Bernoulli **19**(3), 2013, 905–930
DOI: 10.3150/12-BEJ419

Variational estimators for the parameters of Gibbs point process models

ADRIAN BADDELEY[1,2] and DAVID DEREUDRE[3]

[1]*CSIRO Mathematics, Informatics and Statistics, Leeuwin Centre, 65 Brockway Rd, Floreat WA 6014, Australia. E-mail: Adrian.Baddeley@csiro.au*
[2]*School of Mathematics & Statistics, University of Western Australia*
[3]*Laboratoire Paul Painlevé, UMR CNRS 8524, University of Lille, 59655 Villeneuve d'Ascq Cédex, France. E-mail: david.dereudre@univ-lille1.fr*

This paper proposes a new estimation technique for fitting parametric Gibbs point process models to a spatial point pattern dataset. The technique is a counterpart, for spatial point processes, of the variational estimators for Markov random fields developed by Almeida and Gidas. The estimator does not require the point process density to be hereditary, so it is applicable to models which do not have a conditional intensity, including models which exhibit geometric regularity or rigidity. The disadvantage is that the intensity parameter cannot be estimated: inference is effectively conditional on the observed number of points. The new procedure is faster and more stable than existing techniques, since it does not require simulation, numerical integration or optimization with respect to the parameters.

Keywords: Campbell measure; Gibbs point process; non-hereditary interaction; pseudolikelihood; spatial statistics; variational estimator

1. Introduction

Statistical methodology for fitting models to spatial point pattern data has been a subject of intensive research for three decades. Likelihood-based methods were once regarded as computationally prohibitive (Ripley [28], Introduction). Although maximum likelihood and Bayesian methods can now be implemented using Markov chain Monte Carlo (Geyer [19]), this approach is still computationally intensive. Various alternative strategies have been explored, including analytic approximations to the likelihood (Ogata and Tanemura [26]), computationally efficient surrogates such as maximum pseudolikelihood (composite likelihood) (Besag [7]) and Takacs–Fiksel estimators (Fiksel [16], Takacs [30]).

These approaches fail when the point pattern data exhibit a high degree of geometric regularity or rigidity. For example, point patterns which approach a random dense packing of hard spheres, or a randomly perturbed hexagonal lattice, can be constructed as realisations of certain Gibbs models. Existing methods for fitting Gibbs models to the data, such as maximum pseudolikelihood (Baddeley and Turner [2], Besag [7], Billiot *et al.* [8], Coeurjolly and Drouilhet [10], Goulard *et al.* [20]), tend to be numerically unstable in the nearly-rigid case. The likelihood-based procedures require very long computation time for simulations. Other existing methods are generally based on an equilibrium equation involving the addition or removal of points of the process; such transitions may be impossible, or rare, if the model is too rigid. Moreover, most of the existing methods are based on the (Papangelou) conditional intensity; in some recent work (Dereudre [12], Dereudre and Lavancier [15], Dereudre *et al.* [13]), geometrically rigid point patterns are generated using Gibbs models which violate the usual assumption that the probability density is hereditary, so that the conditional intensity may not exist. Although the classical procedures of pseudolikelihood and Takacs–Fiksel estimation have been generalized to the non-hereditary setting (Coeurjolly *et al.* [9], Dereudre and Lavancier [15]), the associated estimators remain unavailable or inefficient if the process is too rigid (see, e.g., the simulations presented in Dereudre and Lavancier [14]).

In this paper, we propose an alternative approach to parameter estimation which is motivated by the variational estimators of Almeida and Gidas [1] for discrete space Markov random fields. This approach does not require the hereditary property. It is based on an equilibrium equation involving infinitesimal perturbation of the local energy and test functionals (see equation (20)). The use of infinitesimal perturbations seems naturally well-adapted for rigid models, since it does not require addition or removal of points. A necessary assumption to obtain identification is that the interaction potential is not constant. Therefore, one of the first consequences is that the intensity parameter z of the Gibbs process cannot be fitted, since it corresponds to a constant point potential $-\ln(z)$. Let us note that the parameters of a Gibbs process (apart from z) can be estimated without knowing the value of z. If necessary, the intensity parameter z may be fitted in a second step using another procedure, such as maximum pseudolikelihood, which performs well when interaction parameters are fixed.

Finally, we note that our procedure is quicker than existing ones, since it does not require simulation, numerical integration or optimization with respect to the parameters. The algorithm is very simple to implement, and requires only the computation of sums and the inversion of a linear system of size equal to the number of parameters (see (26), (30)). The estimator is exact and explicit.

In Section 2, we introduce notation and basic definitions for (grand canonical) Gibbs processes. In Section 3, the variational equilibrium equations are stated in stationary and non-stationary versions. In Section 4, the variational procedure is also presented in two versions, corresponding to the two variational equations. In Section 5, asymptotic properties are investigated. We show that both procedures are strongly consistent and we prove asymptotic normality for one of them. In Section 6, we present a large class of examples for which the procedure is available. We choose two typical examples coming from statistical mechanics and stochastic geometry (Lennard-Jones model, Hard sphere model). Section 7 is devoted to simulation experiments, where the variational procedure is applied in three situations corresponding to non-rigid, rigid and very rigid cases of the Lennard-Jones point process.

2. Definitions and notation

2.1. State spaces and reference measures

Our setting is Euclidean space $\mathbb{R}^d$ of arbitrary dimension $d \geq 1$. An element of $\mathbb{R}^d$ is denoted by $\mathbf{x} = (x^{(1)}, \dots, x^{(d)})$. Lebesgue measure on $\mathbb{R}^d$ is denoted by λ^d.

Definition 1. *A* configuration *is a subset ω of $\mathbb{R}^d$ which is locally finite, meaning that $\omega \cap \Lambda$ has finite cardinality $N_\Lambda(\omega) = \#(\omega \cap \Lambda)$ for every bounded Borel set Λ. The space Ω of all configurations is equipped with the σ-algebra $\mathcal{F}$ generated by the counting variables N_Λ.*

The symbol Λ will always refer to a bounded Borel set in $\mathbb{R}^d$. It will often be convenient to write ω_Λ in place of $\omega \cap \Lambda$. We abbreviate $\omega \cup \{\mathbf{x}\}$ to $\omega \cup \mathbf{x}$ and abbreviate $\omega \backslash \{\mathbf{x}\}$ to $\omega \backslash \mathbf{x}$ for every ω and every $\mathbf{x}$ in ω. For k points $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ in $\mathbb{R}^d$, we denote by $\mathbf{x}_{1\ldots k}$ the configuration $\mathbf{x}_1 \cup \cdots \cup \mathbf{x}_k$ and for $1 \le i \le k$, denote by $\mathbf{x}_{1\ldots k\backslash i}$ the configuration $\mathbf{x}_{1\ldots k} \setminus \mathbf{x}_i$.

As usual, we take the reference measure on $(\Omega, \mathcal{F})$ to be the distribution π^z of the Poisson point process with intensity measure $z\lambda^d$ $(z > 0)$ on $\mathbb{R}^d$. Recall that π^z is the unique probability measure on $(\Omega, \mathcal{F})$ such that the following hold for all subsets Λ: (i) N_Λ is Poisson distributed with parameter $z\lambda^d(\Lambda)$, and (ii) conditional on $N_\Lambda = n$, the n points in Λ are independent with uniform distribution on Λ, for each integer $n \ge 1$. The Poisson point process restricted to Λ will be denoted π^z_Λ.

In this paper, we consider only point process distributions (probability measures P on Ω) such that the intensity measure m on $\mathbb{R}^d$, defined by $m(\Lambda) = \mathbf{E}_P(N_\Lambda)$, for any Λ, is σ-finite. Here $\mathbf{E}_P$ denotes expectation with respect to P. We denote by $C^!_P$ the reduced Campbell measure of P defined on $\mathbb{R}^d \times \Omega$ by

$$C^!_P(g) = \int \sum_{\mathbf{x}\in\omega} g(\mathbf{x}, \omega \setminus \mathbf{x}) P(\mathrm{d}\omega) \tag{1}$$

for any positive measurable function g from $\mathbb{R}^d \times \Omega$ to $\mathbb{R}$.

Translation by a vector $u \in \mathbb{R}^d$ is denoted by τ_u, whether acting on $\mathbb{R}^d$ or on Ω. When P is stationary (i.e., $P = P \circ \tau_u^{-1}$ for any u in $\mathbb{R}^d$) the intensity measure has the form $m = z(P)\lambda^d$ and there exists a unique probability measure $P^{!0}$, called the reduced Palm measure, such that

$$C^!_P(g) = z(P) \int \int g(\mathbf{x}, \tau_{\mathbf{x}}\omega) P^{!0}(\mathrm{d}\omega) \lambda^d(\mathrm{d}\mathbf{x}). \tag{2}$$

See Matthes *et al.* [24] for more details about Campbell and Palm measures.

2.2. Interaction

We shall define the interaction energy in a general setting, along the lines of Preston [27]. Thus, we do *not* assume that the local densities come from a multibody interaction potential. This general viewpoint allows us to deal with the non-hereditary case (see Definition 4 below).

Definition 2. *A* family of energies *is a collection $\mathcal{H} = (H_\Lambda)$, indexed by bounded Borel sets Λ, of measurable functions from Ω to $\mathbb{R} \cup \{+\infty\}$ such that, for every $\Lambda \subset \Lambda'$, there exists a measurable function $\varphi_{\Lambda,\Lambda'}$ from Ω to $\mathbb{R} \cup \{+\infty\}$ such that for every $\omega \in \Omega$*

$$H_{\Lambda'}(\omega) = H_\Lambda(\omega) + \varphi_{\Lambda,\Lambda'}(\omega_{\Lambda^c}). \tag{3}$$

Equation (3) is equivalent to (6.11) and (6.12) in Preston [27], page 92. In physical terms, $H_\Lambda(\omega) = H_\Lambda(\omega_\Lambda \cup \omega_{\Lambda^c})$ represents the potential energy of ω_Λ inside Λ given the configuration ω_{Λ^c} outside Λ.

Definition 3. *A configuration ω has* locally finite energy *with respect to a family of energies $\mathcal{H} = (H_\Lambda)$ if, for every Λ, the energy $H_\Lambda(\omega)$ is finite. We denote by $\Omega_{\neq\infty}(\mathcal{H})$ or simply $\Omega_{\neq\infty}$ the space of configurations which have a locally finite energy.*

Definition 4. *A family of energies $\mathcal{H} = (H_\Lambda)$ is* hereditary *if for all Λ, all $\omega \in \Omega$ and $\mathbf{x} \in \Lambda$*

$$H_\Lambda(\omega) = +\infty \quad \Rightarrow \quad H_\Lambda(\omega \cup \mathbf{x}) = +\infty. \tag{4}$$

The assumption (4) is necessary in many papers, for example, Nguyen and Zessin [25], Ruelle [29]. In this setting, the local energy $h(\mathbf{x}, \omega)$ is defined for every $\omega \in \Omega_{\neq\infty}$ and $\mathbf{x} \notin \omega$ by

$$h(\mathbf{x}, \omega) = H_\Lambda(\omega \cup \mathbf{x}) - H_\Lambda(\omega) \tag{5}$$

for any Λ containing $\mathbf{x}$. Note that by (3), this definition does not depend on Λ.

Some recent work deals with non-hereditary Gibbs models (Dereudre [12], Dereudre and Lavancier [15], Dereudre *et al.* [13]). This setting occurs only if the family of energies has a hardcore part, that is, if $\Omega_{\neq\infty}((H_\Lambda)) \neq \Omega$. Henceforth we will *not* assume that the energy is hereditary. The equations and estimators presented in the following sections are available in the hereditary or non-hereditary setting.

2.3. Gibbs point processes

We are now in a position to define Gibbs measures. Let us make an integrability assumption on the family of energies, equivalent to (6.8) in Preston [27].

