Communications on Stochastic Analysis
Vol. 4, No. 3 (2010) 377-399

Serials Publications
www.serialspublications.com

UPPER BOUNDS ON RUBINSTEIN DISTANCES ON CONFIGURATION SPACES AND APPLICATIONS

LAURENT DECREUSEFOND, ALDÉRIC JOULIN, AND NICOLAS SAVY

Abstract. In this paper, we provide upper bounds on several Rubinstein-type distances on the configuration space equipped with the Poisson measure. Our inequalities involve the two well-known gradients, in the sense of Malliavin calculus, which can be defined on this space. Actually, we show that depending on the distance between configurations which is considered, it is one gradient or the other which is the most effective. Some applications to distance estimates between Poisson and other more sophisticated processes are also provided, and an application of our results to tail and isoperimetric estimates completes this work.

1. Introduction

Let Λ be a σ-compact metric space and Γ_Λ be the space of configurations on Λ equipped with a Poisson measure μ. Defining and evaluating some distances between probability measures on Γ_Λ is an important problem, both theoretical and for applications, since it is equivalent to defining distances between point processes (see for instance Chapters 2 and 3 of [17] for a thorough discussion and references about this topic). Among the large class of distances one may consider, the one we want to study relies on an optimal transportation problem. Letting ρ be a lower semi-continuous distance on Γ_Λ and two configurations $\omega, \eta \in \Gamma_\Lambda$, we understand the quantity $\rho(\omega,\eta)$ as the cost for transporting one unit of mass from ω to η. Hence the optimal transportation cost between μ and some probability measure ν on Γ_Λ is given by

$$\mathcal{T}_\rho(\mu,\nu) = \inf_{\gamma\in\Sigma(\mu,\nu)} \int_{\Gamma_\Lambda}\int_{\Gamma_\Lambda} \rho(\omega,\eta)\ \mathrm{d}\gamma(\omega,\eta),$$

where $\Sigma(\mu,\nu)$ is the set of probability measures on $\Gamma_\Lambda \times \Gamma_\Lambda$ with marginals μ and ν. Such a quantity is called the Rubinstein distance between μ et ν. Being defined by a variational formula, its explicit expression is of difficult access in general but might be estimated from above: the construction of any coupling between μ and ν yields a bound on the Rubinstein distance between μ and ν. In particular, a convenient upper bound ensures its finiteness, which is not guaranteed a priori.

Received 2009-9-10; Communicated by N. Privault.
2000 *Mathematics Subject Classification.* 60G55,60H07,60E15.
Key words and phrases. Configuration space, Poisson measure, Rubinstein distance, Malliavin derivative, Rademacher property, tail estimate, isoperimetry.

Another interesting property of $\mathcal{T}_\rho$ is its rich duality. More precisely, the Kantorovich-Rubinstein duality allows us to rewrite the Rubinstein distance as

$$\mathcal{T}_\rho(\mu,\nu) = \sup_{F\in\rho-\mathrm{Lip}_1} \int_{\Gamma_\Lambda} F \ \mathrm{d}(\mu-\nu),$$

where $\rho - \mathrm{Lip}_1$ denotes the set of 1-Lipschitz functions on Γ_Λ with respect to the distance ρ. This means that $\mathcal{T}_\rho$ depends crucially on the distance on the configuration space as it changes the set of Lipschitz functions, hence incorporates a lot of information on the geometry of Γ_Λ. Using the dual definition of the Rubinstein distance instead of the original one can be very relevant in some cases.

Given a probability measure ν with density L with respect to the Poisson reference measure μ, our purpose in the present paper is to control from above the Rubinstein distance $\mathcal{T}_\rho(\mu,\nu)$ in terms of convenient (and easily computable) quantities involving the density L. Such inequalities belong to the domain of functional inequalities, which is by now a wide field of research with numerous methods of proofs. See for instance the very complete monograph [18] and particularly Chapters 21 and 22 for a large panorama on this topic, with precise references and credit.

To derive our inequalities, the two main ingredients at work are other representations of the Rubinstein distance and the Rademacher property. On the one hand, such representations can be obtained either by embedding the two probability measures into the evolution of a Markov semi-group, or by using the so-called Clark formula. On the other hand, the Rademacher property formally states that given a distance ρ, there exists a notion of gradient such that its domain contains the set $\rho - \mathrm{Lip}_1$ and any function in $\rho - \mathrm{Lip}_1$ has a gradient whose norm is less than 1, i.e., that we can proceed as in finite dimension.

For these two steps, we need a notion of gradient. In the setting of configuration spaces, such a notion does exist within the Malliavin calculus. In fact, we even have two notions of gradient: a "differential" gradient (see [1, 15]) and a gradient expressed as a finite difference operator (see [13]). We show that depending on the distance ρ chosen on the configuration space, one gradient or the other is more convenient, i.e., the Rademacher property holds with one notion of gradient, or the other.

The paper is organized as follows. After the preliminaries of Section 2, we provide in Section 3 various upper bounds on the Rubinstein distance $\mathcal{T}_\rho(\mu,\nu)$, where ρ is the total variation distance, the Wasserstein distance or the trivial distance on the configuration space Γ_Λ. Based on a semi-group approach, the first abstract upper bound involves the gradient associated to our given distance ρ in the sense of the Rademacher property. When dealing with the total variation distance on the one hand, such an estimate has a simplified expression, contained in our first main result, Theorem 3.2, which can be retrieved by using an alternative method, namely the Clark formula. On the other hand, when the configuration space is equipped with the Wasserstein distance, the upper bound we give in our second main result, Theorem 3.4, relies on a time-change argument together with the Girsanov Theorem. Finally, the last Section 4 is devoted to numerous applications of these two inequalities: by choosing the probability measure ν as the distribution of a given process, we are able to estimate from above distances between Poisson processes, between Poisson and Cox processes, between Poisson and Gibbs processes, etc. We thus hope to give a systematic treatment of the various situations one may encounter in applications. We conclude this work by providing another consequence of Theorem 3.2 to tail and isoperimetric estimates. In particular, we obtain sharp deviation inequalities for the total variation distance and also a new estimate of the classical isoperimetric constant, which is asymptotically sharp as the total mass of Λ is small.

2. Preliminaries

Let X be a Polish space and ρ a lower semi-continuous distance on $X\times X$, which does not necessarily generate the topology on X. Given two probability measures μ and ν on X, the optimal transportation problem associated to ρ consists in evaluating the distance

$$\mathcal{T}_\rho(\mu,\nu) = \inf_{\gamma\in\Sigma(\mu,\nu)} \int_X \int_X \rho(x,y)\ \mathrm{d}\gamma(x,y), \tag{2.1}$$

where $\Sigma(\mu,\nu)$ is the set of probability measures on $X\times X$ with first (respectively second) marginal μ (respectively ν). By Theorem 4.1 in [18], there exists at least one probability measure γ for which the infimum is attained. According to the celebrated Kantorovitch-Rubinstein duality theorem, cf. Theorem 5.10 in [18], this minimum is equal to

$$\mathcal{T}_\rho(\mu,\nu) = \sup_{\substack{F\in\rho-\mathrm{Lip}_1\\ F\in L^1(\mu+\nu)}} \int_X F\ \mathrm{d}(\mu-\nu), \tag{2.2}$$

where $\rho-\mathrm{Lip}_m$ is the set of bounded Lipschitz continuous functions F from X to $\mathbb{R}$ with Lipschitz constant m:

$$|F(x)-F(y)| \le m\rho(x,y),\quad x,y\in X.$$

In the context of optimal transportation, $\mathcal{T}_\rho$ is considered as a Rubinstein distance since the cost function is already a distance (see for instance the bibliographical notes at the end of Chapter 6 in [18]).

In this paper, we consider the situation where $X=\Gamma_\Lambda$ is the configuration space on a σ-compact metric space Λ with Borel σ-algebra $\mathcal{B}(\Lambda)$, i.e.,

$$\Gamma_\Lambda = \{\omega\subset\Lambda : \omega\cap K \text{ is a finite set for every compact } K\in\mathcal{B}(\Lambda)\}.$$

Here the σ-compactness means that Λ can be partitioned into the union of countably many compact subspaces. We identify $\omega\in\Gamma_\Lambda$ and the positive Radon measure $\sum_{x\in\omega}\varepsilon_x$, where ε_a is the Dirac measure at point a. Throughout this paper, Γ_Λ is endowed with the vague topology, i.e., the weakest topology such that for all $f\in\mathcal{C}_0(\Lambda)$ (continuous with compact support on Λ), the following maps

$$\omega\mapsto\int_\Lambda f\ \mathrm{d}\omega = \sum_{x\in\omega} f(x)$$

are continuous. When f is the indicator function of a subset B, we will use the shorter notation $\omega(B)$ for the integral of $\mathbf{1}_B$ with respect to ω. We denote by

$\mathcal{B}(\Gamma_\Lambda)$ the corresponding Borel σ-algebra. Let $\mathfrak{M}(\Lambda)$ be the space of positive and diffuse Radon measures on $\mathcal{B}(\Lambda)$ endowed with the corresponding Borel σ-field and equipped with the topology of vague convergence. Given a measure $\sigma \in \mathfrak{M}(\Lambda)$, the probability space under consideration in the remainder of this paper will be the Poisson space $(\Gamma_\Lambda, \mathcal{B}(\Gamma_\Lambda), \mu_\sigma)$, where μ_σ is the Poisson measure of intensity σ, i.e., the probability measure on Γ_Λ fully characterized by

$$\mathbb{E}_{\mu_\sigma}\left[\exp\left(\int_\Lambda f \, \mathrm{d}\omega\right)\right] = \exp\left\{\int_\Lambda (e^f - 1) \, \mathrm{d}\sigma\right\},$$

for all $f \in \mathcal{C}_0(\Lambda)$. Here $\mathbb{E}_{\mu_\sigma}$ stands for the expectation under the measure μ_σ.

2.1. Distances on the configuration space Γ_Λ. Actually, several distance concepts are available between elements of the configuration space Γ_Λ, cf. for instance [17] for a thorough discussion about this topic. We introduce only three of them which will be useful in the sequel. Let ω and η be two configurations in Γ_Λ.

Trivial distance: The trivial distance is simply given by

$$\rho_0(\omega, \eta) = \mathbf{1}_{\{\omega \neq \eta\}}.$$

Total variation distance: The total variation distance is defined as

$$\begin{aligned}\rho_1(\omega, \eta) &= \sum_{x \in \Lambda} |\omega(\{x\}) - \eta(\{x\})| \\ &= \omega \Delta \eta(\Lambda) + \eta \Delta \omega(\Lambda),\end{aligned}$$

where $\omega \Delta \eta = \omega \backslash (\omega \cap \eta)$.

