|t-s|=p$,

$$G\le\sum_{p\ge 2h}2^{2d}(p+1)^d\eta^{p-2}\sum_{s\in S}\min(1,\eta^{|s-s_0|-2p})+\sum_{p<2h}2^{2d}(p+1)^d\sum_{s\in S}\min(1,\eta^{|s-s_0|-p-2})\ .$$

But $\sum_{s\in S}\min(1,\eta^{|s-s_0|-2p})$ is smaller than

$$(2p)^d+\sum_{q\ge 0}2^d(2p+2q)^{d-1}\eta^q$$

and this is smaller than a polynomial in p, so that the first sum in the right-hand term of the above inequality is bounded by a constant, depending on h, $|F|$ and γ_*, and is a increasing function of $\gamma_*<1$. Since it is clearly the same for the second sum, we obtain the fact that G is bounded by a increasing function of γ_* (and thus decreasing in α), which finishes the proof of theorem 5. □

We also can obtain a more precise measure of the difference between ψ_h^α and its limit ψ_h^∞, by the following proposition, which is a refinement of lemma 1. The result is an estimate which is again independent on $|S|$.

Proposition 6. *We have, for all* $\mathbf{u} \in \mathcal{H}_h$,

$$\|\psi_h^\alpha(\mathbf{u}) - \psi_h^\infty(\mathbf{u})\| \leq (4h)^d g(h, 2^k|\mathbf{u}|, \alpha) \tag{33}$$

with the same g *as in theorem 1.*

Proof : We must estimate the quantity

$$Q = \log \frac{\nu_{k,\mathbf{u}}^\alpha(x_B^a)}{\nu_{k,\mathbf{u}}^\alpha(\mathbf{a})} - \log \frac{\pi_{\mathbf{u}}(x_B^a)}{\pi_{\mathbf{u}}(\mathbf{a})}$$

for $\mathrm{diam}(B) < h$.

We have (dropping indices $\alpha, k, \mathbf{u}$),

$$\frac{\nu(x_B^a)\pi(\mathbf{a})}{\pi(x_B^a)\nu(\mathbf{a})} = \frac{\sum_{X\ ;\ x^1 = x_B^a} e^{-\alpha V(X) - \tilde{U}(X) + U(x_B^a)}}{\sum_{X\ ;\ x^1 = \mathbf{a}} e^{-\alpha V(X) - \tilde{U}(X) + U(\mathbf{a})}}\,, \tag{34}$$

where U and $\tilde{U}$ are the energies associated to $\mathbf{u}$ on Ω and Ω^k, as in section 3.1. Since $\mathbf{u}$ is normalized with respect to a, we see from equation 5 that $\tilde{u}_C(x_B^a, x^2, \ldots, x^k) = 0$ for all C such that $C \cap B = \emptyset$. This implies that $\tilde{U}(\mathbf{a}, x^2, \ldots, x^k) = U(\mathbf{a}) = 0$ and that

$$\tilde{U}(x_B^a, x^2, \ldots, x^k) = \sum_{C,\, C \cap B \neq \emptyset} \tilde{u}_C(x_B^a, x^2, \ldots, x^k)\,.$$

Let $\overline{B}$ be the set of all $s \in S$ such that $\mathrm{dist}(s, B) \leq h$. The energies U and $\tilde{U}$ in (34) only depend on x_s^p for $p = 1, \ldots, k$ and $s \in \overline{B}$. For $B' \subset S$, denote by $V_{B'}$ the function

$$\sum_{s \in B'} \sum_{p=1}^{k} d(x_s^p, x_s^{p+1})\,,$$

and for $X = (x^1, \ldots, x^p) \in \Omega^k$, denote by $X_{B'}$ the p-uple formed with the restrictions $(x_{B'}^1, \ldots, x_{B'}^p)$. With this notation, one has

$$\frac{\nu(x_B^a)\pi(\mathbf{a})}{\pi(x_B^a)\nu(\mathbf{a})} = \frac{\sum_{X_{\overline{B}^c}\ ;\ x^1_{\overline{B}^c} = \mathbf{a}_{\overline{B}^c}} e^{-\alpha V_{\overline{B}^c}(X)} \sum_{X_{\overline{B}}\ ;\ x^1_{\overline{B}} = x_{\overline{B}}} e^{-\alpha V_{\overline{B}}(X) - \tilde{U}(X) + U(x_B^a)}}{\sum_{X_{\overline{B}^c}\ ;\ x^1_{\overline{B}^c} = \mathbf{a}_{\overline{B}^c}} e^{-\alpha V_{\overline{B}^c}(X)} \sum_{X_{\overline{B}}\ ;\ x^1_{\overline{B}} = \mathbf{a}_{\overline{B}}} e^{-\alpha V_{\overline{B}}(X)}} \tag{35}$$