Definition 5. *The family of energies $\mathcal{H} = (H_\Lambda)$ is* integrable *if, for every Λ and every ω in $\Omega_{\neq\infty}$, we have*

$$0 < \int \mathrm{e}^{-H_\Lambda(\omega'_\Lambda \cup \omega_{\Lambda^c})} \pi_\Lambda^z(\mathrm{d}\omega'_\Lambda) < +\infty. \tag{6}$$

The second inequality in (6) is in general ensured by the stability of the energy functions. The first inequality is obvious in the classical hereditary setting, while in the non-hereditary case, it remains true under reasonable assumptions (e.g., Dereudre [12], Dereudre *et al.* [13]).

Under this integrability assumption, for every Λ and every ω in $\Omega_{\neq\infty}$, the local conditional density f_Λ is defined by

$$f_\Lambda(\omega) = \frac{1}{Z_\Lambda(\omega_{\Lambda^c})} \mathrm{e}^{-H_\Lambda(\omega)}, \tag{7}$$

where $Z_\Lambda(\omega_{\Lambda^c})$ is the normalization constant defined by $Z_\Lambda(\omega_{\Lambda^c}) = \int \mathrm{e}^{-H_\Lambda(\omega'_\Lambda \cup \omega_{\Lambda^c})} \pi_\Lambda^z(\mathrm{d}\omega'_\Lambda)$. Note that from (6), $0 < Z_\Lambda(\omega_{\Lambda^c}) < +\infty$ and therefore this local density is well-defined.

The usual definition of a "Gibbs point process" is equivalent to the following (Georgii [18], page 28).

Definition 6. *A probability measure P on Ω is a* (grand canonical) Gibbs measure *for the integrable family of energies $\mathcal{H} = (H_\Lambda)$ and the intensity $z > 0$ if $P(\Omega_{\neq\infty}) = 1$ and, for every Λ, for any measurable and integrable function g from Ω to $\mathbb{R}$,*

$$\int g(\omega) P(\mathrm{d}\omega) = \int \int g(\omega'_\Lambda \cup \omega_{\Lambda^c}) f_\Lambda(\omega'_\Lambda \cup \omega_{\Lambda^c}) \pi_\Lambda^z(\mathrm{d}\omega'_\Lambda) P(\mathrm{d}\omega). \tag{8}$$

Equivalently, for P-almost every ω the conditional law of P given ω_{Λ^c} is absolutely continuous with respect to π_Λ^z with the density f_Λ.

The equations (8) are called the Dobrushin–Lanford–Ruelle (DLR) equations. We denote by $\mathcal{G}$ the set of Gibbs measures.

For every Λ, every $k \geq 0$ and every ω in $\Omega_{\neq\infty}$, define

$$Z_{\Lambda,k}(\omega_{\Lambda^c}) = \int_{\Lambda^k} \mathrm{e}^{-H_\Lambda(\mathbf{x}_{1\ldots k} \cup \omega_{\Lambda^c})} \, \mathrm{d}\mathbf{x}_1 \cdots \mathrm{d}\mathbf{x}_k.$$

Under the integrability assumption (6), $Z_{\Lambda,k}(\omega_{\Lambda^c})$ is always finite and there exists at least one k such that $Z_{\Lambda,k}(\omega_{\Lambda^c}) > 0$. Provided $0 < Z_{\Lambda,k}(\omega_{\Lambda^c}) < \infty$ we may define the local conditional density for fixed k,

$$f_{\Lambda,k}(\omega) = \frac{1}{Z_{\Lambda,k}(\omega_{\Lambda^c})} \mathrm{e}^{-H_\Lambda(\omega)} \mathbb{1}\{N_\Lambda(\omega) = k\}. \tag{9}$$

Definition 7. *A probability measure P on Ω is a* (*canonical*) *Gibbs measure for the integrable family of energies $\mathcal{H} = (H_\Lambda)$ if $P(\Omega_{\neq\infty}) = 1$ and, for every Λ, for any measurable and integrable function g from Ω to $\mathbb{R}$,*

$$\begin{aligned} \int g(\omega) P(\mathrm{d}\omega) = \sum_{k=0}^{\infty} \int \int_{\Lambda^k} & \mathbb{1}_{\{N_\Lambda(\omega)=k\}} g(\mathbf{x}_{1\ldots k} \cup \omega_{\Lambda^c}) \\ & \times f_{\Lambda,k}(\mathbf{x}_{1\ldots k} \cup \omega_{\Lambda^c}) \, \mathrm{d}\mathbf{x}_1 \cdots \mathrm{d}\mathbf{x}_k P(\mathrm{d}\omega). \end{aligned} \tag{10}$$

Equivalently, for every $k \geq 0$, for P-almost every ω such that $f_{\Lambda,k}(\omega)$ is well defined, the conditional law of P given ω_{Λ^c} and $N_\Lambda(\omega) = k$ is absolutely continuous with respect to $\pi_\Lambda(\cdot|N_\Lambda = k)$ with density $|\Lambda|^k f_{\Lambda,k}$.

The results in Section 3 below are proved for canonical Gibbs measures. It is obvious that any Gibbs measure P is also a canonical Gibbs measure. Therefore, the results remain true for Gibbs measures.

Let us note that any canonical Gibbs measure admits the mixture representation $P = \int P_z \nu(\mathrm{d}z)$, where ν is a probability measure on $\mathbb{R}^+$ and P_z a Gibbs measure for intensity z (see Georgii [17] and Theorems 2.1 and 2.2 in Preston [27]).

Let us recall the famous Nguyen–Zessin equation (Nguyen and Zessin [25]), which is available only in the hereditary setting. For any Gibbs measure P,

$$C_P^! = z\mathrm{e}^{-h}\lambda^d \otimes P. \tag{11}$$

A generalization in the non-hereditary setting is investigated in Dereudre and Lavancier [15].

3. Variational equation for Gibbs point processes

3.1. Definitions and notation

First, we introduce the concept of a differentiable point in a configuration.

Definition 8. *Let ω be a configuration and $\mathbf{x}$ a point in $\mathbb{R}^d$ such that $\mathbf{x} \notin \omega$. We say that a family $\mathcal{H} = (H_\Lambda)$ is* differentiable *at $(\mathbf{x}, \omega)$ if there exists an open neighborhood $\mathcal{V}$ around $\mathbf{x}$ in $\mathbb{R}^d$ and a bounded set Λ containing $\mathcal{V}$ such that the function $\mathcal{V} \to \mathbb{R} \cup \{\infty\}$ defined by $\mathbf{y} \mapsto H_\Lambda(\omega \cup \mathbf{y})$ is differentiable at $\mathbf{x}$ in the usual sense. By convention, if this function is equal to infinity on $\mathcal{V}$, we take the derivative to be zero. We denote by $\nabla H(\mathbf{x}, \omega)$ the corresponding gradient, and by $\frac{\partial H}{\partial x^{(i)}}(\mathbf{x}, \omega)$ its ith coordinate.*

The notation $\nabla H(\mathbf{x}, \omega)$ is well-defined since this quantity does not depend on Λ by (3). Note that, in the hereditary setting, differentiability of (H_Λ) is equivalent to differentiability of the local energy $h(\mathbf{x}, \omega)$ with respect to the first variable $\mathbf{x}$. Then it is clear that $\nabla H(\mathbf{x}, \omega) = \nabla_{\mathbf{x}} h(\mathbf{x}, \omega)$.

We say that (H_Λ) is *λ^d-a.e. differentiable* if, for every ω, $\nabla H(\mathbf{x}, \omega)$ exists for λ^d-almost every $\mathbf{x}$. Henceforth we assume that (H_Λ) is λ^d-a.e. differentiable.

Next, we introduce classes of functionals used in the sequel.

Definition 9. *Let g be a measurable function from $\mathbb{R}^d \times \Omega$ to $\mathbb{R}$.*

We say g is λ^d-a.e. differentiable *if for any ω the function $\mathbf{x} \mapsto g(\mathbf{x}, \omega)$ is λ^d-a.e. differentiable. We denote the gradient function by $\nabla g(\mathbf{x}, \omega)$ and its ith coordinate by $\frac{\partial g}{\partial x^{(i)}}(\mathbf{x}, \omega)$.*

We say that g is shift invariant *if, for any vector u in $\mathbb{R}^d$, $g(\tau_u(\mathbf{x}), \tau_u(\omega)) = g(\mathbf{x}, \omega)$ for all $\mathbf{x}$ and ω.*

We say that g has compact support *if there exists a compact set K in $\mathbb{R}^d$ such that $g(\mathbf{x}, \omega) = 0$ for all $\mathbf{x}$ outside K and all ω.*

Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is absolutely continuous on $[a, b]$ if it is differentiable λ-a.e. on $[a, b]$ and if, for any $x \in [a, b]$, $f(x) - f(a) = \int_a^x f'(t)\lambda(\mathrm{d}t)$. For example any Lipschitzian function is absolutely continuous.

Definition 10. *Let g be a measurable function from $\mathbb{R}^d \times \Omega$ to $\mathbb{R}$. We say g is* regularizing *with respect to (H_Λ) if for every ω, every $a > 0$, every $1 \le i \le d$ and λ-a.e. every $x^{(1)}, \dots, x^{(i-1)}, x^{(i+1)}, \dots, x^{(d)}$ in $[-a, a]$ the function $[-a, a] \to \mathbb{R} \cup \{\infty\}$ defined by*

$$x^{(i)} \mapsto g(\mathbf{x}, \omega)\mathrm{e}^{-H_{[-a,a]^d}(\omega \cup \mathbf{x})} \tag{12}$$

is absolutely continuous.

If g is λ^d-a.e. differentiable and regularizing, then for λ^d-almost every $\mathbf{x}$ the derivative of (12) is equal to

$$\left(\frac{\partial g}{\partial x^{(i)}}(\mathbf{x},\omega) - g(\mathbf{x},\omega)\frac{\partial H}{\partial x^{(i)}}(\mathbf{x},\omega)\right)\mathrm{e}^{-H_{[-a,a]^d}(\omega\cup\mathbf{x})}.$$

In many situations, it becomes easier to check the regularizing condition, in particular when the energy functions (H_Λ) are regular. For instance, in many examples in Section 6, the function (12) will be continuous and piecewise differentiable which ensures absolute continuity. Nevertheless, we state the most general condition in order to investigate the largest possible class of examples.

We denote by $\mathcal{C}_\mathcal{R} = \mathcal{C}_\mathcal{R}(H_\Lambda)$ the class of functions g which are λ^d-a.e. differentiable and regularizing with respect to (H_Λ). The subclass $\mathcal{C}_{\mathcal{R},K}$ consists of those functions $g \in \mathcal{C}_\mathcal{R}$ which have compact support; the subclass $\mathcal{C}_{\mathcal{R},\tau}$ consists of those g which are shift invariant.

3.2. The general variational equation

Proposition 1. *Let P be a canonical Gibbs measure for the energy functions (H_Λ). Then for every function g in the class $\mathcal{C}_{\mathcal{R},K}$ such that*

$$C^!_P(|\nabla g| + |g\nabla H|) < \infty, \tag{13}$$

we have

$$C^!_P(\nabla g) = C^!_P(g\nabla H). \tag{14}$$

First, we remark that the expectation in (13) is meaningful. Indeed let g be a function in $\mathcal{C}_{\mathcal{R},K}$ with compact support Λ. From (10),

$$\begin{aligned}
&C^!_P(|\nabla g| + |g\nabla H|) \\
&\quad = \int \sum_{\mathbf{x}\in\omega_\Lambda} |\nabla g(\mathbf{x},\omega\setminus\mathbf{x})| + |g\nabla H(\mathbf{x},\omega\setminus\mathbf{x})| P(\mathrm{d}\omega) \\
&\quad = \sum_{k=1}^{+\infty} \int \frac{\mathbb{1}_{\{N_\Lambda(\omega)=k\}}}{Z_{\Lambda,k}(\omega_{\Lambda^c})} \sum_{i=1}^{k} \int_{\Lambda^k} (|\nabla g| + |g\nabla H|)(\mathbf{x}_i, \omega_{\Lambda^c} \cup \mathbf{x}_{1\ldots k\setminus i}) \\
&\qquad\qquad \times \mathrm{e}^{-H_\Lambda(\omega_{\Lambda^c}\cup\mathbf{x}_{1\ldots k})}\,\mathrm{d}\mathbf{x}_1\cdots\mathrm{d}\mathbf{x}_k\,P(\mathrm{d}\omega).
\end{aligned} \tag{15}$$

For every ω, by definition $(\nabla g + g\nabla H)(\mathbf{x},\omega)$ is defined for λ^d-almost every $\mathbf{x}$, so the integral in (15) is well-defined (either finite or infinite).

