The Annals of Probability
2014, Vol. 42, No. 5, 1911–1951
DOI: 10.1214/13-AOP895

GIBBS POINT PROCESS APPROXIMATION: TOTAL VARIATION BOUNDS USING STEIN'S METHOD[1]

By Dominic Schuhmacher and Kaspar Stucki

University of Bern and University of Göttingen

We obtain upper bounds for the total variation distance between the distributions of two Gibbs point processes in a very general setting. Applications are provided to various well-known processes and settings from spatial statistics and statistical physics, including the comparison of two Lennard–Jones processes, hard core approximation of an area interaction process and the approximation of lattice processes by a continuous Gibbs process.

Our proof of the main results is based on Stein's method. We construct an explicit coupling between two spatial birth–death processes to obtain Stein factors, and employ the Georgii–Nguyen–Zessin equation for the total bound.

1. Introduction.

Gibbs processes form one of the most important classes of point processes in spatial statistics that may incorporate dependence between the points [Møller and Waagepetersen (2004), Chapter 6]. They are furthermore, mainly in the special guise of pairwise interaction processes, one of the building blocks of modern statistical physics [Ruelle (1969)].

Up to the somewhat technical condition of hereditarity (see Section 2), a Gibbs process on a compact metric space $\mathcal{X}$ is simply a point process whose distribution is absolutely continuous with respect to a "standard" Poisson process distribution. It is thus a natural counterpart in the point process world to a real-valued random variable that has a density with respect to some natural reference measure. A notorious difficulty with Gibbs processes is that in most cases of interest their densities can only be specified up to normalizing constants, which typically renders explicit calculations, for example, of the total variation distance between two such processes, difficult.

Received July 2012; revised April 2013.

[1]Supported by Swiss National Science Foundation Grant 200021-137527.

AMS 2000 subject classifications. Primary 60G55; secondary 60J75, 82B21.

Key words and phrases. Conditional intensity, pairwise interaction process, birth–death process, Stein's method, total variation distance.

In the current paper we give for the first time a comprehensive theorem about upper bounds on the total variation distance between Gibbs process distributions in a very general setting. These bounds provide natural rates of convergence in many asymptotic settings, and include explicit constants, which are small if one of the Gibbs processes is not too far away from a Poisson process.

For the important special case of bounding the distance between two pairwise interaction processes Ξ_1 and Ξ_2 on $\mathcal{X} \subset \mathbb{R}^D$ with densities proportional to $\beta^{|\xi|}\prod_{\{x,y\}\subset\xi}\varphi_1(x-y)$ and $\beta^{|\xi|}\prod_{\{x,y\}\subset\xi}\varphi_2(x-y)$, respectively, where φ_1 and φ_2 are pairwise interaction functions that are bounded by one (inhibitory case), a consequence of our results is that there is an explicitly computable constant $C = C(\beta, \varphi_2) > 0$ such that

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi_1), \mathscr{L}(\Xi_2)) \leq C\|\varphi_1 - \varphi_2\|_{L^1}. \tag{1}$$

If we relax the condition that the pairwise interaction functions are bounded by one and require suitable stability conditions for Ξ_1 and Ξ_2 instead, we still obtain

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi_1), \mathscr{L}(\Xi_2)) \leq C(\varepsilon)\|\varphi_1 - \varphi_2\|_{L^1} + \varepsilon, \tag{2}$$

where ε can be chosen arbitrarily small, causing a bigger $C(\varepsilon)$. We give more explicit examples for Strauss, bi-scale Strauss, and Lennard–Jones type processes in Sections 3 and 4.

For our proof of the main results we develop Stein's method for Gibbs process approximation. Using the generator approach by Barbour (1988), we re-express the total variation distance in terms of the infinitesimal generator of a spatial birth–death process (SBDP) whose stationary distribution is one of the Gibbs process distributions involved. An upper bound is then obtained by constructing an explicit coupling of such SBDPs in order to obtain the so-called Stein factor and applying the Georgii–Nguyen–Zessin equation.

Previously Stein's method has been applied very successfully for Poisson process approximation; see Barbour and Brown (1992), Chen and Xia (2004) and Schuhmacher (2009). Other notable developments in the domain of point process approximation concentrate on compound Poisson process approximation [Barbour and Månsson (2002)] and on approximation by certain point processes whose points are i.i.d. given their total number [called polynomial birth–death proceses by the authors; see Xia and Zhang (2012)]. In the latter article the authors give substantially improved bounds when replacing approximating Poisson or Compound Poisson processes by their new processes. However, these new processes are by no means flexible enough to approximate processes typically encountered in spatial statistics, where truely local point interactions take place, such as mutual inhibition up to a certain (nonnegligible) distance.

In Barbour and Chen (2005) the editors write in the preface: "Point process approximation, other than in the Poisson context, is largely unexplored." This statement still remains mostly true today, and the present paper makes a substantial contribution in order to change this.

Apart from approaches by Stein's method, the authors are not aware of any publications that give bounds for a probability metric between Gibbs processes in any generality. There is, however, related work by Møller (1989) and Dai Pra and Posta (2013), where convergence rates for distances between an SBDP and its stationary point process distribution were considered.

The plan of the paper is as follows. We start out in Section 2 by giving the necessary definitions and notation, including a somewhat longer introduction to Gibbs and pairwise interaction processes. Section 3 contains the main results. While Section 3.1 treats the common case where the approximating Gibbs process satisfies a stronger stability condition, Sections 3.2 and 3.3 lay out a strategy and give concrete results under very general conditions. Simpler examples are scattered throughout Section 3, while Section 4 looks at the three more involved applications mentioned in the abstract. In Section 5 we discuss spatial birth–death processes and present the coupling needed for obtaining the Stein factors, and in Section 6 we develop Stein's method for Gibbs process approximation and give the proofs of the main results. The paper finishes by an Appendix that justifies the reduction of our main proofs to a state space with diffuse reference measure $\boldsymbol{\alpha}$.

2. Prerequisites. Let $(\mathcal{X}, d)$ be a compact metric space, which serves as the state space for all our point processes. We equip $\mathcal{X}$ with its Borel σ-algebra $\mathcal{B} = \mathcal{B}(\mathcal{X})$. Let $\boldsymbol{\alpha} \neq 0$ be a fixed finite reference measure on $(\mathcal{X}, \mathcal{B})$. If $\mathcal{X}$ has a suitable group structure, $\boldsymbol{\alpha}$ is typically chosen to be the Haar measure. If $\mathcal{X} \subset \mathbb{R}^D$, we tacitly use Lebesgue measure and write $|A| := \mathrm{Leb}^D(A)$. Also, unless specified otherwise, we assume d to be the Euclidean metric in this case and write $\alpha_D = \pi^{D/2}/\Gamma(D/2+1)$ for the volume of the unit ball.

Denote by $(\mathfrak{N}, \mathcal{N})$ the space of finite counting measures ("point configurations") on $\mathcal{X}$ equipped with its canonical σ-algebra; see Kallenberg (1986), Section 1.1. For any $\xi \in \mathfrak{N}$ write $|\xi| = \xi(\mathcal{X})$ for its total number of points. A *point process* is simply a random element of $\mathfrak{N}$.

For a finite measure $\boldsymbol{\lambda}$ on $\mathcal{X}$ recall that a point process Π is called a *Poisson process* with *intensity* measure $\boldsymbol{\lambda}$ if the point counts $\Pi(A_i)$, $1 \leq i \leq n$, are independent $\mathrm{Po}(\boldsymbol{\lambda}(A_i))$-distributed random variables for any $n \geq 1$ and any pairwise disjoint sets $A_1, \ldots, A_n \in \mathcal{B}$. It is a well-known fact that such a Poisson process may be constructed as $\Pi = \sum_{i=1}^N \delta_{X_i}$, where N is a $\mathrm{Po}(\boldsymbol{\lambda}(\mathcal{X}))$-distributed random variable, and X_i are i.i.d. random elements of $\mathcal{X}$ with distribution $\boldsymbol{\lambda}(\cdot)/\boldsymbol{\lambda}(\mathcal{X})$ that are independent of N. We denote the Poisson process distribution with intensity measure $\boldsymbol{\alpha}$ by Po_1. We will make extensive use of the fact that for $\Pi \sim \mathrm{Po}_1$ and any measurable function

$h: \mathfrak{N} \to \mathbb{R}_+$ we have

$$\begin{aligned}\mathbb{E}h(\Pi) &= \int_{\mathfrak{N}} h(\xi)\mathrm{Po}_1(d\xi) \\ &= e^{-\boldsymbol{\alpha}(\mathcal{X})} \sum_{k=0}^{\infty} \frac{1}{k!} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} h\left(\sum_{i=1}^{k} \delta_{x_i}\right) \boldsymbol{\alpha}(dx_1) \cdots \boldsymbol{\alpha}(dx_k),\end{aligned} \tag{3}$$

where we interpret the summand for $k = 0$ as $h(\varnothing)$, writing $\varnothing$ for the empty point configuration. Equation (3) is obtained by conditioning on the total number of points of Π. Note that a similar version may also be found in Møller and Waagepetersen (2004), Proposition 3.1(ii).

2.1. *Gibbs processes.* We give the definition of a Gibbs point process from spatial statistics. Use the natural partial order on $\mathfrak{N}$, by which $\xi \leq \eta$ if and only if $\xi(A) \leq \eta(A)$ for every $A \in \mathcal{B}$. We call a function $u: \mathfrak{N} \to \mathbb{R}_+$ *hereditary* if for any $\xi, \eta \in \mathfrak{N}$ with $\xi \leq \eta$, we have that $u(\xi) = 0$ implies $u(\eta) = 0$.

DEFINITION 1. A point process Ξ on $\mathcal{X}$ is called a *Gibbs process* if it has a hereditary density u with respect to Po_1.

Gibbs processes form important models in spatial statistics. The specification of a density allows us to model point interactions in a simple and intuitive way. Under rather flexible conditions, for example, log-linearity in the parameters, formal inference (partly based on numerical methods) is possible. See Møller and Waagepetersen (2004), Chapter 9.

It will be convenient to identify a Gibbs process by its conditional intensity.

DEFINITION 2. Let Ξ be a Gibbs process with density u. We call the function $\lambda(\cdot|\cdot): \mathcal{X} \times \mathfrak{N} \to \mathbb{R}_+$,

$$\lambda(x|\xi) = \frac{u(\xi + \delta_x)}{u(\xi)}, \tag{4}$$

the *conditional intensity (function)* of Ξ. For this definition we use the convention that $0/0 = 0$.

Note that other definitions of the conditional intensity in the literature may differ at pairs (x, ξ) with $x \in \xi$. It is well known and with the help of equation (3), straightforward to check that the conditional intensity is the $\boldsymbol{\alpha} \otimes \mathscr{L}(\Xi)$-almost everywhere unique product measurable function that satisfies the *Georgii–Nguyen–Zessin equation*

$$\mathbb{E}\left(\int_{\mathcal{X}} h(x, \Xi - \delta_x)\Xi(dx)\right) = \int_{\mathcal{X}} \mathbb{E}(h(x, \Xi)\lambda(x|\Xi))\boldsymbol{\alpha}(dx) \tag{5}$$

for every measurable $h: \mathcal{X} \times \mathfrak{N} \to \mathbb{R}_+$.

A Gibbs process is usually specified via an unnormalized density $\tilde{u}$ that is shown to be Po_1-integrable. Typically the integral and hence the normalized density u cannot be computed explicitly. On the other hand the conditional intensity can be calculated simply as

$$\lambda(x|\xi) = \frac{\tilde{u}(\xi + \delta_x)}{\tilde{u}(\xi)} \tag{6}$$

and has a nice intuitive interpretation as the infinitesimal probability that Ξ produces a (further) point around x given it produces at least the point configuration ξ. Also it determines the Gibbs process distribution completely, since an unnormalized density $\tilde{u}$ can be recovered recursively for increasingly large point configurations by employing (6). We denote by $\mathrm{Gibbs}(\lambda)$ the distribution of the Gibbs process with conditional intensity $\lambda = \lambda(\cdot|\cdot)$.

The measure $\boldsymbol{\lambda}$ given by $\boldsymbol{\lambda}(A) = \mathbb{E}(\Xi(A))$ for any $A \in \mathcal{B}$ is called the *intensity measure* of Ξ, provided that it is finite. By equation (5) we have

$$\boldsymbol{\lambda}(A) = \mathbb{E}\left(\int_{\mathcal{X}} \mathbb{1}\{x \in A\}\Xi(dx)\right) = \int_{\mathcal{X}} \mathbb{1}\{x \in A\}\mathbb{E}(\lambda(x|\Xi))\boldsymbol{\alpha}(dx),$$

that is, $\boldsymbol{\lambda}$ is absolutely continuous with respect to $\boldsymbol{\alpha}$. We call its density $\lambda(x) = \mathbb{E}(\lambda(x|\Xi))$ the *intensity (function)* of Ξ. We use notation $\lambda(\cdot)$ and $\lambda(\cdot|\cdot)$ to distinguish the intensity and the conditional intensity if necessary.

In the main part of the paper we distinguish between the *approximated Gibbs process* Ξ with a general conditional intensity ν, and the *approximating Gibbs process* H, whose conditional intensity λ will typically (except in Sections 3.2–4.1) satisfy the stability condition

$$\sup_{\xi \in \mathfrak{N}} \int_{\mathcal{X}} \lambda(x|\xi)\boldsymbol{\alpha}(dx) < \infty. \tag{S}$$

Note that this condition follows from the *local stability condition*

$$\lambda(x|\xi) \le \psi^*(x)$$

for an integrable function $\psi^*: \mathcal{X} \to \mathbb{R}_+$. Local stability is satisfied for many point process distributions traditionally used in spatial statistics. See Møller and Waagepetersen (2004), page 84ff.

2.2. *Pairwise interaction processes.* A special type of Gibbs processes that are noteworthy both for their relative simplicity and their abundant use in statistical physics are the pairwise interaction processes. We treat distances between such processes in detail in Sections 3 and 4.

DEFINITION 3. A Gibbs process Ξ on $\mathcal{X}$ is called a *pairwise interaction process* (PIP) if there exist $\beta : \mathcal{X} \to \mathbb{R}_+$ and symmetric $\varphi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ such that Ξ has the unnormalized density

$$\tilde{u}(\xi) = \prod_{1 \le i \le n} \beta(x_i) \prod_{1 \le i < j \le n} \varphi(x_i, x_j)$$

for any $\xi = \sum_{i=1}^n \delta_{x_i} \in \mathfrak{N}$. The normalizing constant is usually not analytically computable. We then denote the distribution of Ξ by $\mathrm{PIP}(\beta, \varphi)$. The PIP is called *inhibitory* if $\varphi \le 1$. It is called *hard core* with radius $\delta > 0$ if $\varphi(x, y) = 0$ whenever $d(x, y) \le \delta$.

The conditional intensity of $\Xi \sim \mathrm{PIP}(\beta, \varphi)$ is accordingly given by

$$\lambda(x|\xi) = \beta(x) \prod_{i=1}^n \varphi(x, x_i).$$

For $\tilde{u}$ to be integrable with respect to Po_1, it is by equation (3) necessary that β is integrable. For inhibitory PIPs this is obviously also sufficient. The same holds for hard core PIPs with bounded φ, because by the compactness of $\mathcal{X}$ their total number of points is almost surely bounded. For more general PIPs the situation is not so simple; see Example 21 for a special case. We will then assume the following conditions:

(RS) *Ruelle stability*. There exist a constant c^* and an integrable function ψ^* such that $\tilde{u}(\xi) \le c^* \prod_{i=1}^n \psi^*(x_i)$ for every $\xi = \sum_{i=1}^n \delta_{x_i} \in \mathfrak{N}$.

(UB) *Upper boundedness*. There exists a constant C such that $\varphi(x, y) \le C$ for all $x, y \in \mathcal{X}$.

(RC) *Repulsion condition*. There exist $\delta > 0$ and $0 \le \gamma \le 1$ such that for all $x, y \in \mathcal{X}$ with $d(x, y) \le \delta$ we have $\varphi(x, y) \le \gamma$.

Note that (RS) is the form of Ruelle stability commonly used in spatial statistics; see Møller and Waagepetersen (2004). If we can choose $\psi^*(x)$ as $\beta(x)$ times a constant, we get the classical definition by Ruelle (1969). In any case Ruelle stability ensures that the unnormalized density $\tilde{u}$ is integrable.

If we can write the interaction function as $\varphi(x, y) = e^{-V(x,y)}$, then (UB) is equivalent to requiring that the potential V is bounded from below, which is a commonly used condition in statistical physics; see, for example, Ruelle (1969).

Furthermore, we introduce notation for the inner and outer ranges of attractive interaction.

(IR) *Interaction ranges*. Let $\delta \le r < R$ be constants such that for all $x, y \in \mathcal{X}$ with $d(x, y) \le r$ or $d(x, y) > R$ we have $\varphi(x, y) \le 1$.

Note that such constants always exist due to (RC) and the compactness of $\mathcal{X}$.

Strictly speaking only inhibitory PIPs satisfy (S). However, for our purpose it is actually enough to require finiteness of both the $\mathscr{L}(\Xi)$-ess sup and the $\mathscr{L}(\mathrm{H})$-ess sup instead of the supremum in (S). This would also admit comparisons of arbitrary hard core PIPs. However, for the ease of presentation we deal with hard core PIPs together with the more general PIPs in Sections 3.3 and 4.1.

2.3. *Reduction to a diffuse reference measure* $\boldsymbol{\alpha}$. In the remainder of this paper we will tacitly assume that the reference measure $\boldsymbol{\alpha}$ is diffuse, that is, satisfies $\boldsymbol{\alpha}(\{x\}) = 0$ for any $x \in \mathcal{X}$. This implies that the Po_1-process and the corresponding Gibbs processes are simple, that is, with probability one do not have multiple points at a single location in space. It is then convenient to interpret a point process as a random finite set and use set notation, which is commonly done is spatial statistics. Thus we may write, for example, $\xi \subset \eta$ instead of $\xi \le \eta$, or in the density of a PIP $\prod_{x\in\xi} \beta(x) \prod_{\{x,y\}\subset\xi} \varphi(x, y)$ instead of $\prod_{1\le i\le n} \beta(x_i) \prod_{1\le i<j\le n} \varphi(x_i, x_j)$.