Wasserstein distance: If $\Lambda = \mathbb{R}^k$ and κ is the Euclidean distance, the Wasserstein distance is given by

$$\rho_2(\omega, \eta) = \inf_{\beta \in \Sigma(\omega, \eta)} \sqrt{\int_\Lambda \int_\Lambda \kappa(x, y)^2 \, \mathrm{d}\beta(x, y)},$$

where $\Sigma(\omega, \eta)$ denotes the set of configurations $\beta \in \Gamma_{\Lambda \times \Lambda}$ having marginals ω and η, see [6, 15].

Let us comment on these notions of distance on the configuration space Γ_Λ. First, the total variation distance ρ_1 is nothing but the number of different atoms between two configurations. In particular, we allow them to be infinite so that the total variation distance might take infinite values. Note that our definition is a straightforward generalization of the classical notion of total variation distance between probability measures, since it coincides with the usual definition when the configurations are normalized by their total masses.

As the total variation distance ρ_1, the Wasserstein distance ρ_2 also shares the property that it might takes infinite values. Indeed, if the total masses of two configurations ω and η are finite but differ, then there exists no coupling configuration β in $\Sigma(\omega, \eta)$, hence the distance should be infinite. If $\omega(\Lambda) = \eta(\Lambda) < +\infty$ with

$\omega = \sum_{j=1}^{\omega(\Lambda)} \delta_{x_j}$ and $\eta = \sum_{j=1}^{\eta(\Lambda)} \delta_{y_j}$, we can also write

$$\rho_2(\omega,\eta)^2 = \inf_{\tau\in\mathfrak{S}_{\omega(\Lambda)}} \sum_{j=1}^{\omega(\Lambda)} \kappa(x_j, y_{\tau(j)})^2,$$

where $\mathfrak{S}_{\omega(\Lambda)}$ denotes the symmetric group on the finite set $\{1,2,\dots,\omega(\Lambda)\}$. As such ρ_2 appears as the dimension-free generalization of the Euclidean distance.

In order to use the Kantorovich-Rubinstein duality Theorem, the lower semi-continuity of the distances ρ_i, $i\in\{0,1,2\}$, is required. This is the object of the next lemma.

Lemma 2.1. *For any $i\in\{0,1,2\}$, the distance ρ_i is lower semi-continuous on the product space $\Gamma_\Lambda\times\Gamma_\Lambda$ equipped with the product topology.*

Proof. It is immediate for the trivial distance ρ_0 and it is proved in Lemma 4.1 in [15] for the Wasserstein distance ρ_2. To verify this property for the total variation distance ρ_1, let α be a real number and consider J_α defined by

$$J_\alpha = \{(\omega,\eta)\in\Gamma_\Lambda\times\Gamma_\Lambda :\ \rho_1(\omega,\eta)\le\alpha\}.$$

Let $((\omega_n,\eta_n),\, n\ge 1)$ converge vaguely to (ω,η) and such that for any n, (ω_n,η_n) belongs to J_α. By the triangular inequality, we have for any compact set K and any n:

$$\rho_1(\pi_K\omega,\pi_K\eta) \quad\le\quad \rho_1(\pi_K\omega,\pi_K\omega_n)+\alpha+\rho_1(\pi_K\eta_n,\pi_K\eta),$$

where π_K denotes the restriction to K of a configuration. Hence using the vague convergence, we obtain that $(\pi_K\omega,\pi_K\eta)\in J_\alpha$. Finally, since the metric space Λ is σ-compact, the monotone convergence theorem for an exhaustive sequence of compacts $(K_p)_{p\in\mathbb{N}}$ entails that

$$\rho_1(\omega,\eta) = \lim_{p\to+\infty}\rho_1(\pi_{K_p}\omega,\pi_{K_p}\eta)\le\alpha,$$

hence the set J_α is vaguely closed. □

Let us mention that Lemma 2.1 entails the lower semi-continuity of the Rubinstein distance $\mathcal{T}_{\rho_i}$, $i\in\{0,1,2\}$, with respect to the weak topology on the space of probability measures on Γ_Λ, cf. for instance Remark 6.12 in [18]. In particular, since the space $\mathfrak{M}(\Lambda)$ is equipped with the vague topology, then the application $\sigma\mapsto\mu_\sigma$ is continuous so that the mapping $\sigma\mapsto\mathcal{T}_{\rho_i}(\mu_\sigma,\nu)$, $i\in\{0,1,2\}$, is lower semi-continuous for any given probability measure ν on Γ_Λ. However for $i\in\{1,2\}$, the Rubinstein distances $\mathcal{T}_{\rho_i}$ is not continuous and might be infinite since the distance ρ_i is very often infinite itself, as in the Wiener space situation of [9].

Actually, we mention that our definitions do not coincide with some of the usual definitions of (bounded) distances between point processes, see for instance [2, 3, 17]. As mentioned above, it is customary to use the classical notion of total variation by considering normalized configurations, i.e.,

$$\widetilde{\rho_1}(\omega,\eta) = \rho_1\left(\frac{\omega}{\omega(\Lambda)},\frac{\eta}{\eta(\Lambda)}\right),$$

provided both configurations have finite total masses. It should be noted that since $\widetilde{\rho_1}$ is not lower semi-continuous, the Kantorovich-Rubinstein duality Theorem is

no longer satisfied, so that we cannot use the identity (2.2) in our framework. For instance, let $\Lambda = \mathbb{R}$, $\omega = \varepsilon_0$ and $\eta = \varepsilon_1$. Choose $\omega_n = \varepsilon_0 + \varepsilon_n$ and $\eta_n = \varepsilon_1 + \varepsilon_n$. As n goes to infinity, ω_n and η_n tend vaguely to ω and η respectively. However, we have $\widetilde{\rho_1}(\omega, \eta) = 2$ whereas $\widetilde{\rho_1}(\omega_n, \eta_n) = 1$, for any integer $n \geq 2$.
It is also customary to replace ρ_2 by $\widetilde{\rho_2}$ defined by

$$\widetilde{\rho_2}(\omega, \eta) = \begin{cases} \frac{1}{\omega(\Lambda)}\, \rho_2(\omega, \eta) & \text{if } \omega(\Lambda) = \eta(\Lambda) \neq 0, \\ |\omega(\Lambda) - \eta(\Lambda)| & \text{otherwise.} \end{cases}$$

The normalization by the inverse of $\omega(\Lambda)$ shrinks the ρ_2 distance by a factor roughly equal to the expectation of $\omega(\Lambda)^{-1}$, see [6]. More importantly, the term $|\omega(\Lambda) - \eta(\Lambda)|$ has no dimension (in the sense of dimensional analysis) whereas the term involving ρ_2 has the dimension of a length. Furthermore, the distance ρ_2 has interesting geometric properties of the space Γ_Λ like the Rademacher property (see Lemma 2.5 below), not shared by $\widetilde{\rho_2}$.

2.2. Malliavin derivatives and the Rademacher property. Before introducing the so-called Rademacher property on the configuration space Γ_Λ, we need some additional structure.

Hypothesis 2.2. *Assume now that we have:*

- *A kernel Q on $\Gamma_\Lambda \times \Lambda$, i.e. $Q(\cdot, A)$ is measurable as a function on Γ_Λ for any $A \in \mathcal{B}(\Lambda)$ and $Q(\omega, \cdot)$ is a positive Radon measure on $\mathcal{B}(\Lambda)$ for any $\omega \in \Gamma_\Lambda$. We set $d\alpha(\omega, x) = Q(\omega,\ dx)\ d\mu_\sigma(\omega)$.*
- *A gradient/Malliavin derivative ∇, defined on a dense subset* $\operatorname{Dom} \nabla$ *of $L^2(\mu_\sigma)$, such that for any $F \in \operatorname{Dom} \nabla$,*

$$\int_{\Gamma_\Lambda} \int_\Lambda |\nabla_x F(\omega)|^2 \ d\alpha(\omega, x) < +\infty,$$

i.e., the domain of the gradient is $\operatorname{Dom} \nabla = \{F \in L^2(\mu_\sigma) : \nabla F \in L^2(\alpha)\}$.

We say that a process $u = u(\omega, x)$ belongs to $\operatorname{Dom} \delta$ whenever there exists a constant c such that for any $F \in \operatorname{Dom} \nabla$,

$$\left| \int_{\Gamma_\Lambda} \int_\Lambda \nabla_x F(\omega)\, u(\omega, x)\ \mathrm{d}\alpha(\omega, x) \right| \leq c \|F\|_{L^2(\mu_\sigma)}.$$

For such a process, we define the operator δ by duality:

$$\int_{\Gamma_\Lambda} \int_\Lambda \nabla_x F(\omega)\, u(\omega, x)\ \mathrm{d}\alpha(\omega, x) = \int_{\Gamma_\Lambda} F(\omega)\, \delta u(\omega)\ \mathrm{d}\mu_\sigma(\omega). \tag{2.3}$$

Denote the self-adjoint operator $\mathcal{L} = \delta \nabla$ acting on its domain $\operatorname{Dom} \mathcal{L} \subset \operatorname{Dom} \nabla$ and let $(P_t)_{t \geq 0}$ be the associated Ornstein-Uhlenbeck semi-group, i.e. the semi-group whose infinitesimal generator is $-\mathcal{L}$.

Once the stochastic gradient has been introduced, let us relate it to the geometry of the configuration space Γ_Λ.

Definition 2.3. Given a distance ρ and a gradient ∇ on Γ_Λ, we say that the couple (∇, ρ) has the *Rademacher property* whenever

$$\rho - \mathrm{Lip}_1 \subset \operatorname{Dom} \nabla \qquad \textit{and} \qquad |\nabla_x F(\omega)| \leq 1, \quad \alpha\text{-a.e.} \tag{2.4}$$

To investigate the Rubinstein distance associated to a distance on Γ_Λ, it will be of crucial importance to find the convenient notion of gradient for which the Rademacher property holds.