This implies that the ratio $\nu(x_B^a)\pi(\mathbf{a})/\pi(x_B^a)\nu(\mathbf{a})$ is always smaller than the maximum and larger than the minimum of

$$\frac{\sum_{X_{\overline{B}}\ ;\ x^1_{\overline{B}} = x_{\overline{B}}} e^{-\alpha V_{\overline{B}}(X) - \tilde{U}(X) + U(x_B^a)}}{\sum_{X_{\overline{B}}\ ;\ x^1_{\overline{B}} = \mathbf{a}_{\overline{B}}} e^{-\alpha V_{\overline{B}}(X)}}\,. \tag{36}$$

Letting $\Delta = 2^k|\mathbf{u}| \geq N(\mathbf{u})|\mathbf{u}|$, we see, as in the proof of lemma 1, that the logarithm of the expression in (36) is, in absolute value, smaller than $|\overline{B}|g(k, \Delta, \alpha)$, so that (33) comes from the fact that $|\overline{B}| \leq (4h)^d$. □

This proposition enables us to prove that ψ^h is in fact a diffeomorphism on the compact subsets of $\mathcal{H}_h$ for large enough α independent on $|S|$. Denote by $\mathcal{O}_h(M)$ the set of potential $\mathbf{u} \in \mathcal{H}_h$ with $|\mathbf{u}| < M$.

Theorem 6. *For all $M > 0$, there exists an $\overline{\alpha}_h(M) > \alpha_h(M)$ such that, for all $\alpha > \overline{\alpha}_h(M)$, the restriction of ψ_h^α to $\mathcal{O}_h(M)$ is a diffeomorphism onto its image.*

Indeed, using the inverse mapping theorem for ψ_h^α, we see that there exists $\varepsilon > 0$ (depending on M, h, but not on $|S|$ and α) such that if $\max(|\mathbf{u}|, |\mathbf{u}'|) < M$, $\|\mathbf{u} - \mathbf{u}'\| < \varepsilon$ and $\mathbf{u} \neq \mathbf{u}'$, then $\psi_h^\alpha(\mathbf{u}) \neq \psi_h^\alpha(\mathbf{u}')$. Moreover, since ψ_h^∞ is linear, invertible, and $\|(\psi_h^\infty)^{-1}\| \leq 2^k$, we have

$$\|\mathbf{u} - \mathbf{u}'\| \leq 2^k \|\psi_h^\infty(\mathbf{u}) - \psi_h^\infty(\mathbf{u}')\| .$$

Finally, for large α (independent on $|S|$), we have $\|\psi_h^\alpha(\mathbf{u}) - \psi_h^\infty(\mathbf{u})\| \leq 2^{-k-1}\varepsilon$ for $\|\mathbf{u}\| < M$. These facts together imply that one cannot have, for large α, $\psi_h^\alpha(\mathbf{u}) = \psi_h^\alpha(\mathbf{u}')$ and $\mathbf{u} \neq \mathbf{u}'$.

3.6. **Dobrushin's comparison theorem.** Since the proof of proposition 5 is based on Dobrushin's comparison theorem ([Dobrushin 1968]), as reformalized in ([Föllmer 1982]), we give a brief account of the results. Consider an at most countable set I and the associated configuration space $E = F^I$. If π and $\overline{\pi}$ are two probability distributions over E, an *estimate* for π and $\overline{\pi}$ is a family $(a(i),\, i \in I)$ of positive numbers such that, for all function f on E, which only depends on a finite number of coordinates,

$$\left| \int f d\pi - \int f d\overline{\pi} \right| \leq \sum_i a(i)\omega_i(f) ,$$

where $\omega_i(f)$ denotes the oscillation of f at site i.

If π is a Gibbs distribution on E, denote by $\pi_i(dx_i \,|\, x_j, j \neq i)$ the conditional distribution at site $j \in I$ given the state of all other sites. Define the matrix $\Gamma = (\gamma(i,j))$ by

$$\gamma(i,j) = \|\pi_i(\,.\,|\,x) - \pi_i(\,.\,|\,y)\| = \sup \frac{1}{2} \sum_{\lambda \in F} |\pi_i(\lambda|x) - \pi_i(\lambda|y)| ,$$

the supremum being computed over all $x, y \in E$ with $x_k = y_k$ for $k \neq j$.