In addition to simplifying the notation considerably by making points identifiable by their location in space, assuming a diffuse $\boldsymbol{\alpha}$ also reduces the differences in various definitions of the conditional intensity $\lambda(x|\xi)$ for the case $x \in \xi$ to an $\boldsymbol{\alpha} \otimes \mathrm{Po}_1$-null set.

We show in the Appendix that our results in Section 3 carry over to the nondiffuse case. Essentially this is seen by extending the state space $\mathcal{X}$ to $\mathcal{X} \times [0, 1]$ and considering $\boldsymbol{\alpha} \otimes \mathrm{Leb}|_{[0,1]}$ as a new (always diffuse) reference measure. This is based on an idea used in Chen and Xia (2004).

3. Total variation bounds between Gibbs process distributions.

3.1. *Main results.* Define $\mathcal{F}_{\mathrm{TV}}$ as the set of all measurable functions $f : \mathfrak{N} \to [0, 1]$. Then, for two point processes Ξ and H, the *total variation distance* is defined as

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) = \sup_{f\in\mathcal{F}_{\mathrm{TV}}} |\mathbb{E}f(\Xi) - \mathbb{E}f(\mathrm{H})|. \tag{7}$$

By a simple approximation argument this is equivalent to

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) = \sup_{A\in\mathcal{N}} |\mathbb{P}(\Xi \in A) - \mathbb{P}(\mathrm{H} \in A)|. \tag{8}$$

Denote by $\|\cdot\|$ the total variation norm for signed measures on $\mathcal{X}$. Thus $\|\xi - \eta\|$ for $\xi, \eta \in \mathfrak{N}$ is the total number of points appearing in one of the point configurations, but not in the other.

Our main results are given as Theorems 4 and 8. The principal idea behind our proofs is a suitable variant of Stein's method, which we develop in Section 6. The proofs themselves are deferred to Section 6 as well.

THEOREM 4. *Let* $\Xi \sim \text{Gibbs}(\nu)$ *and* $\mathrm{H} \sim \text{Gibbs}(\lambda)$ *be Gibbs processes. Suppose that* H *satisfies* (S). *Then there is a finite constant* $c_1(\lambda)$ *such that*

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) \leq c_1(\lambda) \int_{\mathcal{X}} \mathbb{E}|\nu(x|\Xi) - \lambda(x|\Xi)| \boldsymbol{\alpha}(dx). \tag{9}$$

More precisely, we have for any $n^* \in \mathbb{N} \cup \{\infty\}$ *that*

$$\begin{aligned} c_1(\lambda) \leq (n^*-1)! \left(\frac{\varepsilon}{c}\right)^{n^*-1} \left(\frac{1}{c} \sum_{i=n^*}^{\infty} \frac{c^i}{i!} + \int_0^c \frac{1}{s} \sum_{i=n^*}^{\infty} \frac{s^i}{i!}\, ds\right) \\ + \frac{1+\varepsilon}{\varepsilon} \sum_{i=1}^{n^*-1} \frac{\varepsilon^i}{i}, \end{aligned} \tag{10}$$

where

$$\varepsilon = \sup_{\|\xi - \eta\| = 1} \int_{\mathcal{X}} |\lambda(x|\xi) - \lambda(x|\eta)| \boldsymbol{\alpha}(dx) < \infty$$

and

$$c = c(n^*) = \sup_{\|\xi - \eta\| \geq n^*} \int_{\mathcal{X}} |\lambda(x|\xi) - \lambda(x|\eta)| \boldsymbol{\alpha}(dx) < \infty.$$

If $n^* = \infty$, *we interpret the long first summand in the upper bound as* 0. *For* $\varepsilon = 0$ *and/or* $c = 0$ *the upper bound is to be understood in the limit sense.*

REMARK 5. The term $c_1(\lambda)$ has a special meaning in our proof and in the theory of Stein's method in general; for its definition see equation (53). It is usually referred to as the *(first) Stein factor*.

REMARK 6 [Special cases for the bound on $c_1(\lambda)$]. (a) We often have a bound for c which does not depend on n^*. In this case we choose $n^* = \lceil c/\varepsilon \rceil$, which turns out to be optimal.

(b) If $\varepsilon < 1$, we can choose $n^* = \infty$ and obtain

$$c_1(\lambda) \leq \frac{1+\varepsilon}{\varepsilon} \log\left(\frac{1}{1-\varepsilon}\right) \leq \frac{1+\varepsilon}{1-\varepsilon}.$$

Conditions of the type $\varepsilon < 1$ are known in the statistical physics literature as "low activity, high temperature" setting; see, for example, Kondratiev and Lytvynov (2005).

(c) If H is a Poisson process, then $\lambda(x|\xi) = \lambda(x)$ does not depend on ξ. We then have $\varepsilon = 0$, and obtain $c_1(\lambda) = 1$. Hence inequality (10) contains the bound on the first Stein factor given in Lemma 2.2(i) of Barbour and Brown (1992) as a special case.

REMARK 7. The assumption that Ξ is a Gibbs process is used only in the proof of Theorem 4 for invoking the Georgii–Nguyen–Zessin equation, and is of course not needed for bounding $c_1(\lambda)$. The Georgii–Nguyen–Zessin equation may be generalized by replacing the kernel $\nu(x|\xi)\boldsymbol{\alpha}(dx)$ with the Papangelou kernel $\boldsymbol{\nu}(dx|\xi)$ of Ξ if the so-called condition (Σ) is satisfied, that is, if $\mathbb{P}(\Xi(A)=0|\Xi|_{A^c})>0$ a.s. for every $A\in\mathcal{B}$. See Kallenberg (1986), Section 13.2, or Daley and Vere-Jones (2008), Section 15.6, for details.

We may therefore generalize Theorem 4 as follows. Let Ξ be a point process that satisfies condition (Σ) and has Papangelou kernel $\boldsymbol{\nu}$, and let $\mathrm{H}\sim\mathrm{Gibbs}(\lambda)$ satisfy condition (S). Then

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi),\mathscr{L}(\mathrm{H}))\leq c_1(\lambda)\mathbb{E}\|\boldsymbol{\nu}(dx|\Xi)-\lambda(x|\Xi)\boldsymbol{\alpha}(dx)\|,$$

where $c_1(\lambda)$ is as above, and $\|\cdot\|$ is the total variation norm for signed measures on $\mathcal{X}$.

For inhibitory PIPs we obtain the following theorem, which relates the total variation distance to the L^1-distance between the interaction functions.

THEOREM 8. *Suppose that $\Xi\sim\mathrm{PIP}(\beta,\varphi_1)$ and $\mathrm{H}\sim\mathrm{PIP}(\beta,\varphi_2)$ are inhibitory. Let $\nu(y)=\mathbb{E}(\nu(y|\Xi))$ denote the intensity of Ξ. Then*

$$\begin{aligned}&d_{\mathrm{TV}}(\mathscr{L}(\Xi),\mathscr{L}(\mathrm{H}))\\&\quad\leq c_1(\lambda)\int_{\mathcal{X}}\int_{\mathcal{X}}\beta(x)\nu(y)|\varphi_1(x,y)-\varphi_2(x,y)|\boldsymbol{\alpha}(dx)\boldsymbol{\alpha}(dy),\end{aligned}\tag{11}$$

where $c_1(\lambda)$ is bounded in inequality (10) with

$$c\leq\int_{\mathcal{X}}\beta(x)\boldsymbol{\alpha}(dx)\quad\text{and}\quad\varepsilon=\sup_{y\in\mathcal{X}}\int_{\mathcal{X}}\beta(x)(1-\varphi_2(x,y))\boldsymbol{\alpha}(dx).$$

In the case where $\mathcal{X}\subset\mathbb{R}^D$, β is constant, and $\varphi_i(x,y)=\varphi_i(x-y)$ depends only on the difference, we obtain

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi),\mathscr{L}(\mathrm{H}))\leq c_1(\lambda)\beta\mathbb{E}(|\Xi|)\int_{\mathbb{R}^D}|\varphi_1(x)-\varphi_2(x)|\,dx.\tag{12}$$

Note that $\nu(\cdot)$ can usually not be calculated explicitly, but at least it can always be bounded by $\beta(\cdot)$. In particular, inequality (12) implies inequality (1) in the Introduction. Note that better bounds on a constant ν have been obtained in Stucki and Schuhmacher (2014).

REMARK 9. Our bounds on the total variation distance in Theorems 4 and 8 may be larger than one, in which case they give no new information. They are small if one of the processes is not too far away from a Poisson process, and the conditional intensities (or the pairwise interaction functions in the case of Theorem 8) are close in an L^1-sense. In what follows we are mainly interested in an asymptotic setting, where, for example, the interaction function of a PIP converges to the interaction function of another PIP.

If one of the processes is a Poisson process, we obtain both a slight improvement and a very substantial generalization of the bounds in Brown and Greig (1994).

EXAMPLE 10. Let Π be a Poisson process with intensity function β, and let $\Xi \sim \mathrm{PIP}(\beta, \varphi)$ be inhibitory, denoting its intensity function by ν. By Theorem 8 and Remark 6(c), we obtain

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\Pi)) \leq \int_{\mathcal{X}} \int_{\mathcal{X}} \beta(x)\nu(y)(1 - \varphi(x,y))\boldsymbol{\alpha}(dx)\boldsymbol{\alpha}(dy). \tag{13}$$

The special case where $\mathcal{X} = [0,1]^D$ with torus convention, $\boldsymbol{\alpha}$ is Lebesgue measure, and Ξ is a stationary hard core process with constant β and $\varphi(x,y) = \mathbb{1}\{\|x-y\| > r\}$ was considered in Barbour and Brown (1992) and Brown and Greig (1994), except that these articles approximate by a Poisson process $\widetilde{\Pi}$ that has the same intensity ν as Ξ. We obtain from (13) that

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\Pi)) \leq \beta\nu\alpha_D r^D \leq \beta^2 \alpha_D r^D. \tag{14}$$

The best bound in Brown and Greig (1994), namely inequality (12), says that under a somewhat complicated additional condition on the parameters, we have

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\widetilde{\Pi})) \leq \left(1 + \frac{1}{2^D}\right)\nu^2\alpha_D r^D. \tag{15}$$

By a straightforward upper bound on β [see Brown and Greig (1994), inequality (11)] our result (14) may be bounded further to obtain

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\Pi)) \leq \frac{1}{1 - \nu\alpha_D r^D}\nu^2\alpha_D r^D$$

for $r < 1/(\nu\alpha_D)^{1/D}$, which holds without the additional condition and is an asymptotic improvement over (15) by a factor of $2^D/(2^D+1)$ as $r \to 0$.

This suggests that for small r it is better to approximate Ξ by Π than by $\widetilde{\Pi}$, both because we get a smaller bound and because the intensity of Π is known explicitly from the parameters of Ξ.

EXAMPLE 11. A PIP is called a *Strauss process* if its interaction function is given by

$$\varphi(x,y) = \begin{cases} \gamma, & \text{if } d(x,y) \leq R, \\ 1, & \text{if } d(x,y) > R \end{cases}$$

for some constants $0 \le \gamma \le 1$ and $R > 0$. Let Ξ and H be Strauss processes with constant β and further parameters γ_1, R_1 and γ_2, R_2, respectively, where $R_1 > R_2$. Denote by $\mathbb{B}(y, R)$ the closed ball in $\mathcal{X}$ with center at y and radius R. Then by Theorem 8,

$$\begin{aligned} & d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) \\ & \quad \le c_1(\lambda)\mathbb{E}(|\Xi|)\beta \\ & \qquad \times \sup_{y\in\mathcal{X}}((1-\gamma_1)\boldsymbol{\alpha}(\mathbb{B}(y,R_1)\setminus\mathbb{B}(y,R_2)) + |\gamma_1-\gamma_2|\boldsymbol{\alpha}(\mathbb{B}(y,R_2))), \end{aligned} \tag{16}$$

where $c_1(\lambda)$ is bounded in inequality (10) with

$$\varepsilon = \beta(1-\gamma_2)\sup_{y\in\mathcal{X}}\boldsymbol{\alpha}(\mathbb{B}(y,R_2)) \quad \text{and} \quad c \le \beta\boldsymbol{\alpha}(\mathcal{X}).$$

3.2. *Processes violating the stability condition* (S). Many Gibbs processes satisfy condition (S), but there are some important exceptions. In the present subsection we provide a technique for treating these exceptions. In Section 3.3 we apply this technique to general PIPs.

We call an event A *hereditary* if the corresponding indicator function is hereditary, that is, if $\eta \in A$ implies $\xi \in A$ for all subconfigurations $\xi \subset \eta$.

Let A be a hereditary event such that $\mathbb{P}(\mathrm{H} \in A) > 0$. Let $\mathrm{H}_A \sim \mathscr{L}(\mathrm{H}|\mathrm{H} \in A)$. For instance, if $A = \{\eta \in \mathfrak{N} : |\eta| \le M\}$ for some $M \in \mathbb{N}$; then H_A has the same distribution as H conditioned on not having more than M points. In many cases H_A then satisfies (S), even if the original process H does not.

The following two lemmas are needed for reducing the problem of approximating by the process H to a problem of approximating by H_A.

LEMMA 12. *The process* H_A *has hereditary density* $u_A(\xi) = u(\xi)\mathbb{1}\{\xi \in A\}/\mathbb{P}(\mathrm{H} \in A)$ *with respect to* Po_1 *and conditional intensity* $\lambda_A(x|\xi) = \lambda(x|\xi)\mathbb{1}\{\xi + \delta_x \in A\}$, *where* u *and* λ *denote the density and conditional intensity of* H, *respectively.*

PROOF. Note that for all measurable $f : \mathfrak{N} \to \mathbb{R}_+$,

$$\begin{aligned} \mathbb{E}f(\mathrm{H}_A) &= \mathbb{E}(f(\mathrm{H})|A) \\ &= \frac{\mathbb{E}(f(\mathrm{H})\mathbb{1}\{\mathrm{H} \in A\})}{\mathbb{P}(\mathrm{H} \in A)} \\ &= \int_{\mathfrak{N}} f(\xi)\frac{u(\xi)\mathbb{1}\{\xi \in A\}}{\mathbb{P}(\mathrm{H} \in A)}\mathrm{Po}_1(d\xi). \end{aligned}$$

Furthermore, by the definition of the conditional intensity, equation (4),

$$\lambda_A(x|\xi) = \frac{u_A(\xi+\delta_x)}{u_A(\xi)} = \frac{u(\xi+\delta_x)\mathbb{1}\{\xi+\delta_x \in A\}}{u(\xi)\mathbb{1}\{\xi \in A\}} = \lambda(x|\xi)\mathbb{1}\{\xi+\delta_x \in A\},$$

where the last equality follows by the hereditarity of A. $\square$

PROPOSITION 13. *Let A be a hereditary event, and let $\mathrm{H}_A \sim \mathscr{L}(\mathrm{H}|A)$. Then*

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) \leq d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H}_A)) + \mathbb{P}(\mathrm{H} \notin A). \tag{17}$$

PROOF. Note that

$$\begin{aligned}
&d_{\mathrm{TV}}(\mathscr{L}(\mathrm{H}_A), \mathscr{L}(\mathrm{H}))\\
&\quad = \sup_{B\in\mathcal{N}} |\mathbb{P}(\mathrm{H} \in B | \mathrm{H} \in A) - \mathbb{P}(\mathrm{H} \in B)|\\
&\quad = \sup_{B\in\mathcal{N}} \left| \frac{\mathbb{P}(\mathrm{H} \in B, \mathrm{H} \in A)}{\mathbb{P}(\mathrm{H} \in A)} - \mathbb{P}(\mathrm{H} \in B, \mathrm{H} \in A) - \mathbb{P}(\mathrm{H} \in B, \mathrm{H} \notin A) \right|\\
&\quad = \sup_{B\in\mathcal{N}} \left| (1 - \mathbb{P}(\mathrm{H} \in A)) \frac{\mathbb{P}(\mathrm{H} \in B, \mathrm{H} \in A)}{\mathbb{P}(\mathrm{H} \in A)} - \mathbb{P}(\mathrm{H} \in B, \mathrm{H} \notin A) \right|\\
&\quad \leq \max\left(\sup_{B\in\mathcal{N}} \mathbb{P}(\mathrm{H} \notin A) \frac{\mathbb{P}(\mathrm{H} \in B, \mathrm{H} \in A)}{\mathbb{P}(\mathrm{H} \in A)}, \sup_{B\in\mathcal{N}} \mathbb{P}(\mathrm{H} \in B, \mathrm{H} \notin A) \right)\\
&\quad = \mathbb{P}(\mathrm{H} \notin A)
\end{aligned}$$

and the triangle inequality yields the claim. □

COROLLARY 14. *For hereditary events A and A' we get*

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) \leq d_{\mathrm{TV}}(\mathscr{L}(\Xi_A), \mathscr{L}(\mathrm{H}_{A'})) + \mathbb{P}(\Xi \notin A) + \mathbb{P}(\mathrm{H} \notin A').$$

3.3. *General pairwise interaction processes (PIP).* For PIPs the following hereditary event is very useful. Let $k \in \mathbb{N}$, $\delta > 0$ and

$$A_k = \left\{ \xi \in \mathfrak{N}: \sup_{y\in\mathcal{X}} \xi(\mathbb{B}(y, \delta/2)) \leq k \right\}, \tag{18}$$

that is, we require that the PIP has at most k points inside any closed ball with radius $\delta/2$. If $k = 1$, this is equivalent to the event that the PIP has a hard core radius δ.