In the hereditary setting this equation characterizes the canonical Gibbs measures if the energy functions (H_Λ) are sufficiently regular (Dereudre [11] Theorem 1). However it is clear that the characterization of canonical Gibbs measures via (14) in the non-hereditary setting is false in general. Similar problems have been noted in Dereudre and Lavancier [15]; see Remark 1 after Proposition 2 below.

Proof of Proposition 1. Let g be a function in $\mathcal{C}_{\mathcal{R},\tau}$ satisfying (13) with compact support contained in $\Lambda = [-a,a]^d$. The following calculations are available thanks to the integrability assumption (13). As in (15), we have

$$C_P^!(\nabla g - g\nabla H) = \sum_{k=1}^{+\infty} \int \frac{\mathbb{1}_{\{N_\Lambda(\omega)=k\}}}{Z_{\Lambda,k}(\omega_{\Lambda^c})} \sum_{i=1}^{k} I_{i,k}^{\Lambda}(\omega_{\Lambda^c}) P(\mathrm{d}\omega), \tag{16}$$

where

$$\begin{aligned} I_{i,k}^{\Lambda}(\omega_{\Lambda^c}) = \int_{\Lambda^k} \big(&\nabla g(\mathbf{x}_i, \omega_{\Lambda^c} \cup \mathbf{x}_{1\ldots k\setminus i}) \\ &- g(\mathbf{x}_i, \omega_{\Lambda^c} \cup \mathbf{x}_{1\ldots k\setminus i}) \nabla H(\mathbf{x}_i, \omega_{\Lambda^c} \cup \mathbf{x}_{1\ldots k\setminus i})\big) \mathrm{e}^{-H_\Lambda(\omega_{\Lambda^c}\cup \mathbf{x}_{1\ldots k})}\, \mathrm{d}\mathbf{x}_1 \cdots \mathrm{d}\mathbf{x}_k. \end{aligned} \tag{17}$$

Next, we show that $I_{i,k}^{\Lambda}(\omega_{\Lambda^c}) = 0$ for every ω_{Λ^c}, every $k \geq 1$ and every $1 \leq i \leq k$. Consider the jth coordinate of $I_{i,k}^{\Lambda}(\omega_{\Lambda^c})$ which is denoted by $[I_{i,k}^{\Lambda}(\omega_{\Lambda^c})]^{(j)}$. By Fubini's theorem, and since g is regularizing, we obtain

$$\begin{aligned} &[I_{i,k}^{\Lambda}(\omega_{\Lambda^c})]^{(j)} \\ &:= \int_\Lambda \cdots \int_\Lambda \left(\frac{\partial g}{\partial x^{(j)}}(\mathbf{x}_i, \omega_{\Lambda^c} \cup \mathbf{x}_{1\ldots k\setminus i}) - g(\mathbf{x}_i, \omega_{\Lambda^c} \cup \mathbf{x}_{1\ldots k\setminus i}) \frac{\partial H}{\partial x^{(j)}}(\mathbf{x}_i, \omega_{\Lambda^c} \cup \mathbf{x}_{1\ldots k\setminus i}) \right) \\ &\qquad \times \mathrm{e}^{-H_\Lambda(\omega_{\Lambda^c}\cup \mathbf{x}_1 \cup \cdots \cup \mathbf{x}_k)}\, \mathrm{d}\mathbf{x}_1 \cdots \mathrm{d}\mathbf{x}_k \\ &= \int_\Lambda \cdots \int_\Lambda \left(\int_{[-a,a]^{d-1}} \left(\int_{-a}^{a} \frac{\partial}{\partial x_i^{(j)}} \big(g(\mathbf{x}_i, \omega_{\Lambda^c} \cup \mathbf{x}_{1\ldots k\setminus i}) \mathrm{e}^{-H_\Lambda(\mathbf{x}_i, \omega_{\Lambda^c}\cup \mathbf{x}_{1\ldots k\setminus i})} \big)\, \mathrm{d}x_i^{(j)} \right) \right. \\ &\qquad \left. \mathrm{d}x_i^{(1)} \cdots \mathrm{d}x_i^{(j-1)}\, \mathrm{d}x_i^{(j+1)} \cdots \mathrm{d}x_i^{(d)} \right) \mathrm{d}\mathbf{x}_1 \cdots \mathrm{d}\mathbf{x}_{i-1}\, \mathrm{d}\mathbf{x}_{i+1}\, \mathrm{d}\mathbf{x}_k \\ &= \int_\Lambda \cdots \int_\Lambda \left(\int_{[-a,a]^{d-1}} \big[g(\mathbf{x}_i, \omega_{\Lambda^c} \cup \mathbf{x}_{1\ldots k\setminus i}) \mathrm{e}^{-H_\Lambda(\mathbf{x}_i, \omega_{\Lambda^c}\cup \mathbf{x}_{1\ldots k\setminus i})} \big]_{x_i^{(j)}=-a}^{a} \right. \\ &\qquad \left. \mathrm{d}x_i^{(1)} \cdots \mathrm{d}x_i^{(j-1)}\, \mathrm{d}x_i^{(j+1)} \cdots \mathrm{d}x_i^{(d)} \right) \mathrm{d}\mathbf{x}_1 \cdots \mathrm{d}\mathbf{x}_{i-1}\, \mathrm{d}\mathbf{x}_{i+1}\, \mathrm{d}\mathbf{x}_k \\ &= 0. \end{aligned} \tag{18}$$

The last equality is due to the fact that g has compact support inside Λ. This proves the proposition. □

3.3. Variational equation in the stationary case

Here we give a counterpart of Proposition 1 in the setting of stationary point processes. As usual in this context, we assume the energy functions (H_Λ) are *shift-invariant*, meaning that for any vector u in $\mathbb{R}^d$, any configuration ω and any Λ, we have $H_\Lambda(\omega) = H_{\tau_u(\Lambda)}(\tau_u(\omega))$.

Proposition 2. *Let P be a stationary canonical Gibbs measure for shift-invariant energy functions (H_Λ). Write $\mathbf{E}^{!0}$ for the expectation with respect to the reduced Palm distribution $P^{!0}$ defined in (2). Then for every function g in the class $\mathcal{C}_{\mathcal{R},\tau}$ such that*

$$\mathbf{E}^{!0}\big(|\nabla g(0,\omega)| + |g(0,\omega)| + |g(0,\omega)\nabla H(0,\omega)|\big) < \infty, \tag{19}$$

we have

$$\mathbf{E}^{!0}(\nabla g(0,\omega)) = \mathbf{E}^{!0}(g(0,\omega)\nabla H(0,\omega)). \tag{20}$$

The most important difference between (14) and (20) is that, in the latter, the function g is not assumed to have compact support with respect to the first variable. This assumption was crucial to the proof of Proposition 1, as it ensures that the integral of the derivative is zero in (18). Here stationarity replaces this assumption. This kind of result was not observed in Almeida and Gidas [1]. A condition like (1.2) in Almeida and Gidas [1] (namely $\int_{\mathbb{R}^n} \nabla \cdot (\mathbf{W}^{(\alpha)}\pi_\theta(\mathbf{x}))\,\mathrm{d}\mathbf{x} = 0$ for $\alpha = 1,\dots,m$) is not required in this setting of stationary Gibbs point processes.

Proof of Proposition 2. Let g be a function in $\mathcal{C}_{\mathcal{R},\tau}$ satisfying (19). For any $n \geq 1$ we denote by Λ_n the set $[-n,n]^d$ and $\partial\Lambda_n$ the border of Λ_n defined by $\Lambda_n \setminus \Lambda_{n-1}$. Let ψ_n be a differentiable partition of unity, that is, a function from $\mathbb{R}^d$ to $\mathbb{R}$ such that ψ is zero outside Λ_n and equal to 1 inside Λ_{n-1}. We assume that $|\psi_n(\mathbf{x})|$ and $|\nabla\psi_n(\mathbf{x})|$ are uniformly bounded with respect to $\mathbf{x}$ and n by a finite positive constant C_ψ. Define the function $g_n(\mathbf{x},\omega) = \psi_n(\mathbf{x})g(\mathbf{x},\omega)$. We claim that g_n satisfies the assumptions of Proposition 1. Indeed it is clear than g_n has compact support, so g_n is in $\mathcal{C}_{\mathcal{R},K}$. It remains to check the integrability assumption (19). From (2) and noting that the functions g, ∇g and ∇H are shift invariant, we have

$$\begin{aligned} C_P^!(|\nabla g_n| + |g_n\nabla H|) &\leq z(P)\int\int(|\nabla\psi_n||g| + |\psi_n||\nabla g| + |\psi_n||g\nabla H|) \\ &\qquad\times(\mathbf{x},\tau_{\mathbf{x}}\omega)P^{!0}(\mathrm{d}\omega)\lambda^d(\mathrm{d}\mathbf{x}) \\ &\leq z(P)\lambda^d(\Lambda_n)C_\psi\int(|g| + |\nabla g| + |g\nabla H|)(0,\omega)P^{!0}(\mathrm{d}\omega) < \infty. \end{aligned} \tag{21}$$

Applying Proposition 1 to g_n we obtain

$$\mathbf{E}_P\Bigg[\sum_{\mathbf{x}\in\Lambda_n}\big(\nabla g_n(\mathbf{x},\omega\setminus\mathbf{x}) - g_n(\mathbf{x},\omega\setminus\mathbf{x})\nabla H(\mathbf{x},\omega\setminus\mathbf{x})\big)\Bigg] = 0.$$

By stationarity, and since g_n and ∇g_n on Λ_{n-1} are equivalent to g and ∇g, respectively, it follows that

$$
\begin{aligned}
0 = \mathbf{E}_P \Bigg[\sum_{\mathbf{x} \in \partial \Lambda_n} & \big(\nabla \psi_n(\mathbf{x}) g_n(\mathbf{x}, \omega \setminus \mathbf{x}) + \nabla g_n(\mathbf{x}, \omega \setminus \mathbf{x}) \psi_n(\mathbf{x}) \\
& - \psi_n(\mathbf{x}, \omega \setminus \mathbf{x}) g(\mathbf{x}, \omega \setminus \mathbf{x}) \nabla H(\mathbf{x}, \omega \setminus \mathbf{x}) \big) \Bigg] \\
& + z(P) \lambda^d(\Lambda_{n-1}) \mathbf{E}^{!0} \big(\nabla g(0, \omega) - g(0, \omega) \nabla H(0, \omega) \big).
\end{aligned} \tag{22}
$$

By a similar calculation as for (21), we deduce that the first term in (22) is bounded by $K\lambda^d(\partial \Lambda_n)$ for $0 < K < \infty$. Dividing (22) by n^d and letting n tend to infinity, the first term vanishes and we obtain (20). □

Remark 1. For many choices of the function g in Proposition 2, the equation (20) is the trivial identity $0 = 0$. For, suppose that the energy functions (H_Λ) are symmetric, in the sense that $H_\Lambda(\mathbf{x}, \omega) = H_{\varsigma(\Lambda)}(\varsigma(\mathbf{x}), \varsigma(\omega))$ where ς is the symmetric transformation in $\mathbb{R}^d$ defined by $\varsigma(\mathbf{x}) = -\mathbf{x}$. Suppose that the measure P is also symmetric, $P = P \circ \varsigma^{-1}$. This situation applies to all examples in Section 6. Therefore, for any symmetric function g in $\mathcal{C}_{\mathcal{R},\tau}$, it follows that ∇g and ∇H are anti-symmetric (i.e., $\nabla g(\mathbf{x}, \omega) = -\nabla g(\varsigma(\mathbf{x}), \varsigma(\omega))$) and we deduce

$$
\mathbf{E}^{!0}(\nabla g(0, \omega)) = -\mathbf{E}^{!0}(\nabla g(0, -\omega)) = -\mathbf{E}^{!0}(\nabla g(0, \omega)) = 0.
$$

A similar calculation also gives $\mathbf{E}^{!0}(g(0, \omega) \nabla H(0, \omega)) = 0$.