Finally, π and $\overline{\pi}$ being given, define the family $b(i)\,, i \in I$ by

$$b(i) = \int \|\pi_i(\,.\,|\,x) - \overline{\pi}(\,.\,|\,x)\| \overline{\pi}(dx) .$$

Lemma 5 ([Dobrushin 1968], [Föllmer 1982])**.**

1. *If a is an estimate for π and $\overline{\pi}$, the the vector $a\Gamma + b$ is also an estimate.*
2. *Assume that $\sum_j \gamma(i,j) \leq \gamma_* < 1$ for all i. Let $\Xi = (\xi(i,j)) = \sum_{p\geq 0} \Gamma^p$ Then, for any two functions f and g depending on a finite number of coordinates,*

$$cov_\pi(f,g) \leq \frac{1}{4} \sum_{i,j} \xi(i,j)\omega_i(j)\omega_j(g). \tag{37}$$

We now can give the proof of proposition 5

3.7. **Proof of proposition 6.** Denote by μ^1 (resp. μ^2) the distribution $\mu(\,.\,|\,x^1 = x)$ (resp. $x^1 = y$). Both may be seen as Gibbs distributions on Ω^{k-1}, and we shall apply lemma 5 with $I = \{2,\ldots,k\}\times S$, $\pi = \mu^1$ and $\overline{\pi} = \mu^2$. We must compute the matrix Γ and the vector b for these distributions.

For the vector b, fix a site $i = (l,s) \in I$. By construction of the potential $\tilde{\vec{u}}$, the conditional distributions for μ at i given all other sites in $\{1,\ldots,k\}\times S$ only depends on sites (l',s') with $|s-s'| \leq h$. This implies in particular that $b(i) = 0$ whenever $s \notin \tilde{B}_2$, where $\tilde{B}_2$ is the union of all $C \subset S$ such that $\mathrm{diam}(C) \leq h$ and $C \cap B_2 \neq \emptyset$. If $s \in \tilde{B}_2$, we have $b(i) \leq 1$.

We have a straightforward estimate of $\gamma(i,j)$ for μ^1. Indeed, At site (l,s), the conditional expectation only depends on sites (l',t) such that $|s-t| \leq h$, so that $\gamma(i,j) = 0$ if $|s-t| > h$. Moreover, whatever the external condition is, the conditional distribution for μ^1 at site $i = (l,s)$ is almost equal, for large α, to the Dirac measure at state x_s. Applying estimates such as the ones of lemma 2 in the case of $|S| = 1$, we have

$$\gamma(i,j) \leq 2e^{-\alpha + N(u)|u|}\,.$$

when $|s-t| \leq h$.

Finally, if f is as in proposition 5, one has $\omega_i(f) = 0$ for all $i = (l,s)$ with $s \notin B_1$.

Start with the initial estimate $a_i \equiv 1$ for μ^1 and μ^2, and let $D = \int f d\mu^1 - \int f d\mu^2$. Iterating lemma 5 one has, for all integer $p > 0$:

$$D \leq \sum_{l=2}^{k}\sum_{s\in B_1}\left(a\Gamma^p + \sum_{q=0}^{p-1} b\Gamma^q\right)(l,s)\,.$$

Let $\gamma_* = \max_i \sum_j \gamma_(i,j)$: $\gamma_* \leq 2k^2 e^{-\alpha+N(u)|u|}$. Then, one has $(\Gamma^p)(i,j) \leq \gamma_*^p$ if $i = (l,s)$ and $j = (i',s')$ with $|s-s'| \leq ph$, and $(\Gamma^p)(i,j) = 0$ if $|s-s'| > ph$. This implies that, if p is the largest integer such that one cannot have $s \in B_1$, $s' \in \tilde{B}_2$ and $|s-s'| \leq ph$, ie p is the integer part of $d(B_1,\tilde{B}_2)/h$, then,

$$D \leq (k-1)|B_1|\gamma_*^p\,,$$

which proves the first part of the proposition. The second part is almost straightforward from estimate (37). □

3.8. **Case of stationary local potentials.** In this paragraph, we assume that S is a d-dimensional torus $S = \prod_{i=1}^d \mathbf{Z}/n_i\mathbf{Z}$. In addition to the locality assumptions, we may introduce the constraint that a potential $\mathbf{u} \in \mathcal{H}_h$ is stationary, that is, for all $s \in S$, $C \subset S$, $x \in \Omega$,

$$u_{C+s}(T_s x) = u_C(x)\,, \tag{38}$$

where $T_s x$ is the configuration $y \in \Omega$ such that $y_t = x_{s+t}$. Denote by $\mathcal{H}_h^s(F,S)$ the set of all stationary potentials with radius h, which is a linear subspace of $\mathcal{H}_h(F,S)$.