LEMMA 15. *Suppose that $\mathrm{H} \sim \mathrm{PIP}(\beta, \varphi)$ satisfies the conditions* (RS), (UB), (RC) *and* (IR) *with the constants C, δ, γ, r and R. Then*

$$\begin{aligned}
\mathbb{P}(\mathrm{H} \notin A_k) &= \mathbb{P}(\exists y \in \mathcal{X}: \mathrm{H}(\mathbb{B}(y, \delta/2)) \geq k + 1)\\
&\leq \frac{\gamma^{(k(k+1))/2} B_\delta^k}{(k+1)! C^k} \mathbb{E}(|\mathrm{H}| C^{k|\mathrm{H}|}),
\end{aligned} \tag{19}$$

where $B_\delta = \sup_{y\in\mathcal{X}} \int_{\mathbb{B}(y,\delta)} \beta(x) \boldsymbol{\alpha}(dx)$.

PROOF. Note that by equation (3)

$$\begin{aligned}&\mathbb{P}(\exists y\in\mathcal{X}\colon \mathrm{H}(\mathbb{B}(y,\delta/2))\geq k+1)\\ &\quad= e^{-\boldsymbol{\alpha}(\mathcal{X})}\sum_{n=k+1}^{\infty}\frac{1}{n!}\int_{\mathcal{X}}\cdots\int_{\mathcal{X}} g_{\delta,k+1}(x_1,\dots,x_n)\\ &\qquad\qquad\times u(\{x_1,\dots,x_n\})\boldsymbol{\alpha}(dx_1)\cdots\boldsymbol{\alpha}(dx_n),\end{aligned} \tag{20}$$

where

$$\begin{aligned}&g_{\delta,k+1}(x_1,\dots,x_n)\\ &\quad=\mathbb{1}\{\exists y\in\mathcal{X},\exists\{i_1,\dots,i_{k+1}\}\subset\{1,\dots,n\}\colon x_{i_1},\dots,x_{i_{k+1}}\in\mathbb{B}(y,\delta/2)\}\\ &\quad\leq\sum_{\{i_1,\dots,i_{k+1}\}\subset\{1,\dots,n\}}\mathbb{1}\{\exists y\in\mathcal{X},x_{i_1},\dots,x_{i_{k+1}}\in\mathbb{B}(y,\delta/2)\}.\end{aligned} \tag{21}$$

For any permutation $(i_1,\dots,i_n)$ of $(1,\dots,n)$ the density can be rewritten as

$$\begin{aligned}u(\{x_1,\dots,x_n\})&=\Bigg(\prod_{j=1}^{k}\beta(x_{i_j})\Bigg)\Bigg(\prod_{1\leq j<l\leq k+1}\varphi(x_{i_j},x_{i_l})\Bigg)\\ &\quad\times u(\{x_{i_{k+1}},\dots,x_{i_n}\})\prod_{j=1}^{k}\prod_{l=k+2}^{n}\varphi(x_{i_j},x_{i_l}).\end{aligned}$$

Thus by (RC) and (UB) we obtain

$$\begin{aligned}&u(\{x_1,\dots,x_n\})\mathbb{1}\{\exists y\in\mathcal{X},x_{i_1},\dots,x_{i_{k+1}}\in\mathbb{B}(y,\delta/2)\}\\ &\quad\leq\Bigg(\prod_{j=1}^{k}\beta(x_{i_j})\Bigg)\gamma^{\binom{k+1}{2}}u(\{x_{i_{k+1}},\dots,x_{i_n}\})C^{k(n-k-1)}\\ &\qquad\times\mathbb{1}\{\exists y\in\mathcal{X},x_{i_1},\dots,x_{i_{k+1}}\in\mathbb{B}(y,\delta/2)\}\\ &\quad\leq\Bigg(\prod_{j=1}^{k}\beta(x_{i_j})\Bigg)\gamma^{\binom{k+1}{2}}C^{k(n-k-1)}u(\{x_{i_{k+1}},\dots,x_{i_n}\})\\ &\qquad\times\mathbb{1}\{x_{i_1},\dots,x_{i_k}\in\mathbb{B}(x_{i_{k+1}},\delta)\},\end{aligned}$$

where the last line follows by the triangle inequality. Thus in total from equation (20)

$$\begin{aligned}&\mathbb{P}(\mathrm{H}\notin A_k)\\ &\quad\leq e^{-\boldsymbol{\alpha}(\mathcal{X})}\sum_{n=k+1}^{\infty}\frac{1}{n!}\end{aligned}$$

$$\times \sum_{\{i_1,\dots,i_{k+1}\}\subset\{1,\dots,n\}} \underbrace{\int_{\mathcal{X}}\cdots\int_{\mathcal{X}}}_{n-k} \underbrace{\int_{\mathbb{B}(x_{i_{k+1}},\delta)}\cdots\int_{\mathbb{B}(x_{i_{k+1}},\delta)}}_{k} \left(\prod_{j=1}^{k}\beta(x_{i_j})\right)\gamma^{\binom{k+1}{2}}$$

$$\times C^{k(n-k-1)}u(\{x_{i_{k+1}},\dots,x_{i_n}\})\boldsymbol{\alpha}(dx_{i_1})\cdots\boldsymbol{\alpha}(dx_{i_k})\boldsymbol{\alpha}(dx_{i_{k+1}})\cdots\boldsymbol{\alpha}(dx_{i_n})$$

$$\le \frac{\gamma^{(k(k+1))/2}B_\delta^k}{(k+1)!C^k}e^{-\boldsymbol{\alpha}(\mathcal{X})}$$

$$\times \sum_{n=k+1}^{\infty}\frac{n-k}{(n-k)!}C^{k(n-k)}\underbrace{\int_{\mathcal{X}}\cdots\int_{\mathcal{X}}}_{n-k}u(\{x_{k+1},\dots,x_n\})\boldsymbol{\alpha}(dx_{k+1})\cdots\boldsymbol{\alpha}(dx_n)$$

$$= \frac{\gamma^{(k(k+1))/2}B_\delta^k}{(k+1)!C^k}\mathbb{E}(|\mathrm{H}|C^{k|\mathrm{H}|}),$$

by equation (3). □

LEMMA 16. *Consider the process* H_{A_k}, *where* $\mathrm{H}\sim\mathrm{PIP}(\beta,\varphi)$ *satisfies the conditions* (RS), (UB), (RC) *and* (IR) *with the constants* C, δ, γ, r *and* R. *Define* $M_k = C^{m_k} < \infty$ *with*

$$m_k = \sup_{x\in\mathcal{X},\xi\in A_k}\xi(A(x,r,R)),$$

where $A(x,r,R) = \mathbb{B}(x,R)\setminus\mathbb{B}(x,r)$. *Then*

$$\lambda_{A_k}(x|\xi) = \lambda(x|\xi)\mathbb{1}\{\xi+\delta_x\in A_k\} \le \beta(x)M_k.$$

This means that the new process H_{A_k} *is locally stable and hence satisfies condition* (S).

In the Euclidean setting $m_k \le mk$, *where*

$$m = \alpha_D D^{D/2}\left(\left(\frac{R}{\delta}+1\right)^D - \left(\frac{r}{\delta}-1\right)^D\right).$$

PROOF. By Lemma 12 and conditions (UB) and (IR) we see that $\lambda_{A_k}(x|\xi)$ can be bounded by $\beta(x)C^{\xi(A(x,r,R))}\mathbb{1}\{\xi+\delta_x\in A_k\}\le\beta(x)C^{m_k}$, and m_k is finite, since it can be bounded by k times the minimal number of balls with radius $\delta/2$ needed to cover $A(x,r,R)$.

In the Euclidean case, consider a partition $\{Q_i\}_{i=1}^N$ of $\mathcal{X}$ by cubes of edge length $\delta/\sqrt{D}$. Since the diameter of each cube is δ, one can cover each cube by a ball with radius $\delta/2$. Furthermore, a cube can intersect $A(x,r,R)$ if and only if it is contained in $A(x,r-\delta,R+\delta)$. Thus the number of cubes intersecting $A(x,r,R)$ can be bounded by the volume of $A(x,r-\delta,R+\delta)$ divided by the volume of a cube, that is,

$$\sup_{x\in\mathcal{X}}\frac{|A(x,r-\delta,R+\delta)|}{\delta^D/D^{D/2}}\le\alpha_D D^{D/2}\biggl(\biggl(\frac{R}{\delta}+1\biggr)^D-\biggl(\frac{r}{\delta}-1\biggr)^D\biggr)=m.$$

Thus $m_k\le mk$. $\square$

The next theorem is a generalization of Theorem 8 that includes noninhibitory PIPs.

THEOREM 17. *Assume that $\Xi\sim\mathrm{PIP}(\beta,\varphi_1)$ and $\mathrm{H}\sim\mathrm{PIP}(\beta,\varphi_2)$. Furthermore, assume that they satisfy the conditions* (RS), (UB), (RC) *and* (IR) *with the same constants C, δ, γ, r and R. Let $\nu_{A_k}(y)=\mathbb{E}(\nu_{A_k}(y|\Xi_{A_k}))$ denote the intensity of Ξ_{A_k}. Then we have for any $k\in\mathbb{N}$,*

$$\begin{aligned}&d_{\mathrm{TV}}(\mathscr{L}(\Xi),\mathscr{L}(\mathrm{H}))\\ &\quad\le c_1(\lambda)M_k\int_{\mathcal{X}}\int_{\mathcal{X}}\beta(x)\nu_{A_k}(y)|\varphi_1(x,y)-\varphi_2(x,y)|\boldsymbol{\alpha}(dx)\boldsymbol{\alpha}(dy)\\ &\qquad+\frac{\gamma^{(k(k+1))/2}B_\delta^k}{(k+1)!C^k}\mathbb{E}(|\Xi|C^{k|\Xi|}+|\mathrm{H}|C^{k|\mathrm{H}|}),\end{aligned}\tag{22}$$

where M_k is defined in Lemma 16, and $c_1(\lambda)$ is given in inequality (10) with

$$c\le M_k\int_{\mathcal{X}}\beta(x)\boldsymbol{\alpha}(dx)$$

and

$$\varepsilon\le M_k\sup_{y\in\mathcal{X}}\biggl(\int_{\mathcal{X}\setminus\mathbb{B}(y,\delta)}\beta(x)|\varphi_2(x,y)-1|\boldsymbol{\alpha}(dx)+\int_{\mathbb{B}(y,\delta)}\beta(x)\boldsymbol{\alpha}(dx)\biggr).$$

Note that $k\in\mathbb{N}$ can be chosen such that the two terms in (22) are best balanced. There is also some freedom in the choice of δ in (IR). In particular it is always possible to choose a lower δ at no cost for γ, that is, the last term in (22) can be made arbitrarily small by letting $\delta\to 0$, which on the other hand leads to an explosion of M_k.

REMARK 18. Usually the expectations in (22) are not easy to compute, but at least Ruelle stability guarantees their finiteness. Let u be the density of Ξ, let ψ^* be as in (RS), and write $\boldsymbol{\alpha}(\psi^*)=\int_{\mathcal{X}}\psi^*(x)\boldsymbol{\alpha}(dx)$. Then for a constant $c^{**}>0$

$$\begin{aligned}\mathbb{E}(|\Xi|C^{k|\Xi|})&\le c^{**}e^{-\boldsymbol{\alpha}(\mathcal{X})}\sum_{n=0}^{\infty}\frac{nC^{kn}\boldsymbol{\alpha}(\psi^*)^n}{n!}\\ &=c^{**}C^k\boldsymbol{\alpha}(\psi^*)e^{C^k\boldsymbol{\alpha}(\psi^*)-\boldsymbol{\alpha}(\mathcal{X})}<\infty.\end{aligned}$$

REMARK 19. If one is interested in a very specific Gibbs processes model, then the estimates in Lemmas 15 and 16 may be improved. See, for instance, Section 4.1, where we treat Lennard–Jones type processes.

By a slight adaptation in the proof of Theorem 8, we can get a nicer result for hard core PIPs.

THEOREM 20. *Let $\Xi \sim \mathrm{PIP}(\beta_1, \varphi_1)$ and $\mathrm{H} \sim \mathrm{PIP}(\beta_2, \varphi_2)$. Assume that both processes have a hard core radius of $\delta > 0$, and satisfy* (UB) *and* (IR). *Then*

$$\begin{aligned} & d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) \\ & \quad \leq c_1(\lambda) M_1 \int_{\mathcal{X}} \int_{\mathcal{X}} \beta(x) \nu(y) |\varphi_1(x,y) - \varphi_2(x,y)| \boldsymbol{\alpha}(dx) \boldsymbol{\alpha}(dy). \end{aligned} \tag{23}$$

PROOF OF THEOREM 17. We adapt the proof of Theorem 8 to compute a bound for $d_{\mathrm{TV}}(\mathscr{L}(\Xi_{A_k}), \mathscr{L}(\mathrm{H}_{A_k}))$. Let $\xi = \sum_{i=1}^n \delta_{y_i}$. Since both φ_1 and φ_2 are bounded by the same C,

$$\begin{aligned} & \left(\prod_{i=1}^{j-1} \varphi_1(x, y_i)\right) \left(\prod_{i=j+1}^{n} \varphi_2(x, y_i)\right) \mathbb{1}\{\xi + \delta_x \in A_k\} \\ & \qquad \leq C^{\xi(\mathbb{B}(x,R) \setminus \mathbb{B}(x,r))} \mathbb{1}\{\xi + \delta_x \in A_k\} \leq C^{m_k} = M_k. \end{aligned}$$

Thus

$$\begin{aligned} & d_{\mathrm{TV}}(\mathscr{L}(\Xi_{A_k}), \mathscr{L}(\mathrm{H}_{A_k})) \\ & \quad \leq c_1(\lambda) M_k \int_{\mathcal{X}} \int_{\mathcal{X}} \beta(x) \nu_{A_k}(y) |\varphi_1(x,y) - \varphi_2(x,y)| \boldsymbol{\alpha}(dx) \boldsymbol{\alpha}(dy). \end{aligned} \tag{24}$$

Since by Lemma 16 $|\lambda_{A_k}(x|\xi) - \lambda_{A_k}(x|\eta)| \leq \beta(x) M_k$ for all $x \in \mathcal{X}$ and $\xi, \eta \in \mathfrak{N}$, we have $c \leq M_k \int_{\mathcal{X}} \beta(x) \boldsymbol{\alpha}(dx)$. By Remark 41 the supremum in the formula for ε can be replaced by an essential supremum with respect to $\mathscr{L}(\Xi_{A_k}) + \mathscr{L}(\mathrm{H}_{A_k})$. This implies that it is enough to take the supremum only over $\xi, \eta \in A_k$. Note that if $d(x,y) > \delta$, $\xi + \delta_x \in A_k$ and $\xi + \delta_y \in A_k$, then also $\xi + \delta_x + \delta_y \in A_k$. Therefore, by using $\lambda(x|\xi + \delta_y) = \lambda(x|\xi)\varphi_2(x,y)$, we obtain

$$\begin{aligned} \varepsilon &= \sup_{\xi,\eta \in A_k, \|\xi - \eta\| = 1} \int_{\mathcal{X}} |\lambda_{A_k}(x|\eta) - \lambda_{A_k}(x|\xi)| \boldsymbol{\alpha}(dx) \\ &= \sup_{y \in \mathcal{X}, \xi + \delta_y \in A_k} \int_{\mathcal{X}} |\lambda(x|\xi + \delta_y) \mathbb{1}\{\xi + \delta_x + \delta_y \in A_k\} \\ & \qquad\qquad - \lambda(x|\xi) \mathbb{1}\{\xi + \delta_x \in A_k\}| \boldsymbol{\alpha}(dx) \end{aligned}$$

$$
\begin{aligned}
&\le \sup_{y\in\mathcal{X},\xi+\delta_y\in A_k}\biggl(\int_{\mathcal{X}\setminus\mathbb{B}(y,\delta)}\lambda(x|\xi)|\varphi_2(x,y)-1|\mathbb{1}\{\xi+\delta_x+\delta_y\in A_k\}\boldsymbol{\alpha}(dx)\\
&\qquad\qquad+\int_{\mathbb{B}(y,\delta)}|\lambda(x|\xi+\delta_y)\mathbb{1}\{\xi+\delta_x+\delta_y\in A_k\}\\
&\qquad\qquad\qquad\qquad-\lambda(x|\xi)\mathbb{1}\{\xi+\delta_x\in A_k\}|\boldsymbol{\alpha}(dx)\biggr)\\
&\le M_k\sup_{y\in\mathcal{X}}\biggl(\int_{\mathcal{X}\setminus\mathbb{B}(y,\delta)}\beta(x)|\varphi_2(x,y)-1|\boldsymbol{\alpha}(dx)+\int_{\mathbb{B}(y,\delta)}\beta(x)\boldsymbol{\alpha}(dx)\biggr).
\end{aligned}
$$

The claim now follows by applying Corollary 14 with $A=A'=A_k$ and Lemma 15. □

Proof of Theorem 20. Since $\mathscr{L}(\Xi)=\mathscr{L}(\Xi_{A_1})$ and $\mathscr{L}(\mathrm{H})=\mathscr{L}(\mathrm{H}_{A_1})$, the statement follows from inequality (24). □

Example 21. Let $\mathcal{X}\subset\mathbb{R}^D$. A PIP is called a *multi-scale Strauss process*, if its interaction function $\varphi(x,y)$ depends only on $\|x-y\|$, is piecewise constant and takes only finitely many values. We restrict ourselves to bi-scale Strauss processes, that is, the interaction function is given by

$$
\varphi(x,y)=\begin{cases}\gamma, & \text{if } \|x-y\|\le r,\\ C, & \text{if } r<\|x-y\|\le R,\\ 1, & \text{if } \|x-y\|>R\end{cases}
$$

for some constants $0\le\gamma\le 1$, $C\ge 0$ and $0<r<R$. To ensure (RS), we furthermore require that $C\le\gamma^{-1/(2m)}$ with

$$
m=m(r,R;D)=\alpha_D D^{D/2}\biggl(\frac{R}{r}+1\biggr)^D.
$$

Note that m is the same as in Lemma 16 with $\delta=r$. (RS) then follows by a criterion of Kondratiev, Pasurek and Röckner (2012), Section 2.3. The authors use m as a bound on the maximal number of cubes with edge length $\delta/\sqrt{D}$ that intersect the annulus $A(0,r-\delta,R+\delta)$, as we did in Lemma 16.

