



Upper bounds on Rubinstein distances on configuration spaces and applications

Laurent Decreusefond, Aldéric Joulin, Nicolas Savy

► To cite this version:

Laurent Decreusefond, Aldéric Joulin, Nicolas Savy. Upper bounds on Rubinstein distances on configuration spaces and applications. 2010. hal-00347899v1

HAL Id: hal-00347899

<https://hal.science/hal-00347899v1>

Preprint submitted on 17 Dec 2008 (v1), last revised 22 Mar 2010 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RUBINSTEIN DISTANCES ON CONFIGURATION SPACES

L. DECREUSEFOND, A. JOULIN, AND N. 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 investigation of our results to functional inequalities completes this work.

1. INTRODUCTION

Let Λ be a Lusin 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 define distances between point processes (see for instance Chapters 2 and 3 of [15] for a thorough discussion and references about this topic). Once a lower-semi-continuous distance ρ on the configuration space is defined, the Rubinstein distance between μ and some probability measure ν on Γ_Λ is given by

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

where $\Sigma(\mu, \nu)$ is the set of probability measures on $\Gamma_\Lambda \times \Gamma_\Lambda$ with first (respectively second) marginal μ (respectively ν). According to the Kantorovich-Rubinstein duality theorem, the distance rewrites as

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

where $\rho - \text{Lip}_1$ denotes the set of 1-Lipschitz functions on Γ_Λ with respect to the distance ρ . This means that the distance between two fixed probability measures depends crucially on the distance on the base space as it changes the set of Lipschitz functions. Since $\mathcal{T}_\rho(\mu, \nu)$ is given by a variational formula, its explicit expression is of difficult access and one might estimate it from above. In particular, a convenient upper bound ensures its finiteness, which is not guaranteed a priori. Such inequalities involving the Rubinstein distance like the ones we are intended to prove is part of 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 [16] and particularly Chapters 21 and 22 for a large

Date: December 17, 2008.

1991 Mathematics Subject Classification. 60G55, 60H07, 60E15.

Key words and phrases. Configuration space, Poisson measure, Rubinstein distance, Malliavin derivative, Rademacher property, functional inequalities.

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, one way to get such representations is either to embed the two probability measures into the evolution of a Markov semi-group, or to use 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 $\rho - \text{Lip}_1$ and any function in $\rho - \text{Lip}_1$ 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 Malliavin calculus. In fact, we even have two notions of gradient: a “differential” gradient (see [1, 13]) and a gradient expressed as a finite difference operator (see [11]). 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 an abstract upper bound on the Rubinstein distance $\mathcal{T}_\rho(\mu, \nu)$ via a semi-group approach, where ρ is the total variation distance, the Wasserstein distance or the trivial distance on the configuration space Γ_Λ . When dealing with the total variation distance, such an estimate has a simplified expression, which can be retrieved in Section 4 by using an alternative method, namely the Clark formula. When the configuration space is equipped with the Wasserstein distance, cf. Section 5, the upper bound we give on the Rubinstein distance relies on a time-change argument together with the Girsanov Theorem. Finally, Section 6 is devoted to numerous applications: estimating 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 consequences of our results to various functional inequalities such as log-Sobolev, concentration or isoperimetric inequalities.

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

$$(1) \quad \mathcal{T}_\rho(\mu, \nu) = \inf_{\gamma \in \Sigma(\mu, \nu)} \int_X \int_X \rho(x, y) \, d\gamma(x, y),$$

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 [16], 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 [16],

this minimum is equal to

$$(2) \quad \mathcal{T}_\rho(\mu, \nu) = \sup_{\substack{F \in \rho - \text{Lip}_1 \\ F \in L^1(\mu + \nu)}} \int_X F \, d(\mu - \nu),$$

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

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

In the context of optimal transportation, \mathcal{T}_ρ 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 [16]).

In this paper, we consider the situation where $X = \Gamma_\Lambda$ is the configuration space on a Lusin space Λ , i.e.,

$$\Gamma_\Lambda = \{\omega \subset \Lambda; \omega \cap K \text{ is a finite set for every compact } K \subset \Lambda\}.$$

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 $f \in \mathcal{C}_0$ (continuous with compact support on Λ), the following maps

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

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. Given σ a σ -finite measure on the Borel σ -algebra $\mathcal{B}(\Lambda)$, the probability space under consideration in the remainder of this paper will then 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

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

for any measurable function $f : \Lambda \rightarrow (-\infty, 0)$. Here \mathbb{E}_{μ_σ} 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 [15] 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 $\omega \Delta \eta = \omega \setminus (\omega \cap \eta)$.