This remark shows that, in order to obtain useful instances of the equation (20), the choice of function g is delicate. In the next section, we will see that an interesting choice will be $g = \operatorname{div} H$ provided this belongs to $\mathcal{C}_{\mathcal{R},\tau}$. In this situation, g is anti-symmetric.

4. Variational estimator procedure for exponential models

Now we assume the energy functions depend on parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$ in $\mathbb{R}^p$ in the linear form

$$
H_\Lambda^{\boldsymbol{\theta}} = \boldsymbol{\theta} \cdot \mathbf{H}_\Lambda := \theta_1 H_\Lambda^1 + \cdots + \theta_p H_\Lambda^p
$$

where $(\mathbf{H}_\Lambda) = ((H_\Lambda^1), \dots, (H_\Lambda^p))$ is a suite of p families of energy functions. The resulting point process model is an exponential family in the sense of Barndorff-Nielsen [5], Küchler and Sørensen [22].

In the remainder of the paper, we assume that $(\mathbf{H}_\Lambda)$ is shift invariant and that, for some $\boldsymbol{\theta} \in \mathbb{R}^p$, there exists a stationary canonical Gibbs measure P for the energy functions $(H_\Lambda^{\boldsymbol{\theta}})$. In the following, we use $\mathcal{C}_{\mathcal{R},K}$ to denote the intersection of all classes $\mathcal{C}_{\mathcal{R},K}((H_\Lambda^i))$ for $1 \le i \le p$, and similarly for $\mathcal{C}_{\mathcal{R},\tau}$.

In this setting, we present two estimation procedures for $\boldsymbol{\theta}$. The first, called *shift-invariant estimation* is based on the equilibrium equation (20) in Proposition 2. This estimator is very natural and exploits the stationarity of the process. The second procedure, called *grid estimation*, is based on the equilibrium equation (14) in Proposition 1. In this case, we subdivide the observation window using a grid, and apply the equilibrium equation (14) in each cell of the grid. This procedure is less natural than the first, but seems to enjoy better asymptotic properties. In the next section, we show that both procedures are strongly consistent, but we are only able to prove asymptotic normality for the second procedure.

4.1. Shift-invariant estimation procedure

From Proposition 2, for any function g in the class $\mathcal{C}_{\mathcal{R},\tau}$ satisfying the integrability assumption (19) for any energy functions (H_Λ^k), $1 \leq k \leq p$, we obtain

$$\theta_1 \mathbf{E}^{!0}[g(0,\omega)\nabla H^1(0,\omega)] + \cdots + \theta_p \mathbf{E}^{!0}[g(0,\omega)\nabla H^p(0,\omega)] = \mathbf{E}^{!0}[\nabla g(0,\omega)]. \tag{23}$$

This vectorial equation in dimension d with p parameters gives a linear system of equations for $\boldsymbol{\theta}$. However, in many situations, the symmetry properties of the functions g and the energy functions (H_Λ^k) (see examples in Section 6) give d identical equations in (23). Consequently, we keep only one equation in summing the d equations in (23). This seems to be the best strategy for extracting maximum possible information from the data. Therefore, in the following, the divergence operator $\operatorname{div} = \frac{\partial}{\partial x^{(1)}} + \cdots + \frac{\partial}{\partial x^{(d)}}$ is used in place of the gradient operator. Now a system of p equations is obtained by choosing p functions $g_1, \ldots, g_p$ in $\mathcal{C}_{\mathcal{R},\tau}$ satisfying the integrability assumption (19).

Denoting by A the $p \times p$ matrix

$$A_{i,j} = \mathbf{E}^{!0}[g_i(0,\omega) \operatorname{div} H^j(0,\omega)], \tag{24}$$

and by $b = (b_1, \ldots, b_p)$ the p-vector

$$b_i = \mathbf{E}^{!0}[\operatorname{div} g_i(0,\omega)], \tag{25}$$

the system of linear equations to determine $\boldsymbol{\theta}$ is then $A\boldsymbol{\theta} = b$.

Classically, we approximate A and b by the empirical average. So for every $n \geq 1$, denote by Λ_n the set $[-n, n]^d$ and for P-almost every realization ω define the matrix $\hat{A}^{(n)}$ with entries

$$\hat{A}_{i,j}^{(n)} = \sum_{\mathbf{x} \in \omega_{\Lambda_n}} g_i(\mathbf{x},\omega) \operatorname{div} H^j(\mathbf{x},\omega),$$

and the vector $\hat{b}^{(n)}$ with entries

$$\hat{b}_i^{(n)} = \sum_{\mathbf{x} \in \omega_{\Lambda_n}} \operatorname{div} g_i(\mathbf{x},\omega).$$

These are unnormalised empirical sums; $(1/N_{\Lambda_n}(\omega))\hat{A}^{(n)}$ and $(1/N_{\Lambda_n}(\omega))\hat{b}^{(n)}$ are consistent estimates of A and b, respectively. If the matrix $\hat{A}^{(n)}$ is invertible, we define the estimator $\hat{\boldsymbol{\theta}}^{(n)}$ by

$$\hat{\boldsymbol{\theta}}^{(n)} = (\hat{A}^{(n)})^{-1}\hat{b}^{(n)}. \tag{26}$$

Under suitable assumptions, the invertibility of $\hat{A}^{(n)}$ for sufficiently large n and the strong consistency of this estimator are proved in Proposition 3 of Section 5.

4.2. Grid estimation procedure

In this section, the cells of the grid are the cubes $\Delta_u = \tau_u([0,1]^d)$ for any u in $\mathbb{Z}^d$. By a classical rescaling procedure, it is always possible to consider a grid with cubes of side length $a > 0$. For any function g in $\mathcal{C}_{\mathcal{R},K}$ with compact support in Δ_0 and satisfying integrability assumption (13) for any energy functions (H^i_Λ), $1 \le i \le p$, we obtain by Proposition 1

$$\begin{aligned} &E_P\Biggl[\sum_{\mathbf{x}\in\omega_{\Delta_0}} \nabla g(\mathbf{x},\omega\setminus\mathbf{x})\Biggr] \\ &\quad= \theta_1 E_P\Biggl[\sum_{\mathbf{x}\in\omega_{\Delta_0}} g(\mathbf{x},\omega\setminus\mathbf{x})\nabla H^1(\mathbf{x},\omega\setminus\mathbf{x})\Biggr] + \cdots \\ &\qquad + \theta_p E_P\Biggl[\sum_{\mathbf{x}\in\omega_{\Delta_0}} g(\mathbf{x},\omega\setminus\mathbf{x})\nabla H^p(\mathbf{x},\omega\setminus\mathbf{x})\Biggr]. \end{aligned} \tag{27}$$

As in the first estimation procedure, we sum these equations and we use the divergence operator in place of the gradient. So by choosing p such functions $g_1, \ldots, g_p$, we denote by A the $p \times p$ matrix

$$A_{i,j} = E_P\Biggl[\sum_{\mathbf{x}\in\omega_{\Delta_0}} g_i(\mathbf{x},\omega\setminus\mathbf{x}) \operatorname{div} H^j(\mathbf{x},\omega\setminus\mathbf{x})\Biggr] \tag{28}$$

and by $b = (b_1, \ldots, b_p)$ the p-vector

$$b_i = E_P\Biggl[\sum_{\mathbf{x}\in\omega_{\Delta_0}} \operatorname{div} g_i(\mathbf{x},\omega\setminus\mathbf{x})\Biggr]. \tag{29}$$

As in the first estimation procedure, the system of linear equations to determine $\boldsymbol{\theta}$ is then $A\boldsymbol{\theta} = b$.

For any function g in $\mathcal{C}_{\mathcal{R},K}$ with compact support in Δ_0, we denote by $\bar{g}$ its periodic version defined by $\bar{g}(\mathbf{x},\omega) = \sum_{u\in\mathbb{Z}^d} g(\tau_u(\mathbf{x}), \tau_u(\omega))$. Then the unnormalised empirical approximations of A and b are defined for every $n \ge 1$ and P-almost every realization ω by

$$\begin{aligned} \hat{A}^{(n)}_{i,j} &= \sum_{\mathbf{x}\in\omega_{\Lambda_n}} \bar{g}_i(\mathbf{x},\omega\setminus\mathbf{x}) \operatorname{div} H^j(\mathbf{x},\omega\setminus\mathbf{x}), \\ \hat{b}^{(n)}_i &= \sum_{\mathbf{x}\in\omega_{\Lambda_n}} \operatorname{div} \bar{g}_i(\mathbf{x},\omega\setminus\mathbf{x}). \end{aligned}$$

As in the first procedure, if the matrix $\hat{A}^{(n)}$ is invertible, $\hat{\boldsymbol{\theta}}^{(n)}$ is defined by

$$\hat{\boldsymbol{\theta}}^{(n)} = \bigl(\hat{A}^{(n)}\bigr)^{-1}\hat{b}^{(n)}. \tag{30}$$

Under suitable assumptions, the invertibility of $\hat{A}^{(n)}$ for sufficiently large n, strong consistency and asymptotic normality of this estimator are proved in Propositions 3 and 5 of Section 5.

5. Asymptotic properties of the estimators

5.1. Strong consistency

Proposition 3. *Let P be a stationary canonical Gibbs measure for shift invariant energy functions $H_\Lambda^{\boldsymbol{\theta}}$. Let $g_1, \dots, g_p$ be functions in the class $\mathcal{C}_{\mathcal{R},\tau}$ (resp., in $\mathcal{C}_{\mathcal{R},K}$ with compact support in Δ_0) satisfying the integrability assumption (19) (resp., (13)) for all (H_Λ^i), $1 \le i \le p$. We assume also that, for any Gibbs measure $\tilde{P}$ with energy functions $H_\Lambda^{\boldsymbol{\theta}}$, the matrix A defined in (24) (resp., in (28)) via $\tilde{P}$ is invertible. Then for P-almost every realization ω, for sufficiently large n the matrix $\hat{A}^{(n)}$ is invertible and the estimator $\hat{\boldsymbol{\theta}}^{(n)}$ defined in (26) (resp., in (30)) converges to $\boldsymbol{\theta}$. That is, $\hat{\boldsymbol{\theta}}^{(n)}$ is strongly consistent.*

Proof. Let P be a stationary canonical Gibbs measure and $P = \int P_z \nu(\mathrm{d}z)$ its mixture representation mentioned after Definition 7. Each Gibbs measure P_z is itself a mixture of ergodic Gibbs measures and therefore P is also a mixture of ergodic Gibbs measures (Georgii [17,18]).