Given a stationary potential $\mathbf{u}$ in $\mathcal{H}_h^s$, we can construct an associated potential on Ω^k, $\tilde{\vec{u}}$, which is stationary too. For this, a little care must be taken in the ordering of the sets C which is assumed in equation (5), since this ordering must be invariant by translation, ie. if the elements of C are numbered in $c_1,\ldots,c_l$, and those of $C' = C+s$ in $c'_1,\ldots,c'_l$, one must have $c'_p = c_p + s$ for all p (this can always be achieved).

We have the proposition, in which everything is obvious, by construction or by theorem 6.

Proposition 7. *Assume that the ordering of the subsets of S is invariant by translation. Then, the application ψ_h^α defined in section 3.5 leaves $\mathcal{H}_h^s$ invariant.*

Therefore, for all $M > 0$, for all $\alpha > \alpha_h(M)$, the set of all potential $\mathbf{u}$ in $\mathcal{H}_h^s$ such that $|\mathbf{u}| < M$ is diffeomorphic to its image by ψ^α.

This proposition will be used in the next section.

4. Case of infinite S.

Let S be an infinite, countable lattice. Gibbs fields in this case may be defined by means of an infinite potential, which is a family $\mathbf{u} = (u_C)$ of functions indexed by the finite subsets of S. In the following, we assume that the reader is acquainted with the basic definitions and properties of Gibbs fields over countable lattices (see [Georgii 1988]).

Definition 2.1 of $\mathcal{D}_k(F,S)$ and $\mathcal{S}_k(F,S)$ remains valid in this case, since points a. and b. are meaningful also in the infinite dimensional case. One may ask the same question as asked in ([Geman et al 1993]) for hidden Markov random fields: is $\bigcup_k \mathcal{S}_k$ dense (for convergence in distribution) in the set of Gibbs fields over S, (and thus in the set of probabilities on Ω). The answer is positive. Indeed, let π be a Gibbs field, or more generally any field which has strictly positive marginals on finite subsets of S. Consider a increasing sequence S_n of finite subsets of S, such that $\bigcup_n S_n = S$. Let π_n be the random field given by:

$$\pi_n = \pi_{|S_n} \otimes \bigotimes_{s \notin S_n} \eta,$$

where η is the uniform probability measure on F and $\pi_{|S_n}$ is the marginal of π on S_n. In other terms, π_n coincides with π for events which only depends on configurations over S_n, whereas the states of sites in $S \setminus S_n$ are mutually independent, with law η, and independent of what happens on S_n. Then π_n converge in distribution to π (note that, since F is finite, the set of probability distributions on F^S is compact). Moreover, since $\mathcal{P}^+(F,S_n) = \mathcal{S}_{|S_n|}(F,S_n)$, $\pi_{|S_n}$ is the first marginal of a distribution $\mu^n \in \mathcal{D}_{|S_n|}(F,S_n)$. Now, π_n is the marginal of the distribution μ_n on $\Omega^{|S_n|}$ equal to μ^n on $F^{|S_n|^2}$, (ie. $|S_n|$ copies of $|F^{|S_n|}|$), and such that the states are all other sites are mutually independent with law η. This last distribution is in $\mathcal{D}_{|S_n|}(F,S)$, and thus $\pi_n \in \mathcal{S}_{|S_n|}(F,S)$. We therefore have proved the theorem:

Theorem 7. *Synchronous random fields are dense in the set of probability measures on Ω.*

We now specialize to the case of the integer lattice and local potentials. In this case, our estimates were independent of the size of the set S, and we can carry some of our results to the infinite dimensional case. From now on, let $S = \mathbb{Z}^d$.

We still denote by $\mathcal{H}_h^s(F,S)$ the set of all stationary potentials on S, the definitions in section 3.8 being obvioully valid in the case $S = \mathbb{Z}^d$. We also assume that the ordering of the subsets of S is translation invariant, so that the potential $\tilde{\vec{u}}$ on Ω^k which is built from $\mathbf{u}$ in $\mathcal{H}_h^s$ remains stationary.