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

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

where $\Sigma(\omega, \eta)$ denotes the set of configurations $\beta \in \Gamma_{\Lambda \times \Lambda}$ having marginals ω and η , see [6, 13].

Let us comment 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 is required. This is the object of the next lemma.

Lemma 1. *The distances ρ_i are 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 [13] 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

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

Let $((\omega_n, \eta_n), n \geq 1)$ 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 :

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

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, the monotone convergence theorem for an exhaustive sequence of compacts $(K_p)_{p \in \mathbb{N}}$ entails that

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

hence the set J_{α} is vaguely closed. \square

In particular, Lemma 1 entails the lower-semi-continuity of the Rubinstein distances \mathcal{T}_{ρ_i} with respect to the weak topology on the space of probability measures on Γ_Λ , cf. for instance Remark 6.12 in [16]. However for $i \in \{1, 2\}$, the Rubinstein distances \mathcal{T}_{ρ_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 [8].

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, 15]. As mentioned above, it is customary to use the classical notion of total variation by considering normalized configurations, i.e.,

$$\tilde{\rho}_1(\omega, \eta) = \rho_1 \left(\frac{\omega}{\omega(\Lambda)}, \frac{\eta}{\eta(\Lambda)} \right),$$

provided both configurations have finite total masses. It should be noted that since $\tilde{\rho}_1$ is not lower-semi-continuous, the Kantorovich-Rubinstein duality theorem is no longer satisfied, so that we cannot use the identity (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 $\tilde{\rho}_1(\omega, \eta) = 2$ whereas $\tilde{\rho}_1(\omega_n, \eta_n) = 1$, for any integer $n \geq 2$.

It is also customary to replace ρ_2 by $\tilde{\rho}_2$ defined by

$$\tilde{\rho}_2(\omega, \eta) = \begin{cases} \frac{1}{\omega(\Lambda)} \rho_2(\omega, \eta) & \text{if } \omega(\Lambda) = \eta(\Lambda) \neq 0, \\ |\omega(\Lambda) - \eta(\Lambda)| & \text{otherwise.} \end{cases}$$

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 important, 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 3 below), not shared by $\tilde{\rho}_2$.

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 I. *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 σ -finite 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 $\text{Dom } \nabla$ of $L^2(\mu_\sigma)$, such that for any $F \in \text{Dom } \nabla$,

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

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

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

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

$$(3) \quad \int_{\Gamma_\Lambda} \int_{\Lambda} \nabla_x F(\omega) u(\omega, x) d\alpha(\omega, x) = \int_{\Gamma_\Lambda} F(\omega) \delta u(\omega) d\mu_\sigma(\omega).$$

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

Definition 1. *Given a distance ρ and a gradient ∇ on Γ_Λ , we say that the couple (∇, ρ) has the Rademacher property whenever $\rho - \text{Lip}_1 \subset \text{Dom } \nabla$ and*

$$|\nabla_x F(\omega)| \leq 1, \quad \alpha\text{-a.e.}$$

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 $\nabla^\# F$, is defined by

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

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

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

We set $Q^\#(\omega, dx) = d\sigma(x)$ so that $\alpha^\# = \mu_\sigma \otimes \sigma$. 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

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

where $\Delta_n = \{(x_1, \dots, x_n) \in \Lambda^n, x_i \neq x_j, i \neq j\}$, and we have the isometry formula

$$(4) \quad \mathbb{E}_{\mu_\sigma} [J_n(f_n) J_m(f_m)] = n! \mathbf{1}_{\{n=m\}} \int_{\Lambda^n} f_n f_m d\sigma^{\otimes n}.$$

According to [14, 11], 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 the discrete gradient acts on multiple stochastic integrals as

$$\nabla_x^\# F = \sum_{n=1}^{+\infty} n J_{n-1}(f_n(\cdot, x)), \quad \mu_\sigma \otimes \sigma\text{-a.e.}$$