So, for any P-a.e realization ω, ω is also a realization of an ergodic Gibbs measure $\tilde{P}$. By ergodic theorem, the normalized matrix $\hat{A}^{(n)}$ and vector $\hat{b}^{(n)}$ converge $\tilde{P}$-a.s. to the matrix A and vector b defined via $\tilde{P}$. Thanks to the equilibrium equations (14), (20) and the assumption that A is invertible for any Gibbs measure, the rest of the proof is clear. □

As in Almeida and Gidas [1], it is not easy to find general conditions on $g_1, \dots, g_p$ which ensure that the matrix A is invertible. Nevertheless, there is one interesting choice for which it is easy to prove it. Let us develop this situation in the rest of the section.

In Almeida and Gidas [1], the authors propose to define $g_i = \operatorname{div} H^i$. We follow this idea with a small modification. Indeed, in general, there is no reason why $\operatorname{div} H^i$ should be regularizing, satisfy the integrability assumption (19) or have compact support in Δ_0. So we propose the following variant. Let us choose a fixed nonnegative function Ψ from $\mathbb{R}^d \times \Omega$ to $\mathbb{R}$ and define

$$g_i = \Psi \operatorname{div} H^i, \qquad 1 \le i \le p. \tag{31}$$

In the setting of shift invariant estimation, we will assume that Ψ is shift invariant. In the setting of grid estimation, we will assume that Ψ has compact support contained in Δ_0.

Proposition 4. *Suppose that the functions $(g_i)_{1 \le i \le p}$ defined as in (31) satisfy the following identifiability assumption: for any $X \in \mathbb{R}^p$,*

$$\text{the function} \sum_{\mathbf{x} \in \Delta_0} \Psi(\mathbf{x}, \omega \setminus \mathbf{x}) |X^\top \cdot \operatorname{div} \mathbf{H}(\mathbf{x}, \omega \setminus \mathbf{x})| = 0, \qquad P\text{-a.s.} \quad \text{iff} \quad X = 0. \tag{32}$$

Then the associated matrix A defined in (24) or (28) is invertible as soon as all the terms are integrable.

Let us remark that in the case of shift-invariant estimation, the identification assumption (32) can be reformulated by: for any $X \in \mathbb{R}^p$,

$$\text{the function } \Psi(0,\omega) X^\top \cdot \operatorname{div} \mathbf{H}(0,\omega) = 0, \qquad P^{!0}\text{-a.s.} \quad \text{iff} \quad X = 0. \tag{33}$$

Proof of Proposition 4. In the shift invariant setting, from (2) we observe that the expressions for A in (24) and (28) are equivalent up to a multiplicative scalar $z(P)$. We shall prove that A defined in (28) is invertible, by showing it is positive-definite. Let X be a vector in $\mathbb{R}^p$. We have

$$X^\top A X = X^\top \cdot E_P\left[\sum_{\mathbf{x}\in\omega_{\Delta_0}} \Psi(\mathbf{x}, \omega \setminus \mathbf{x}) \operatorname{div} \mathbf{H}(\mathbf{x}, \omega \setminus \mathbf{x}) \big(\operatorname{div} \mathbf{H}(\mathbf{x}, \omega \setminus \mathbf{x})\big)^\top\right] \cdot X$$

$$= E_P\left[\sum_{\mathbf{x}\in\omega_{\Delta_0}} \Psi(\mathbf{x}, \omega \setminus \mathbf{x}) \big(X^\top \cdot \operatorname{div} \mathbf{H}(\mathbf{x}, \omega \setminus \mathbf{x})\big)^2\right].$$

Since Ψ is nonnegative, this quantity is nonnegative and thanks to the identification assumption (32) it is positive as soon as $X \neq 0$. □

Remark 2. When $g_i = \psi \operatorname{div} H^i$, some terms in the matrix A and the vector b can be simplified if the point process P is symmetric in each direction. Indeed, following the arguments in Remark 1, it is easy to show that $\mathbf{E}^{!0}(\frac{\partial H^i}{\partial x^{(k)}} \frac{\partial H^j}{\partial x^{(l)}}) = 0$ and $\mathbf{E}^{!0}(\frac{\partial^2 H^i}{\partial x^{(k)} \partial x^{(l)}}) = 0$ as soon as $k \neq l$. Therefore, in the setting of shift invariant estimator, the matrix A and vector b have the following simpler expression,

$$A_{i,j} = \mathbf{E}^{!0}\left[\Psi(0,\omega)\left(\sum_{k=1}^d \frac{\partial H^i}{\partial x^{(k)}}(0,\omega) \frac{\partial H^j}{\partial x^{(k)}}(0,\omega)\right)\right], \tag{34}$$

and

$$b_i = \mathbf{E}^{!0}[\Psi(0,\omega) \Delta H^i(0,\omega) + \operatorname{div} \Psi(0,\omega) H^i(0,\omega)], \tag{35}$$

where Δ denotes the classical Laplacian operator $\sum_{k=1}^d \frac{\partial^2}{\partial^2 x^{(k)}}$. Obviously an analogue simplification occurs for the grid estimator and these modifications should be incorporated in the computation of the empirical matrix $\hat{A}_n$ and vector $\hat{b}_n$.

Note that the variational estimator of Almeida and Gidas [1] is an example of a time-invariance estimator Baddeley [4], that is, it can be derived from properties of the infinitesimal generator of a certain diffusion. In our case, the variational estimator can again be viewed as a time-invariance estimator, associated with the diffusion with drift ∇H.

5.2. Asymptotic normality

We have seen above that shift invariant estimation seems more natural than the grid estimation in the context of stationary processes. Nevertheless, in this section we prove asymptotic normality for the grid estimator, while we did not succeed in showing it for the shift invariant estimator. However, in the simulations presented in Section 7, we do not notice difference between the asymptotic properties of these both estimators.

A function g on $\mathbb{R}^d \times \Omega$ is said to have a *finite range* R with $0 < R < \infty$ if for all $\mathbf{x}$ and ω

$$g(\mathbf{x}, \omega) = g\big(\mathbf{x}, \omega_{B(\mathbf{x},R)}\big).$$

Proposition 5. *Let P be a stationary ergodic Gibbs measure for shift invariant energy functions $H_\Lambda^{\boldsymbol{\theta}}$. Let $g_1, \dots, g_p$ be functions in the class in $\mathcal{C}_{\mathcal{R},K}$ satisfying the assumptions of Proposition 3. Moreover, we assume that the functions (g_i) and (∇H^i) have finite range $R > 0$, and that for every $1 \le i, j \le p$*

$$\mathbf{E}_P\Bigg(\bigg|\sum_{\mathbf{x}\in\omega_{\Delta_0}} \big(\nabla g_i(\mathbf{x}, \omega \setminus \mathbf{x}) + g_i \nabla H^j(\mathbf{x}, \omega \setminus_x x)\big)\bigg|^3\Bigg) < \infty. \tag{36}$$

Then the estimator $\hat{\boldsymbol{\theta}}^{(n)}$ is asymptotically normal,

$$\lambda^d(\Lambda_n)^{1/2}\big(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\big) \Rightarrow \mathcal{N}(0, A^{-1}\Sigma A), \tag{37}$$

where $\Rightarrow$ denotes convergence in distribution as $n \to \infty$, and Σ is the matrix defined in (41).

Let us notice that in Proposition 5, we assume that P is ergodic which ensures, in general, that P is an extremal Gibbs measure. If it is not the case, then by the classical mixture argument, we prove that the left term in (37) converges in distribution to a mixture of normal laws.

Proof. From the definition (30) we have

$$\hat{A}^{(n)}\big(\hat{\boldsymbol{\theta}}^{(n)} - \boldsymbol{\theta}\big) = \hat{b}^{(n)} - \hat{A}^{(n)}\boldsymbol{\theta}.$$

Since the assumptions of Proposition 3 hold and since P is ergodic, the matrix $\lambda^d(\Lambda_n)^{-1}\hat{A}^{(n)}$ converges almost surely to A, which is invertible. The proof will be complete if we can show that the vector $Z_n = \lambda^d(\Lambda_n)^{-1/2}(b^{(n)} - \hat{A}^{(n)}\boldsymbol{\theta})$ converges in distribution to $\mathcal{N}(0, \Sigma)$. Denote by Z_n^i the ith coordinate of Z_n, we have

$$Z_n^i = \lambda^d(\Lambda_n)^{-1/2} \sum_{u\in\mathbb{Z}^d\cap[-n,n-1]^d} Y_u^i,$$

where

$$Y_u^i = \sum_{\mathbf{x}\in\omega_{\Delta_u}} \Bigg(\operatorname{div} \bar{g}_i(\mathbf{x}, \omega \setminus \mathbf{x}) - \sum_{j=1}^{p} \theta_j \bar{g}_i(\mathbf{x}, \omega \setminus \mathbf{x}) \operatorname{div} H^j(\mathbf{x}, \omega \setminus \mathbf{x})\Bigg)$$

where again $\Delta_u = \tau_u([0, 1]^d)$.

We apply Theorem 2.1 in Jensen and Künsch [21] to obtain a Central Limit theorem for Z_n. Let us check the three fundamental assumptions (38), (39) and (40) below. For every u in $\mathbb{Z}^d$, consider the neighborhood of u defined by

$$\mathcal{V}_u = \{v \in \mathbb{Z}^d, \text{ there exist } \mathbf{x} \in \Delta_u, y \in \Delta_v \text{ such that } |\mathbf{x} - y| \leq R\}.$$

From the finite range of (g_i) and (∇H_i), it is easy to remark that

$$Y_u^i(\omega) = Y_u^i\left(\bigcup_{v \in \mathcal{V}_u} \omega_{\Delta_v}\right). \tag{38}$$

From stationarity of P, shift invariance of $\mathbf{H}$, definition of $\bar{g}_i$ and integrability assumption (36), we obtain that for every $u \in \mathbb{Z}^d$ and every $1 \leq i \leq p$

$$\mathbf{E}_P(|Y_u^i|^3) = \mathbf{E}_P(|Y_0^i|^3) < \infty. \tag{39}$$

In Proposition 1, the equation (14) remains valid if P is replaced by $P(\cdot|\omega_{\Lambda^c})$ where Λ is the compact support of g. For, in the proof we show that $I_{i,k}^{\Lambda}(\omega_{\Lambda^c}) = 0$ for every ω_{Λ^c}, every $k \geq 1$ and every $1 \leq i \leq k$. Since the function g_i has a compact support in Δ_0, we deduce that for every $1 \leq i \leq p$ and $u \in \mathbb{Z}^d$

$$\mathbf{E}_P(Y_u^i|\omega_{\Delta_u^c}) = 0. \tag{40}$$

Applying Theorem 2.1 in Jensen and Künsch [21], we conclude that Z_n converges in distribution to $\mathcal{N}(0, \Sigma)$ where Σ is the $p \times p$ matrix defined by

$$\Sigma_{i,j} = \mathbf{E}_P\left(\sum_{v \in \mathcal{V}_0} Y_0^i Y_v^i\right). \tag{41}$$

The proposition is proved. □

To prove asymptotic normality for the shift invariant estimator, we can apply the same argument as in the proof in Proposition 5. Nevertheless it is not possible to show that Z_n satisfies a Central Limit theorem via Theorem 2.1 in Jensen and Künsch [21] because the fundamental property (40) fails in this situation. An alternative solution would be to substitute (40) by some mixing properties of the Gibbs measure P. We do not investigate this solution here since we think that our estimators are interesting in the setting of rigid point processes with strong interaction. It is well known that mixing properties are not established in this setting. Asymptotic normality of the shift invariant estimator is not established.

6. Examples

In this section, we present examples which are amenable to the estimation procedures described in Section 4. The first example is a model with unbounded and finite range pairwise potential without hardcore part. The second involves a model of hardcore spheres with interaction.

For the shift-invariant estimators, we will always choose the functions g_i with the form

$$g_i(\mathbf{x}, \omega) = \Psi(\mathbf{x}, \omega) \operatorname{div} H^i(\mathbf{x}, \omega), \qquad 1 \le i \le p, \tag{42}$$

where Ψ is a shift invariant function from $\mathbb{R}^d \times \Omega$ to $\mathbb{R}$ which we will determine for each model.