Discrete gradient on configuration space. Given a functional $F \in L^2(\mu_\sigma)$, the discrete gradient of F, denoted by $\nabla^\sharp F$, is defined by

$$\nabla^\sharp_x F(\omega) = F(\omega + \varepsilon_x) - F(\omega), \quad (\omega, x) \in \Gamma_\Lambda \times \Lambda.$$

In particular, $\operatorname{Dom} \nabla^\sharp$ is the subspace of $L^2(\mu_\sigma)$ random variables such that

$$\mathbb{E}_{\mu_\sigma}\left[\int_\Lambda |\nabla^\sharp_x F|^2 \, \mathrm{d}\sigma(x)\right] < +\infty.$$

We set $Q^\sharp(\omega, \mathrm{d}x) = \mathrm{d}\sigma(x)$ so that $\alpha^\sharp = \mu_\sigma \otimes \sigma$. The n-th multiple stochastic integral of a real-valued square-integrable symmetric function $f_n \in L^2(\sigma^{\otimes n})$ is defined as

$$J_n(f_n) = \int_{\Delta_n} f_n(x_1, \dots, x_n) \, \mathrm{d}(\omega - \sigma)(x_1) \dots \mathrm{d}(\omega - \sigma)(x_n),$$

where $\Delta_n = \{(x_1, \dots, x_n) \in \Lambda^n, \ x_i \neq x_j, \, i \neq j\}$. As a convention, we identify $L^2(\sigma^{\otimes 0})$ to $\mathbb{R}$ and let $J_0(f_0) = f_0$, $f_0 \in L^2(\sigma^{\otimes 0}) \simeq \mathbb{R}$. We have the isometry formula

$$\mathbb{E}_{\mu_\sigma}\left[J_n(f_n) J_m(f_m)\right] = n! \, \mathbf{1}_{\{n=m\}} \int_{\Lambda^n} f_n \, f_m \, \mathrm{d}\sigma^{\otimes n}. \tag{2.5}$$

According to [16, 13], the Chaotic Representation Property holds on the configuration space, i.e., every functional $F \in L^2(\mu_\sigma)$ can be written as

$$F = \mathbb{E}_{\mu_\sigma}[F] + \sum_{n=1}^{+\infty} J_n(f_n).$$

Moreover, if $F \in \operatorname{Dom} \nabla^\sharp$, then the discrete gradient acts on multiple stochastic integrals as

$$\nabla^\sharp_x F = \sum_{n=1}^{+\infty} n J_{n-1}(f_n(\cdot, x)), \quad \alpha^\sharp\text{-a.e.}$$

Denote $\delta^\sharp$ the adjoint operator of $\nabla^\sharp$ in the sense of (2.3). Then the self-adjoint number operator $\mathcal{L}^\sharp = \delta^\sharp \nabla^\sharp$ has the following expression in terms of chaos:

$$\mathcal{L}^\sharp F = \sum_{n=1}^{+\infty} n J_n(f_n),$$

whenever $F \in \operatorname{Dom} \mathcal{L}^\sharp$, and the associated Ornstein-Uhlenbeck semi-group $(P^\sharp_t)_{t \geq 0}$ is given by

$$P^\sharp_t F = \mathbb{E}_{\mu_\sigma}[F] + \sum_{n=1}^{+\infty} e^{-nt} J_n(f_n).$$

Hence the invariance property of the Poisson measure μ_σ with respect to the semi-group reads as $\mathbb{E}_{\mu_\sigma}[P^\sharp_t F] = \mathbb{E}_{\mu_\sigma}[F]$. Moreover, we have the commutation property between gradient and semi-group, which will be useful in the sequel: if $F \in \mathrm{Dom}\, \nabla^\sharp$,

$$\nabla_x^\sharp P_t^\sharp F = e^{-t} P_t^\sharp \nabla_x^\sharp F, \quad x \in \Lambda, \quad t \geq 0. \tag{2.6}$$

By the isometry formula (2.5), the semi-group is exponentially ergodic in $L^2(\mu_\sigma)$ with respect to the Poisson measure μ_σ, i.e., for any $t \geq 0$,

$$\begin{aligned} \| P_t F - \mathbb{E}_{\mu_\sigma}[F] \|^2_{L^2(\mu_\sigma)} &= \sum_{n \geq 1} e^{-2nt} \, \mathbb{E}_{\mu_\sigma}\left[J_n(f_n)^2 \right] \\ &\leq e^{-2t} \, \| F - \mathbb{E}_{\mu_\sigma}[F] \|^2_{L^2(\mu_\sigma)}. \end{aligned}$$

Using the discrete gradient, the distances of interest on Γ_Λ are the trivial distance ρ_0 and the total variation distance ρ_1, as illustrated by the following Lemma.

Lemma 2.4. *Assume that the intensity measure σ is finite on Λ. Then the couples $(\nabla^\sharp, \rho_0)$ and $(\nabla^\sharp, \rho_1)$ satisfy the Rademacher property (2.4).*

Proof. Letting $F \in \rho_i - \mathrm{Lip}_1$, $i \in \{0, 1\}$, we have by the very definition of the discrete gradient:

$$|\nabla_x^\sharp F(\omega)| = |F(\omega + \varepsilon_x) - F(\omega)| \leq \rho_i(\omega + \varepsilon_x, \omega) \leq 1.$$

Since σ is finite, it follows that

$$\int_\Lambda |\nabla_x^\sharp F(\omega)|^2 \, \mathrm{d}\sigma(x) \leq \sigma(\Lambda),$$

hence that F belongs to $\mathrm{Dom}\, \nabla^\sharp$. The proof is achieved. □

Note that the converse direction holds for the total variation distance ρ_1. Indeed, consider two configurations ω and η. If $\rho_1(\omega, \eta) = +\infty$, there is nothing to prove. If $\rho_1(\omega, \eta)$ is finite, then since $|\nabla_x^\sharp F(\omega)| \leq 1$, $\alpha^\sharp$-a.e., we get

$$\begin{aligned} |F(\eta) - F(\omega)| &\leq |F(\eta \cap \omega \,\cup\, \eta \Delta \omega) - F(\eta \cap \omega)| + |F(\eta \cap \omega \,\cup\, \omega \Delta \eta) - F(\eta \cap \omega)| \\ &\leq (\eta \Delta \omega)(\Lambda) + (\omega \Delta \eta)(\Lambda) \\ &= \rho_1(\eta, \omega). \end{aligned}$$

Differential gradient on configuration space. Let us introduce another stochastic gradient on the configuration space Γ_Λ which is a derivation, see [1, 15]. Given the Euclidean space $\Lambda = \mathbb{R}^k$, let $V(\Lambda)$ be the space of $\mathcal{C}^\infty$ vector fields on Λ and $V_0(\Lambda) \subset V(\Lambda)$, the subspace consisting of all vector fields with compact support. For $v \in V_0(\Lambda)$, for any $x \in \Lambda$, the curve

$$t \mapsto \mathcal{V}_t^v(x) \in \Lambda$$

is defined as the solution of the following Cauchy problem

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} \mathcal{V}_t^v(x) &= v(\mathcal{V}_t^v(x)), \\ \mathcal{V}_0^v(x) &= x. \end{cases} \tag{2.7}$$

The associated flow $(\mathcal{V}_t^v, t \in \mathbb{R})$ induces a curve $(\mathcal{V}_t^v)^* \omega = \omega \circ (\mathcal{V}_t^v)^{-1}$, $t \in \mathbb{R}$, on Γ_Λ: if $\omega = \sum_{x \in \omega} \varepsilon_x$ then $(\mathcal{V}_t^v)^* \omega = \sum_{x \in \omega} \varepsilon_{\mathcal{V}_t^v(x)}$. We are then in position to define a notion of differentiability on Γ_Λ. We take $Q^c(\omega, \, \mathrm{d}x) = \, \mathrm{d}\omega(x) = \sum_{y \in \omega} \, \mathrm{d}\varepsilon_y(x)$

and $\mathrm{d}\alpha^c(\omega, x) = \mathrm{d}\omega(x)\,\mathrm{d}\mu_\sigma(\omega)$. A measurable function $F : \Gamma_\Lambda \to \mathbb{R}$ is said to be differentiable if for any $v \in V_0(\Lambda)$, the following limit exists:

$$\lim_{t\to 0} \frac{F(\mathcal{V}_t^v(\omega)) - F(\omega)}{t}.$$

We denote $\nabla_v^c F(\omega)$ the preceding quantity. The domain of ∇^c is then the set of integrable and differentiable functions such that there exists a process $(\omega, x) \mapsto \nabla_x^c F(\omega)$ which belongs to $L^2(\alpha^c)$ and satisfies

$$\nabla_v^c F(\omega) = \int_\Lambda \nabla_x^c F(\omega) v(x)\, \mathrm{d}\omega(x).$$

We denote by δ^c the adjoint operator of ∇^c in the sense of (2.3). Note that the integration in the left-hand-side of the duality formula (2.3) is made with respect to a configuration ω, whereas the intensity measure σ is involved in the case of the discrete gradient. Given the self-adjoint operator $\mathcal{L}^c = \delta^c \nabla^c$, the associated Ornstein-Uhlenbeck semi-group $(P_t^c)_{t\geq 0}$ is ergodic in $L^2(\mu_\sigma)$ with respect to the Poisson measure μ_σ, cf. Theorem 4.3 in [1]. However, in contrast to the case of the discrete gradient, there is no known commutation relationship between the gradient ∇^c and the semi-group P_t^c.

The distance we focus on in this part is the Wasserstein distance ρ_2. We have the following lemma.

Lemma 2.5. *The couple (∇^c, ρ_2) satisfies the Rademacher property* (2.4)*.*

Proof. The proof is straightforward. Indeed, letting $F \in \rho_2 - \mathrm{Lip}_1$, we know from Theorem 1.3 in [15] that $F \in \mathrm{Dom}\,\nabla^c$ and that

$$\sum_{x\in\omega} |\nabla_x^c F(\omega)|^2 = \int_\Lambda |\nabla_x^c F(\omega)|^2\, \mathrm{d}\omega(x) \leq 1, \quad \mu_\sigma\text{-a.s.}$$

Hence we obtain $|\nabla_x^c F(\omega)| \leq 1$, α^c-a.e., in other words the Rademacher property (2.4) is satisfied. □

3. Upper Bounds on Rubinstein Distances

3.1. An abstract upper bound on Rubinstein distances. Let us establish first an abstract upper bound on the Rubinstein distance by using a semi-group method, provided the associated couple gradient/distance satisfies the Rademacher property (2.4). Denote ρ a lower semi-continuous distance on the configuration space Γ_Λ and assume that Hypothesis 2.2 is fulfilled.