Denote $\delta^\#$ the adjoint operator of $\nabla^\#$ in the sense of (3). Then the self-adjoint number operator $\mathcal{L}^\# = \delta^\# \nabla^\#$ has the following expression in terms of chaos:

$$\mathcal{L}^\# F = \sum_{n=1}^{+\infty} n J_n(f_n), \quad \text{if } F = \mathbb{E}_{\mu_\sigma} [F] + \sum_{n=1}^{+\infty} J_n(f_n).$$

Moreover, the associated Ornstein-Uhlenbeck semi-group $(P_t^\#)_{t \geq 0}$, whose infinitesimal generator is $-\mathcal{L}$, acts on the multiple stochastic integrals as follows:

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

Using the Chaotic Representation Property, we have the commutation relation between gradient and semi-group:

$$(5) \quad \nabla_x^\# P_t^\# F = e^{-t} P_t^\# \nabla_x^\# F, \quad x \in \Lambda, \quad t \geq 0.$$

Such a property will be useful in the sequel.

By the isometry formula (4), the semi-group is ergodic with respect to the Poisson measure μ_σ , in the sense that the semi-group converges in $L^2(\mu_\sigma)$ to μ_σ .

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

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

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

Since σ is finite, it follows that

$$\int_\Lambda |\nabla_x^\# F(\omega)|^2 d\sigma(x) \leq \sigma(\Lambda),$$

hence that F belongs to $\text{Dom } \nabla^\#$. \square

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 $|\nabla_x^\# F(\omega)| \leq 1$, $\alpha^\#$ -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 Γ_Λ which is a derivation, see [1, 13]. Let $V(\Lambda)$ be the space of \mathcal{C}^∞ 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

$$(6) \quad \begin{cases} \frac{d}{dt} \mathcal{V}_t^v(x) &= v(\mathcal{V}_t^v(x)), \\ \mathcal{V}_0^v(x) &= x. \end{cases}$$

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

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

We then 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

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

We denote by δ^c the adjoint operator of ∇^c in the sense of (3). 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 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 3. *The couple (∇^c, ρ_2) satisfies the Rademacher property.*

Proof. The proof is straightforward. Indeed, letting $F \in \rho_2 - \text{Lip}_1$, we know from Theorem 1.3 in [13] that $F \in \text{Dom } \nabla^c$ and that

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

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

3. BOUNDS ON \mathcal{T}_{ρ_1} AND \mathcal{T}_{ρ_2} RUBINSTEIN DISTANCES VIA SEMI-GROUPS

Let us establish first an abstract upper bound on the Rubinstein distance by using a semi-group method, provided the associated pair gradient/distance satisfies the Rademacher property. Denote ρ one of the distances ρ_i associated to the gradient ∇ in the sense of the Rademacher property of Lemmas 2 or 3. To unify the notation, we use the notation of Hypothesis I, with also $\mathcal{L} = \delta \nabla$, and P_t denotes either the semi-group P_t^\sharp or P_t^c .

Theorem 1. *Let L be the density of an absolutely continuous probability measure ν with respect to μ_σ . Then the following upper bound on the Rubinstein distance holds:*

$$(7) \quad \mathcal{T}_\rho(\mu_\sigma, \nu) \leq \mathbb{E}_{\mu_\sigma} \left[\int_{\Lambda} \left| \int_0^{+\infty} \nabla_x P_t L \, dt \right| \, d\sigma(x) \right],$$

provided the inequality makes sense.

Proof. Since the σ -finite case might be established by a simple limiting procedure (use the lower-semi-continuity of the Rubinstein distance), let us assume that σ is finite for simplicity, so that Lemma 2 is verified if $(\nabla, \rho) = (\nabla^\sharp, \rho_0)$ or (∇^\sharp, ρ_1) , as Lemma 3 in full generality if $(\nabla, \rho) = (\nabla^c, \rho_2)$.

Letting $F \in \rho - \text{Lip}_1$, we have by reversibility and using Fubini's theorem:

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

Using then the Rademacher property, the result holds by taking the supremum over all functions $F \in \rho - \text{Lip}_1$. \square

The upper bound in the inequality (7) is interesting in its own right, but seems to be somewhat difficult to compute in full generality. One notices that such an estimate might be simplified once a commutation relation is obtained between gradient and semi-group. As mentioned above, it is only verified in the case of the discrete gradient, so that we focus now on the pair (∇^\sharp, ρ_1) .