Denote by $\tilde{\vec{u}}^\alpha$ the potential on Ω^k containing all functions $(\tilde{u}_C, C \subset S)$, and to which are added functions corresponding to the term $\alpha V(X)$, ie. functions $tu^\alpha_{s,l,l+1}(X) = \alpha d(x_s^l, x_s^{l+1})$. Since $\tilde{\vec{u}}^\alpha$ is local on Ω^k, there exist Gibbs distributions on Ω^k which are associated to it. They all satisfy point b. and among them, some of them satisfy the permutation constraint a. and are stationary. For $\alpha > 0$ and $\mathbf{u} \in \mathcal{H}_h^s(F, S)$, we therefore can define $\mathcal{D}(\alpha, \mathbf{u})$ as the nonempty set of stationary, permutation invariant, Gibbs distributions associated to the potential $\tilde{\vec{u}}^{\alpha}$, and $\mathcal{S}(\alpha, \mathbf{u})$ as the set of first marginals of distributions in $\mathcal{D}(\alpha, \mathbf{u})$. We therefore have built a parametrization of synchronous fields by means of a potential. That the union of the sets $\mathcal{S}(\alpha, \mathbf{u})$ is dense in the set of stationary and permutation invariant Gibbs fields may be proved by more or less the same arguments as before. We shall show that this parametrization furthermore satisfies the following identifiability theorem:

Theorem 8. *For all $h > 0$, and for all $M > 0$, there exists an $\alpha_h^s(M) > 0$ such that, if $\mathbf{u}, \mathbf{u}' \in \mathcal{H}_h^s$ and $\max(|\mathbf{u}|, |\mathbf{u}'|) < M$, for all $\alpha > \alpha_h^s(M)$,*

$$\mathcal{S}(\alpha, \mathbf{u}) \cap \mathcal{S}(\alpha, \mathbf{u}') = \emptyset\,.$$

Before proving this, we wish to note that identifiability is a very important result in the context of parameter estimation. For example, it is required for consistency of maximum likelihood estimator for Gibbs distribution, and our results allow us to apply the results of Comets and Gidas ([Comets and Gidas 1992]).

Non identifiability may occur for some α. Take, for example, the "synchronous Ising model", which is a 2-synchonous random field, with $F = \{-1, 1\}$, defined as the first marginal of

$$\mu(x^1, x^2) = \frac{1}{Z} \exp(\alpha \sum_s x_s^1 x_s^2 + \beta \sum_{s \neq t} x_s^1 x_t^2)\,.$$

In this case, when $\alpha = 0$, it is easy to check that β and $-\beta$ yield the same marginal distribution for x^1. In this particular case, one may show, at the cost of some lengthy computation, that the model is globally identifiable when $\alpha \neq 0$.

Proof of theorem 8: Let $\mathbf{u} \in \mathcal{H}_h^s$, $\alpha > 0$ and $\nu \in \mathcal{S}(\alpha, \mathbf{u})$. Denote by μ the associated element of $\mathcal{D}(\alpha, \mathbf{u})$. Consider an increasing sequence of hypercubes $S_n = [-c_n h, c_n h]^d$, where $c_n = 2d_n + 1$ is an increasing sequence of odd numbers tending to infinity. We denote by μ_n the associated Gibbs distribution on $\{1, \ldots, k\} \times S_n$ with periodic boundary conditions. It is the element of $\mathcal{D}(S_n, F)$ associated to the potential $\tilde{\vec{u}}_n^{\alpha} = (\tilde{u}_C^{\alpha}(\hat{x}_C), C \cap S_n \neq \emptyset)$, where $\hat{x}$ is the configuration in Ω obtained from $x \in \Omega_{S_n}$ by periodic replication outside S_n. Let ν_n be the first marginal of μ_n. Define in the same way, for $\mathbf{u}' \in \mathcal{H}_h$, the distributions μ'_n and ν'_n on S_n. From the results of the previous section, we know that there exists an $\alpha_0 > 0$, $\varepsilon > 0$ and $\tau > 0$ such that, for all $\alpha > \alpha_0$, for all n, there exists $B_n \subset S_n$ with $\mathrm{diam}(B_n) \leq h$, and a configuration x_{B_n} over B_n such that

$$|\log \frac{\nu_n(x_{B_n}^a)}{\nu_n(\mathbf{a})} - \log \frac{\nu'_n(x_{B_n}^a)}{\nu'_n(\mathbf{a})}| > \min\left[\varepsilon, \tau \max(|u_C(x_C) - u'_C(x_C)|)\right]\,.$$