To illustrate Theorem 17, let Ξ and H be bi-scale Strauss processes with constant β, the same γ, r, R, and with $1\le C_{\mathrm{H}}\le C_\Xi$. The ingredients for computing $c_1(\lambda)$ are

$$
\varepsilon=\alpha_D\beta C_{\mathrm{H}}^{mk}(r^D+(C_{\mathrm{H}}-1)(R^D-r^D))\quad\text{and}\quad c=\beta C_{\mathrm{H}}^{mk}|\mathcal{X}|.
$$

Since $\gamma\le C_\Xi^{-2m}$, Theorem 17 yields

$$
\begin{aligned}
d_{\mathrm{TV}}(\mathscr{L}(\Xi),\mathscr{L}(\mathrm{H}))&\le c_1(\lambda)\alpha_D\beta C_\Xi^{mk}\mathbb{E}(|\Xi_{A_k}|)(C_\Xi-C_{\mathrm{H}})(R^D-r^D)\\
&\qquad+\frac{\alpha_D\beta^k r^{Dk}}{(k+1)!}C_\Xi^{-mk^2-(m+1)k}\mathbb{E}(|\Xi|C_\Xi^{k|\Xi|}+|\mathrm{H}|C_\Xi^{k|\mathrm{H}|}).
\end{aligned}
$$

4. Applications.

4.1. *Lennard–Jones type processes.* In this subsection let $(\mathcal{X}, \boldsymbol{\alpha})$ be a compact subset of $\mathbb{R}^D$ equipped with Lebesgue measure. We say a PIP is of *Lennard–Jones type* [see Ruelle (1969)], if its interaction function can be written as $\varphi(x,y) = \exp(-bV(\|x-y\|))$, and the pair potential V satisfies the following conditions:

(1) There exist $r \le R$ and a $\varrho > D$ such that

$$V(x) \ge \|x\|^{-\varrho} \qquad \text{for } \|x\| \le r,$$
$$V(x) \ge -\|x\|^{-\varrho} \qquad \text{for } \|x\| \ge R.$$

(2) $V(x) \ge -M$ for a $M \ge 0$ and for all $x \ge 0$, that is, the interaction function φ is bounded from above by e^{bM}.

The technique used in Section 3.3 stands and falls with a good estimate on the term $\sup_{\xi \in \mathfrak{N}} \lambda_{A_k}(x|\xi)$. The next lemma gives a neat replacement of Lemma 16 for Lennard–Jones type processes.

LEMMA 22. *Assume that* H *is a* PIP *of Lennard–Jones type with constants* ϱ, r, R, M. *Choose a positive* $\delta \le r$ *such that also* $\delta < R/2$. *Define*

$$M_k = \exp\left(bk\left(mM + \frac{\alpha_D D}{\varrho - D}\left(\frac{\sqrt{D}}{\delta}\right)^D \frac{(R-\delta)^{D-1}}{(R-2\delta)^{\varrho-1}}\right)\right), \tag{25}$$

where

$$m = m(r,R;\delta) = \alpha_D D^{D/2}\left(\left(\frac{R}{\delta}+1\right)^D - \left(\frac{r}{\delta}-1\right)^D\right)\mathbb{1}\{r<R\}.$$

Then for all $x \in \mathcal{X}$ *and for all* $\xi \in \mathfrak{N}$ *we have*

$$\lambda_{A_k}(x|\xi) \le \beta(x) M_k.$$

PROOF. Since $\varphi \le e^{bM}$, we obtain analogously as in the proof of Lemma 16 and by using the translation invariance of V that

$$\begin{aligned}\sup_{\xi \in A_k} \lambda(x|\xi) &\le \beta(x) \sup_{\xi \in A_k} \exp\left(bM\xi(A(x,r,R)) - b\sum_{y\in\xi, \|y\|\ge R} V(\|y\|)\right) \\ &\le \beta(x) e^{bMmk} \sup_{\xi \in A_k} \exp\left(-b\sum_{y\in\xi,\|y\|\ge R} V(\|y\|)\right).\end{aligned} \tag{26}$$

Let $\{Q_z\}_{z\in\mathbb{Z}^D}$ denote the partition of $\mathbb{R}^D$ into cubes of edge length $\delta/\sqrt{D}$ and centre points $\delta/\sqrt{D}\mathbb{Z}^D$. Since $\xi \in A_k$, each cube contains at most k

points. A cube intersects $\mathbb{B}(0,R)^c$ if and only if its center point is contained in $\mathbb{B}(0,R-\delta/2)^c$. For $x=(x_1,\dots,x_D)\in\mathbb{R}^D$, denote $\|x\|_{\max}=\max_{i=1,\dots,D}|x_i|$. Thus

$$\sup_{\xi\in A_k}\sum_{y\in\xi,\|y\|\geq R}-V(\|y\|)$$

$$\leq \sum_{\substack{z\in\mathbb{Z}^D\\ \|z\|\geq\sqrt{D}(R/\delta-1/2)}} k \sup_{\|x-z\|_{\max}\leq 1/2}\left(\frac{\delta}{\sqrt{D}}\|x\|\right)^{-\varrho} \tag{27}$$

$$= k\left(\frac{\delta}{\sqrt{D}}\right)^{-\varrho}\sum_{\substack{z\in\mathbb{Z}^D\\ \|z\|\geq\sqrt{D}(R/\delta-1/2)}}\sup_{\|x-z\|_{\max}\leq 1/2}\|x\|^{-\varrho}.$$

Consider the function $g(x)=(\|x\|-\sqrt{D})^{-\varrho}$. Since

$$\inf_{\|x-z\|_{\max}\leq 1/2} g(x)\geq\left(\|z\|-\frac{\sqrt{D}}{2}\right)^{-\varrho}\geq\sup_{\|x-z\|_{\max}\leq 1/2}\|x\|^{-\varrho},$$

the last sum in (27) can be bounded by an integral over the function g. For any $a>0$ we get

$$\int_{\mathbb{B}(0,a)^c}(\|x\|-\sqrt{D})^{-\varrho}\,dx$$

$$=\alpha_D D\int_a^\infty (r-\sqrt{D})^{-\varrho}r^{D-1}\,dr=\alpha_D D\int_{a-\sqrt{D}}^\infty\left(\frac{r+\sqrt{D}}{r}\right)^{D-1}r^{D-1-\varrho}\,dr$$

$$\leq\alpha_D D\left(\frac{a}{a-\sqrt{D}}\right)^{D-1}\int_{a-\sqrt{D}}^\infty r^{D-\varrho-1}\,dr$$

$$=\alpha_D D\left(\frac{a}{a-\sqrt{D}}\right)^{D-1}\frac{1}{\varrho-D}\frac{1}{(a-\sqrt{D})^{\varrho-D}}.$$

In order to catch all cubes in (27), the integration must begin at $a=\sqrt{D}(R/\delta-1)$, which together with (26) yields the claim. □

We may also give a more explicit bound on $\mathbb{P}(\mathrm{H}\notin A_k)$ than inequality (19) in the case of a Lennard–Jones type process.

LEMMA 23. *Assume that* H *is a* PIP *of Lennard–Jones type with constants* ϱ,r,R,M. *Then for any* $\delta<\min(r,R/2)$

$$\mathbb{P}(\mathrm{H}\notin A_k)\leq\left(\int_{\mathcal{X}}\beta(x)\,dx\right)\sum_{j=k+1}^{\infty}\frac{1}{j!}B_\delta^{j-1}\exp(-j^2bL(\delta)),\tag{28}$$

where B_δ is given in Lemma 15 and

$$L(\delta) = \frac{1}{4}\delta^{-\varrho} - \delta^{-D}\biggl(M\alpha_D D^{D/2}((R+\delta)^D - (r-\delta)^D)\mathbb{1}\{r<R\} + \frac{\alpha_D D^{D/2+1}}{\varrho - D}\frac{(R-\delta)^{D-1}}{(R-2\delta)^{\varrho-1}}\biggr),$$

which is positive for reasonably small δ, since $\varrho > D$.

Proof. Let $\widetilde{A}_k = A_k \setminus A_{k-1}$ for $k \geq 2$. The sets $\widetilde{A}_k$, $k \geq 2$, are pairwise disjoint and $\mathfrak{N} \setminus A_l = \bigcup_{k=l+1}^\infty \widetilde{A}_k$ for all $l \geq 1$. Then by (3)

$$\begin{aligned}\mathbb{P}(\mathrm{H} \in \widetilde{A}_k) &= e^{-|\mathcal{X}|}\sum_{n=k}^\infty \frac{1}{n!}\int_{\mathcal{X}}\cdots\int_{\mathcal{X}} \mathbb{1}\{\{x_1,\ldots,x_n\} \in A_k\} \\ &\qquad\times g_{\delta,k}(x_1,\ldots,x_n)u(\{x_1,\ldots,x_n\})\,dx_1\cdots dx_n, \end{aligned} \tag{29}$$

where $g_{\delta,k}$ has been defined in (21). For any permutation $(i_1,\ldots,i_n)$ of $(1,\ldots,n)$ the density can be rewritten in a similar way as in the proof of Lemma 15 as

$$\begin{aligned}&u(\{x_1,\ldots,x_n\}) \\ &\quad= \biggl(\prod_{1\leq j<l\leq k}\varphi(x_{i_j},x_{i_l})\biggr)\biggl(\prod_{j=1}^k \lambda(x_{i_j}|\{x_{i_{k+1}},\ldots,x_{i_n}\})\biggr)u(\{x_{i_{k+1}},\ldots,x_{i_n}\}).\end{aligned}$$

Since $\delta \leq r$, for all x,y with $\|x-y\| \leq \delta$, we have $\varphi(x,y) \leq \exp(-b\delta^{-\varrho})$. The term $g_{\delta,k}(x_1,\ldots,x_n)$ can be bounded as in (21), and by using Lemma 22 and the triangle inequality, we get

$$\begin{aligned}&\mathbb{P}(\mathrm{H} \in \widetilde{A}_k) \\ &\quad\leq e^{-|\mathcal{X}|}\sum_{n=k}^\infty \frac{1}{n!}\sum_{\{i_1,\ldots,i_k\}\subset\{1,\ldots,n\}} \underbrace{\int_{\mathcal{X}}\cdots\int_{\mathcal{X}}}_{n-k}\int_{\mathcal{X}}\underbrace{\int_{\mathbb{B}(x_{i_k},\delta)}\cdots\int_{\mathbb{B}(x_{i_k},\delta)}}_{k-1} e^{-b\delta^{-\varrho}\binom{k}{2}} \\ &\qquad\times\biggl(\prod_{j=1}^k \beta(x_{i_j})\biggr)M_k^k u(\{x_{i_{k+1}},\ldots,x_{i_n}\})\,dx_{i_1}\cdots dx_{i_n} \\ &\quad\leq \frac{1}{k!}\biggl(\int_{\mathcal{X}}\beta(x)\,dx\biggr)B_\delta^{k-1}e^{-b\delta^{-\varrho}\binom{k}{2}}M_k^k e^{-|\mathcal{X}|} \\ &\qquad\times\sum_{n=k}^\infty \frac{1}{(n-k)!}\underbrace{\int_{\mathcal{X}}\cdots\int_{\mathcal{X}}}_{n-k} u(\{x_{k+1},\ldots,x_n\})\,dx_{k+1}\cdots dx_n\end{aligned}$$

$$= \frac{1}{k!}\left(\int_{\mathcal{X}} \beta(x)\,dx\right) B_\delta^{k-1} e^{-b\delta^{-\varrho}\binom{k}{2}} M_k^k.$$

For $k \geq 2$ we have $\binom{k}{2} = (k^2/2)(1-1/k) \geq k^2/4$. Thus Lemma 22 yields $\exp(-b\delta^{-\varrho}\binom{k}{2})M_k^k \leq \exp(-bk^2 L(\delta))$. The statement then follows by $\mathbb{P}(\mathrm{H} \notin A_l) = \sum_{k=l+1}^{\infty} \mathbb{P}(\mathrm{H} \in \widetilde{A}_k)$. $\square$

Putting the pieces together as we did for the proof of Theorem 17 we obtain the corresponding upper bound. For the sake of simplicity we formulate the following result for the special case of the classical Lennard–Jones process in three dimensions, where the pair potential is of the form

$$V_R(x) = \left(\frac{R}{\|x\|}\right)^{12} - \left(\frac{R}{\|x\|}\right)^{6}$$

for $R > 0$ and $x \in \mathbb{R}^3$. In this case we can choose $r = R$ and $\varrho = 6$.

Theorem 24. *Let $\Xi \sim \mathrm{PIP}(\beta, \varphi_1)$ and $\mathrm{H} \sim \mathrm{PIP}(\beta, \varphi_2)$ be classical Lennard–Jones processes with interaction functions $\varphi_i(x) = \exp(-b_i V_{R_i}(x))$ for $i = 1, 2$. Let $\nu_{A_k}(y) = \mathbb{E}(\nu_{A_k}(y|\Xi_{A_k}))$ denote the intensity of Ξ_{A_k}. Then we have for any $k \in \mathbb{N}$ and $\delta < \min(R_1, R_2)/2$ that*

$$\begin{aligned} &d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) \\ &\quad\leq c_1(\lambda) \exp\left(b_2 k 4\pi\sqrt{3}\frac{(R_2-\delta)^2}{\delta^3(R_2-2\delta)^5}\right) \\ &\qquad\times \int_{\mathcal{X}}\int_{\mathcal{X}} \beta(x)\nu_{A_k}(y)|\varphi_1(\|x-y\|) - \varphi_2(\|x-y\|)|\,dx\,dy \\ &\qquad+ \left(\int_{\mathcal{X}} \beta(x)\,dx\right) \sum_{j=k+1}^{\infty} \frac{1}{j!} B_\delta^{j-1}\left(e^{-j^2 b_1 L_1(\delta)} + e^{-j^2 b_2 L_2(\delta)}\right), \end{aligned} \tag{30}$$

where

$$L_i(\delta) = \frac{1}{4}\delta^{-6} - 4\pi\sqrt{3}\frac{(R_i-\delta)^2}{\delta^3(R_i-2\delta)^5}$$

for $i = 1, 2$. Note that ν_{A_k} may be bounded in a crude manner by

$$\nu_{A_k}(y) \leq \beta(y)\exp\left(b_1 k 4\pi\sqrt{3}\frac{(R_1-\delta)^2}{\delta^3(R_1-2\delta)^5}\right).$$

Remark 25. Typically any endeavors to make the second summand in inequality (30) small make the exponential factor in the first summand quite large so that the bounds are mainly useful in an asymptotic setting where the interaction functions are very close. Note, however, that if R_1 and R_2 are large, so that δ may be chosen quite a bit larger than 1 but still substantially smaller than R_i, choosing a large k results in a situation where the second summand is close to zero and the exponential factor is close to one.

4.2. *The hard core process as limit of area interaction processes.* For simplicity let again $(\mathcal{X}, \boldsymbol{\alpha})$ be a compact subset of $\mathbb{R}^D$ with Lebesgue measure. In this subsection let $\mathbb{B}(x,R)$ always denote the closed ball in $\mathbb{R}^D$ rather than in $\mathcal{X}$. Let H be a Strauss process with parameters $R, \beta_0 > 0$ and $\gamma_0 = 0$ (hard core case). Let furthermore $\Xi := \Xi_{\beta,\gamma}$ be an area interaction process with parameters $R/2, \beta, \gamma$, where $\gamma \in (0,1]$. The unnormalized density of such a process is given by

$$\tilde{u}(\xi) = \beta^{|\xi|}\gamma^{-|\bigcup_{y\in\xi}\mathbb{B}(y,R/2)|}$$

and the conditional intensity is therefore

$$\nu(x|\xi) = \beta\gamma^{-|\mathbb{B}(x,R/2)\setminus\bigcup_{y\in\xi}\mathbb{B}(y,R/2)|};$$

see Baddeley and van Lieshout (1995) for more details. The authors show that $\mathscr{L}(\Xi_{\beta,\gamma}) \to \mathscr{L}(\mathrm{H})$ weakly as $\beta, \gamma \to 0$ in such a way that $\beta\gamma^{-\alpha_D(R/2)^D} \to \beta_0$ (it is easily seen that the hard core process referred to by Baddeley and van Lieshout is in fact the Strauss hard core process we use). We derive a rate for this convergence.