Proposition 3.1. *Assume that the couple (∇, ρ) satisfies the Rademacher property (2.4). Let L be the density of an absolutely continuous probability measure ν with respect to μ_σ. Then provided the inequality makes sense, the following upper bound on the Rubinstein distance holds:*

$$\mathcal{T}_\rho(\mu_\sigma, \nu) \leq \int_{\Gamma_\Lambda} \int_\Lambda \left| \int_0^{+\infty} \nabla_x P_t L(\omega)\ dt \right|\ d\alpha(\omega, x). \tag{3.1}$$

Proof. The proof follows the approach emphasized by Houdré and Privault in [11] to derive covariance identities and then concentration inequalities. Letting $F \in \rho - \mathrm{Lip}_1$, we have by reversibility and using Fubini's Theorem:

$$\begin{aligned}
\int_{\Gamma_\Lambda} F \,\mathrm{d}(\mu_\sigma - \nu) &= \int_{\Gamma_\Lambda} \left(\int_{\Gamma_\Lambda} F \,\mathrm{d}\mu_\sigma - F \right) L \,\mathrm{d}\mu_\sigma \\
&= \int_{\Gamma_\Lambda} \left(\int_0^{+\infty} \frac{\mathrm{d}}{\mathrm{d}t} P_t F \,\mathrm{d}t \right) L \,\mathrm{d}\mu_\sigma \\
&= -\int_{\Gamma_\Lambda} \int_0^{+\infty} P_t \mathcal{L} F \, L \,\mathrm{d}t \,\mathrm{d}\mu_\sigma \\
&= -\int_{\Gamma_\Lambda} \int_0^{+\infty} \delta \nabla F \, P_t L \,\mathrm{d}t \,\mathrm{d}\mu_\sigma \\
&= -\int_{\Gamma_\Lambda} \int_\Lambda \nabla_x F \int_0^{+\infty} \nabla_x P_t L \,\mathrm{d}t \,\mathrm{d}\alpha(\cdot, x).
\end{aligned}$$

Using then the Rademacher property (2.4), the result holds by taking the supremum over all functions $F \in \rho - \mathrm{Lip}_1$. □

Note that the upper bound in the inequality (3.1) is interesting in its own right, but seems to be somewhat difficult to compute in full generality. Hence we turn in the sequel to more concrete situations, i.e., when the gradient of interest is the discrete gradient $\nabla^\sharp$ or the differential one ∇^c and is associated to the convenient distance ρ_i, $i \in \{0, 1, 2\}$, in the sense of the Rademacher property (2.4).

3.2. A qualitative upper bound on $\mathcal{T}_{\rho_1}$. Once the abstract estimate (3.1) has been obtained, one notices that it might be simplified whenever a commutation relation between gradient and semi-group holds. To the knowledge of the authors, such a property is only verified in the case of the discrete gradient, so that we focus in this part on the couple $(\nabla^\sharp, \rho_1)$. Here is one of the two main results of the paper.

Theorem 3.2. *Let L be the density of an absolutely continuous probability measure ν with respect to μ_σ, and assume that $L \in \mathrm{Dom}\,\nabla^\sharp$ and $\nabla^\sharp L \in L^1(\mu_\sigma \otimes \sigma)$. Then we get the following estimate:*

$$\mathcal{T}_{\rho_1}(\mu_\sigma, \nu) \leq \mathbb{E}_{\mu_\sigma}\left[\int_\Lambda |\nabla_x^\sharp L| \, d\sigma(x) \right]. \tag{3.2}$$

The same inequality also holds under the distance ρ_0.

Proof. Since the case of a general intensity measure $\sigma \in \mathfrak{M}(\Lambda)$ might be established by a simple limiting procedure (use the σ-compactness of the metric space Λ and the lower semi-continuity of the application $\sigma \mapsto \mathcal{T}_{\rho_1}(\mu_\sigma, \nu)$), let us assume that σ is finite, so that the Rademacher property stated in Lemma 2.4 is satisfied by the couple $(\nabla^\sharp, \rho_1)$. Hence Proposition 3.1 above entails the inequality

$$\mathcal{T}_{\rho_1}(\mu_\sigma, \nu) \quad \leq \quad \mathbb{E}_{\mu_\sigma}\left[\int_\Lambda \left| \int_0^{+\infty} \nabla_x^\sharp P_t^\sharp L \,\mathrm{d}t \right| \,\mathrm{d}\sigma(x) \right].$$

Using now the commutation relation (2.6), we have:

$$\begin{aligned} \mathcal{T}_{\rho_1}(\mu_\sigma,\nu) &\leq \mathbb{E}_{\mu_\sigma}\left[\int_\Lambda \left|\int_0^{+\infty} e^{-t} P_t^\sharp \nabla_x^\sharp L \ \mathrm{d}t\right| \ \mathrm{d}\sigma(x)\right] \\ &\leq \mathbb{E}_{\mu_\sigma}\left[\int_\Lambda \int_0^{+\infty} e^{-t} P_t^\sharp |\nabla_x^\sharp L| \ \mathrm{d}t \ \mathrm{d}\sigma(x)\right] \\ &= \mathbb{E}_{\mu_\sigma}\left[\int_\Lambda \int_0^{+\infty} e^{-t} |\nabla_x^\sharp L| \ \mathrm{d}t \ \mathrm{d}\sigma(x)\right] \\ &= \mathbb{E}_{\mu_\sigma}\left[\int_\Lambda |\nabla_x^\sharp L| \ \mathrm{d}\sigma(x)\right], \end{aligned} \tag{3.3}$$

where we have used Jensen's inequality and the invariance property of the Poisson measure μ_σ with respect to the semi-group $P_t^\sharp$. The desired inequality (3.2) is thus established.

Finally, the case of the trivial distance ρ_0 is similar since the couple $(\nabla^\sharp, \rho_0)$ also satisfies the Rademacher property, cf. Lemma 2.4. The proof is achieved in full generality. □

Actually, the well-known relationship between semi-group and generator states that for any $G \in L^2(\mu_\sigma)$,

$$\int_0^{+\infty} e^{-t} P_t^\sharp G \ \mathrm{d}t = (\mathrm{Id} + \mathcal{L}^\sharp)^{-1} G.$$

Applying then such an identity in the inequality (3.3) above gives the following bound:

$$\mathcal{T}_{\rho_1}(\mu_\sigma,\nu) \leq \mathbb{E}_{\mu_\sigma}\left[\int_\Lambda |(\mathrm{Id} + \mathcal{L}^\sharp)^{-1} \nabla_x^\sharp L| \ \mathrm{d}\sigma(x)\right]. \tag{3.4}$$

It seems theoretically slightly better than the upper bound of Theorem 3.2 but often yields to intractable computations, except when the chaos representation of L is given, as noticed in Section 4.1 below. Note that the very analog of (3.4) on Wiener space was proved by a different though related way in Theorem 3.2 of [9].

Let us provide another method leading to Theorem 3.2 which is based on the so-called Clark formula. Instead of considering configurations in Γ_Λ, the idea is to use multivariate Poisson processes, i.e., point processes on $[0,1]$ with marks in the σ-compact metric space Λ. Borrowing an idea of [19], we first explain how to embed a Poisson process into a multivariate Poisson process.

Let $\widehat{\mu}$ be the Poisson measure of intensity $\lambda \otimes \sigma$ on the new configuration space $\Gamma_{\widehat{\Lambda}}$, where the enlarged state space is $\widehat{\Lambda} = [0,1] \times \Lambda$, and λ denotes the Lebesgue measure on $[0,1]$. Any generic element $\widehat{\omega} \in \Gamma_{\widehat{\Lambda}}$ has the form $\widehat{\omega} = \sum_{(t,x)\in\widehat{\omega}} \varepsilon_{t,x}$. The canonical filtration is defined for any $t \in [0,1]$ as

$$\mathfrak{F}_t = \sigma\left\{\widehat{\omega}([0,s] \times B), \ 0 \leq s \leq t, \ B \in \mathcal{B}(\Lambda)\right\}.$$

Let us recall the Clark formula, cf. for instance [7] or Lemma 1.3 in [19], which states that every functional $G : \Gamma_{\widehat{\Lambda}} \to \mathbb{R}$ belonging to $\mathrm{Dom}\, \nabla^\sharp$ might be written as

$$G = \mathbb{E}_{\widehat{\mu}}[G] + \int_0^1 \int_\Lambda \mathbb{E}_{\widehat{\mu}}\left[\nabla_{t,x}^\sharp G \,|\, \mathfrak{F}_{t-}\right] \ \mathrm{d}(\widehat{\omega} - \lambda \otimes \sigma)(t,x), \tag{3.5}$$

where $\nabla^\sharp_{t,x}$ denotes the discrete gradient on the enlarged configuration space $\Gamma_{\widehat{\Lambda}}$.

For an element $\widehat{\omega} \in \Gamma_{\widehat{\Lambda}}$, we define by $\pi\widehat{\omega}$ its projection on Γ_Λ, i.e.,

$$\pi\widehat{\omega}(B) = \widehat{\omega}([0,1] \times B), \quad B \in \mathcal{B}(\Lambda),$$

and given $F : \Gamma_\Lambda \to \mathbb{R}$, we define the functional $\widehat{F}$ as

$$\begin{aligned} \widehat{F} : \Gamma_{\widehat{\Lambda}} &\longrightarrow \mathbb{R} \\ \widehat{\omega} &\longmapsto F(\pi\widehat{\omega}). \end{aligned}$$

In particular, we have clearly $\nabla^\sharp_{t,x}\widehat{F}(\widehat{\omega}) = \nabla^\sharp_x F(\pi\widehat{\omega})$ for any $(t,x) \in \widehat{\Lambda}$. Moreover, we have $\mathbb{E}_{\widehat{\mu}}[\widehat{F}] = \mathbb{E}_{\mu_\sigma}[F]$ since the image measure of $\widehat{\mu}$ by π is μ_σ.

The total variation distance on $\Gamma_{\widehat{\Lambda}}$ is defined as

$$\widehat{\rho_1}(\widehat{\omega}, \widehat{\eta}) = \sum_{(t,x)\in\widehat{\Lambda}} |\widehat{\omega}(\{t,x\}) - \widehat{\eta}(\{t,x\})|.$$

The key point is the following lemma.

Lemma 3.3. *For any $F \in \rho_1 - \mathrm{Lip}_1$, the functional $\widehat{F}$ belongs to $\widehat{\rho_1} - \mathrm{Lip}_1$.*

Proof. Given $F \in \rho_1 - \mathrm{Lip}_1$, we have for any $\widehat{\omega}, \widehat{\eta} \in \Gamma_{\widehat{\Lambda}}$:

$$\begin{aligned} |\widehat{F}(\widehat{\omega}) - \widehat{F}(\widehat{\eta})| &= |F(\pi\widehat{\omega}) - F(\pi\widehat{\eta})| \\ &\leq \rho_1(\pi\widehat{\omega}, \pi\widehat{\eta}) \\ &= \sum_{x\in\Lambda} |\pi\widehat{\omega}(\{x\}) - \pi\widehat{\eta}(\{x\})| \\ &= \sum_{x\in\Lambda} \left| \sum_{t\in[0,1]} \widehat{\omega}(\{t,x\}) - \widehat{\eta}(\{t,x\}) \right| \\ &\leq \sum_{(t,x)\in\widehat{\Lambda}} |\widehat{\omega}(\{t,x\}) - \widehat{\eta}(\{t,x\})| \\ &= \widehat{\rho_1}(\widehat{\omega}, \widehat{\eta}). \end{aligned}$$

The proof is complete. □

Now we are able to give a second proof of Theorem 3.2 by means of the Clark formula (3.5) and Lemma 3.3.