But, since each ν_n is stationnary, the left hand term of the preceding inequality is invariant by tranlation, so that we may assume that $0 \in B_n$, which implies that B_n (and thus x_{B_n}) may only take a finite number of values. Thus, replacing c_n by a subsequence if necessary, we may take $B_n = B$ and x_B independent of n, which yields the result: if $\mathbf{u} \neq \mathbf{u}'$,

$$\liminf |\log \frac{\nu_n(x_B^a)}{\nu_n(\mathbf{a})} - \log \frac{\nu'_n(x_B^a)}{\nu'_n(\mathbf{a})}| > 0\,. \tag{39}$$

It remains to show that this cannot happen (at least for large α) when $\nu = \nu'$. Note that

$$\frac{\nu_n(x_B^a)}{\nu_n(\mathbf{a})} = \frac{\nu_n(x_B \,|\, x_t = a, t \in S_n \setminus B)}{\nu_n(x_s = a, s \in B \,|\, x_t = a, t \in S_n \setminus B)} .$$

The end of the proof comes with the help of the next lemma which implies, together with (39) that when $\mathbf{u}' \neq \mathbf{u}$,

$$\left| \log \frac{\nu(x_B \,|\, x_t = a, t \in S_n \setminus B)}{\nu(x_s = a, s \in B \,|\, x_t = a, t \in S_n \setminus B)} - \log \frac{\nu'(x_B \,|\, x_t = a, t \in S_n \setminus B)}{\nu'(x_s = a, s \in B \,|\, x_t = a, t \in S_n \setminus B)} \right| > 0 .$$

which implies of course $\nu \neq \nu'$.

Lemma 6. *If $0 \in B$ and $\operatorname{diam}(B) \leq h$, for any configuration x_s, $s \in B$*

$$|\nu_n(x_B \,|\, x_t = a, t \in S_n \setminus B) - \nu(x_B \,|\, x_t = a, t \in S_n \setminus B)| \leq K(h, |F|, c_n)\eta^{c_n} \tag{40}$$

with $\eta < 1$ for $\alpha > \alpha_0'(\mathbf{u})$, and K is at most polynomial in c_n.

Moreover $\nu(x_B \,|\, x_t = a, t \in S_n \setminus B) > 0$ for all x_B.

The proof of lemma 6 relies again on Dobrushin's comparison results. Fix an arbitrary configuration $Z \in \Omega^k$, let $\overline{\mu}$ denote the conditional distribution for μ given that $x_t^l = z_t^l$ for $t \notin S_n$ and $l \in \{1, \dots, k\}$. It is well defined, since it only depends on z_t^l for $c_n < |t| \leq c_n + h$. We apply lemma 5 with $\pi = \mu_n(\,.\,|\, x_t = a, t \in S_n \setminus B)$ and $\overline{\pi} = \overline{\mu}(\,.\,|\, x_t = a, t \in S_n \setminus B)$. As in the proof of proposition 5, the contraction coefficients $\gamma(i,j)$ for π verify, for $i = (l,s)\,, s \in S_n \setminus B$ and $j = (l', t)$, $|s - t| \leq h$

$$\gamma(i,j) \leq 2e^{N(u)|u|}e^{-\alpha}.$$

We shall take α_0' such that $\gamma_* = (2h)^{2d} 2e^{N(u)|u|}e^{-\alpha_0'} < 1$.

If $s \in B$, $|s - t| \leq h$, we take $\gamma(i,j) \leq 1$ and $\gamma(i,j) = 0$ when $|s - t| > h$. Concerning the coefficients b_i, we have $b_i = 0$ for $i = (l,s)$, $s \in [-c_n h + h, c_n h - h]^d$.

We have taken $c_n = 2d_n + 1$, let $D_n = [-d_n h - h, d_n h + h]^d \setminus [-d_n h, d_n h]^d$, so that $B \cap D_n = \emptyset$ and $D_n \subset S_n$. Moreover, we have

$$E_\pi \left[\mathbf{1}_{x_B}(x_B^1)\right] = E_\pi \left\{E_\pi \left[\mathbf{1}_{x_B}(x_B^1) \,|\, x_t^l,\, t \in D_n\, l = 1, \dots, k\right]\right\} ,$$

and similarly for $\overline{\pi}$. But, the conditional expectations on B given D_n for π and $\overline{\pi}$ coincide. Let

$$f = E_\pi \left[\mathbf{1}_{x_B}(x_B^1) \,|\, x_t^l,\, t \in D_n\, l = 1, \dots, k\right] .$$