THEOREM 26. *Let* Ξ *and* H *be as above. Then*

$$\begin{aligned} &d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) \\ &\qquad \le c_1(\lambda)(|\beta\gamma^{-\alpha_D(R/2)^D} - \beta_0||\mathcal{X}| + \beta\gamma^{-\alpha_D(R/2)^D}\mathbb{E}|\Xi| I_D(R,\gamma)), \end{aligned}$$

where

$$I_D(R,\gamma) := \int_{\mathbb{B}(0,R)} \gamma^{|\mathbb{B}(x,R/2)\cap\mathbb{B}(0,R/2)|}\,dx \le 2\alpha_D D R^{D-1}\log(\gamma^{-\alpha_D})^{-1/D}.$$

PROOF. For the difference between the conditional intensities we obtain

$$\nu(x|\xi) - \lambda(x|\xi) = \begin{cases} \beta\gamma^{-|\mathbb{B}(x,R/2)|} - \beta_0, & \text{if } \mathrm{dist}(x,\xi) > R, \\ \beta\gamma^{-|\mathbb{B}(x,R/2)\setminus\bigcup_{y\in\xi}\mathbb{B}(y,R/2)|}, & \text{if } \mathrm{dist}(x,\xi) \le R, \end{cases}$$

where $\mathrm{dist}(x,A) = \inf_{y\in A}\|x-y\|$ for any $A \subset \mathbb{R}^D$. Therefore

$$\begin{aligned} &\int |\nu(x|\xi) - \lambda(x|\xi)|\,dx \\ &\quad \le \int |\beta\gamma^{-|\mathbb{B}(x,R/2)|} - \beta_0|\,dx \\ &\qquad + \beta\gamma^{-\alpha_D(R/2)^D}\int_{\{\tilde{x}:\mathrm{dist}(\tilde{x},\xi)\le R\}} \gamma^{|\mathbb{B}(x,R/2)\cap\bigcup_{y\in\xi}\mathbb{B}(y,R/2)|}\,dx. \end{aligned}$$

The last integral may be bounded further by

$$
\begin{aligned}
&|\xi| \int_{\mathbb{B}(0,R)} \gamma^{|\mathbb{B}(x,R/2)\cap\mathbb{B}(0,R/2)|}\,dx \\
&\qquad \le |\xi| \int_{\mathbb{B}(0,R)} \gamma^{\alpha_D(R-\|x\|)^D/2^D}\,dx = |\xi|\alpha_D D \int_0^R \gamma^{\alpha_D(R-r)^D/2^D} r^{D-1}\,dr.
\end{aligned}
$$

By the substitution $y = \log(\gamma^{-\alpha_D})(\frac{R-r}{2})^D$ this is equal to

$$
\begin{aligned}
&|\xi|\alpha_D D \log(\gamma^{-\alpha_D})^{-1/D} \frac{2}{D} \int_0^{\log(\gamma^{-\alpha_D})(R/2)^D} e^{-y} y^{(1/D)-1} \\
&\qquad\qquad\qquad\qquad \times \left(R - 2\left(\frac{y}{\log(\gamma^{-\alpha_D})}\right)^{1/D}\right)^{D-1} dy \\
&\qquad \le 2|\xi|\alpha_D D R^{D-1} \log(\gamma^{-\alpha_D})^{-1/D} \frac{1}{D} \int_0^\infty e^{-y} y^{(1/D)-1}\,dy \\
&\qquad \le 2|\xi|\alpha_D D R^{D-1} \log(\gamma^{-\alpha_D})^{-1/D}.
\end{aligned}
$$

The last inequality holds since the integral is equal to $\Gamma(1/D)$ and the functional equality of the Gamma function yields $(1/D)\Gamma(1/D) = \Gamma(1+1/D) \le 1$ for all $D \ge 1$. The result follows now from Theorem 4. □

The next proposition shows that in the case $\beta\gamma^{-\alpha_D(R/2)^D} = \beta_0$ the above rate is sharp. Define $\mathcal{X}^{(-R)} := \{x \in \mathcal{X} : \operatorname{dist}(x, \mathcal{X}^c) \ge R\}$ and choose R_0 such that $|\mathcal{X}^{(-R_0)}| > 0$.

PROPOSITION 27. *Let Ξ and H be as above. Assume that $\beta\gamma^{-\alpha_D(R/2)^D} = \beta_0$ and $R \le R_0$. Then there exists a positive constant κ such that*

$$
d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) \ge \kappa I_D(R, \gamma). \tag{31}
$$

PROOF. Define $A = \{\xi \in \mathfrak{N} : \exists\{x, y\} \subset \xi, \|x - y\| \le R\}$. Note that $\mathbb{P}(\mathrm{H} \in A) = 0$. Hence

$$
\begin{aligned}
&d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) \\
&\qquad = \sup_{B\in\mathcal{N}} |\mathbb{P}(\Xi \in B) - \mathbb{P}(\mathrm{H} \in B)| \ge |\mathbb{P}(\Xi \in A) - \mathbb{P}(\mathrm{H} \in A)| = \mathbb{P}(\Xi \in A).
\end{aligned}
$$

Denote by c_Ξ the normalizing constant of the density of Ξ. Then

$$
\begin{aligned}
&\mathbb{P}(\Xi \in A) \\
&\qquad \ge \mathbb{P}(\Xi \in A, |\Xi| = 2) \\
&\qquad = c_\Xi \frac{e^{-|\mathcal{X}|}}{2} \int_{\mathcal{X}} \int_{\mathcal{X}} \beta^2 \gamma^{-|\mathbb{B}(x_1,R/2)\cup\mathbb{B}(x_2,R/2)|} \mathbb{1}\{\|x_1 - x_2\| \le R\}\,dx_1\,dx_2
\end{aligned}
$$

$$= c_\Xi \frac{e^{-|\mathcal{X}|}}{2} \int_{\mathcal{X}} \int_{\mathbb{B}(x_2,R)\cap\mathcal{X}} \beta^2 \gamma^{-|\mathbb{B}(x_1,R/2)|} \times \gamma^{-|\mathbb{B}(x_2,R/2)|} \gamma^{|\mathbb{B}(x_1,R/2)\cap\mathbb{B}(x_2,R/2)|}\, dx_1\, dx_2$$

$$\geq c_\Xi \frac{e^{-|\mathcal{X}|}}{2} \beta_0^2 |\mathcal{X}^{(-R)}| \int_{\mathbb{B}(0,R)} \gamma^{|\mathbb{B}(x,R/2)\cap\mathbb{B}(0,R/2)|}\, dx,$$

where we used that $\mathbb{B}$ denotes a ball in $\mathbb{R}^D$ and the translation invariance of the Lebesgue measure. Since $\gamma \leq 1$, we have $u_\Xi(\xi) \leq c_\Xi \beta_0^{|\xi|}$ for all $\xi \in \mathfrak{N}$. Integrating with respect to Po_1 yields $c_\Xi \geq \exp(|\mathcal{X}|(1-\beta_0))$. Thus one may choose $\kappa = e^{-\beta_0|\mathcal{X}|}\beta_0^2|\mathcal{X}^{(-R_0)}|/2$. $\square$

4.3. *Discrete processes.* Let $(\mathcal{X}, \boldsymbol{\alpha})$ be a general space with a diffuse measure $\boldsymbol{\alpha}$. Our aim is to compare Gibbs processes which live on a finite subset $\Lambda = \Lambda_n = \{y_i\}_{i=1}^n$ of $\mathcal{X}$ with Gibbs processes on $\mathcal{X}$. Let $V = \{V_i\}_{i=1}^n$ be a partition of $\mathcal{X}$ such that $y_i \in V_i$ for all $i = 1, \ldots, n$. A natural choice is the Voronoi tesselation, provided that $\boldsymbol{\alpha}(\partial V_i) = 0$ for all $i = 1, \ldots, n$. Define $r_V = \max_{i=1,\ldots,n} \sup_{x\in V_i} d(x, y_i)$, the maximal radius of the cells in V. Furthermore, let $(\mathfrak{N}_\Lambda, \mathcal{N}_\Lambda)$ denote the space of point measures on Λ with its natural σ-algebra, which coincides with the power set of $\mathfrak{N}$.

Define the reference measure $\boldsymbol{\alpha}_\Lambda$ on Λ by $\boldsymbol{\alpha}_\Lambda(\{y_i\}) = \boldsymbol{\alpha}(V_i)$ for all $i = 1, \ldots, n$ and let Po_Λ denote the Poisson process distribution on Λ with intensity measure $\boldsymbol{\alpha}_\Lambda$. The Gibbs point processes on Λ are then defined in the obvious way, that is, as the point processes that have a hereditary density with respect to Po_Λ.

Let $\Xi_\Lambda \sim \mathrm{Gibbs}(u_\Lambda)$ be a Gibbs process on Λ. Define a point process Ξ_U on $\mathcal{X}$ in the following manner. Each point of Ξ_Λ is replaced by a $\boldsymbol{\alpha}(\cdot)|_{V_i}/\boldsymbol{\alpha}(V_i)$-distributed point in the corresponding cell V_i. More formally, if $\Xi_\Lambda = \sum_{i=1}^n N_i \delta_{y_i}$, then

$$\Xi_U = \sum_{i=1}^n \sum_{l=1}^{N_i} \delta_{U_{il}}, \tag{32}$$

where the U_{il} are all independent and $U_{il} \sim \boldsymbol{\alpha}(\cdot)|_{V_i}/\boldsymbol{\alpha}(V_i)$.

Define a function $t: \mathcal{X} \to \Lambda$ which maps each point in $\mathcal{X}$ to its lattice point in Λ, that is, $t(x) = y_i$ if $x \in V_i$ for $i = 1, \ldots, n$. In the same spirit set $t(\xi) = t(\sum_{x\in\xi} \delta_x) = \sum_{x\in\xi} \delta_{t(x)}$ for every $\xi \in \mathfrak{N}$.

LEMMA 28. *Let $\Xi_\Lambda \sim \mathrm{Gibbs}(u_\Lambda)$. Then the corresponding point process Ξ_U on $\mathcal{X}$ has density $u_U(\xi) = u_\Lambda(t(\xi))$ with respect to* Po_1.

PROOF. For $j_1, \ldots, j_n \in \mathbb{Z}_+$, set $k = \sum_{i=1}^n j_i$ and

$$A_{j_1,\ldots,j_n} = \{\xi \in \mathfrak{N}: \xi(V_1) = j_1, \ldots, \xi(V_n) = j_n\}.$$

Then, writing

$$\binom{k}{j_1,\dots,j_n} = \frac{k!}{j_1!\cdots j_n!}$$

for the multinomial coefficient, we have

$$\begin{aligned}
&\mathbb{P}(\Xi_U \in A_{j_1,\dots,j_n})\\
&\quad = \mathbb{P}(\Xi_\Lambda \in A_{j_1,\dots,j_n})\\
&\quad = \frac{e^{-\boldsymbol{\alpha}(\mathcal{X})}}{k!} u_\Lambda\left(\sum_{i=1}^n j_i \delta_{y_i}\right)\binom{k}{j_1,\dots,j_n}\boldsymbol{\alpha}(V_1)^{j_1}\cdots\boldsymbol{\alpha}(V_n)^{j_n}\\
&\quad = \frac{e^{-\boldsymbol{\alpha}(\mathcal{X})}}{k!}\int_{\mathcal{X}}\cdots\int_{\mathcal{X}} u_\Lambda\left(\sum_{r=1}^k \delta_{t(x_r)}\right) \qquad (33)\\
&\qquad\qquad \times \mathbb{1}\{\#\{r: x_r \in V_1\} = j_1,\dots,\\
&\qquad\qquad\qquad \#\{r: x_r \in V_n\} = j_n\}\boldsymbol{\alpha}(dx_1)\cdots\boldsymbol{\alpha}(dx_k)\\
&\quad = \mathbb{P}(\widetilde{\Xi} \in A_{j_1,\dots,j_n}),
\end{aligned}$$

where $\widetilde{\Xi} \sim \mathrm{Gibbs}(u_\Lambda \circ t)$.

Note that the density $u_\Lambda(t(\cdot))$ is constant on any $A_{j_1,\dots,j_n}$. Hence, given that $\widetilde{\Xi}(V_1) = j_1,\dots,\widetilde{\Xi}(V_n) = j_n$, we may write $\widetilde{\Xi}$ as $\sum_{i=1}^n \sum_{l=1}^{j_i} \delta_{U_{il}}$, where the U_{il} are all independent and $U_{il} \sim \boldsymbol{\alpha}(\cdot)|_{V_i}/\boldsymbol{\alpha}(V_i)$. Thus, for every measurable $h: \mathfrak{N} \to \mathbb{R}_+$ we get $\mathbb{E}(h(\Xi_U)|A_{j_1,\dots,j_n}) = \mathbb{E}(h(\widetilde{\Xi})|A_{j_1,\dots,j_n})$, which together with equation (33) and the formula of total expectation yields the claim. □

Many Gibbs processes Ξ on $\mathcal{X}$ with density u have a discrete analogon, which is obtained by restricting the density to Λ and renormalizing, that is, by using the unnormalized density $\tilde{u}_\Lambda = u|_{\mathfrak{N}_\Lambda}$ on $\mathfrak{N}_\Lambda$, provided that it is $\boldsymbol{\alpha}_\Lambda$-integrable. Some special care is required when evaluating u at point configurations $\xi \in \mathfrak{N}$ that have multi-points. For the continuous Gibbs process such ξ form a null set, whereas for the discrete analogon the values of $\tilde{u}_\Lambda$ at such ξ become important. We avoid this problem by assuming that $u(\xi) = 0$ for any ξ with multi-points, which leads to discrete analoga Ξ_Λ that may be represented as collections of Bernoulli random variables $(I_y)_{y\in\Lambda}$. Consequently Ξ_U may have no more than one point in any cell V_i.

By Theorem 4 and Lemma 28 we immediately obtain the following proposition.

PROPOSITION 29. *Suppose that $\Xi \sim \mathrm{Gibbs}(\nu)$ satisfies* (S). *Let Ξ_Λ be its discrete analogon, and let Ξ_U be given by (32). Then*

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi_U), \mathscr{L}(\Xi)) \le c_1(\lambda)\int_{\mathcal{X}} \mathbb{E}|\nu(t(x)|t(\Xi_U)) - \nu(x|\Xi_U)|\boldsymbol{\alpha}(dx). \qquad (34)$$

Consider the special case where $\Xi \sim \mathrm{PIP}(\beta, \varphi)$ with a constant β and with $\varphi \leq 1$ (inhibitory case). Our process Ξ_Λ is then an *auto-logistic process* in the terminology of Besag (1974). In Besag, Milne and Zachary (1982) convergence of the corresponding Ξ_U-process density toward the Ξ-process density is studied under a continuity condition on φ, without providing rates.

We have

$$\nu(t(x)|t(\Xi_U)) = \beta \prod_{y \in \Xi_U} \varphi(t(x), t(y)) \qquad \text{a.s.}$$

Imitating the proof of Theorem 8, we obtain from Proposition 29 the following result.

PROPOSITION 30. *Let $\Xi \sim \mathrm{PIP}(\beta, \varphi)$ with constant β and $\varphi \leq 1$. Then*

$$\begin{aligned} &d_{\mathrm{TV}}(\mathscr{L}(\Xi_U), \mathscr{L}(\Xi)) \\ &\quad \leq c_1(\lambda)\mathbb{E}|\Xi_U|\beta \sup_{y \in \mathcal{X}} \int_{\mathcal{X}} |\varphi(t(x), t(y)) - \varphi(x, y)| \boldsymbol{\alpha}(dx). \end{aligned} \tag{35}$$

COROLLARY 31. *Let $\Xi \sim \mathrm{PIP}(\beta, \varphi)$ with constant β and a $\varphi \leq 1$, that is, Lipschitz continuous with constant L in both components. Then*

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi_U), \mathscr{L}(\Xi)) \leq 2c_1(\lambda)\mathbb{E}|\Xi_U|\beta L \boldsymbol{\alpha}(\mathcal{X}) r_V. \tag{36}$$

PROOF. Note that by the triangle inequality,

$$\begin{aligned} &|\varphi(t(x), t(y)) - \varphi(x, y)| \\ &\quad \leq |\varphi(t(x), t(y)) - \varphi(t(x), y)| + |\varphi(t(x), y) - \varphi(x, y)| \leq 2L r_V. \qquad \square \end{aligned}$$

EXAMPLE 32. Let Ξ be a Strauss process, that is, $\varphi(x, y) = \gamma^{1\{d(x,y) \leq R\}}$ for a $\gamma \in [0, 1]$. Write $A(y, R_1, R_2) = \{x \in \mathcal{X} : R_1 < d(x, y) \leq R_2\}$. For $x \notin A(y, R - 2r_V, R + 2r_V)$, we have

$$\gamma^{1\{d(t(x),t(y)) \leq R\}} - \gamma^{1\{d(x,y) \leq R\}} = 0$$

and for $x \in A(y, R - 2r_V, R + 2r_v)$ the modulus of the above difference is at most $(1 - \gamma)$. Hence by Proposition 30,

$$\begin{aligned} &d_{\mathrm{TV}}(\mathscr{L}(\Xi_U), \mathscr{L}(\Xi)) \\ &\quad \leq c_1(\lambda)\mathbb{E}|\Xi_U|\beta(1 - \gamma) \sup_{y \in \mathcal{X}} \boldsymbol{\alpha}(A(y, R - 2r_V, R + 2r_V)). \end{aligned} \tag{37}$$

In the Euclidean case we get a linear rate in r_V, since $\boldsymbol{\alpha}(A(y, R - 2r_V, R + 2r_V)) \leq 4\alpha_D D(R + 2r_V)^{D-1} r_V$.

REMARK 33. By combining the techniques of Corollary 31 and Example 32, we obtain linear rates in r_V for any inhibitory interaction function φ that is piecewise Lipschitz continuous.

To compute the distance between Ξ and its discrete analogon Ξ_Λ we need another distance than the total variation. This is because $\mathbb{P}(\Xi \subset \Lambda) = \mathbb{P}(|\Xi| = 0)$ and thus $d_{\mathrm{TV}}(\mathscr{L}(\Xi_\Lambda), \mathscr{L}(\Xi)) \geq 1 - \mathbb{P}(|\Xi| = 0)$, whereas one would like to have a distance that vanishes as $r_V \to 0$. We use the following Wasserstein metric; see Barbour and Brown (1992), Section 3, for details. Let $\xi = \sum_{i=1}^n \delta_{x_i}$ and $\eta = \sum_{i=1}^m \delta_{y_i}$. Define a metric d_1 on $\mathfrak{N}$ by

$$d_1(\xi, \eta) = \begin{cases} 1, & \text{if } n \neq m, \\ \dfrac{1}{n} \min_{\sigma \in S_n} \displaystyle\sum_{i=1}^n \min(d(x_i, y_{\sigma(i)}), 1), & \text{if } n = m, \end{cases}$$

where S_n denotes the permutation group of order n. Denote by $\mathcal{F}_2$ the set of functions $f : \mathfrak{N} \to [0, 1]$ such that $|f(\xi) - f(\eta)| \leq d_1(\xi, \eta)$ for all $\xi, \eta \in \mathfrak{N}$. Our Wasserstein distance is then defined by

$$d_2(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) = \sup_{f \in \mathcal{F}_2} |\mathbb{E} f(\Xi) - \mathbb{E} f(\mathrm{H})|. \tag{38}$$

We obtain the following theorem.