For the grid estimator, we will choose the functions g_i with the form

$$g_i(\mathbf{x}, \omega) = \psi(\mathbf{x})\Psi(\mathbf{x}, \omega) \operatorname{div} H^i(\mathbf{x}, \omega), \qquad 1 \le i \le p, \tag{43}$$

where ψ is a continuous and piecewise differentiable function from $\mathbb{R}^d$ to $\mathbb{R}$ with a compact support exactly equals to $[0, 1]^d$ and such that $\nabla\psi$ is bounded. This function is fixed for all the models. An example of such a function ψ is

$$\psi(\mathbf{x}) = \mathbb{1}_{[0,1]^d}(\mathbf{x}) \prod_{i=1}^{d} x^{(i)}\bigl(1 - x^{(i)}\bigr). \tag{44}$$

6.1. Pairwise interaction model

In this section we study a general, unbounded, pairwise potential with finite range. The infinite range case could be also investigated, but it is more complicated to present: tempered configurations have to be introduced, and the integrability assumptions are much more difficult to obtain. Moreover in statistical applications, the infinite range case has limited interest because the observation window is typically bounded.

We assume also that the interaction does not include a hard core; this setting is addressed in the next section. Let $(\varphi_i)_{1\le i\le p}$ be p twice differentiable functions from $]0, +\infty[$ to $\mathbb{R} \cup \{+\infty\}$ with compact support. We denote by R_0 the common range of all potentials (φ_i) (i.e., a real $R_0 > 0$ such that every function φ_i is null on $[R_0, +\infty[$). The energy functions (H^i_Λ) are defined by

$$H^i_\Lambda(\omega) = \sum_{\substack{\{\mathbf{x},\mathbf{y}\}\in\omega \\ \{\mathbf{x},\mathbf{y}\}\cap\Lambda\neq\varnothing}} \varphi_i\bigl((\mathbf{x} - \mathbf{y})^2\bigr). \tag{45}$$

The infinite sum in (45) is well defined for every ω in Ω since the functions (φ_i) have compact supports. The classical way to define a pairwise interaction is in general via the quantities $\varphi(|\mathbf{x} - \mathbf{y}|)$ but in our setting it is simpler to use the equivalent form $\varphi((\mathbf{x} - \mathbf{y})^2)$ since we will compute derivatives with respect to the coordinates.

The global energy functions of the system are then defined by the linear combination $H^{\boldsymbol{\theta}}_\Lambda = \theta_1 H^1_\Lambda + \cdots + \theta_p H^p_\Lambda$ for $\boldsymbol{\theta} \in \mathbb{R}^p$. Define the potential $\varphi^{\boldsymbol{\theta}} = \theta_1\varphi_1 + \cdots + \theta_p\varphi_p$. We assume that the potential $r \to \varphi^{\boldsymbol{\theta}}(r^2)$ is superstable and lower regular (conditions (SS) and (LR) in Ruelle [29], pages 131–132). These conditions assume that double sums of pair potential terms $\sum_{i\neq j} \varphi^{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)$ can be bounded from below by suitable expressions: the details are not required for this paper. These conditions ensure the existence of a stationary Gibbs measure P (Ruelle

[29], Theorem 5.8) for any intensity $z > 0$ and provide also the following property (Ruelle [29], Corollary 2.9): for every $R > 0$, there exists $\alpha > 0$ such that

$$\mathbf{E}_P(\mathrm{e}^{\alpha N^2_{B(0,R)}}) < 0. \tag{46}$$

We emphasise that the component functions $\varphi_i(r^2)$ are not assumed to be superstable and lower regular for each i.

For this example, the function Ψ in (42) is the constant function equal to 1 and so $g_i = \operatorname{div} H^i$ for $1 \le i \le p$. Observe that the energy functions $(H^i)_\Lambda$ are hereditary, so the local energy $h^i(\mathbf{x}, \omega)$ exists and is defined by

$$h^i(\mathbf{x}, \omega) = \sum_{\mathbf{y} \in \omega} \varphi_i\big((\mathbf{x} - \mathbf{y})^2\big). \tag{47}$$

For every $\omega \in \Omega$, the local energy (47) is twice differentiable at every $\mathbf{x} \notin \omega$ and so

$$g_i(\mathbf{x}, \omega) = \operatorname{div} H^i(\mathbf{x}, \omega) = \operatorname{div} h^i(\mathbf{x}, \omega) = 2 \sum_{\mathbf{y} \in \omega} \left(\varphi_i'\big((\mathbf{x} - \mathbf{y})^2\big) \sum_{k=1}^{d} \big(x^{(k)} - y^{(k)}\big) \right), \tag{48}$$

and

$$\begin{aligned} \operatorname{div} g_i(\mathbf{x}, \omega) &= \operatorname{div} \circ \operatorname{div} H^i(\mathbf{x}, \omega) \\ &= 2 \sum_{\mathbf{y} \in \omega} \left(d\varphi_i'\big((\mathbf{x} - \mathbf{y})^2\big) + 2\varphi_i''\big((\mathbf{x} - \mathbf{y})^2\big) \left(\sum_{k=1}^{d} \big(x^{(k)} - y^{(k)}\big) \right)^2 \right). \end{aligned} \tag{49}$$

Let us give a collection of assumptions that will ensure the shift invariant estimator $\hat{\boldsymbol{\theta}}^{(n)}$ is strongly consistent.

Definition 11. *The function $\varphi^{\boldsymbol{\theta}}$ satisfies property* (LB) *if it has a lower bound $C_\varphi \in \mathbb{R}$,*

$$\inf_{r>0} \varphi^{\boldsymbol{\theta}}(r) \ge C_\varphi.$$

It satisfies the explosion control property (EX) *if there exists $C_b > 0$ such that for every $1 \le i \le p$*

$$\sup_{r>0} |\varphi_i'(r)|^2 \mathrm{e}^{-\varphi^{\boldsymbol{\theta}}(r)} \le C_b \quad \text{and} \quad \sup_{r>0} |\varphi_i''(r)| \mathrm{e}^{-\varphi^{\boldsymbol{\theta}}(r)} \le C_b.$$

It satisfies the linear independence property (I) *if the functions $(\varphi_i')_{1 \le i \le p}$ are linearly independent in the vectorial space of continuous functions from $]0, +\infty] \to \mathbb{R}$.*

Proposition 6. *Let (φ_i) be a family of potentials and $\boldsymbol{\theta} \in \mathbb{R}^p$ such that $\varphi^{\boldsymbol{\theta}}$ is superstable and lower regular, and such that properties* (LB), (EX) *and* (I) *hold. Then for any Gibbs measure associated to $\varphi^{\boldsymbol{\theta}}$, the shift invariant estimator or the grid estimator $\hat{\boldsymbol{\theta}}^{(n)}$ is strongly consistent.*

Proof. Thanks to Propositions 3 and 4, it remains to check that (g_i) are in $\mathcal{C}_{\mathcal{R},\tau}$ and that assumptions (19) and (33) hold for any Gibbs measure P. The expression (48) shows that g_i is in $\mathcal{C}_{\mathcal{R},\tau}$ for every $1 \le i \le p$.

Concerning (33), from (2) and (11) we deduce that

$$P^{!0}(\mathrm{d}\omega) = \frac{1}{z(P)} \mathrm{e}^{-h^{\boldsymbol{\theta}}(0,\omega)} P(\mathrm{d}\omega). \tag{50}$$

By the DLR equations (8), it follows that $P^{!0}$ is locally absolutely continuous with respect to the Poisson point Process π^z with positive density. We deduce that the probability under $P^{!0}$ that the configuration $\omega_{B(0,R_0)}$ is reduced to a point $\{\mathbf{x}\}$ is positive and that the law of this single point $\mathbf{x}$ is absolutely continuous with respect to the Lebesgue measure on $B(0, R_0)$. In this situation, for any X in $\mathbb{R}^d$ the left term in (33) is nothing more than

$$\Psi(0,\omega) X^\top \cdot \operatorname{div} \mathbf{H}(0,\omega) = -2\left(\sum_{k=1}^{d} x^{(k)}\right)\left(\varphi_1'(x^2), \dots, \varphi_p'(x^2)\right) \cdot X = 0. \tag{51}$$

From assumption (I), this implies that $X = 0$ and (33) is proved.

It remains to check (19). From expressions (48), (49) and formula (50), we have

$$\begin{aligned}
&\mathbf{E}^{!0}\left(|\nabla g_i(0,\omega)| + |g_i(0,\omega)| + |g_i(0,\omega)\nabla H^j(0,\omega)|\right) \\
&\quad\le K\mathbf{E}^{!0}\left[\sum_{y\in\omega_{B(0,R_0)}} \left(|\varphi_i'(y^2)| + |\varphi_i''(y^2)|\right) + \left(\sum_{y\in\omega_{B(0,R_0)}} |\varphi_i'(y^2)|\right)\left(\sum_{y\in\omega_{B(0,R_0)}} |\varphi_j'(y^2)|\right)\right] \\
&\quad\le \frac{K}{z(P)} E_P\left[\sum_{y\in\omega_{B(0,R_0)}} \left(|\varphi_i'(y^2)| + |\varphi_i''(y^2)|\right)\mathrm{e}^{-h^{\boldsymbol{\theta}}(0,\omega)}\right. \\
&\qquad\qquad \left. + N_{B(0,R_0)}\mathrm{e}^{-h^{\boldsymbol{\theta}}(0,\omega)}\left(\left(\sum_{y\in\omega_{B(0,R_0)}} \varphi_i'(y^2)^2\right) + \left(\sum_{y\in\omega_{B(0,R_0)}} \varphi_j'(y^2)^2\right)\right)\right],
\end{aligned}$$

where $K > 0$ is a constant. Using (LB) and (EX), we obtain

$$\begin{aligned}
&\mathbf{E}^{!0}\left(|\nabla g_i(0,\omega)| + |g_i(0,\omega)| + |g_i(0,\omega)\nabla H^j(0,\omega)|\right) \\
&\quad\le \frac{K}{z(P)} E_P\left[\sum_{y\in\omega_{B(0,R_0)}} \left(1 + \varphi_i'(y^2)^2 + |\varphi_i''(y^2)|\right)\mathrm{e}^{-\varphi^{\boldsymbol{\theta}}(y^2)}\mathrm{e}^{-h^{\boldsymbol{\theta}}(0,\omega\setminus y)}\right. \\
&\qquad\qquad \left. + N_{B(0,R_0)} \sum_{y\in\omega_{B(0,R_0)}} \left(\varphi_i'(y^2)^2 + \varphi_j'(y^2)^2\right)\mathrm{e}^{-\varphi^{\boldsymbol{\theta}}(y^2)}\mathrm{e}^{-h^{\boldsymbol{\theta}}(0,\omega\setminus y)}\right] \\
&\quad\le K E_P\left((1 + 4C_b) N_{B(0,R_0)}^2 \mathrm{e}^{C_\varphi N_{B(0,R_0)}}\right).
\end{aligned} \tag{52}$$

From (46) this quantity is finite. $\square$

For the grid estimator, we need a stronger assumption to obtain asymptotic normality. Assumption (EX) is replaced by (EXb):

$$\sup_{r>0} |\varphi_i'(r)|^6 \mathrm{e}^{-\varphi^{\boldsymbol{\theta}}(r)} \leq C_b \quad \text{and} \quad \sup_{r>0} |\varphi_i''(r)|^3 \mathrm{e}^{-\varphi^{\boldsymbol{\theta}}(r)} \leq C_b.$$

Proposition 7. *Let* (φ_i) *be a family of potentials and* $\boldsymbol{\theta} \in \mathbb{R}^p$ *such that* $\varphi^{\boldsymbol{\theta}}$ *is superstable and lower regular, and such that properties* (LB), (EXb) *and* (I) *hold. Then for any Gibbs measure associated to* $\varphi^{\boldsymbol{\theta}}$, *the grid estimator* $\hat{\boldsymbol{\theta}}^{(n)}$ *is strongly consistent and asymptotically normal.*

Proof. Thanks to Propositions 3, 4 and 5, it remains to check that (g_i) are in $\mathcal{C}_{\mathcal{R},K}$ and that they satisfy (13), (36) and (32). Expression (48) and the assumptions on ψ show that g_i is in $\mathcal{C}_{\mathcal{R},K}$ for every $1 \leq i \leq p$. Since the function ψ in (42) is bounded and has bounded gradient, by a calculation similar to that in (52) and from (EXb) we prove that (13) and (36) hold. The new assumption (EXb) plays a crucial role in the proof of (36). □

An example of a potential $\varphi^{\boldsymbol{\theta}}$ satisfying all assumptions (SS), (LR), (LB), (EXb) and (I) is the Lennard-Jones potential

$$\varphi^{\boldsymbol{\theta}}(r) = \mathbb{1}_{[0,R_0]}(r)\left(\theta_1 \frac{1}{r^6} + \theta_2 \frac{1}{r^3}\right), \tag{53}$$

with $\theta_1 > 0$ and $\theta_2 \in \mathbb{R}$. The Lennard-Jones potential was first introduced in Lennard-Jones [23] with the parametrization $\theta_1 = 4\varepsilon\sigma^{12}$ and $\theta_2 = -4\varepsilon\sigma^6$. In Section 7, we show some simulations of this model and investigate the estimation of ε and σ. The result shows that our estimator is very efficient and in particular, when the model is very rigid, it seems better than the classical Likelihood procedures.