Theorem 2. *Let L be the density of an absolutely continuous probability measure ν with respect to μ_σ , and assume that $L \in \text{Dom } \nabla^\sharp$ and $\nabla^\sharp L \in L^1(\mu_\sigma \otimes \sigma)$. Then we get the following estimate:*

$$(8) \quad \mathcal{T}_{\rho_1}(\mu_\sigma, \nu) \leq \mathbb{E}_{\mu_\sigma} \left[\int_\Lambda |\nabla_x^\sharp L| \, d\sigma(x) \right].$$

The same inequality also holds under the distance ρ_0 .

Proof. Using the commutation relation (5) in Theorem 1 above, we obtain

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

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 . \square

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

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

Applying then such an identity in the first inequality of the proof of Theorem 2 above gives the following bound:

$$(9) \quad \mathcal{T}_{\rho_1}(\mu_\sigma, \nu) \leq \mathbb{E}_{\mu_\sigma} \left[\int_{\Lambda} |(\text{Id} + \mathcal{L}^\sharp)^{-1} \nabla_x^\sharp L| \, d\sigma(x) \right].$$

It seems theoretically slightly better than the upper bound of Theorem 2 but often yields to intractable computations, except when the chaos representation of L is given, as noticed in Section 6.1 below. Note that the very analog of (9) on Wiener space was proved by a different though related way in Theorem 3.2 of [8].

4. CLARK FORMULA AND THE \mathcal{T}_{ρ_1} RUBINSTEIN DISTANCE

In this part, we provide another method leading to Theorem 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 the line with marks in an abstract Lusin space. Borrowing an idea of [17, 12], we first explain how to embed a Poisson process into a multivariate Poisson process.

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

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

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

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

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

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

In particular, if $\nabla_{t,x}^\sharp$ is the discrete gradient on the enlarged configuration space $\Gamma_{\hat{\Lambda}}$, then we have clearly $\nabla_{t,x}^\sharp \hat{F}(\hat{\omega}) = \nabla_x^\sharp F(\pi\hat{\omega})$ for any $(t, x) \in \hat{\Lambda}$. Moreover, we have $\mathbb{E}_{\hat{\mu}}[\hat{F}] = \mathbb{E}_{\mu_\sigma}[F]$ since the image measure of $\hat{\mu}$ by π is μ_σ .

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

$$\hat{\rho}_1(\hat{\omega}, \hat{\eta}) = \sum_{(t,x) \in \hat{\Lambda}} |\hat{\omega}(\{t, x\}) - \hat{\eta}(\{t, x\})|.$$

The key point is the following lemma.

Lemma 4. *For any $F \in \rho_1 - \text{Lip}_1$, the functional \hat{F} belongs to $\hat{\rho}_1 - \text{Lip}_1$.*

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

$$\begin{aligned}
|\hat{F}(\hat{\omega}) - \hat{F}(\hat{\eta})| &= |F(\pi\hat{\omega}) - F(\pi\hat{\eta})| \\
&\leq \rho_1(\pi\hat{\omega}, \pi\hat{\eta}) \\
&= \sum_{x \in \Lambda} |\pi\hat{\omega}(\{x\}) - \pi\hat{\eta}(\{x\})| \\
&= \sum_{x \in \Lambda} \left| \sum_{t \in [0,1]} \hat{\omega}(\{t, x\}) - \hat{\eta}(\{t, x\}) \right| \\
&\leq \sum_{(t,x) \in \hat{\Lambda}} |\hat{\omega}(\{t, x\}) - \hat{\eta}(\{t, x\})| \\
&= \hat{\rho}_1(\hat{\omega}, \hat{\eta}).
\end{aligned}$$

The proof is complete. \square

Let us recall the Clark formula, cf. for instance [7] or Lemma 1.3 in [17], which states that every functional $G : \Gamma_{\hat{\Lambda}} \rightarrow \mathbb{R}$ belonging to $\text{Dom } \nabla^\sharp$ might be written as

$$(10) \quad G = \mathbb{E}_{\hat{\mu}}[G] + \int_0^1 \int_{\Lambda} \mathbb{E}_{\hat{\mu}} \left[\nabla_{t,x}^\sharp G \mid \mathfrak{F}_{t-} \right] d(\hat{\omega} - \lambda \otimes \sigma)(t, x).$$

Now we are able to give a second proof of Theorem 2 by means of the Clark formula.