THEOREM 34. *Suppose that* $\Xi \sim \mathrm{Gibbs}(\nu)$ *satisfies* (S). *Let* Ξ_Λ *be the discrete analogon. Then*

$$d_2(\mathscr{L}(\Xi_\Lambda), \mathscr{L}(\Xi)) \leq r_V + c_1(\lambda) \int_{\mathcal{X}} \mathbb{E} |\nu(t(x)|t(\Xi_U)) - \nu(x|\Xi_U)| \boldsymbol{\alpha}(dx). \tag{39}$$

PROOF. We have

$$d_2(\mathscr{L}(\Xi_\Lambda), \mathscr{L}(\Xi)) \leq d_2(\mathscr{L}(\Xi_\Lambda), \mathscr{L}(\Xi_U)) + d_2(\mathscr{L}(\Xi_U), \mathscr{L}(\Xi)).$$

For the first summand we obtain by the Lipschitz continuity of $f \in \mathcal{F}_2$ that

$$\sup_{f \in \mathcal{F}_2} |\mathbb{E} f(\Xi_\Lambda) - \mathbb{E} f(\Xi_U)| \leq \mathbb{E}\, d_1(\Xi_\Lambda, \Xi_U) \leq r_V,$$

where we used that the distance between any point in Ξ_Λ and its replacement point in Ξ_U is at most r_V.

Since $\mathcal{F}_2 \subset \mathcal{F}_{\mathrm{TV}}$, the Wasserstein distance is always bounded by the total variation distance. The second summand above may therefore be bounded according to Proposition 29, which yields the claim. □

5. Couplings of spatial birth–death processes. Let $b(\cdot|\cdot), d(\cdot|\cdot)\colon \mathcal{X} \times \mathfrak{N} \to \mathbb{R}_+$ be measurable functions such that $\bar{b}(\xi) := \int b(x|\xi)\boldsymbol{\alpha}(dx) < \infty$ for every $\xi \in \mathfrak{N}$, and set $\bar{d}(\xi) = \sum_{x\in\xi} d(x|\xi)$ and $\bar{a}(\xi) = \bar{b}(\xi) + \bar{d}(\xi)$. A *spatial birth–death process* (SBDP) $(Z(t))_{t\geq 0}$ with *birth rate* b and *death rate* d is a pure-jump Markov process on $\mathfrak{N}$ that can be described as follows: given it is in state $\xi \in \mathfrak{N}$ and $\bar{a}(\xi) > 0$ it stays there for an $\mathrm{Exp}(\bar{a}(\xi))$-distributed time, after which a point is added to ξ ("birth") with probability $\bar{b}(\xi)/\bar{a}(\xi)$ or deleted from ξ with probability $\bar{d}(\xi)/\bar{a}(\xi)$ ("death"). If a birth occurs, the new point is positioned according to the density $b(x|\xi)/\bar{b}(\xi)$. If a death occurs, the point $x \in \xi$ is omitted with probability $d(x|\xi)/\bar{d}(\xi)$. In the case $\bar{a}(\xi) = 0$ the SBDP is absorbed in ξ, that is, stays there indefinitely. Preston (1975) and Møller and Waagepetersen (2004), Chapter 11 and Appendix G, give more formal definitions of general SBDP and a wealth of other results, including conditions to assure that the SBDP is *nonexplosive*, that is, that with probability 1 only finitely many jumps can occur in any bounded time interval. Denote by $(Z_\xi(t))_{t\geq 0}$ the process with deterministic starting configuration ξ, that is, $Z_\xi(0) = \xi$.

We concentrate here on the case, where $b(x|\xi) := \lambda(x|\xi)$ and $d(x|\xi) = 1$ ("unit per-capita death rate") and where condition (S) holds. Thus we obtain a nonexplosive SBDP that is time-reversible with respect to $\mathrm{Gibbs}(\lambda)$ and converges in distribution to $\mathrm{Gibbs}(\lambda)$ [see Møller and Waagepetersen (2004), Propositions G.2–G.4]. Time-reversibility with respect to $\mathrm{Gibbs}(\lambda)$ means that if $Z(0) \sim \mathrm{Gibbs}(\lambda)$, we have that $(Z(t))_{t\in[0,T]}$ and $(Z(T-t))_{t\in[0,T]}$ have the same distribution for any $T > 0$. Since condition (S) holds, we may also characterize the SBDP with birth rate λ and unit per-capita death rate as the unique Markov process with infinitesimal generator

$$\begin{aligned} \mathcal{A}h(\xi) &= \int_{\mathcal{X}} [h(\xi + \delta_x) - h(\xi)]\lambda(x|\xi)\boldsymbol{\alpha}(dx) \\ &\quad + \int_{\mathcal{X}} [h(\xi - \delta_x) - h(\xi)]\xi(dx) \end{aligned} \tag{40}$$

for all bounded measurable $h\colon \mathcal{X} \to \mathbb{R}$; see Ethier and Kurtz (1986), Sections 4.2 and 4.11, Problem 5. The domain $\mathscr{D}(\mathcal{A})$ of $\mathcal{A}$ is the set of all measurable functions h for which the right-hand side above is well definied. $\mathscr{D}(\mathcal{A})$ contains at least all the functions h for which $\sup_{\xi\in\mathfrak{N}, x\in\mathcal{X}} |h(\xi + \delta_x) - h(\xi)|$ is finite.

In what follows we construct a coupling $(Z_\xi(t), Z_\eta(t))_{t\geq 0}$ of two SBDPs with identical birth rate λ started at individual configurations $\xi, \eta \in \mathfrak{N}$. We introduce the notation

$$\lambda_{\max}(x|\xi,\eta) = \max(\lambda(x|\xi), \lambda(x|\eta)), \qquad \bar{\lambda}_{\max}(\xi,\eta) = \int_{\mathcal{X}} \lambda_{\max}(x|\xi,\eta)\boldsymbol{\alpha}(dx),$$

$$\lambda_{\min}(x|\xi,\eta) = \min(\lambda(x|\xi), \lambda(x|\eta)), \qquad \bar{\lambda}_{\min}(\xi,\eta) = \int_{\mathcal{X}} \lambda_{\min}(x|\xi,\eta)\boldsymbol{\alpha}(dx).$$

Define (Z_ξ, Z_η) as a pure-jump Markov process with right-continuous paths, holding intervals $D_1, D_2, \ldots,$ start time $T_0 := 0$, and jump times $T_j := \sum_{i=1}^{j} D_i$ for all $j \geq 1$. Given $Z_\xi(T_{j-1}) = \xi', Z_\eta(T_{j-1}) = \eta'$ the distribution of the next jump is described by the following random variables, which are assumed to be independent of one another unless specified otherwise. Let

$$\begin{aligned}
D_j &\sim \mathrm{Exp}(\bar{\lambda}_{\max}(\xi', \eta') + |\xi' \cup \eta'|), \\
G_j &\sim \mathrm{Bernoulli}\biggl(\frac{\bar{\lambda}_{\max}(\xi', \eta')}{\bar{\lambda}_{\max}(\xi', \eta') + |\xi' \cup \eta'|}\biggr), \\
Y_j &\sim \frac{\lambda_{\max}(\cdot|\xi', \eta')}{\bar{\lambda}_{\max}(\xi', \eta')}, \\
U_j &\sim \mathrm{Unif}(\xi' \cup \eta'), \\
B_{\xi,j}|Y_j &\sim \mathrm{Bernoulli}\biggl(\frac{\lambda(Y_j|\xi')}{\lambda_{\max}(Y_j|\xi', \eta')}\biggr), \\
B_{\eta,j}|Y_j &\sim \mathrm{Bernoulli}\biggl(\frac{\lambda(Y_j|\eta')}{\lambda_{\max}(Y_j|\xi', \eta')}\biggr),
\end{aligned}$$

where $B_{\xi,j}$ and $B_{\eta,j}$ are maximally coupled given Y_j, that is,

$$\mathbb{P}(B_{\xi,j} = B_{\eta,j} = 1|Y_j) = \frac{\lambda_{\min}(Y_j|\xi', \eta')}{\lambda_{\max}(Y_j|\xi', \eta')}.$$

If $G_j = 1$, set $Z_\xi(T_j) = Z_\xi(T_{j-1}) + B_{\xi,j}\delta_{Y_j}$ and $Z_\eta(T_j) = Z_\eta(T_{j-1}) + B_{\eta,j}\delta_{Y_j}$. If $G_j = 0$, set $Z_\xi(T_j) = Z_\xi(T_{j-1}) - \mathbb{1}\{U_j \in Z_\xi(T_{j-1})\}\delta_{U_j}$ and $Z_\eta(T_j) = Z_\eta(T_{j-1}) - \mathbb{1}\{U_j \in Z_\eta(T_{j-1})\}\delta_{U_j}$.

In the special case where $\lambda(x|\xi)$ does not depend on the configuration ξ, that is, the time-reversible distribution is the Poisson distribution with intensity function λ, our construction reduces to the coupling used in Barbour and Brown (1992).

PROPOSITION 35. *Both components Z_ξ, Z_η of the coupling are SBDPs with generator (40).*

PROOF. Since by condition (S) the rate $a(\xi, \eta) = \bar{\lambda}_{\max}(\xi, \eta) + |\xi \cup \eta|$ is bounded by $c + |\xi \cup \eta|$, for a constant $c > 0$, we get

$$\begin{aligned}
\mathbb{P}(D_1 > t) &= 1 - a(\xi, \eta)t + O(t^2), \\
\mathbb{P}(D_1 \leq t, D_1 + D_2 > t) &= a(\xi, \eta)t + O(t^2)
\end{aligned}$$

and

$$\mathbb{P}(D_1 + D_2 \leq t) = O(t^2)$$

as $t \to 0$. Thus for a bounded function h

$$\begin{aligned}
&\mathbb{E}h(Z_\xi(t)) \\
&\quad = t(\bar{\lambda}_{\max}(\xi,\eta) + |\xi \cup \eta|) \\
&\qquad \times (\mathbb{E}h(\xi + B_{\xi,1}\delta_{Y_1})\mathbb{P}(G_1 = 1) + \mathbb{E}h(\xi - \mathbb{1}\{U_1 \in \xi\}\delta_{U_1})\mathbb{P}(G_1 = 0)) \\
&\qquad + (1 - t(\bar{\lambda}_{\max}(\xi,\eta) + |\xi \cup \eta|))h(\xi) + O(t^2).
\end{aligned}$$

Since

$$\mathbb{P}(G_1 = 1) = 1 - \mathbb{P}(G_1 = 0) = \frac{\bar{\lambda}_{\max}(\xi,\eta)}{\bar{\lambda}_{\max}(\xi,\eta) + |\xi \cup \eta|},$$

we obtain

$$\begin{aligned}
&\lim_{t\to 0} \frac{\mathbb{E}h(Z_\xi(t)) - h(\xi)}{t} \\
&\quad = \bar{\lambda}_{\max}(\xi,\eta)\mathbb{E}(h(\xi + B_{\xi,1}\delta_{Y_1}) - h(\xi)) \\
&\qquad + |\xi \cup \eta|\mathbb{E}(h(\xi - \mathbb{1}\{U_1 \in \xi\}\delta_{U_1}) - h(\xi)) \\
&\quad = \bar{\lambda}_{\max}(\xi,\eta) \int_{\mathcal{X}} [h(\xi + \delta_y) - h(\xi)] \frac{\lambda(y|\xi)}{\lambda_{\max}(y|\xi,\eta)} \frac{\lambda_{\max}(y|\xi,\eta)}{\bar{\lambda}_{\max}(\xi,\eta)} \boldsymbol{\alpha}(dy) \\
&\qquad + |\xi \cup \eta| \int_{\mathcal{X}} [h(\xi - \mathbb{1}\{u \in \xi\}\delta_u) - h(\xi)] \frac{(\xi \cup \eta)(du)}{|\xi \cup \eta|} \\
&\quad = \int_{\mathcal{X}} [h(\xi + \delta_y) - h(\xi)]\lambda(y|\xi)\boldsymbol{\alpha}(dy) + \int_{\mathcal{X}} [h(\xi - \delta_u) - h(\xi)]\xi(du) \\
&\quad = \mathcal{A}h(\xi). \qquad \square
\end{aligned}$$

Define the coupling time as $\tau = \tau_{\xi,\eta} = \inf\{t \geq 0 : Z_\xi(t) = Z_\eta(t)\}$. In order to investigate the coupling time, it is convenient to use the stopping times $\tau_0 = 0$, $\tau_k = \inf\{t > \tau_{k-1} : Z_\xi(t) - Z_\eta(t) \neq Z_\xi(\tau_{k-1}) - Z_\eta(\tau_{k-1})\}$. These times are the times when something interesting happens, that is, one of the noncommon points of Z_ξ and Z_η dies or there is a birth in just one of the processes.

Let us call the event that a noncommon point dies a "good death" and the event that only one process has a birth a "bad birth." Assume that there are n noncommon points in ξ and η, that is, $\|\xi - \eta\| = n$, where $\|\cdot\|$ denotes the total variation norm for signed measures. Define the event $A_n = \{\|Z_\xi(\tau_1) - Z_\eta(\tau_1)\| = n - 1\}$ and the filtration $\mathcal{F}_t = \sigma((Z_\xi(s), Z_\eta(s)); s \leq t)$. Note that by construction (Z_ξ, Z_η) has the strong Markov property; see, for example, Kallenberg (2002), Theorem 12.14. An easy calculation then gives us the following probabilities:

$$\begin{aligned}
&\mathbb{P}(\text{“good death” at time } T_j|\mathcal{F}_{T_{j-1}})\\
&\quad = \frac{\|Z_\xi(T_{j-1}) - Z_\eta(T_{j-1})\|}{\bar\lambda_{\max}(Z_\xi(T_{j-1}), Z_\eta(T_{j-1})) + |Z_\xi(T_{j-1}) \cup Z_\eta(T_{j-1})|},\\
&\mathbb{P}(\text{“bad birth” at time } T_j|\mathcal{F}_{T_{j-1}})\\
&\quad = \frac{\bar\lambda_{\max}(Z_\xi(T_{j-1}), Z_\eta(T_{j-1})) - \bar\lambda_{\min}(Z_\xi(T_{j-1}), Z_\eta(T_{j-1}))}{\bar\lambda_{\max}(Z_\xi(T_{j-1}), Z_\eta(T_{j-1})) + |Z_\xi(T_{j-1}) \cup Z_\eta(T_{j-1})|}.
\end{aligned}$$

LEMMA 36. *The probability of the event A_n is bounded from below as*

$$\mathbb{P}(A_n) \ge \left(1 + \frac{1}{n} \sup_{\|\xi'-\eta'\|=n} \int_{\mathcal{X}} |\lambda(x|\xi') - \lambda(x|\eta')| \boldsymbol{\alpha}(dx)\right)^{-1} > 0. \tag{41}$$

PROOF. We argue in terms of the discrete time Markov chains $(Z_\xi(T_j))_{j\in\mathbb{Z}_+}$ and $(Z_\eta(T_j))_{j\in\mathbb{Z}_+}$ and refer to them as the *jump chains* of our SBDPs. Define the $\mathbb{N}$-valued random variable J by $\tau_1(\omega) = T_{J(\omega)}(\omega)$ as the index of the first interesting jump. Then

$$\begin{aligned}
\mathbb{P}(A_n) &= \sum_{j=1}^{\infty} \int_{\mathfrak{N}^2} \mathbb{P}(A_n|J = j, Z_\xi(T_{j-1}) = \xi', Z_\eta(T_{j-1}) = \eta')\\
&\qquad\qquad \times \mathbb{P}(J = j, Z_\xi(T_{j-1}) \in d\xi', Z_\eta(T_{j-1}) \in d\eta')\\
&= \sum_{j=1}^{\infty} \int_{\mathfrak{N}^2} \mathbb{P}(\|Z_\xi(T_j) - Z_\eta(T_j)\| = n-1|\\
&\qquad\qquad \|Z_\xi(T_i) - Z_\eta(T_i)\| = n,\ \forall i \le j-1,\\
&\qquad\qquad \|Z_\xi(T_j) - Z_\eta(T_j)\| \in \{n-1, n+1\},\\
&\qquad\qquad Z_\xi(T_{j-1}) = \xi', Z_\eta(T_{j-1}) = \eta')\\
&\qquad\qquad \times \mathbb{P}(J = j, Z_\xi(T_{j-1}) \in d\xi', Z_\eta(T_{j-1}) \in d\eta')\\
&= \sum_{j=1}^{\infty} \int_{\mathfrak{N}^2} \frac{n}{n + (\bar\lambda_{\max}(\xi', \eta') - \bar\lambda_{\min}(\xi', \eta'))}\\
&\qquad\qquad \times \mathbb{P}(J = j, Z_\xi(T_{j-1}) \in d\xi', Z_\eta(T_{j-1}) \in d\eta')\\
&= \mathbb{E}\left(\frac{n}{n + (\bar\lambda_{\max}(Z_\xi(T_{J-1}), Z_\eta(T_{J-1})) - \bar\lambda_{\min}(Z_\xi(T_{J-1}), Z_\eta(T_{J-1})))}\right)\\
&\ge \frac{n}{n + \sup_{\|\xi'-\eta'\|=n}(\bar\lambda_{\max}(\xi', \eta') - \bar\lambda_{\min}(\xi', \eta'))}.
\end{aligned}$$

The claim follows since

$$\bar{\lambda}_{\max}(\xi,\eta) - \bar{\lambda}_{\min}(\xi,\eta) = \int_{\mathcal{X}} |\lambda(x|\xi) - \lambda(x|\eta)| \boldsymbol{\alpha}(dx),$$

which is uniformly bounded in ξ and η by condition (S). □

THEOREM 37. *For all configurations ξ,η the coupling time $\tau_{\xi,\eta}$ is integrable. In particular if ξ and η differ in only one point, we have for any $n^* \in \mathbb{N} \cup \{\infty\}$,*

$$\begin{aligned}\mathbb{E}\tau_{\xi,\eta} \leq (n^*-1)!\left(\frac{\varepsilon}{c}\right)^{n^*-1}\left(\frac{1}{c}\sum_{i=n^*}^{\infty}\frac{c^i}{i!} + \int_0^c \frac{1}{s}\sum_{i=n^*}^{\infty}\frac{s^i}{i!}\,ds\right) \\ + \frac{1+\varepsilon}{\varepsilon}\sum_{i=1}^{n^*-1}\frac{\varepsilon^i}{i},\end{aligned} \tag{42}$$

where

$$\varepsilon = \sup_{\|\xi-\eta\|=1} \int_{\mathcal{X}} |\lambda(x|\xi) - \lambda(x|\eta)| \boldsymbol{\alpha}(dx)$$

and

$$c = c(n^*) = \sup_{\|\xi-\eta\|\geq n^*} \int_{\mathcal{X}} |\lambda(x|\xi) - \lambda(x|\eta)| \boldsymbol{\alpha}(dx)$$

with the interpretations detailed in Theorem 4. The constants ε and c are finite by condition (S).