Proof. Letting $\widehat{\nu}$ be the measure with density $\widehat{L}$ with respect to $\widehat{\mu}$, we obtain:

$$\begin{aligned} \mathcal{T}_{\rho_1}(\mu_\sigma, \nu) &= \sup_{F\in\rho_1-\mathrm{Lip}_1} \mathbb{E}_{\mu_\sigma}[F(L-1)] \\ &= \sup_{F\in\rho_1-\mathrm{Lip}_1} \mathbb{E}_{\widehat{\mu}}[\widehat{F}(\widehat{L}-1)] \\ &= \sup_{F\in\rho_1-\mathrm{Lip}_1} \mathbb{E}_{\widehat{\nu}}[\widehat{F}] - \mathbb{E}_{\widehat{\mu}}[\widehat{F}]. \end{aligned}$$

Now using the Clark formula (3.5) and taking expectation with respect to $\widehat{\nu}$,

$$\begin{aligned}\mathbb{E}_{\widehat{\nu}}[\widehat{F}] &= \mathbb{E}_{\widehat{\mu}}[\widehat{F}] + \mathbb{E}_{\widehat{\nu}}\left[\int_0^1 \int_\Lambda \mathbb{E}_{\widehat{\mu}}\left[\nabla^\sharp_{t,x}\widehat{F} \,|\, \mathfrak{F}_{t^-}\right] \, \mathrm{d}(\widehat{\omega} - \lambda\otimes\sigma)(t,x)\right] \\ &= \mathbb{E}_{\widehat{\mu}}[\widehat{F}] + \mathbb{E}_{\widehat{\mu}}\left[\widehat{L}\int_0^1 \int_\Lambda \mathbb{E}_{\widehat{\mu}}\left[\nabla^\sharp_{t,x}\widehat{F} \,|\, \mathfrak{F}_{t^-}\right] \, \mathrm{d}(\widehat{\omega} - \lambda\otimes\sigma)(t,x)\right] \\ &= \mathbb{E}_{\widehat{\mu}}[\widehat{F}] + \mathbb{E}_{\widehat{\mu}}\left[\int_0^1 \int_\Lambda \mathbb{E}_{\widehat{\mu}}\left[\nabla^\sharp_{t,x}\widehat{F} \,|\, \mathfrak{F}_{t^-}\right] \nabla^\sharp_{t,x}\widehat{L} \; \mathrm{d}t \; \mathrm{d}\sigma(x)\right],\end{aligned}$$

where in the second line we also used the Clark formula (3.5) applied to the functional $\widehat{L}$. By Lemma 2.4, the couple $(\nabla^\sharp, \widehat{\rho_1})$ satisfies the Rademacher property (2.4) on $\Gamma_{\widehat{\Lambda}}$. Hence Lemma 3.3 implies that for $F \in \rho_1 - \mathrm{Lip}_1$, the quantity $\left|\mathbb{E}_{\widehat{\mu}}\left[\nabla^\sharp_{t,x}\widehat{F} \,|\, \mathfrak{F}_{t^-}\right]\right|$ is bounded by 1, $\widehat{\mu}\otimes\lambda\otimes\sigma$-a.e., so that we obtain finally

$$\begin{aligned}\mathcal{T}_{\rho_1}(\mu_\sigma, \nu) &\le \mathbb{E}_{\widehat{\mu}}\left[\int_0^1 \int_\Lambda |\nabla^\sharp_{t,x}\widehat{L}| \; \mathrm{d}t \; \mathrm{d}\sigma(x)\right] \\ &= \mathbb{E}_{\mu_\sigma}\left[\int_\Lambda |\nabla^\sharp_x L| \; \mathrm{d}\sigma(x)\right].\end{aligned}$$

The second proof of Theorem 3.2 is thus complete. □

3.3. A qualitative upper bound on $\mathcal{T}_{\rho_2}$ by time-change. Recall that by Lemma 2.5, the couple (∇^c, ρ_2) satisfies the Rademacher property (2.4). Hence Proposition 3.1 entails an upper bound on the $\mathcal{T}_{\rho_2}$ Rubinstein distance as follows: if L denotes the density of an absolutely continuous probability measure ν with respect to μ_σ, then we have

$$\mathcal{T}_{\rho_2}(\mu_\sigma, \nu) \quad \le \quad \int_{\Gamma_\Lambda}\int_\Lambda \left| \int_0^{+\infty} \nabla^c_x P^c_t L(\omega) \; \mathrm{d}t \right| \; \mathrm{d}\omega(x) \; \mathrm{d}\mu_\sigma(\omega),$$

provided the inequality makes sense. However, despite its theoretical interest, such an inequality is not really tractable in practise, since no commutation relation has been established yet between the differential gradient ∇^c and the semi-group P^c_t. Hence the purpose of this section is to provide another estimate on $\mathcal{T}_{\rho_2}$ through a different approach relying on a time-change argument together with the Girsanov Theorem.

We consider the notation of Section 3.2 above, with the difference that the state space is now $\widehat{\Lambda} := [0,\infty)\times\Lambda$, where Λ is the space $\mathbb{R}^k$ equipped with the Euclidean distance κ. In this part, the distance of interest on the enlarged configuration space $\Gamma_{\widehat{\Lambda}}$ is the Wasserstein distance:

$$\widehat{\rho_2}(\widehat{\omega}, \widehat{\eta})^2 = \inf_{\beta\in\Sigma(\widehat{\omega},\widehat{\eta})} \int_{\widehat{\Lambda}}\int_{\widehat{\Lambda}} (\kappa(x,y)^2 + |t-s|^2) \; \mathrm{d}\beta((s,x),(t,y)).$$

The following theorem is our second main result.

Theorem 3.4. *Let L be the (positive) density of an absolutely continuous probability measure $\widehat{\nu}$ with respect to $\widehat{\mu}$. Then provided the inequality makes sense, we* *get the following upper bound on the Rubinstein distance* $\mathcal{T}_{\widehat{\rho_2}}(\widehat{\mu},\widehat{\nu})$*:*

$$\begin{aligned}\mathcal{T}_{\widehat{\rho_2}}(\widehat{\mu},\widehat{\nu})^2 &\le \mathbb{E}_{\widehat{\mu}}\left[L\int_\Lambda\int_0^{+\infty}\left|\int_0^t u(s,z)\ ds\right|^2(1+u(t,z))\ dt\ d\sigma(z)\right]\\ &= \mathbb{E}_{\widehat{\mu}}\left[L\int_\Lambda\int_0^{+\infty}\left|r-v^{-1}(r,z)\right|^2\ dr\ d\sigma(z)\right],\end{aligned} \tag{3.6}$$

where $u(t,z)>-1$ *is the following predictable process:*

$$u(t,z)=\frac{\mathbb{E}\left[\nabla^\sharp_{t,z}L|\mathfrak{F}_{t^-}\right]}{\mathbb{E}\left[L|\mathfrak{F}_{t^-}\right]},\quad v(t,z):=t+\int_0^t u(s,z)\ ds,\quad z\in\Lambda,$$

and $v^{-1}(\cdot,z)$ *is the inverse of the increasing mapping* $t\mapsto v(t,z)$*.*

Remark 3.5. Note that for $z\in\Lambda$ fixed, the term $\int_0^{+\infty}\left|r-v^{-1}(r,\,z)\right|^2\ \mathrm{d}r$ can be interpreted as a generalized Wassertein distance between the infinite measures $\mathrm{d}r$ and $(1+u(r,z))\ \mathrm{d}r$, see [18]. Then, the $\mathcal{T}_{\widehat{\rho_2}}$ distance is bounded from above by the expectation under $\widehat{\nu}$ of this generalized distance integrated over Λ according to the marks distribution.

Proof. By the Girsanov Theorem, there exists a predictable process u such that for any compact set $K\in\mathcal{B}(\Lambda)$, the process

$$t\mapsto\widehat{\omega}([0,t]\times K)-\int_0^t\int_K(1+u(s,z))\ \mathrm{d}s\ \mathrm{d}\sigma(z),$$

is a $\widehat{\nu}$-martingale. Moreover, the conditional expectation $L_t:=\mathbb{E}\left[L|\mathfrak{F}_t\right]$ might be identified as follows:

$$\begin{aligned}L_t &= \exp\left\{\int_0^t\int_\Lambda\ln(1+u(s,z))\ \mathrm{d}\widehat{\omega}(s,z)-\int_0^t\int_\Lambda u(s,z)\ \mathrm{d}s\ \mathrm{d}\sigma(z)\right\}\\ &= \mathcal{E}\left(\int_0^t\int_\Lambda u(s,z)\ \mathrm{d}(\widehat{\omega}-\lambda\otimes\sigma)(s,z)\right)\\ &= 1+\int_0^t\int_\Lambda L_{s^-}u(s,z)\ \mathrm{d}(\widehat{\omega}-\lambda\otimes\sigma)(s,z),\end{aligned}$$

where $\mathcal{E}$ denotes the classical Doléans-Dade exponential. On the other hand, the Clark formula (3.5) extended to the set $(0,+\infty)$ induces that

$$L_t=1+\int_0^t\int_\Lambda\mathbb{E}\left[\nabla^\sharp_{s,z}L_t|\mathfrak{F}_{s^-}\right]\ \mathrm{d}(\widehat{\omega}-\lambda\otimes\sigma)(s,z).$$

By identification, we obtain:

$$u(s,z)=\frac{\mathbb{E}\left[\nabla^\sharp_{s,z}L_t|\mathfrak{F}_{s^-}\right]}{L_{s^-}}=\frac{\mathbb{E}\left[\nabla^\sharp_{s,z}L|\mathfrak{F}_{s^-}\right]}{L_{s^-}},$$

since for any $s\in(0,t)$ a commutation relation holds between the discrete gradient $\nabla^\sharp_{s,z}$ and the conditional expectation knowing $\mathfrak{F}_t$, cf. for instance Lemma 3.2 in [13]. Define on $\Gamma_{\widehat{\Lambda}}$ the time-change configuration $\tau\widehat{\omega}$ by

$$\tau\widehat{\omega}=\sum_{(t_i,z_i)\in\widehat{\omega}}\varepsilon_{v(t_i,z_i),z_i},$$

where $v(t,z)$ is given above. By Theorem 3 in [5], the distribution of $\tau\widehat{\omega}$ under $\widehat{\nu}$ is nothing but the law of the configuration $\widehat{\omega}$ under $\widehat{\mu}$. Hence using Cauchy-Schwarz' inequality in the second line below, we obtain:

$$\begin{aligned}\mathcal{T}_{\widehat{\rho_2}}(\widehat{\mu},\widehat{\nu}) &\leq \mathbb{E}_{\widehat{\nu}}\left[\widehat{\rho_2}(\widehat{\omega},\tau\widehat{\omega})\right] \\ &\leq \mathbb{E}_{\widehat{\nu}}\left[\int_\Lambda \int_0^{+\infty} |t-v(t,z)|^2 \ \mathrm{d}\widehat{\omega}(t,z)\right]^{1/2} \\ &= \mathbb{E}_{\widehat{\nu}}\left[\int_\Lambda \int_0^{+\infty} |t-v(t,z)|^2 \ \frac{\mathrm{d}v}{\mathrm{d}t}(t,z) \ \mathrm{d}t \ \mathrm{d}\sigma(z)\right]^{1/2},\end{aligned}$$

where we used the classical compensation formula for stochastic integrals with respect to Poisson random measures. Finally, the change of variable $r = v(t,z)$ for $z \in \Lambda$ being fixed allows us to obtain the desired inequality (3.6). □

4. Applications

4.1. Distance estimates between processes. The purpose of the present part is to apply our main results Theorems 3.2 and 3.4 to provide distance estimates between a Poisson process and several other more sophisticated processes, such as Cox or Gibbs processes. See for instance the pioneer monograph [3] or also [2, 17] for similar results with respect to another (bounded) distances on the configuration space Γ_Λ. The three first examples below rely on the total variation distance ρ_1, whereas in the last one the Wasserstein distance ρ_2 is considered.