6.2. Interacting hard sphere model

In this section, we consider the classical model of hard spheres with pairwise interaction. We assume in this section that the pair potential has finite interaction range, is smooth, and does not explode (i.e., is bounded near 0). Exploding potentials were studied in the previous section; a mixture of both models could be investigated without additional difficulties.

Let us consider $R_0 > r_0 > 0$ and p functions $(\varphi_i)_{1 \leq i \leq p}$ from $\mathbb{R}^+$ to $\mathbb{R} \cup \{+\infty\}$ assumed to be twice differentiable on $[r_0, +\infty[$ with continuous second derivative. We assume that φ_i is equal to infinity on $[0, r_0[$ and is null on $[R_0, +\infty[$. The energy functions (H_Λ^i) are defined as in (45) and the definition of $H^{\boldsymbol{\theta}}$ and $\varphi^{\boldsymbol{\theta}}$ follows. In this setting, $\varphi^{\boldsymbol{\theta}}$ is necessarily superstable and lower regular, so a stationary Gibbs measure P exists.

In this example, the energy functions are hereditary and so the local energy $h^i(\mathbf{x}, \omega)$ exists and is defined in (47). This local energy is differentiable for λ^d-almost every $\mathbf{x}$ and so expression (48) holds with the convention $\varphi_i'(r) = 0$ if $r < r_0$.

It is clear that the function $\operatorname{div} H^i(\mathbf{x}, \omega)$ is not regularizing because the function defined in (12) is not continuous. Therefore, the choice $g_i = \operatorname{div} H^i$ is not available in this setting and an expression of type $g_i = \Psi \operatorname{div} H^i$ is necessary. We define the function Ψ by

$$\Psi(\mathbf{x}, \omega) = \mathbb{1}_{\Omega_{\neq\infty}}(\omega) \prod_{\mathbf{y}\in\omega} \chi_{r_0,r_1}\big((\mathbf{x}-\mathbf{y})^2\big), \tag{54}$$

where χ_{r_0,r_1} is the real function defined for any $r_1 > r_0$ by

$$\chi_{r_0,r_1}(r) = \begin{cases} 0 & \text{if } r \le r_0, \\ \dfrac{r-r_0}{r_1-r_0} & \text{if } r_0 \le r \le r_1, \\ 1 & \text{if } r \ge r_1. \end{cases} \tag{55}$$

The product in (54) in fact involves only a finite number of terms and so is well defined. It is clear that the function Ψ is λ^d-a.e. differentiable with

$$\begin{aligned} \operatorname{div} \Psi(\mathbf{x}, \omega) = {} & \frac{2}{r_1 - r_0} \mathbb{1}_{\Omega_{\neq\infty}}(\omega) \\ & \times \sum_{\mathbf{y}\in\omega} \left(\mathbb{1}_{[r_0,r_1]}\big((\mathbf{x}-\mathbf{y})^2\big) \left(\sum_{k=1}^{d} x^{(k)} - y^{(k)} \right) \prod_{\mathbf{z}\in\omega\setminus\mathbf{y}} \chi_{r_0,r_1}\big((\mathbf{x}-\mathbf{z})^2\big) \right). \end{aligned} \tag{56}$$

Proposition 8. *If the potentials* (φ_i') *are linearly independent* (*assumption* (I)) *then the shift invariant estimator and the grid estimator* $\hat{\boldsymbol{\theta}}^{(n)}$ *are strongly consistent. Moreover, the grid estimator is asymptotically normal.*

Proof. Thanks to Propositions 3, 4 and 5, it remains to check that (g_i) are in $\mathcal{C}_{\mathcal{R},\tau}$ and that they satisfy (13), (32) and (36). First, from expression (48) and (54) let us note that g_i is regularising since the function defined in (12) is always continuous and piecewise differentiable with continuous derivatives. Since g_i is clearly shift invariant it follows that g_i is in $\mathcal{C}_{\mathcal{R},\tau}$. For integrability assumptions (13), (36), it is sufficient to remark that g_i, ∇g_i and $g_i \nabla H_j$ are uniformly bounded. Finally for identification assumption (32), the proof is exactly the same as in Proposition 6. □

Let us remark that, in the case of hard packing (i.e., when z is very large), the Maximum Likelihood and Maximum Pseudolikelihood procedures are not really available to estimate $\boldsymbol{\theta}$. Our procedure should be efficient.

The variational estimator can also be applied to models with rigid geometric constraints, such as the rigid Voronoi models of Dereudre and Lavancier [15].

7. Simulations

In this section, we present the results of simulation experiments assessing the performance of parameter estimation in the Lennard-Jones model (53). We consider four cases where ε takes the values 0.1, 0.5, 1 and 2, respectively, which we call the cases of low, moderate, high and extreme rigidity. We chose intensity $z = 100$ and characteristic range $\sigma = 0.1$. For each model,

Table 1. Sample statistics for 1000 estimates of the parameters ε, σ for the Lennard-Jones model using the variational estimators (GRID and INVARIANT) and the approximate maximum pseudolikelihood estimator (MPL)

		Low		Moderate		High		Extreme	
		σ	ε	σ	ε	σ	ε	σ	ε
TRUE		0.1	0.1	0.1	0.5	0.1	1	0.1	2
GRID	mean	0.094	0.789	0.101	0.966	0.101	1.277	0.100	1.987
	median	0.091	0.570	0.098	0.882	0.099	1.244	0.100	1.946
	sd	0.016	2.253	0.011	0.715	0.007	0.691	0.003	0.822
	E	0.099	0.124	0.100	0.511	0.100	1.019	0.100	1.807
INVARIANT	mean	0.093	0.636	0.100	0.970	0.100	1.333	0.099	2.210
	median	0.091	0.576	0.098	0.870	0.099	1.300	0.099	2.164
	sd	0.013	4.491	0.010	0.663	0.006	0.653	0.003	0.704
	E	0.097	0.149	0.099	0.558	0.099	1.106	0.099	2.069
MPL	mean	0.068	−614.696	0.102	0.311	0.102	0.307	0.103	0.327
	median	0.094	0.050	0.101	0.314	0.102	0.299	0.103	0.273
	sd	0.040	763.422	0.004	0.129	0.002	0.096	0.002	0.168

we simulated 1000 realizations in the window $[0,2]^2$ and estimated the parameters ε and σ (via $\theta_1 = 4\varepsilon\sigma^{12}$ and $\theta_2 = -4\varepsilon\sigma^6$) using the shift invariant estimator (26) and the grid estimator (30) where the window was subdivided into a 10×10 grid of squares. The terms in (26) were computed using (48) and (49) with $\varphi_1(s) = s^{-6}$ and $\varphi_2(s) = s^{-3}$. Similarly for (30) using the periodic function ψ in (44). For comparison, we also computed the maximum pseudolikelihood estimator (MPLE) using the method of Baddeley and Turner [2].

Table 1 summarises the 1000 estimated values for each experiment. For each estimation method, we report the sample mean, median and standard deviation of the individual estimates $\hat{\sigma}$ and $\hat{\varepsilon}$ for each of the three estimation methods. As a cross-check on the validity of the method, the row marked E in Table 1 denotes the 'pooled' estimate of the parameters obtained by computing the averages A^*, b^* of the 1000 empirical matrices and vectors $\hat{A}, \hat{b}$, then solving $A^*\boldsymbol{\theta} = b^*$ and transforming to obtain (σ, ε). Table 2 summarises the corresponding estimates of the canonical parameters θ_1, θ_2 for the same experiment.

Figure 1 displays typical realisations of the model in each case, and a scatterplot of a subsample of 100 estimated values giving an impression of the probability distribution of the estimates.

The most striking feature of these simulations is the bias in the maximum pseudolikelihood estimator. The bias increases with ε. We computed the MPLE using the Berman–Turner device Baddeley and Turner [2], Berman and Turner [6] with a 256×256 grid of sample points. To minimise numerical problems (overflow, instability, slow convergence), we rescaled the interpoint distances to a unit equal to the true value of σ. This is unrealistic with respect to applications (since the true σ would not be known) but gives the most optimistic assessment of performance from the MPLE algorithm. Although the MPLE is known to be biased in the presence of strong interaction, we conjecture that the very large bias observed here may be due to discretization

Table 2. Summary of estimates of canonical parameters θ_1, θ_2 in the same experiments as previous table. Values of θ_1 multiplied by 10^{12}; values of θ_2 multiplied by 10^6

		Low		Moderate		High		Extreme	
		θ_1	θ_2	θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
TRUE		0.4	−0.4	2	−2	4	−4	8	−8
GRID	mean	0.777	−1.371	2.705	−3.107	4.496	−4.718	7.462	−7.636
	median	0.671	−1.163	2.552	−2.990	4.392	−4.695	7.364	−7.538
	sd	0.511	1.296	1.210	1.945	1.370	2.036	1.553	2.387
	E	0.416	−0.454	2.018	−2.031	4.020	−4.048	7.123	−7.176
INVARIANT	mean	0.758	−1.362	2.682	−3.127	4.535	−4.86	7.668	−8.199
	median	0.663	−1.212	2.540	−2.947	4.463	−4.814	7.608	−8.127
	sd	0.470	1.178	1.147	1.805	1.290	1.911	1.379	2.033
	E	0.421	−0.502	2.057	−2.143	4.116	−4.267	7.386	−7.819
MPL	mean	0.274	−0.096	1.435	−1.317	1.526	−1.360	1.715	−1.487
	median	0.291	−0.240	1.410	−1.332	1.509	−1.355	1.602	−1.309
	sd	0.265	0.575	0.350	0.419	0.266	0.316	0.471	0.557

error (Baddeley and Turner [2], equation (17), page 290) and numerical problems related to the pair potential.

The parameter space for the canonical parameters is $\Theta = \{(\theta_1, \theta_2)\colon \theta_1 > 0, \theta_2 < 0\}$. Our implementation of the three parameter estimation methods is unconstrained, so that the algorithms may yield estimates $(\hat{\theta}_1, \hat{\theta}_2)$ that are "invalid" in the sense that they lie outside Θ. Invalid estimates are plotted as crosses in Figure 1. The observed proportion of invalid estimates is shown in Table 3. In the case of an invalid estimate, the model should be refitted with θ constrained to lie on the boundary of Θ (i.e., constraining either $\theta_1 = 0$ or $\theta_2 = 0$).