The following lemma treats the case $n^* = 1$ and will be useful for the proof of Theorem 37.

LEMMA 38. *For all configurations ξ,η the coupling time $\tau_{\xi,\eta}$ is integrable. In particular if ξ and η differ in only one point, we have*

$$\mathbb{E}\tau_{\xi,\eta} \leq \frac{e^c - 1}{c} + \int_0^c \frac{e^s - 1}{s}\,ds, \tag{43}$$

where $c = \sup_{\xi,\eta\in\mathfrak{N}} \int |\lambda(x|\xi) - \lambda(x|\eta)| \boldsymbol{\alpha}(dx)$, which is finite by (S).

PROOF. Let $p_n = (1+c/n)^{-1}$, $n \geq 1$. Construct a new pure-jump Markov process $(Y(t))_{t\geq 0}$ on $\mathbb{Z}_+$ by the following rule. Given $Y(t)$ is in state $n \in \mathbb{Z}_+$, after an exponentially distributed time with mean $1/n$, it jumps to $n-1$ with probability p_n and to $n+1$ with probability $1-p_n$. Define stopping times $\tilde{\tau}_n = \inf\{t \geq 0 : Y_n(t) = 0\}$ for all $n \geq 0$, where $(Y_n(t))_{t\geq 0}$ denotes the process started at n.

We show that $\tau_{\xi,\eta}$ is stochastically dominated by $\tilde{\tau}_n$, and therefore $\mathbb{E}\tau_{\xi,\eta} \leq \mathbb{E}\tilde{\tau}_n$. Denote by $X = \|Z_\xi - Z_\eta\|$ the process counting the noncommon points, and define a new jump process Y' by the following construction. Set $Y'(0) =$

$X(0)$. If X and Y' are on the same level, say n, they move together. If X jumps, then Y' jumps with probability p_n to $n-1$ and with probability $1-p_n$ to $n+1$. Since $\mathbb{P}(A_n) \geq p_n$, the jumps can be coupled such that Y' stays above X. If they are separated, let Y' behave like Y until they meet again. Thus we have $Y'(t) \geq X(t)$ for all $t \geq 0$, and hence $\tau_{\xi,\eta}$ is stochastically dominated by $\tilde{\tau}'_n = \inf\{t \geq 0 : Y'_n(t) = 0\}$, where $(Y'_n(t))_{t\geq 0}$ denotes the process Y' started at n.

Note that Y' has the same transition probabilities as Y, but its holding times are sometimes those of X. Since the processes Z_ξ and Z_η have unit per-capita death rates, each of their noncommon points dies independently after a standard exponentially distributed time. If there are n such points, the minimum of these times is exponentially distributed with mean $1/n$. Hence the holding times of X at n, and therefore also all of the holding times of Y' at n, may be coupled with exponentially distributed random variables with mean $1/n$ that are almost surely larger or equal. This yields a coupling of Y' and Y with exactly matched jump chains where Y is just a slower version of Y'. Hence $\tilde{\tau}'_n$ is stochastically dominated by $\tilde{\tau}_n$.

Define now $e_n = \mathbb{E}\tilde{\tau}_n$, $n \geq 0$. Then, by conditioning on the next jump in the Y-chain, we obtain

$$\begin{aligned} e_n &= \left(e_{n-1} + \frac{1}{n}\right)p_n + \left(e_{n+1} + \frac{1}{n}\right)(1-p_n) \\ &= e_{n-1}p_n + e_{n+1}(1-p_n) + \frac{1}{n}. \end{aligned} \tag{44}$$

If $c = 0$, we have $e_n = e_{n-1} + 1/n = \sum_{i=1}^n 1/i$, in particular $e_1 = 1$. For $c > 0$, define $a_n = e_n - e_{n-1}$, $n \geq 1$, assuming that the e_n are finite. Then $p_n = (1 + c/n)^{-1}$ yields

$$a_{n+1} = \frac{n}{c}a_n - \frac{1}{c} - \frac{1}{n} \qquad \text{for all } n \geq 1. \tag{45}$$

Since $e_0 = 0$, the starting point is $a_1 = e_1$. The general solution of (45) is given by

$$a_n = \sum_{i=0}^{\infty} \frac{c^i}{\prod_{k=0}^{i}(n+k)}\left(1 + \frac{c}{n+i}\right) + C\frac{(n-1)!}{c^{n-1}} \tag{46}$$

for an arbitrary constant $C \in \mathbb{R}$.

A result in Grimmett and Stirzaker (2001) [Exercise 6, page 265] states that the sequence of expected return times is the smallest nonnegative solution of (44), which yields that the e_n are in fact finite and given by setting $C = 0$. We obtain for all $n \geq 1$ that

$$e_n = \sum_{i=1}^{n} a_n \leq n a_1 = n e_1 \tag{47}$$

and

$$e_1 = a_1 = \sum_{i=0}^{\infty} \frac{c^i}{(1+i)!}\left(1 + \frac{c}{1+i}\right) = \frac{e^c - 1}{c} + \int_0^c \frac{e^s - 1}{s}\, ds.$$

Note that e_1 converges to 1 for c going to 0. □

Proof of Theorem 37. Let ξ and η be point configurations differing in n points, that is, they can be written as

$$\xi = \zeta + \sum_{i=1}^{k} \delta_{y_i} \quad \text{and} \quad \eta = \zeta + \sum_{i=k+1}^{n} \delta_{y_i},$$

where $0 \le k \le n$, $y_1, \dots, y_n$ are the noncommon points of ξ and η, and $\zeta = \xi \cap \eta$. Then

$$\begin{aligned} |\lambda(x|\xi) - \lambda(x|\eta)| = \Bigg| &\sum_{j=1}^{k} \left[\lambda\left(x \,\middle|\, \zeta + \sum_{i=1}^{j} \delta_{y_i}\right) - \lambda\left(x \,\middle|\, \zeta + \sum_{i=1}^{j-1} \delta_{y_i}\right)\right] \\ &- \sum_{j=k+1}^{n} \left[\lambda\left(x \,\middle|\, \zeta + \sum_{i=k+1}^{j} \delta_{y_i}\right) - \lambda\left(x \,\middle|\, \zeta + \sum_{i=k+1}^{j-1} \delta_{y_i}\right)\right]\Bigg|. \end{aligned}$$

By the triangle inequality we obtain

$$\begin{aligned} &\sup_{\|\xi-\eta\|=n} \int_{\mathcal{X}} |\lambda(x|\xi) - \lambda(x|\eta)| \boldsymbol{\alpha}(dx) \\ &\qquad \le n \sup_{\|\xi-\eta\|=1} \int_{\mathcal{X}} |\lambda(x|\xi) - \lambda(x|\eta)| \boldsymbol{\alpha}(dx) = n\varepsilon. \end{aligned}$$

Thus by Lemma 36 we have $\mathbb{P}(A_n) \ge (1+\varepsilon)^{-1}$ for $n \ge 1$ and $\mathbb{P}(A_n) \ge (1 + c/n)^{-1}$ for $n \ge n^*$. Assume $\varepsilon > 0$, $c > 0$ and $n^* \in \mathbb{N}$. Replace the jump-down probabilities p_n of Y in the proof of Lemma 38 by the above bounds. We then obtain for the differences $a_n = e_n - e_{n-1}$ of the expected return times to zero the recursion equations

$$\begin{cases} a_{n+1} = \dfrac{1}{\varepsilon} a_n - \dfrac{1+\varepsilon}{n\varepsilon}, & \text{for } 1 \le n < n^*, \\ a_{n+1} = \dfrac{n}{c} a_n - \dfrac{1}{c} - \dfrac{1}{n}, & \text{for } n \ge n^*. \end{cases} \tag{48}$$

The differences for larger n must still be the same. Hence the proof of Lemma 38 gives

$$a_n = \sum_{i=0}^{\infty} \frac{c^i}{\prod_{k=0}^{i}(n+k)}\left(1 + \frac{c}{n+i}\right)$$

for all $n \geq n^*$. The second recursion equation in (48) is best solved backwards, which yields the solution

$$a_{n^*-k} = \varepsilon^k a_{n^*} + (1+\varepsilon)\sum_{i=1}^{k} \frac{\varepsilon^{k-i}}{n^*-i}$$

for all $1 \leq k \leq n^* - 1$. Thus

$$e_1 = a_1 = \varepsilon^{n^*-1}\sum_{i=0}^{\infty} \frac{c^i}{\prod_{k=0}^{i}(n^*+k)}\left(1+\frac{c}{n^*+i}\right) + \frac{1+\varepsilon}{\varepsilon}\sum_{i=1}^{n^*-1}\frac{\varepsilon^{n^*-i}}{n^*-i}, \tag{49}$$

which can be rewritten as (42).

Letting $n^* \to \infty$ in the inequality $\mathbb{E}\tau_{\xi,\eta} \leq e_1$, we obtain $\mathbb{E}\tau_{\xi,\eta} \leq \frac{1+\varepsilon}{\varepsilon}\sum_{i=1}^{\infty}\frac{\varepsilon^i}{i}$ if $\varepsilon < 1$ irrespective of c, which justifies setting the long first summand in (42) to zero. Analogously as in the proof of Lemma 38 we get for $\varepsilon = 0$ (which implies $c = 0$) that $e_1 = 1$, and for $\varepsilon > 0, c = 0$ that $a_{n^*} = 1/n^*$, which justifies the interpretation of the upper bound in (42) in the limit sense.

If c does not depend on n^*, we choose n^* such that $(1+\varepsilon)^{-1} > (1+c/n)^{-1}$ for $n < n^*$ and $(1+\varepsilon)^{-1} \leq (1+c/n)^{-1}$ for $n \geq n^*$. This is obviously the optimal choice and leads to $n^* = \lceil c/\varepsilon \rceil$. $\square$

One can also couple SBDPs with random starting configurations. It is convenient to use notation of the form Z_Ξ also if Ξ is a point process with the obvious meaning that $(Z_\Xi(t))_{t\geq 0}$ is an SBDP with generator (40) and $Z(0) = \Xi$ almost surely. Since this may lead to confusing notation when dealing with two processes, we always distinguish the processes by adding a prime, thus writing $(Z_\Xi(t))$ and $(Z'_{\widetilde{\Xi}}(t))$ for point processes Ξ and $\widetilde{\Xi}$.

PROPOSITION 39. *Assume that Ξ and $\widetilde{\Xi}$ are Gibbs processes satisfying* (S). *Consider the coupling $(Z_\Xi(t), Z'_{\widetilde{\Xi}}(t))$, $t \geq 0$. Then the coupling time $\tau_{\Xi,\widetilde{\Xi}} = \inf\{t \geq 0 : Z_\Xi(t) = Z'_{\widetilde{\Xi}}(t)\}$ is integrable.*

PROOF. The Georgii–Nguyen–Zessin equation and condition (S) yield

$$\mathbb{E}|\Xi| = \mathbb{E}\int_{\mathcal{X}} 1\Xi(dx) = \mathbb{E}\int_{\mathcal{X}} \nu(x|\Xi)\boldsymbol{\alpha}(dx) < \infty,$$

where ν is the conditional intensity of Ξ, and analogously $\mathbb{E}|\widetilde{\Xi}| < \infty$. Then (47) implies

$$\begin{aligned}\mathbb{E}\tau_{\Xi,\widetilde{\Xi}} &= \mathbb{E}(\mathbb{E}(\tau_{\Xi,\widetilde{\Xi}}||\Xi|,|\widetilde{\Xi}|)) \\ &\leq \mathbb{E}(\mathbb{E}(e_{|\Xi|+|\widetilde{\Xi}|}||\Xi|,|\widetilde{\Xi}|)) \leq e_1\mathbb{E}(|\Xi|+|\widetilde{\Xi}|) < \infty. \end{aligned}$$ $\square$

In particular the coupling time $\tau_{\xi,\mathrm{H}}$ for (nonrandom) $\xi \in \mathfrak{N}$ and H is integrable.

6. Stein's method for Gibbs process approximation. Stein's method, originally conceived for normal approximation [Stein (1972)], has evolved over the last forty years to become an important tool in many areas of probability theory and for a wide range of approximating distributions. See Barbour and Chen (2005) for an overview.

A milestone in the evolution of Stein's method was the discovery in Barbour (1988) that a natural Stein equation may often be set up by choosing as a right-hand side the infinitesimal generator of a Markov process whose stationary distribution is the approximating distribution of interest. Many important developments stem from this so-called *generator approach* to Stein's method, and several of them concern point process approximation, such as Barbour and Brown (1992), Barbour and Månsson (2002), Schuhmacher (2009) or Xia and Zhang (2012).

In this section we develop the generator approach for Gibbs process approximation. Let $\mathrm{H} \sim \mathrm{Gibbs}(\lambda)$ be our approximating Gibbs processes satisfying (S). Define the generator

$$\mathcal{A}h(\xi) = \int_{\mathcal{X}} [h(\xi + \delta_x) - h(\xi)] \lambda(x|\xi) \boldsymbol{\alpha}(dx) + \int_{\mathcal{X}} [h(\xi - \delta_x) - h(\xi)] \xi(dx) \tag{50}$$

for all $h: \mathfrak{N} \to \mathbb{R}$ in its domain $\mathscr{D}(\mathcal{A})$. In Section 5 we noted that $\mathcal{A}$ is the generator of a spatial birth–death process Z with stationary distribution $\mathrm{Gibbs}(\lambda)$, and that its domain contains at least all functions h with bounded first differences.

For any measurable $f: \mathfrak{N} \to \mathbb{R}$ set up the so-called *Stein equation* formally as

$$f(\xi) - \mathbb{E}f(\mathrm{H}) = \mathcal{A}h(\xi). \tag{51}$$

A first goal is to find a function $h = h_f$ that satisfies this equation. By analogy to the Poisson process case, a natural candidate is given by

$$h_f(\xi) = -\int_0^\infty [\mathbb{E}f(Z_\xi(t)) - \mathbb{E}f(\mathrm{H})]\, dt. \tag{52}$$

The following lemma shows that h_f is indeed a solution to equation (51) if $f \in \mathcal{F}_{\mathrm{TV}}$.

LEMMA 40. *Assume that f is bounded and* H *satisfies* (S). *Then h_f is well defined, that is, the integral exists for all $\xi \in \mathfrak{N}$, and it is a solution to (51).*

PROOF. We use the coupling (Z_ξ, Z'_{H}) from Section 5, where Z'_{H} is started in the random configuration H, as explained at the end of that section. Since $\mathscr{L}(\mathrm{H})$ is the stationary measure of the SBDP Z, we have

$\mathbb{E}f(\mathrm{H}) = \mathbb{E}f(Z'_{\mathrm{H}}(t))$ for all $t \geq 0$. Thus

$$\begin{aligned}\int_0^\infty |\mathbb{E}f(Z_\xi(t)) - \mathbb{E}f(\mathrm{H})|\,dt &= \int_0^\infty \mathbb{E}|f(Z_\xi(t)) - f(Z'_{\mathrm{H}}(t))|\mathbb{1}\{\tau_{\xi,\mathrm{H}} > t\}\,dt \\ &\leq 2\|f\|_\infty \int_0^\infty \mathbb{P}(\tau_{\xi,\mathrm{H}} > t)\,dt \\ &= 2\|f\|_\infty \mathbb{E}(\tau_{\xi,\mathrm{H}}) < \infty\end{aligned}$$

by Proposition 39. Hence h_f is well defined. The Markov property of the SBDP implies $\mathscr{L}(Z'_{Z_\xi(s)}(t)) = \mathscr{L}(Z_\xi(t+s))$. Thus by the substitution $v = t+s$

$$\begin{aligned}&\frac{1}{s}(\mathbb{E}h_f(Z_\xi(s)) - h_f(\xi)) \\ &\quad = \frac{1}{s}\left(-\int_0^\infty \mathbb{E}[f(Z_\xi(t+s)) - f(\mathrm{H})]\,dt + \int_0^\infty \mathbb{E}[f(Z_\xi(t)) - f(\mathrm{H})]\,dt\right) \\ &\quad = \frac{1}{s}\int_0^s \mathbb{E}[f(Z_\xi(v)) - f(\mathrm{H})]\,dv.\end{aligned}$$

By condition (S) we have $\mathbb{P}(Z_\xi(v) \neq \xi) = O(v)$ and since f is bounded, $\mathbb{E}f(Z_\xi(v)) = f(\xi) + O(v)$, which implies

$$\frac{1}{s}\int_0^s \mathbb{E}f(Z_\xi(v))\,dv = f(\xi) + O(s).$$

Thus

$$\mathcal{A}h_f(\xi) = \lim_{s\to 0}\frac{1}{s}(\mathbb{E}h_f(Z_\xi(s)) - h_f(\xi)) = f(\xi) - \mathbb{E}f(\mathrm{H}). \qquad \square$$

Define then the *Stein factor* as

$$c_1(\lambda) = \sup_{f\in\mathcal{F}_{\mathrm{TV}}} \sup_{x\in\mathcal{X},\xi\in\mathfrak{N}} |h_f(\xi+\delta_x) - h_f(\xi)|. \tag{53}$$

We are now ready to give proofs for the results in Section 3.1.