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

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

Now using the Clark formula (10) and taking expectation with respect to $\hat{\nu}$,

$$\begin{aligned}
\mathbb{E}_{\hat{\nu}} [\hat{F}] &= \mathbb{E}_{\hat{\mu}} [\hat{F}] + \mathbb{E}_{\hat{\nu}} \left[\int_0^1 \int_{\Lambda} \mathbb{E}_{\hat{\mu}} \left[\nabla_{t,x}^\sharp \hat{F} \mid \mathfrak{F}_{t-} \right] d(\hat{\omega} - \lambda \otimes \sigma)(t, x) \right] \\
&= \mathbb{E}_{\hat{\mu}} [\hat{F}] + \mathbb{E}_{\hat{\mu}} \left[\hat{L} \int_0^1 \int_{\Lambda} \mathbb{E}_{\hat{\mu}} \left[\nabla_{t,x}^\sharp \hat{F} \mid \mathfrak{F}_{t-} \right] d(\hat{\omega} - \lambda \otimes \sigma)(t, x) \right] \\
&= \mathbb{E}_{\hat{\mu}} [\hat{F}] + \mathbb{E}_{\hat{\mu}} \left[\int_0^1 \int_{\Lambda} \mathbb{E}_{\hat{\mu}} \left[\nabla_{t,x}^\sharp \hat{F} \mid \mathfrak{F}_{t-} \right] \nabla_{t,x}^\sharp \hat{L} dt d\sigma(x) \right].
\end{aligned}$$

By Lemma 2, the pair $(\nabla^\sharp, \hat{\rho}_1)$ satisfies the Rademacher property on $\Gamma_{\hat{\Lambda}}$. Hence Lemma 4 implies that for $F \in \rho_1 - \text{Lip}_1$, the quantity $\left| \mathbb{E}_{\hat{\mu}} \left[\nabla_{t,x}^\sharp \hat{F} \mid \mathfrak{F}_{t-} \right] \right|$ is bounded by 1, $\lambda \otimes \sigma$ -a.e., so that we obtain finally

$$\begin{aligned}
\mathcal{T}_{\rho_1}(\mu_\sigma, \nu) &\leq \mathbb{E}_{\hat{\mu}} \left[\int_0^1 \int_{\Lambda} |\nabla_{t,x}^\sharp \hat{L}| dt d\sigma(x) \right] \\
&= \mathbb{E}_{\mu_\sigma} \left[\int_{\Lambda} |\nabla_x^\sharp L| d\sigma(x) \right].
\end{aligned}$$

The second proof of Theorem 2 is thus complete. \square

5. TIME-CHANGING AND THE \mathcal{T}_{ρ_2} RUBINSTEIN DISTANCE

As we have seen in Section 3, Theorem 1 entails an upper bound on the \mathcal{T}_{ρ_2} Rubinstein distance which is not really tractable, since no commutation relation has been established yet between the differential gradient ∇^c and the semi-group P_t^c . Hence the purpose of this section is to provide another estimate on \mathcal{T}_{ρ_2} through a different approach relying on a time-change argument together with the Girsanov theorem.

We consider the notation of Section 4, with the slight difference that the state space is now $\widehat{\Lambda} := [0, \infty) \times \Lambda$. In this part, the distance of interest on the enlarged configuration space $\Gamma_{\widehat{\Lambda}}$ is the Wasserstein distance:

$$\widehat{\rho}_2(\widehat{\omega}, \widehat{\eta})^2 = \inf_{\beta \in \Sigma(\widehat{\omega}, \widehat{\eta})} \int_{\widehat{\Lambda}} \int_{\widehat{\Lambda}} (\kappa(x, y)^2 + |t - s|^2) d\beta((s, x), (t, y)).$$