Then, $0 \leq f \leq 1$, f depends only on coordinates x_s^l for $s \in D_n$, and

$$\pi(x_B^1 = x_B) - \overline{\pi}(x_B^1 = x_B) = E_\pi(f) - E\overline{\pi}(f) .$$

Thus, if $(a_i\,, i \in \{1, \dots, k\} \times S_n)$ is an estimate for π and $\overline{\pi}$, we have

$$|\pi(x_B^1 = x_B) - \overline{\pi}(x_B^1 = x_B)| \leq \sum_{l=1}^{k} \sum_{s \in D_n} a_{l,s}.$$

We shall now iterate lemma 5, which says that if $(a_i^{(p)})$ is an estimate for π and $\overline{\pi}$, then $(a_i^{(p+1)})$ is also an estimate, where $a^{(p+1)} = a^{(p)}\Gamma + b$.

Let $\gamma_p(i,j)$ denote the (i,j)-coefficient of Γ^p. Since $\gamma(i,j) = 0$ if $|i - j| > h$, one has $\gamma_p(i,j) = 0$ for $i = (l,s)$, $j = (l',t)$, and $|s - t| > ph$. Moreover, we have chosen α_0 so that for $i = (l,s)$ and $s \notin B$, $\sum_j \gamma(i,j) < \gamma_* < 1$. This implies that, if $d(s,B) > ph$, $\gamma_p(i,j) < \gamma_*^{p+1}$ for all j. Thus, taking $p = d_n$, we see that $a_{l,s}^{(p)} < \gamma_* p$ for $s \in D_n$, which yields:

$$|\pi(x_B^1 = x_B) - \overline{\pi}(x_B^1 = x_B)| \le k(2d_n h)^d \gamma_*^{d_n}.$$

Recall that $\overline{\pi}$ is the conditional distribution for μ on S_n given an external condition z outside S_n, and the fact that $x_s^1 = a$ for $s \in S_n \setminus B$. Taking the mean of the preceding inequality over z implies that

$$|\nu_n(x_B^1 \,|\, x_t = a, t \in S_n \setminus B) - \nu(x_B^1 \,|\, x_t = a, t \in S_n \setminus B)| \le k(2d_n h)^d \gamma_*^{d_n} ,$$

which finishes the first part of lemma 6.

To prove the last assertion, we only need to show that $\overline{\pi}(x, \ldots, x)$ is bounded away from zero whatever to external condition z is. But this is obvious, since $\overline{\pi}$ is positive and only depends on a finite number of variables.

Thus, lemma 6 is proved, which also finishes the proof of theorem 8. □

References

[Ackley et al. 1985] D.H. Ackley, G. Hinton, T. Sejnowski (1985) : A learning algorithm for Boltzmann machines. *Cognitive Sci.* 9 147-169.

[Azencott 1987] R. Azencott (1987): Image analysis and Markov random fields. Proc. of the int. Conf. on Ind. and Appl. Math. SIAM, Paris.

[Azencott 1990] R. Azencott (1990) : Synchronous Boltzmann Machines and Gibbs Fields in *Neurocomputing edit.* Folgerman-Hierault, NATO ASI series, vol.F 68, Spinger Verlag.

[Azencott 1991] R. Azencott (1991) : High-Order Interactions and Synchronous Learning. *Proceedings of "stochastic models, statistical methods and algorithms in image analysis"*, Ed. P. Barone and A. Frigessi, Lecture Notes in Statistics 74, Springer.

[Azencott et al 1993] R. Azencott, A. Doutriaux, L. Younes (1993) : Synchronous Boltzmann machines and curve identification tasks. *Network : computations in Neural Networks.* 4 461-480.

[Besag 1974] J. Besag (1974) : Spatial Interaction and the Statistical Analysis of Lattice Systems. *J. of Roy. Stat. Soc.* B-36 pp 192-236.

[Besag and Green 1993] J. Besag and P. Green (1993) : Spatial Statistics and Bayesian Computation. *J. of Roy. Stat. Soc.* B-55.

[Comets and Gidas 1992] F. Comets and B. Gidas (1992) : Parameter estimation for Gibbs distributions from dartially dbserved data *Annals of Applied Probability* Vol 2. No 1 142-170.

[Dawson 1975] D.A. Dawson (1975) : Synchronous and asynchronous reversible Markov systems *Canad. Math. Bull.* 17 633-649.

[Dobrushin 1968] R.L. Dobrushin (1968) : The description of a random field by means of conditional probabilities and conditions of its regularity. *Thry. Prob. Appl.*, 13 197-224.