Proof of Theorem 4. By the Stein equation (51)

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) = \sup_{f\in\mathcal{F}_{\mathrm{TV}}} |\mathbb{E}f(\Xi) - \mathbb{E}f(\mathrm{H})| = \sup_{f\in\mathcal{F}_{\mathrm{TV}}} |\mathbb{E}\mathcal{A}h_f(\Xi)|,$$

where h_f is the Stein solution given in (52). Then the Georgii–Nguyen–Zessin equation (5) yields

$$\begin{aligned}&|\mathbb{E}\mathcal{A}h_f(\Xi)| \\ &\quad = \left|\mathbb{E}\int_{\mathcal{X}} [h_f(\Xi+\delta_x) - h_f(\Xi)]\lambda(x|\Xi)\boldsymbol{\alpha}(dx)\right.\end{aligned}$$

$$
\begin{aligned}
&\qquad + \mathbb{E}\int_{\mathcal{X}} [h_f(\Xi - \delta_x) - h_f(\Xi)]\Xi(dx)\bigg| \\
&= \bigg|\mathbb{E}\int_{\mathcal{X}} [h_f(\Xi + \delta_x) - h_f(\Xi)](\lambda(x|\Xi) - \nu(x|\Xi))\boldsymbol{\alpha}(dx)\bigg| \qquad (54)\\
&\le \sup_{\xi,\eta\in\mathfrak{N}, \|\xi-\eta\|=1} |h_f(\xi) - h_f(\eta)| \int_{\mathcal{X}} \mathbb{E}|\nu(x|\Xi) - \lambda(x|\Xi)|\boldsymbol{\alpha}(dx) \\
&\le c_1(\lambda)\int_{\mathcal{X}} \mathbb{E}|\nu(x|\Xi) - \lambda(x|\Xi)|\boldsymbol{\alpha}(dx).
\end{aligned}
$$

Consider the coupling $(Z_{\xi+\delta_x}(t), Z_\xi(t))_{t\ge 0}$ in the sense of Section 5. Then by equation (52)

$$
\begin{aligned}
|h_f(\xi+\delta_x) - h_f(\xi)| &= \bigg|\mathbb{E}\int_0^\infty [f(Z_{\xi+\delta_x}(t)) - f(Z_\xi(t))]\mathbb{1}\{\tau_{\xi+\delta_x,\xi} > t\}\,dt\bigg| \\
&\le \sup_{\xi,\eta\in\mathfrak{N}} |f(\xi) - f(\eta)| \int_0^\infty \mathbb{P}(\tau_{\xi+\delta_x,\xi} > t)\,dt \le \mathbb{E}\tau_{\xi+\delta_x,\xi},
\end{aligned}
$$

where we used that $0 \le f \le 1$. The upper bound on $c_1(\lambda)$ follows now from Theorem 37. $\square$

REMARK 41. In the proof above the $\sup_{\xi,\eta\in\mathfrak{N},\|\xi-\eta\|=1}$ could actually be replaced by an essential supremum with respect to $\mathscr{L}(\Xi) + \mathscr{L}(\mathrm{H})$. This can be seen as follows. Let $N \in \mathcal{N}$ be a null set with respect to both $\mathscr{L}(\Xi)$ and $\mathscr{L}(\mathrm{H})$. Without loss of generality we may set the densities of Ξ and H to zero on N. By the hereditarity of the densities we have

$$
\begin{aligned}
&[h_f(\Xi + \delta_x) - h_f(\Xi)](\lambda(x|\Xi) - \nu(x|\Xi)) \\
&\qquad = [h_f(\Xi + \delta_x) - h_f(\Xi)]\mathbb{1}\{\Xi + \delta_x \in N^c, \Xi \in N^c\}(\lambda(x|\Xi) - \nu(x|\Xi)),
\end{aligned}
$$

whence it follows that the first inequality of (54) holds also for the essential supremum.

A consequence of this replacement is that it suffices to take the essential supremum with respect to $\mathscr{L}(\Xi) + \mathscr{L}(\mathrm{H})$ for the computation of the constants ε and c in Theorem 4.

PROOF OF THEOREM 8. Let ν and λ denote the conditional intensities of Ξ and H. Then for $\xi = \sum_{i=1}^n \delta_{y_i} \in \mathfrak{N}$ we have

$$
\begin{aligned}
&\nu(x|\xi) - \lambda(x|\xi) \\
&\qquad = \beta(x)\bigg(\prod_{y\in\xi}\varphi_1(x,y) - \prod_{y\in\xi}\varphi_2(x,y)\bigg)
\end{aligned}
$$

$$= \beta(x) \sum_{j=1}^{n} \left(\left(\prod_{i=1}^{j} \varphi_1(x, y_i) \right) \left(\prod_{i=j+1}^{n} \varphi_2(x, y_i) \right) - \left(\prod_{i=1}^{j-1} \varphi_1(x, y_i) \right) \left(\prod_{i=j}^{n} \varphi_2(x, y_i) \right) \right)$$

$$= \beta(x) \sum_{j=1}^{n} \left((\varphi_1(x, y_j) - \varphi_2(x, y_j)) \left(\prod_{i=1}^{j-1} \varphi_1(x, y_i) \right) \left(\prod_{i=j+1}^{n} \varphi_2(x, y_i) \right) \right).$$

Therefore it follows by Theorem 4, $\varphi_i \leq 1$ for $i = 1, 2$ and Campbell's formula [see Daley and Vere-Jones (2008), Section 9.5] that

$$d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) \leq c_1(\lambda) \mathbb{E}\left(\int_{\mathcal{X}} \int_{\mathcal{X}} \beta(x) |\varphi_1(x, y) - \varphi_2(x, y)| \boldsymbol{\alpha}(dx) \Xi(dy) \right)$$

$$= c_1(\lambda) \int_{\mathcal{X}} \int_{\mathcal{X}} \beta(x) \nu(y) |\varphi_1(x, y) - \varphi_2(x, y)| \boldsymbol{\alpha}(dx) \boldsymbol{\alpha}(dy).$$

Regarding the bound on $c_1(\lambda)$ we have for all $x, y \in \mathcal{X}$ and $\xi \in \mathfrak{N}$ that $\lambda(x|\xi + \delta_y) = \lambda(x|\xi)\varphi_2(x, y)$ and $\lambda(x|\xi) \leq \beta(x)$, with equality if $\xi = \varnothing$. Thus by Theorem 37,

$$\begin{aligned} \varepsilon &= \sup_{\|\xi - \eta\| = 1} \int_{\mathcal{X}} |\lambda(x|\xi) - \lambda(x|\eta)| \boldsymbol{\alpha}(dx) \\ &= \sup_{y \in \mathcal{X}, \xi \in \mathfrak{N}} \int_{\mathcal{X}} |\lambda(x|\xi + \delta_y) - \lambda(x|\xi)| \boldsymbol{\alpha}(dx) \\ &= \sup_{y \in \mathcal{X}, \xi \in \mathfrak{N}} \int_{\mathcal{X}} \lambda(x|\xi) |\varphi_2(x, y) - 1| \boldsymbol{\alpha}(dx) \\ &= \sup_{y \in \mathcal{X}} \int_{\mathcal{X}} \beta(x)(1 - \varphi_2(x, y)) \boldsymbol{\alpha}(dx). \end{aligned}$$

Furthermore, $|\lambda(x|\xi) - \lambda(x|\eta)| \leq \beta(x)$ for all $x \in \mathcal{X}$ and for all $\xi, \eta \in \mathfrak{N}$. Thus $c \leq \int_{\mathcal{X}} \beta(x) \boldsymbol{\alpha}(dx)$. $\square$

APPENDIX: THE CASE OF A NONDIFFUSE REFERENCE MEASURE α

In the main part we restrict ourselves to a diffuse reference measure $\boldsymbol{\alpha}$; see Section 2.3. This appendix shows that our results remain true for general $\boldsymbol{\alpha}$ (always finite).

Suppose that $\boldsymbol{\alpha}$ is not diffuse. Consider then instead of $\mathcal{X}$ the extended space $\widetilde{\mathcal{X}} = \mathcal{X} \times [0, 1]$ and equip it with the pseudometric $\tilde{d}((x, u), (y, u)) =$

$d(x,y)$. In the main part the metric d serves the double purpose of inducing the σ-algebra on $\mathcal{X}$ and allowing us to define balls when it comes to more detailed considerations of conditional intensities. For the first purpose, which does not require an explicit metric, we just use the product topology on $\widetilde{\mathcal{X}}$; for the second purpose we use the pseudometric $\tilde{d}$. Note that the topology induced by $\tilde{d}$ is coarser than the product topology, so there are no measurability problems. Define $\tilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha} \otimes \mathrm{Leb}|_{[0,1]}$, and denote by $\widetilde{\mathrm{Po}}_1$ the distribution of the Poisson process with intensity measure $\tilde{\boldsymbol{\alpha}}$. Note that $\tilde{\boldsymbol{\alpha}}$ is always a diffuse measure, and that a $\widetilde{\mathrm{Po}}_1$-process is simple, that is, almost surely free of multi-points.

Transform a point process $\Xi = \sum_{i=1}^N \delta_{X_i}$ on $\mathcal{X}$ into a point process $\widetilde{\Xi}$ on $\widetilde{\mathcal{X}}$ by randomizing its points in the new coordinate. More precisely let $\widetilde{\Xi} = \sum_{i=1}^N \delta_{(X_i,U_i)}$, where $U_1, U_2, \ldots$ are i.i.d. $\mathrm{Leb}|_{[0,1]}$-distributed random variables that are independent of Ξ. Following Kallenberg (1986), we refer to $\widetilde{\Xi}$ as the *uniform randomization* of Ξ. Writing $\widetilde{\mathfrak{N}}$ for the space of finite counting measures on $\widetilde{\mathcal{X}}$, we introduce the projection $\pi_{\mathcal{X}}\colon \widetilde{\mathfrak{N}} \to \mathfrak{N}$, $\pi_{\mathcal{X}}(\sum_{i=1}^n \delta_{(x_i,u_i)}) = \sum_{i=1}^n \delta_{x_i}$. Note that the image measure of $\widetilde{\mathrm{Po}}_1$ under $\pi_{\mathcal{X}}$ is $\widetilde{\mathrm{Po}}_1 \pi_{\mathcal{X}}^{-1} = \mathrm{Po}_1$. If Ξ is a Gibbs process with density u with respect to Po_1, then a short calculation shows that $\widetilde{\Xi}$ is a Gibbs process with density

$$u_{\widetilde{\Xi}}(\tilde{\xi}) = u(\pi_{\mathcal{X}}(\tilde{\xi})) \qquad \text{for any } \tilde{\xi} \in \widetilde{\mathfrak{N}}$$

with respect to $\widetilde{\mathrm{Po}}_1$.

Hence the conditional intensity of $\widetilde{\Xi}$ is given by

$$\tilde{\lambda}((x,u)|\tilde{\xi}) = \frac{u(\pi_{\mathcal{X}}(\tilde{\xi}) + \delta_x)}{u(\pi_{\mathcal{X}}(\tilde{\xi}))} = \lambda(x|\pi_{\mathcal{X}}(\tilde{\xi})).$$

Let $\widetilde{\mathcal{F}}_{\mathrm{TV}}$ be the class of measurable functions $f\colon \widetilde{\mathfrak{N}} \to [0,1]$ for which $f(\tilde{\xi}) = f(\tilde{\eta})$ whenever $\pi_{\mathcal{X}}(\tilde{\xi}) = \pi_{\mathcal{X}}(\tilde{\eta})$. Furthermore, let $\widetilde{\mathcal{A}}$ be the generator of an SBDP on $\widetilde{\mathfrak{N}}$ with birth rate $\tilde{\lambda}(\cdot|\cdot)$ and unit per-capita death rate. Note that

$$\tilde{c}_1(\tilde{\lambda}) = \sup_{f \in \widetilde{\mathcal{F}}_{\mathrm{TV}}} \sup_{\tilde{x} \in \widetilde{\mathcal{X}}, \tilde{\xi} \in \widetilde{\mathfrak{N}}} |h_f(\tilde{\xi} + \delta_{\tilde{x}}) - h_f(\tilde{\xi})| = c_1(\lambda).$$

Then for any two-point processes Ξ and H on $\mathcal{X}$ we can show by slightly adapting the proof of Theorem 4 that

$$\begin{aligned} d_{\mathrm{TV}}(\mathscr{L}(\Xi), \mathscr{L}(\mathrm{H})) &= \sup_{f \in \widetilde{\mathcal{F}}_{\mathrm{TV}}} |\mathbb{E} f(\widetilde{\Xi}) - \mathbb{E} f(\widetilde{\mathrm{H}})| \\ &= \sup_{f \in \widetilde{\mathcal{F}}_{\mathrm{TV}}} |\mathbb{E} \widetilde{\mathcal{A}} h_f(\widetilde{\Xi})| \\ &\le \tilde{c}_1(\tilde{\lambda}) \int_{\widetilde{\mathcal{X}}} \mathbb{E} |\tilde{\nu}((x,u)|\widetilde{\Xi}) - \tilde{\lambda}((x,u)|\widetilde{\Xi})| \tilde{\boldsymbol{\alpha}}(d(x,u)) \end{aligned}$$

$$= \tilde{c}_1(\tilde{\lambda}) \int_{\mathcal{X}} \int_0^1 \mathbb{E}|\nu(x|\pi_{\mathcal{X}}(\widetilde{\Xi})) - \lambda(x|\pi_{\mathcal{X}}(\widetilde{\Xi}))| \, du \boldsymbol{\alpha}(dx)$$

$$= c_1(\lambda) \int_{\mathcal{X}} \mathbb{E}|\nu(x|\Xi) - \lambda(x|\Xi)| \boldsymbol{\alpha}(dx),$$

where for the last equality the two expectations are the same by the transformation theorem.

The upper bound for $c_1(\lambda) = \tilde{c}_1(\tilde{\lambda})$ in inequality (10) can be obtained analogously as before by bounding the expected coupling time between two SBDPs with generator $\widetilde{\mathcal{A}}$ whose starting configuration differs in only one point. The expected coupling time will not be larger than before, because we may match the additional components $u \in [0, 1]$ of any new born points perfectly in the two processes.

For the more general statement in Remark 7 to hold, we have to replace condition (Σ) by condition (Σ'); see Kallenberg (1986), Section 13.2.

For the other results in Section 3, which are essentially corollaries of Theorem 4, it is easy to verify that we did not use the fact that α is diffuse, in particular not that $\xi, \eta \in \mathfrak{N}$ are multi-point free, except for notational purposes.

Acknowledgements. We thank the referees for their pertinent comments, which have led to an improvement of this article.

REFERENCES

BADDELEY, A. J. and VAN LIESHOUT, M. N. M. (1995). Area-interaction point processes. *Ann. Inst. Statist. Math.* **47** 601–619. MR1370279

BARBOUR, A. D. (1988). Stein's method and Poisson process convergence. *J. Appl. Probab.* **25A** 175–184. MR0974580

BARBOUR, A. D. and BROWN, T. C. (1992). Stein's method and point process approximation. *Stochastic Process. Appl.* **43** 9–31. MR1190904

BARBOUR, A. D. and CHEN, L. H. Y., eds. (2005). *An Introduction to Stein's Method*. Singapore Univ. Press, Singapore. MR2235447

BARBOUR, A. D. and MÅNSSON, M. (2002). Compound Poisson process approximation. *Ann. Probab.* **30** 1492–1537. MR1920275

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236. MR0373208

BESAG, J., MILNE, R. and ZACHARY, S. (1982). Point process limits of lattice processes. *J. Appl. Probab.* **19** 210–216. MR0644433

BROWN, T. C. and GREIG, D. (1994). Correction to: "Stein's method and point process approximation" [*Stochastic Process. Appl.* **43** (1992), no. 1, 9–31; MR1190904 (93k:60120)] by A. D. Barbour and Brown. *Stochastic Process. Appl.* **54** 291–296. MR1307341

CHEN, L. H. Y. and XIA, A. (2004). Stein's method, Palm theory and Poisson process approximation. *Ann. Probab.* **32** 2545–2569. MR2078550

DAI PRA, P. and POSTA, G. (2013). Entropy decay for interacting systems via the Bochner–Bakry–Émery approach. *Electron. J. Probab.* **18** no. 52, 21. MR3065862

Daley, D. J. and Vere-Jones, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II*, 2nd ed. Springer, New York. MR2371524

Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes*. Wiley, New York. MR0838085

Grimmett, G. R. and Stirzaker, D. R. (2001). *Probability and Random Processes*, 3rd ed. Oxford Univ. Press, New York. MR2059709

Kallenberg, O. (1986). *Random Measures*, 4th ed. Akademie-Verlag, Berlin. MR0854102

Kallenberg, O. (2002). *Foundations of Modern Probability: Probability and Its Applications*, 2nd ed. Springer, New York. MR1876169

Kondratiev, Y. and Lytvynov, E. (2005). Glauber dynamics of continuous particle systems. *Ann. Inst. Henri Poincaré Probab. Stat.* **41** 685–702. MR2144229

Kondratiev, Y., Pasurek, T. and Röckner, M. (2012). Gibbs measures of continuous systems: An analytic approach. *Rev. Math. Phys.* **24** 1250026, 54. MR2999411

Møller, J. (1989). On the rate of convergence of spatial birth-and-death processes. *Ann. Inst. Statist. Math.* **41** 565–581. MR1032600

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman & Hall/CRC, Boca Raton, FL. MR2004226

Preston, C. (1975). Spatial birth-and-death processes. In *Proceedings of the 40th Session of the International Statistical Institute (Warsaw, 1975)* 371–391. MR0474532

Ruelle, D. (1969). *Statistical Mechanics: Rigorous Results*. W. A. Benjamin, New York. MR0289084

Schuhmacher, D. (2009). Distance estimates for dependent thinnings of point processes with densities. *Electron. J. Probab.* **14** 1080–1116. MR2506126

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, CA, 1970/1971), Vol. II: Probability Theory* 583–602. Univ. California Press, Berkeley, CA. MR0402873

Stucki, K. and Schuhmacher, D. (2014). Bounds for the probability generating functional of a Gibbs point process. *Adv. in Appl. Probab.* **46** 21–34. MR3189046

Xia, A. and Zhang, F. (2012). On the asymptotics of locally dependent point processes. *Stochastic Process. Appl.* **122** 3033–3065. MR2946435

Institute for Mathematical Stochastics
Georg-August-Universität Göttingen
Goldschmidtstrasse 7
37077 Göttingen
Germany
E-mail: dominic.schuhmacher@mathematik.uni-goettingen.de
kaspar.stucki@mathematik.uni-goettingen.de