Theorem 3. *Let L be the (positive) density of an absolutely continuous probability measure $\widehat{\nu}$ with respect to $\widehat{\mu}$. Then we get the following upper bound on the Rubinstein distance $\mathcal{T}_{\widehat{\rho}_2}(\widehat{\mu}, \widehat{\nu})$, provided the inequality makes sense:*

$$(11) \quad \begin{aligned} \mathcal{T}_{\widehat{\rho}_2}(\widehat{\mu}, \widehat{\nu})^2 &\leq \mathbb{E}_{\widehat{\mu}} \left[L \int_{\Lambda} \int_0^{+\infty} \left| \int_0^t u(s, z) ds \right|^2 (1 + u(t, z)) dt d\sigma(z) \right] \\ &= \mathbb{E}_{\widehat{\mu}} \left[L \int_{\Lambda} \int_0^{+\infty} |r - v^{-1}(r, z)|^2 dr d\sigma(z) \right], \end{aligned}$$

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

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

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

Note that for $z \in \Lambda$ fixed, the term $\int_0^{+\infty} |r - v^{-1}(r, z)|^2 dr$ can be interpreted as a (generalized) Wasserstein distance between the infinite measures dr and $(1 + u(r, z)) dr$, see [16]. Then, the $\widehat{\rho}_2$ distance is bounded by the expectation under $\widehat{\nu}$, of these distances, 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

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

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

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

where \mathcal{E} denotes the classical Doleans-Dade exponential. On the other hand, the Clark formula (10) - extended to the set $(0, +\infty)$ - induces that

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

By identification, we obtain:

$$u(s, z) = \frac{1}{L_{s-}} \mathbb{E} \left[\nabla_{s,z}^{\sharp} L_t | \mathfrak{F}_{s-} \right] = \frac{1}{L_{s-}} \mathbb{E} \left[\nabla_{s,z}^{\sharp} L | \mathfrak{F}_{s-} \right],$$

since for any $s \in (0, t)$, we have a commutation relation between the discrete gradient $\nabla_{s,z}^{\sharp}$ and the conditional expectation knowing \mathfrak{F}_t , cf. for instance Lemma 3.2 in [11].

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:

$$\begin{aligned} \mathcal{T}_{\widehat{\rho}_2}(\widehat{\mu}, \widehat{\nu}) &\leq \mathbb{E}_{\widehat{\nu}} [\widehat{\rho}_2(\widehat{\omega}, \tau\widehat{\omega})] \\ &\leq \mathbb{E}_{\widehat{\nu}} \left[\int_{\Lambda} \int_0^{+\infty} |t - v(t, z)|^2 d\widehat{\omega}(t, z) \right]^{1/2} \\ &= \mathbb{E}_{\widehat{\mu}} \left[L \int_{\Lambda} \int_0^{+\infty} |t - v(t, z)|^2 \frac{dv}{dt}(t, z) dt d\sigma(z) \right]^{1/2} \\ &= \mathbb{E}_{\widehat{\mu}} \left[L \int_{\Lambda} \int_0^{+\infty} |r - v^{-1}(r, z)|^2 dr d\sigma(z) \right]^{1/2}, \end{aligned}$$

by the change of variable $r = v(t, z)$ for $z \in \Lambda$ being fixed. The proof of inequality (11) is finished. \square

6. APPLICATIONS

6.1. Distances estimates between processes. The purpose of the present part is to apply our main results Theorems 2 and 3 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, 15] 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 on a Lusin space. Here the probability measure ν is supposed to be another Poisson measure on Γ_{Λ} .

Theorem 4. *Let μ_{τ} be a Poisson measure on Γ_{Λ} of intensity τ . We assume that τ admits an integrable density p with respect to σ . Then we have*

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

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

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

It is then straightforward that $\nabla_x^\# L = L(p(x) - 1)$, hence by Theorem 2,

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

The proof is achieved. \square

Note that in this very simple situation, the inequality (9) yields to the same bound. Indeed, since p is deterministic, the density L has the following chaos representation

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

cf. identity (7) in [14], so that we have

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

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. Denote by $\mathfrak{M}(\Lambda)$ the space of diffuse Radon measures on Λ endowed with the vague topology and the corresponding Borel σ -field. We denote by M the canonical random variable on $\mathfrak{M}(\Lambda)$ and P_M its distribution. On the space $\Gamma_\Lambda \times \mathfrak{M}(\Lambda)$, we consider $d\mu'_M(\omega, m) = d\mu_m(\omega) dP_M(m)$ and $d\mu'_\sigma(\omega, m) = d\mu_\sigma(\omega) dP_M(m)$. We assume that the application $m \mapsto \mu_m$ is measurable from $\mathfrak{M}(\Lambda)$ into the set of probability measures on Γ_Λ . It then follows that the application $m \mapsto \mathcal{T}_{\rho_i}(\mu_m, \mu_\sigma)$ is lower-semi-continuous, hence measurable. The distribution μ'_M on Γ_Λ is said to be Cox whenever