[Dobrushin 1970] R.L. Dobrushin (1970) : Prescribing a system of random variables by conditional distributions. *Thry. Prob. Appl.*, 15 458-486.

[Föllmer 1982] : H. Föllmer (1982) : A covariance estimate for Gibbs measures. *J. of Funct. Analysis* 46 387-395.

[François 1990] O. François : Ergodicité des processus neuronaux (1990). *C. Rendus. Acad. Sci. Paris t310 Série 1* 435-440.

[Geman 1991] D. Geman (1991) :*Random Fields and Inverse Problems in Imaging,* In *Proceedings of the Ecole d'été de Saint-Flour*, Lecture Notes in Mathematics 1427, Springer Verlag, New York.

[Geman 1984] D. and S. Geman (1984): Stochastic Relaxation, Gibbs Distribution and Bayesian Restoration of Images. *IEEE TPAMI.* Vol PAMI-6 pp 721-741.

[Geman et al 1993] S. Geman, A. Kehagias and H. Künsch (1993): Hidden Markov random fields. (preprint).

[Georgii 1988] H. O. Georgii (1988) : *Gibbs measures and phase transitions* De Gruyter.

[Goutsias 1991] J. Goutsias (1991) : Unilateral approximation of Gibbs random field images. *CVGIP: graphical models and image processing* Vol. 53, No 3, 240-257.

[Koslov and Vasilyev 1980] O. Koslov and N. Vasilyev (1980) : Reversible Markov chains with local interactions. In *Multicomponent Random Systems*, R.L. Dobrushin and Ya. G. Sinai Editors. (Dekker New York). 415-469.

[Künsch 1984] H. Künsch (1984) : Time reversal and stationary Gibbs measures. *Stochastic Processes and Appl.* 17 159-166.

[Lebowitz et al 1990] J.L. Lebowitz, C. Maes and E. Speer (1990) : Statistical mechanics of probabilistic cellular automata. *J. Stat. Physics* 59, 117-170.

[Little 1974] W.A. Little (1974) : The existence of persistent states in the brain. *Math. Biosci.* 19, 101,120.

[Maes and Shlosman 1989] C. Maes and S.B. Shlosman (1989) : Ergodicity of Probabilistic Cellular Automata : a constructive criterion *Comm. Stat. Phys.* 135, 233-251.

[Peretto and Niez 1886] P. Peretto and J.J. Niez (1986) : Collective properties of Neural Networks. in *Disordered systems and biological organisation*, eds. Bienenstock et al., Springer-Verlag.

[Poggio 1985] T. Poggio (1985) : Early vision: from computational structure to algorithms and parallel hardware. *Comp. Vision, Graphics, and Image Processing* vol. 31.

[Preston 1976] C. Preston (1976) : *Random Fields.* Lecture notes in mathematics 534. Springer.

[Prum 1986] B. Prum (1986) : *Processus sur un réseau et mesures de Gibbs ; applications.* Techniques stochastiques, Masson.

[Ruelle 1978] D. Ruelle (1978) : *Thermodynamic formalism.* In Encyclopedia of mathematics and its applications. Vol 5. Addisson-Wesley.

[Smith and Roberts 1993] A.F.M. Smith and G.O. Roberts (1993) : Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. J.R. Stat. Soc. B, 55.

[Sokal 1989] A. Sokal (1989) : Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms, *Cours de troisième cycle de la physique en Suisse romande*, Lausanne.

[Sussmann 1988] H.J. Sussmann (1988) : On the convergence of learning algorithms for Boltzmann Machines (preprint Rutgers University).

[Trouvé 1992] A. Trouvé (1992) : Partially parallel simulated annealing : low and high temperature approach of the invariant measure. In *Proceedings of the US-French Workshop on applied stochatic analysis,* I. Karatzas and D. Ocone eds., Lecture notes, Springer Verlag.

[Younes 1989] L. Younes (1989) : Parameter estimation for imperfectly observed Gibbsian fields. *Prob. Theory and Rel. fields* 82, 625–645.

[Younes 1992] L. Younes (1992) : Parameter estimation for imperfectly observed Gibbs fields and some comments on Chalmond's EM Gibssian algorithm. *Proceedings of "stochastic models, statistical methods and algorithms in image analysis"*, Ed. P. Barone and A. Frigessi, Lecture Notes in Statistics, Springer.

[Younes 1993a] L. Younes (1993) : Synchronous Boltzmann machines can be universal estimators. To appear in *Applied math. Letters.*

[Younes 1993b] L. Younes (1993) : Synchronous random fields and Image restoration. (preprint).