Poisson processes. Here the probability measure ν is supposed to be another Poisson measure on Γ_Λ, where Λ is a σ-compact metric space.

Proposition 4.1. *Let μ_τ be a Poisson measure on Γ_Λ of intensity τ. We assume that τ admits a density p with respect to σ such that $p-1 \in L^1(\sigma)$. Then we have*

$$\mathcal{T}_{\rho_1}(\mu_\sigma,\mu_\tau) \leq \int_\Lambda |p(x)-1| \ d\sigma(x). \tag{4.1}$$

Proof. Since μ_τ is a Poisson measure on Γ_Λ of intensity τ, it is well known that it is absolutely continuous with respect to μ_σ and the density L is given by

$$L(\omega) = \exp\left\{\int_\Lambda \log p(x) \ \mathrm{d}\omega(x) + \int_\Lambda (1-p(x)) \ \mathrm{d}\sigma(x)\right\}.$$

It is then straightforward that $\nabla_x^\sharp L = L(p(x)-1)$, hence by Theorem 3.2,

$$\mathcal{T}_{\rho_1}(\mu_\sigma,\mu_\tau) \leq \mathbb{E}_{\mu_\sigma}\left[L\int_\Lambda |p(x)-1| \ \mathrm{d}\sigma(x)\right] = \int_\Lambda |p(x)-1| \ \mathrm{d}\sigma(x).$$

The proof is achieved. □

Note that in this very simple situation, the inequality (3.4) yields the same bound. Indeed, since p is deterministic, the density L has the following chaos representation

$$L = 1 + \sum_{n=1}^\infty \frac{1}{n!} J_n\left((p-1)^{\otimes n}\right),$$

cf. identity (7) in [16], so that we have

$$((\mathrm{Id}+\mathcal{L}^\sharp)^{-1}\nabla_x^\sharp L = (p(x)-1)\sum_{n=1}^{\infty}\frac{1}{(n-1)!}J_{n-1}\left((p-1)^{\otimes n-1}\right) = (p(x)-1)L.$$

Actually, one might obtain the inequality (4.1) by using another very intuitive approach. Indeed, let ω_0, ω_1 and ω_2 be three independent configurations in Γ_Λ with respective intensities

$$\mathrm{d}\sigma_0 := (p\wedge 1)\ \mathrm{d}\sigma, \quad \sigma_1 := \sigma-\sigma_0, \quad \sigma_2 := \tau-\sigma_0.$$

Then $\omega_0+\omega_1$ and $\omega_0+\omega_2$ have respective distribution μ_σ and μ_τ. Hence we have

$$\begin{aligned}\mathcal{T}_{\rho_1}(\mu_\sigma,\mu_\tau) &= \inf\left\{\mathbb{E}\left[\rho_1(\omega,\bar\omega)\right] : \omega\sim\mu_\sigma,\ \bar\omega\sim\mu_\tau\right\}\\ &\le \mathbb{E}\left[\rho_1(\omega_0+\omega_1,\omega_0+\omega_2)\right]\\ &= \mathbb{E}\left[(\omega_1+\omega_2)(\Lambda)\right]\\ &= \int_\Lambda |p(x)-1|\ \mathrm{d}\sigma(x).\end{aligned}$$

Cox processes. A Cox process is a Poisson process with a random intensity. To construct a Cox process, we need to enlarge our probability space. Recall that $\mathfrak{M}(\Lambda)$ is the space of positive and diffuse Radon measures on Λ endowed with the vague topology and the corresponding Borel σ-field. Given an arbitrary probability measure P_M on $\mathfrak{M}(\Lambda)$, we denote by M the canonical random variable on $(\mathfrak{M}(\Lambda),\mathrm{P}_M)$, i.e. M given by $M(m)=m$ has distribution P_M. On the space $\Gamma_\Lambda\times\mathfrak{M}(\Lambda)$, we consider the probability measures

$$\mathrm{d}\mu'_M(\omega,m) :=\ \mathrm{d}\mu_m(\omega)\ \mathrm{dP}_M(m) \quad \text{and} \quad \mathrm{d}\mu'_\sigma(\omega,m) :=\ \mathrm{d}\mu_\sigma(\omega)\ \mathrm{dP}_M(m).$$

Note that the second one is the distribution of the independent couple (N,M), where N is the canonical random variable on Γ_Λ with distribution μ_σ.

As noticed in Section 2.1, the application $m\mapsto\mathcal{T}_{\rho_1}(\mu_m,\mu_\sigma)$ is lower semi-continuous, hence measurable. The distribution μ'_M on Γ_Λ is said to be Cox whenever for any function $f\in\mathcal{C}_0(\Lambda)$,

$$\mathbb{E}_{\mu'_M}\left[\exp\left(\int_\Lambda f\ \mathrm{d}\omega\right)\ \Big|\ M\right] = \exp\left\{\int_\Lambda (e^f-1)\ \mathrm{d}M\right\}.$$

In the definition of the distance between μ'_M and μ'_σ, we do not include any information on M, so that the distance ρ_1 remains the same and we have:

$$\begin{aligned}\mathcal{T}_{\rho_1}(\mu'_\sigma,\mu'_M) &= \sup_{F\in\rho_1-\mathrm{Lip}_1}\int_{\Gamma_\Lambda\times\mathfrak{M}(\Lambda)}F(\omega)\ \mathrm{d}\mu'_\sigma(\omega,m) - \int_{\Gamma_\Lambda\times\mathfrak{M}(\Lambda)}F(\omega)\ \mathrm{d}\mu'_M(\omega,m)\\ &= \sup_{F\in\rho_1-\mathrm{Lip}_1}\int_{\mathfrak{M}(\Lambda)}\left(\int_{\Gamma_\Lambda}F(\omega)\ \mathrm{d}(\mu_\sigma-\mu_m)(\omega)\right)d\mathrm{P}_M(m).\end{aligned}$$

Proposition 4.2. *Assume that μ'_σ-a.s., the measure M is absolutely continuous with respect to σ and that there exists a measurable version of $dM/\ d\sigma$ and such that $dM/\ d\sigma-1\in L^1(\mu'_\sigma\otimes\sigma)$. Then we have*

$$\mathcal{T}_{\rho_1}(\mu'_\sigma,\mu'_M) \le \mathbb{E}_{\mu'_\sigma}\left[\int_\Lambda\left|\frac{dM}{d\sigma}(x)-1\right|\ d\sigma(x)\right].$$

Proof. We have:

$$\begin{aligned}\mathcal{T}_{\rho_1}(\mu'_\sigma,\mu'_M) &\le \int_{\mathfrak{M}(\Lambda)} \sup_{F\in\rho_1-\mathrm{Lip}_1}\left(\int_{\Gamma_\Lambda} F(\omega)\ \mathrm{d}(\mu_\sigma-\mu_m)(\omega)\right)\ \mathrm{dP}_M(m)\\ &= \int_{\mathfrak{M}(\Lambda)} \mathcal{T}_{\rho_1}(\mu_\sigma,\mu_m)\ \mathrm{dP}_M(m)\\ &\le \int_{\mathfrak{M}(\Lambda)}\int_\Lambda \left|\frac{\mathrm{d}m}{\mathrm{d}\sigma}(x)-1\right|\ \mathrm{d}\sigma(x)\ \mathrm{dP}_M(m),\end{aligned}$$

where the last inequality follows from Proposition 4.1. □

Gibbs processes. Let $\Lambda = \mathbb{R}^k$ and assume that the measure ν is a Gibbs measure on Γ_Λ with respect to the reference measure μ_σ, i.e. the density of ν with respect to μ_σ is of the form $L = e^{-V}$, where

$$V(\omega) := \int_\Lambda\int_\Lambda \phi(x-y)\ \mathrm{d}\omega(x)\ \mathrm{d}\omega(y) < +\infty, \quad \mu_\sigma - a.s.,$$

and where the potential $\phi : \Lambda \to (0,+\infty)$ is such that $\phi(x) = \phi(-x)$ and

$$\int_\Lambda\int_\Lambda \phi(x-y)\ \mathrm{d}\sigma(x)\ \mathrm{d}\sigma(y) < +\infty.$$

We have the following result.

Proposition 4.3. *The Rubinstein distance $\mathcal{T}_{\rho_1}$ between the Poisson measure μ_σ and the Gibbs measure ν is bounded as follows:*

$$\mathcal{T}_{\rho_1}(\mu_\sigma,\nu) \le 2\int_\Lambda\int_\Lambda \phi(x-y)\ d\sigma(x)\ d\sigma(y).$$

Proof. Since V is μ_σ-a.s. finite, so does $\int_\Lambda \phi(x-y)\ \mathrm{d}\omega(y)$ for any x. We have:

$$\nabla_x^\sharp L(\omega) = -L(\omega)\left(1-\exp\left\{-2\int_\Lambda \phi(x-y)\ \mathrm{d}\omega(y)\right\}\right), \quad x\in\Lambda.$$

Since $0 \le L \le 1$, Theorem 3.2 together with the inequality $1-e^{-u}\le u$ imply:

$$\begin{aligned}\mathcal{T}_{\rho_1}(\mu_\sigma,\nu) &\le \mathbb{E}_{\mu_\sigma}\left[L\int_\Lambda\left(1-\exp\left\{-2\int_\Lambda \phi(x-y)\ \mathrm{d}\omega(y)\right\}\right)\ \mathrm{d}\sigma(x)\right]\\ &\le \mathbb{E}_{\mu_\sigma}\left[L\int_\Lambda 2\int_\Lambda \phi(x-y)\ \mathrm{d}\omega(y)\ \mathrm{d}\sigma(x)\right]\\ &\le 2\,\mathbb{E}_{\mu_\sigma}\left[\int_\Lambda\int_\Lambda \phi(x-y)\ \mathrm{d}\omega(y)\ \mathrm{d}\sigma(x)\right]\\ &= 2\int_\Lambda\int_\Lambda \phi(x-y)\ \mathrm{d}\sigma(x)\ \mathrm{d}\sigma(y).\end{aligned}$$

The proof is complete. □

Poisson processes on the half-line. In this example, we give a bound on the Rubinstein distance between Poisson processes, with respect to the Wasserstein distance ρ_2. Consider to simplify Poisson processes on $\mathbb{R}_+$ (the generalization to multivariate Poisson processes is straightforward). Letting $U : \mathbb{R}_+ \to \mathbb{R}$ be a continuously differentiable function vanishing at infinity and with $U(0) = 0$, we also assume that $U \in L^2(\lambda)$, where λ is the Lebesgue measure, and that its derivative U' is valued in $(-1, +\infty)$. A typical example of such a function is $U(t) = t/(1+t^3)$, $t \geq 0$. Then we obtain by Theorem 3.4 the following result.