The variational estimator seems to work better when the model is rigid. In the low rigidity case, the estimation of ε is severely biased. We interpret the bias as a consequence of the nonlinear relationship between $\boldsymbol{\theta}$ and (σ, ε) combined with high variability in the low rigidity case.

The standard deviation of $\hat{\varepsilon}$ remains high in all cases. This is unsurprising since the estimation of ε is a difficult problem for all standard procedures even if the model is not rigid.

The difference between the shift invariant estimator and the grid estimator seems rather slight overall. The better asymptotic properties proved for the grid estimator seem to be irrelevant in these simulations. The estimates of σ and ε are strongly negatively correlated in all cases. In the low rigidity case, the empirical distribution of $\hat{\varepsilon}$ appears to deviate substantially from a Normal distribution, suggesting that the behaviour predicted by the Central Limit theorem has not yet set in.

However, the *pooled* grid estimate of ε appears to be biased in the extremely rigid case. We attribute this to the absence of edge correction in our algorithm.

In these experiments, the maximum pseudolikelihood estimator had substantially greater computational cost than the variational estimators. All algorithms were implemented in the R language using the spatial statistics package `spatstat` Baddeley and Turner [3]. Average com-

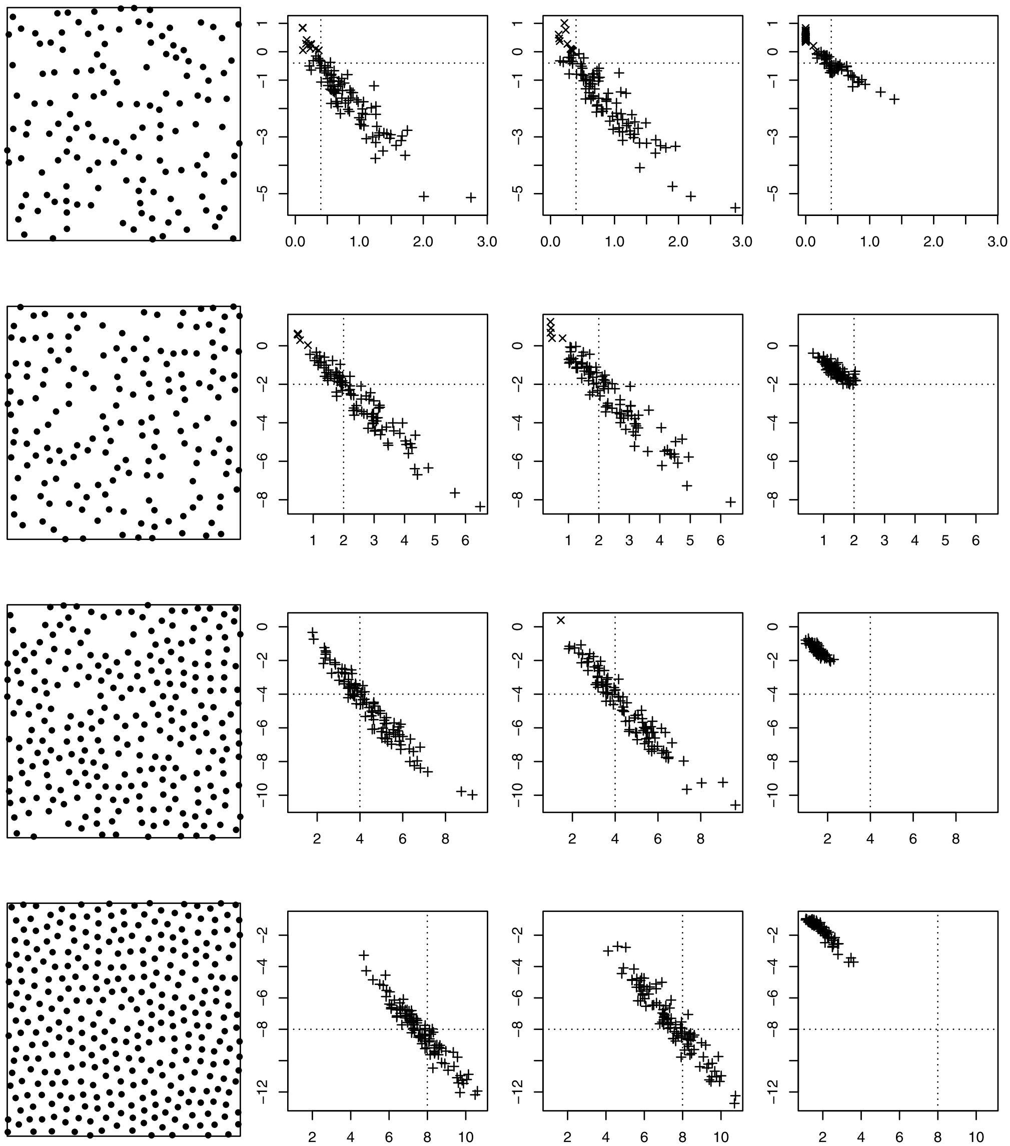

Figure 1. Distribution of parameter estimates for the Lennard-Jones model. Each row corresponds to a different experiment, obtained by simulating the Lennard-Jones model with (from top to bottom) $\varepsilon = 0.1, 0.5, 1, 2$, respectively. The left column shows a typical realization. The remaining columns show estimates of (θ_1, θ_2) from 100 realizations, obtained using (from left to right) the shift invariant estimator, grid estimator and maximum pseudolikelihood estimator respectively. Dotted lines show the true parameter values.

Table 3. Proportion of estimates that were invalid, in each of the four experiments

	Low	Moderate	High	Extreme
GRID	0.116	0.031	0.006	0.000
INVARIANT	0.105	0.023	0.004	0.000
MPL	0.413	0.001	0.000	0.000

putation times on a 2.5 GHz laptop were about 7 seconds to generate one simulated realisation, about 0.5 seconds each for the variational estimators, and for the MPLE, about 2 sec, 4 sec, 7 sec and 60 sec for the low, medium, high and extreme rigidity cases respectively. A supercomputing cluster was used to conduct the simulations in the extreme-rigidity case.

Acknowledgements

We thank Mark Westcott and two anonymous referees for insightful comments. Lawrence Murray and John Taylor provided invaluable advice on supercomputing.

References

[1] Almeida, M.P. and Gidas, B. (1993). A variational method for estimating the parameters of MRF from complete or incomplete data. *Ann. Appl. Probab.* **3** 103–136. MR1202518

[2] Baddeley, A. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns (with discussion). *Aust. N. Z. J. Stat.* **42** 283–322. MR1794056

[3] Baddeley, A. and Turner, R. (2005). Spatstat: An R package for analyzing spatial point patterns. *J. Statist. Softw.* **12** 1–42.

[4] Baddeley, A.J. (2000). Time-invariance estimating equations. *Bernoulli* **6** 783–808. MR1791902

[5] Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. Chichester: Wiley. MR0489333

[6] Berman, M. and Turner, T.R. (1992). Approximating point process likelihoods with GLIM. *Applied Statistics* **41** 31–38.

[7] Besag, J. (1977). Some methods of statistical analysis for spatial data. *Bull. Inst. Internat. Statist.* **47** 77–91, 138–147. MR0617572

[8] Billiot, J.M., Coeurjolly, J.F. and Drouilhet, R. (2008). Maximum pseudolikelihood estimator for exponential family models of marked Gibbs point processes. *Electron. J. Stat.* **2** 234–264. MR2399195

[9] Coeurjolly, J.F., Dereudre, D., Drouilhet, R. and Lavancier, F. (2010). Takacs–Fiksel method for stationary marked Gibbs point processes. Preprint.

[10] Coeurjolly, J.F. and Drouilhet, R. (2010). Asymptotic properties of the maximum pseudo-likelihood estimator for stationary Gibbs point processes including the Lennard-Jones model. *Electron. J. Stat.* **4** 677–706. MR2678967

[11] Dereudre, D. (2002). Une caractérisation de champs de Gibbs canoniques sur $\mathbb{R}^d$ et $\mathcal{C}([0,1],\mathbb{R}^d)$. *C. R. Math. Acad. Sci. Paris* **335** 177–182. MR1920016

[12] Dereudre, D. (2008). Gibbs Delaunay tessellations with geometric hardcore conditions. *J. Stat. Phys.* **131** 127–151. MR2394701

[13] Dereudre, D., Drouilhet, R. and Georgii, H.O. (2012). Existence of Gibbsian point processes with geometry-dependent interactions. *Probab. Theory Related Fields* **153** 643–670. MR2948688

[14] Dereudre, D. and Lavancier, F. (2011). Practical simulation and estimation for Gibbs Delaunay–Voronoi tessellations with geometric hardcore interaction. *Comput. Statist. Data Anal.* **55** 498–519. MR2736572

[15] Dereudre, D. and Lavancier, F. (2009). Campbell equilibrium equation and pseudo-likelihood estimation for non-hereditary Gibbs point processes. *Bernoulli* **15** 1368–1396. MR2597597

[16] Fiksel, T. (1984). Estimation of parametrized pair potentials of marked and nonmarked Gibbsian point processes. *Elektron. Informationsverarb. Kybernet.* **20** 270–278. MR0760146

[17] Georgii, H.O. (1979). *Canonical Gibbs Measures: Some Extensions of de Finetti's Representation Theorem for Interacting Particle Systems. Lecture Notes in Math.* **760**. Berlin: Springer. MR0551621

[18] Georgii, H.O. (1988). *Gibbs Measures and Phase Transitions. de Gruyter Studies in Mathematics* **9**. Berlin: de Gruyter. MR0956646

[19] Geyer, C. (1999). Likelihood inference for spatial point processes. In *Stochastic Geometry* (*Toulouse*, 1996) (O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout, eds.). *Monogr. Statist. Appl. Probab.* **80** 79–140. Chapman & Hall/CRC, Boca Raton, FL. MR1673118

[20] Goulard, M., Särkkä, A. and Grabarnik, P. (1996). Parameter estimation for marked Gibbs point processes through the maximum pseudolikelihood method. *Scand. J. Stat.* **23** 365–379.

[21] Jensen, J.L. and Künsch, H.R. (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Ann. Inst. Statist. Math.* **46** 475–486. MR1309718

[22] Küchler, U. and Sørensen, M. (1997). *Exponential Families of Stochastic Processes. Springer Series in Statistics*. New York: Springer. MR1458891

[23] Lennard-Jones, J.E. (1924). On the determination of molecular fields. *Proc. R. Soc. Lond. Ser. A* **106** 463–477.

[24] Matthes, K., Kerstan, J. and Mecke, J. (1978). *Infinitely Divisible Point Processes*. Chichester: Wiley. MR0517931

[25] Nguyen, X.X. and Zessin, H. (1979). Integral and differential characterizations of the Gibbs process. *Math. Nachr.* **88** 105–115. MR0543396

[26] Ogata, Y. and Tanemura, M. (1984). Likelihood analysis of spatial point patterns. *J. Roy. Statist. Soc. Ser. B* **46** 496–518. MR0790635

[27] Preston, C. (1976). *Random Fields. Lecture Notes in Mathematics* **534**. Berlin: Springer. MR0448630

[28] Ripley, B.D. (1988). *Statistical Inference for Spatial Processes*. Cambridge: Cambridge Univ. Press. MR0971986

[29] Ruelle, D. (1970). Superstable interactions in classical statistical mechanics. *Comm. Math. Phys.* **18** 127–159. MR0266565

[30] Takacs, R. (1986). Estimator for the pair-potential of a Gibbsian point process. *Statistics* **17** 429–433. MR0849741

Received August 2010 and revised September 2011