Proposition 4.4. *Let μ_λ be the Poisson measure of Lebesgue intensity λ on the configuration space $\Gamma_{\mathbb{R}_+}$, and consider the Poisson measure ν of intensity $(1 + U')\, d\lambda$. Then we have the upper bound on $\mathcal{T}_{\rho_2}(\mu_\lambda, \nu)$:*

$$\mathcal{T}_{\rho_2}(\mu_\lambda, \nu) \quad \leq \quad \|U\|_{L^2(\lambda)}.$$

4.2. Tail and isoperimetric estimates. The aim of this final part is to derive several consequences of Theorem 3.2 above in terms of tail estimates and isoperimetric inequalities.

Tail estimates. Our main result Theorem 3.2 allows us to obtain a first tail estimate as follows. Let $F \in \rho_1 - \mathrm{Lip}_1$ be centered and let $\lambda > 0$. Denote $Z_\lambda = \mathbb{E}_{\mu_\sigma}\left[e^{\lambda F}\right]$ and consider ν^λ the absolutely continuous probability measure with density $e^{\lambda F}/Z_\lambda$ with respect to μ_σ. Using a somewhat similar argument as in [11], we have:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\lambda} \log Z_\lambda &= \int_{\Gamma_\Lambda} F \; \mathrm{d}\nu^\lambda \\ &\leq \mathcal{T}_{\rho_1}(\mu_\sigma, \nu^\lambda) \\ &\leq \mathbb{E}_{\mu_\sigma}\left[\int_\Lambda |\nabla_x^\sharp e^{\lambda F}| \; \mathrm{d}\sigma(x)\right] \\ &\leq (e^\lambda - 1)\, \|\nabla^\sharp F\|_{1,\infty}, \end{aligned}$$

where in the last inequality we used the fact that the function $x \mapsto (e^x - 1)/x$ is non-decreasing on $(0, +\infty)$. Here the notation $\|\nabla^\sharp F\|_{1,\infty}$ stands for

$$\|\nabla^\sharp F\|_{1,\infty} := \mu_\sigma - \mathrm{esssup} \int_\Lambda |\nabla_x^\sharp F| \; \mathrm{d}\sigma(x).$$

Hence we obtain the following bound on the Laplace transform:

$$\mathbb{E}_{\mu_\sigma}\left[e^{\lambda F}\right] = Z_\lambda \leq \exp\left\{\|\nabla^\sharp F\|_{1,\infty}\,(e^\lambda - \lambda - 1)\right\}, \quad \lambda > 0.$$

Finally using Chebychev's inequality, we get the deviation inequality available for any $r \geq 0$:

$$\mu_\sigma\,(F \geq r) \leq \exp\left\{r - (r + \|\nabla^\sharp F\|_{1,\infty})\, \log\left(1 + \frac{r}{\|\nabla^\sharp F\|_{1,\infty}}\right)\right\}. \tag{4.2}$$

Note that such a tail estimate is somewhat similar to that established for instance by Wu and Houdré-Privault in [19, 11]. However, in contrast to their results, we do not exhibit at the denominator the sharp variance term

$$\|\nabla^\sharp F\|_{2,\infty}^2 := \mu_\sigma - \mathrm{esssup} \int_\Lambda |\nabla_x^\sharp F|^2 \; \mathrm{d}\sigma(x),$$

since our method relies on the L^1-inequality (3.2). In particular, if we apply (4.2) for instance to the centered function $F \in \rho_1 - \mathrm{Lip}_1$ given by $F(\omega) = (\omega - \sigma)(K)$, where K is some compact subset of Λ, we obtain the inequality

$$\mu_\sigma\,(\omega(K) \geq \sigma(K) + r) \leq e^{r - (r + \sigma(K))\, \log\left(1 + \frac{r}{\sigma(K)}\right)}.$$

Unfortunately, neither (4.2) nor the results emphasized in [19, 11] are sharp in terms of the deviation level r since the following asymptotic estimate holds, cf. for instance p.1225 of Houdré [10]:

$$\begin{aligned}\mu_\sigma\left(\omega(K) \geq \sigma(K)+r\right) &= \mu_\sigma\left(\omega(K) \geq [\sigma(K)+r]\right)\\ &\underset{r\to+\infty}{\sim} \frac{e^{[\sigma(K)+r]-\sigma(K)-[\sigma(K)+r]\,\log\left(\frac{[\sigma(K)+r]}{\sigma(K)}\right)}}{\sqrt{2\pi[\sigma(K)+r]}},\end{aligned}$$

where $[R] := \inf\{N \in \mathbb{N}_* : N \geq R\}$ denotes the upper integer part of any positive real number R. Hence the purpose of this part is to recover this multiplicative polynomial factor by means of a simple use of Theorem 3.2. We proceed as follows. Let ν be the absolutely continuous probability measure with density with respect to μ_σ:

$$L := \frac{1}{\mu_\sigma\left(\omega(K) \geq [\sigma(K)+r]\right)}\,\mathbf{1}_{\{\omega(K)\geq[\sigma(K)+r]\}}, \quad r>0.$$

Using Theorem 3.2, we compute as follows:

$$\begin{aligned}&\mu_\sigma\left(\omega(K) \geq \sigma(K)+r\right)\\ &= \mu_\sigma\left(\omega(K) \geq [\sigma(K)+r]\right)\\ &\leq \frac{1}{[\sigma(K)+r]}\left(\int_{\Gamma_\Lambda}\omega(K)\,L(\omega)\,\mathrm{d}\mu_\sigma(\omega)\right)\mu_\sigma\left(\omega(K)\geq[\sigma(K)+r]\right)\\ &\leq \frac{1}{[\sigma(K)+r]}\left(\mathcal{T}_{\rho_1}(\mu_\sigma,\nu)+\sigma(K)\right)\mu_\sigma\left(\omega(K)\geq[\sigma(K)+r]\right)\\ &\leq \frac{1}{[\sigma(K)+r]}\left(\mathbb{E}_{\mu_\sigma}\left[\int_\Lambda|\nabla_x^\sharp L|\;\mathrm{d}\sigma(x)\right]+\sigma(K)\right)\mu_\sigma\left(\omega(K)\geq[\sigma(K)+r]\right)\\ &= \frac{\sigma(K)}{[\sigma(K)+r]}\left(\mu_\sigma\left(\omega(K)=[\sigma(K)+r]-1\right)+\mu_\sigma\left(\omega(K)\geq[\sigma(K)+r]\right)\right),\end{aligned}$$

so that we obtain

$$\begin{aligned}\mu_\sigma\left(\omega(K)\geq\sigma(K)+r\right) &\leq \frac{[\sigma(K)+r]}{[\sigma(K)+r]-\sigma(K)}\,e^{-\sigma(K)}\,\frac{\sigma(K)^{[\sigma(K)+r]}}{[\sigma(K)+r]!}\\ &\leq \frac{[\sigma(K)+r]}{r}\,e^{-\sigma(K)}\,\frac{\sigma(K)^{[\sigma(K)+r]}}{[\sigma(K)+r]!}.\end{aligned}$$

Hence using the lower bound below on the factorial function of any positive integer N, cf. for instance [8]:

$$\sqrt{2\pi}\,N^{N+\frac{1}{2}}\,e^{-N}\;\leq\;N!\;\leq\;\sqrt{2\pi}\,N^{N+\frac{1}{2}}\,e^{-N+\frac{1}{12N}}, \tag{4.3}$$

we obtain the following result.

Proposition 4.5. *Given any compact set $K\subset\Lambda$ and any $r>0$, we have the tail estimate:*

$$\mu_\sigma\left(\omega(K)\geq\sigma(K)+r\right) \;\leq\; \frac{[\sigma(K)+r]}{r}\,\frac{e^{[\sigma(K)+r]-\sigma(K)-[\sigma(K)+r]\,\log\left(\frac{[\sigma(K)+r]}{\sigma(K)}\right)}}{\sqrt{2\pi[\sigma(K)+r]}}.$$

To the knowledge of the authors, although the latter non-asymptotic tail estimate is straightforward to establish via Theorem 3.2 as we have seen above, it seems to be new and recovers exactly the asymptotic regime emphasized above. Note that Paulauskas obtained a somewhat similar deviation inequality in Proposition 3 in [14], but with a constant which is however not sharp, in contrast to ours.

Now we aim at extending this tail estimate to a more general context. Given a fixed configuration $\eta \in \Gamma_\Lambda$, we provide in the sequel a deviation inequality from its mean of the total variation distance ρ_1 between η and random configurations. Assume that σ is a finite measure. Denoting the function $\rho_\eta := \rho_1(\cdot, \eta)$ which clearly belongs to the set $\rho_1 - \mathrm{Lip}_1$ and using the same argument as above, we have

$$\begin{aligned}
&\mu_\sigma\left(\rho_\eta \geq \mathbb{E}_{\mu_\sigma}[\rho_\eta] + r\right) \\
&\quad = \mu_\sigma\left(\rho_\eta \geq [\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r]\right) \\
&\quad \leq \frac{1}{[\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r]} \, \mathbb{E}_{\mu_\sigma}\left[\rho_\eta \, \mathbf{1}_{\{\rho_\eta \geq [\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r]\}}\right] \\
&\quad \leq \frac{1}{[\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r]} \left(\mathbb{E}_{\mu_\sigma}\left[\int_\Lambda |\nabla_x^\sharp \mathbf{1}_{\{\rho_\eta \geq [\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r]\}}| \; d\sigma(x)\right] \right. \\
&\qquad\qquad \left. + \mathbb{E}_{\mu_\sigma}[\rho_\eta] \, \mu_\sigma\left(\rho_\eta \geq [\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r]\right) \right) \\
&\quad \leq \frac{[\sigma(\Lambda) + r] - r}{[\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r]} \left(\mu_\sigma\left(\rho_\eta \geq [\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r - 1]\right) - \mu_\sigma\left(\rho_\eta \geq [\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r]\right) \right) \\
&\qquad\qquad + \frac{1}{[\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r]} \, \mathbb{E}_{\mu_\sigma}[\rho_\eta] \; \mu_\sigma\left(\rho_\eta \geq [\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r]\right),
\end{aligned}$$

since the intensity measure σ is diffuse. Hence we obtain for any $r > 0$:

$$\mu_\sigma\left(\rho_\eta \geq [\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r]\right) \leq \frac{[\sigma(\Lambda) + r] - r}{[\sigma(\Lambda) + r]} \, \mu_\sigma\left(\rho_\eta \geq [\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r - 1]\right),$$

and iterating the procedure entails the inequality

$$\mu_\sigma\left(\rho_\eta \geq [\mathbb{E}_{\mu_\sigma}[\rho_\eta] + r]\right) \leq \frac{\left([\sigma(\Lambda) + r] - r\right)^r [\sigma(\Lambda)]!}{[\sigma(\Lambda) + r]!}.$$

Finally using the estimates (4.3) yield the following result.

Proposition 4.6. *Given any fixed configuration $\eta \in \Gamma_\Lambda$ and provided the intensity measure σ is finite, we have for any $r > 0$:*

$$\begin{aligned}
&\mu_\sigma\left(\rho_\eta \geq \mathbb{E}_{\mu_\sigma}[\rho_\eta] + r\right) \\
&\qquad \leq \frac{\sqrt{2\pi[\sigma(\Lambda)]}\,[\sigma(\Lambda)]^{[\sigma(\Lambda)]} e^{\frac{1}{12[\sigma(\Lambda)]}}}{\sigma(\Lambda)^{\sigma(\Lambda)}} \, \frac{e^{[\sigma(\Lambda)+r] - [\sigma(\Lambda)] - [\sigma(\Lambda)+r] \, \log\left(\frac{[\sigma(\Lambda)+r]}{[\sigma(\Lambda)+r]-r}\right)}}{\sqrt{2\pi[\sigma(\Lambda)+r]}},
\end{aligned}$$

where ρ_η denotes the total variation distance $\rho_1(\cdot, \eta)$.

Hence one deduces that the tail behavior of the total variation distance is comparable to the previous ones, up to constant multiplicative factors depending on the total mass $\sigma(\Lambda)$.

Isoperimetric inequality. Here the distance of interest is the trivial distance ρ_0. In the sequel, we assume that the intensity measure σ is finite, so that the domain $\mathrm{Dom}\,\nabla^\sharp$ contains the indicator functions $\mathbf{1}_A$, $A \in \mathcal{B}(\Gamma_\Lambda)$.

Given a Borel set $A \in \mathcal{B}(\Gamma_\Lambda)$, we define its surface measure as

$$\mu_\sigma(\partial A) := \mathbb{E}_{\mu_\sigma}\left[\int_\Lambda |\nabla_x^\sharp \mathbf{1}_A| \ \mathrm{d}\sigma(x)\right].$$

Denote h_{μ_σ} the classical isoperimetric constant that we aim at estimating:

$$h_{\mu_\sigma} = 2 \inf_{0<\mu_\sigma(A)<1} \frac{\mu_\sigma(\partial A)}{\mu_\sigma(A)(1-\mu_\sigma(A))}.$$

By the following co-area formula, available for any $F \in \mathrm{Dom}\,\nabla^\sharp$:

$$\mathbb{E}_{\mu_\sigma}\left[\int_\Lambda |\nabla_x^\sharp F| \ \mathrm{d}\sigma(x)\right] = \mathbb{E}_{\mu_\sigma}\left[\int_\Lambda \int_{-\infty}^{+\infty} |\nabla_x^\sharp \mathbf{1}_{\{F>t\}}| \ \mathrm{d}t \ \mathrm{d}\sigma(x)\right],$$

which might be deduced from the identity $|a-b| = \int_{-\infty}^{+\infty} |\mathbf{1}_{\{a>t\}} - \mathbf{1}_{\{b>t\}}| \ \mathrm{d}t$, the constant h_{μ_σ} is also the best constant h in the L^1-type functional inequality

$$h\,\mathbb{E}_{\mu_\sigma}\left[|F - \mathbb{E}_{\mu_\sigma}[F]|\right] \le 2\,\mathbb{E}_{\mu_\sigma}\left[\int_\Lambda |\nabla_x^\sharp F| \ \mathrm{d}\sigma(x)\right], \quad F \in \mathrm{Dom}\,\nabla^\sharp. \tag{4.4}$$

We have the following result, which is convenient for small total mass $\sigma(\Lambda)$.

Proposition 4.7. *Assume that the measure σ is finite. Then we have*

$$1 \le h_{\mu_\sigma} \le \frac{\sigma(\Lambda)}{1 - e^{-\sigma(\Lambda)}}. \tag{4.5}$$

In particular, we have the asymptotic for small total mass:

$$\lim_{\sigma(\Lambda)\to 0} h_{\mu_\sigma} = 1.$$

Remark 4.8. Note that Houdré and Privault established first the inequality $h_{\mu_\sigma} \ge 1$ by using Poincaré inequality, cf. Proposition 6.4 in [12]. Hence we recover their result via another approach. On the other hand, our estimate in the right-hand-side of (4.5) is sharp for small values of $\sigma(\Lambda)$, but is worse than their estimate for large $\sigma(\Lambda)$ since their upper bound is $8 + 8\sqrt{\sigma(\Lambda)}$.

Proof. In order to show $h_{\mu_\sigma} \ge 1$, let us establish the inequality (4.4) with $h = 1$. By homogeneity, it is sufficient to prove the result for functionals $F \in \mathrm{Dom}\,\nabla^\sharp$ such that $\mathbb{E}_{\mu_\sigma}[F] = 1$. Denote by ν the absolutely continuous probability measure with density F with respect to the Poisson measure μ_σ. Using duality,

$$\begin{aligned}
\mathcal{T}_{\rho_0}(\mu_\sigma, \nu) &= \sup_{G \in \rho_0-\mathrm{Lip}_1} \mathbb{E}_{\mu_\sigma}[G(F-1)] \\
&= \frac{1}{2} \sup_{\mu_\sigma-\mathrm{esssup}\,|G|\le 1} \mathbb{E}_{\mu_\sigma}[G(F-1)] \\
&= \frac{1}{2}\,\mathbb{E}_{\mu_\sigma}[|F-1|].
\end{aligned}$$

Hence using Theorem 3.2 with the trivial distance ρ_0, we get the inequality (4.4) with $h = 1$, thus obtaining the desired inequality $h_{\mu_\sigma} \geq 1$. On the other hand, to provide the upper bound in (4.5), note that we have by the very definition of h_{μ_σ}:

$$\begin{aligned} h_{\mu_\sigma} &\leq \frac{2\,\mu_\sigma(\partial\{\omega(\Lambda)=0\})}{\mu_\sigma(\omega(\Lambda)=0)\,(1-\mu_\sigma(\omega(\Lambda)=0))} \\ &= \frac{\sigma(\Lambda)}{1-e^{-\sigma(\Lambda)}}. \end{aligned}$$

The proof is achieved. □

References

1. Albeverio, S., Kondratiev, Y. G., and Röckner, M.: Analysis and geometry on configuration spaces, *J. Funct. Anal.* **154** (1998) 444–500.
2. Barbour, A. D., Brown, T. C., and Xia, A.: Point processes in time and Stein's method, *Stochastics Stochastics Rep.* **65** (1998) 127–151.
3. Barbour, A. D., Holst, L., and Janson, S.: *Poisson Approximation*, Oxford Studies in Probability, The Clarendon Press Oxford University Press, New York, 1992.
4. Bobkov, S. G. and Ledoux, M.: On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures, *J. Funct. Anal.* **156** (1998) 347–365.
5. Decreusefond, L.: Perturbation Analysis and Malliavin Calculus, *Ann. Appl. Probab.* **8** (1998) 496–523.
6. Decreusefond, L.: Wasserstein distance on configurations space, *Potential Anal.* **28** (2008) 283–300.
7. Dermoune, A., Krée, P., and Wu, L.: Calcul stochastique non adapté par rapport à la mesure aléatoire de Poisson, Séminaire de Probabilités XXII, *Lecture Notes in Math.* **1321** (1988) 477–484, Springer, Berlin.
8. Feller, W.: *An Introduction to Probability Theory and its Applications*, Third edition, Wiley, New York, 1968.
9. Feyel, D. and Üstünel, A. S.: Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space, *Probab. Theory Related Fields* **128** (2004) 347–385.
10. Houdré, C.: Remarks on deviation inequalities for functions of infinitely divisible random vectors, *Ann. Probab.* **30** (2002) 1223–1237.
11. Houdré, C. and Privault, N.: Concentration and deviation inequalities in infinite dimensions via covariance representations, *Bernoulli* **8** (2002) 697–720.
12. Houdré, C. and Privault, N.: Isoperimetric and related bounds on configuration spaces, *Statist. Probab. Lett.* **78** (2008) 2154–2164.
13. Nualart, D. and Vives, J.: Anticipative calculus for the Poisson process based on the Fock space, Séminaire de probabilités XXIV, *Lecture Notes in Math.* **1426** (1988) 154–165, Springer, Berlin.
14. Paulauskas, V.: Some comments on inequalities for deviations for infinitely divisible random vectors, *Lithuanian Math. J.* **42** (2002) 394–410.
15. Röckner, M. and Schied, A.: Rademacher's theorem on configuration spaces and applications, *J. Funct. Anal.* **169** (1999) 325–356.
16. Ruiz de Chavez, J.: Espaces de Fock pour les processus de Wiener et de Poisson, Séminaire de probabilités XIX, *Lecture Notes in Math.* **1123** (1985) 230–241, Springer, Berlin.
17. Schuhmacher, D.: *Estimation of Distances Between Point Process Distributions*, Ph.D. Thesis, Universität Zürich, 2005.
18. Villani, C.: *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften, Springer, Berlin, 2009.
19. Wu, L.: A new modified logarithmic Sobolev inequality for Poisson point processes and several applications, *Probab. Theory Related Fields* **118** (2000) 427–438.

Laurent Decreusefond: Institut TELECOM, TELECOM ParisTech, CNRS LTCI, Paris, France
E-mail address: Laurent.Decreusefond@telecom-paristech.fr

Aldéric Joulin: Université de Toulouse, Institut National des Sciences Appliquées, Institut de Mathématiques de Toulouse, F-31077 Toulouse, France
E-mail address: alderic.joulin@math.univ-toulouse.fr

Nicolas Savy: Université de Toulouse, Université Paul Sabatier, Institut de Mathématiques de Toulouse, F-31062 Toulouse, France
E-mail address: nicolas.savy@math.univ-toulouse.fr