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

for any measurable function $f : \Lambda \rightarrow (-\infty, 0)$. 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:

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

Theorem 5. Assume that μ'_σ -a.s., the measure M is absolutely continuous with respect to σ and that there exists a measurable version of $dM/d\sigma$. Then we have

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

Proof. We have:

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

where the last inequality follows from Theorem 4. \square

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

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

and where the potential $\phi : \Lambda \rightarrow (0, +\infty)$ is such that $\phi(x) = \phi(-x)$. We have the following result.

Theorem 6. *The Rubinstein distance \mathcal{T}_{ρ_1} between the Poisson measure μ_σ and the Gibbs measure ν is bounded as follows:*

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

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

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

Since $0 \leq L \leq 1$, Theorem 2 together with the inequality $1 - e^{-u} \leq u$ imply:

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

The proof is finished. \square

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 be a continuous function on \mathbb{R}_+ such that $u(t) \sim t^{-\alpha}$ for t going to infinity and $\alpha > 3/2$, we have clearly that

$$\int_0^{+\infty} \left(\int_0^t u(s) \, ds \right)^2 (1 + u(t)) \, dt < +\infty.$$

Thus we obtain by Theorem 3 the following result.

Theorem 7. *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 $\mathcal{T}_{\rho_2}(\mu_\lambda, \nu)$:*

$$\mathcal{T}_{\rho_2}(\mu_\lambda, \nu)^2 \leq \int_0^{+\infty} \left(\int_0^t u(s) ds \right)^2 (1+u(t)) dt.$$

6.2. Applications to functional inequalities. The aim of this final part is to derive several functional inequalities such as log-Sobolev, concentration and isoperimetric inequalities by means of Theorem 2 above.

Modified log-Sobolev inequality and concentration. Define the entropy of a smooth functional G with respect to the Poisson measure μ_σ by

$$\text{Ent}_{\mu_\sigma}(G) = \int_{\Gamma_\Lambda} G \log \left(\frac{G}{\int_{\Gamma_\Lambda} G d\mu_\sigma} \right) d\mu_\sigma.$$

Theorem 8. *Letting $\lambda > 0$ and $F \in \rho_1 - \text{Lip}_1$ such that $e^{\lambda F} \in \text{Dom } \nabla^\sharp$, we have the L^1 -modified log-Sobolev inequality:*

$$(12) \quad \text{Ent}_{\mu_\sigma}(e^{\lambda F}) \leq \lambda^2 e^\lambda \mathbb{E}_{\mu_\sigma} \left[\int_\Lambda |\nabla_x^\sharp F| e^{\lambda F} d\sigma(x) \right],$$

Proof. By homogeneity, it is sufficient to establish the inequality for functionals $F \in \rho_1 - \text{Lip}_1$ such that $\mathbb{E}_{\mu_\sigma}[e^{\lambda F}] = 1$. Note that we have in this case $\mathbb{E}_{\mu_\sigma}[F] \leq 0$ by Jensen's inequality. Denoting ν^λ the absolutely continuous probability measure with density $e^{\lambda F}$ with respect to μ_σ , we have by Theorem 2:

$$\begin{aligned} \text{Ent}_{\mu_\sigma}(e^{\lambda F}) &= \lambda \int_{\Gamma_\Lambda} F d\nu^\lambda \\ &\leq \lambda \mathcal{T}_\rho(\mu_\sigma, \nu^\lambda) + \lambda \mathbb{E}_{\mu_\sigma}[F] \\ &\leq \lambda \mathcal{T}_\rho(\mu_\sigma, \nu^\lambda) \\ &\leq \lambda \mathbb{E}_{\mu_\sigma} \left[\int_\Lambda |\nabla_x^\sharp e^{\lambda F}| d\sigma(x) \right] \\ &\leq \lambda^2 e^\lambda \mathbb{E}_{\mu_\sigma} \left[\int_\Lambda |\nabla_x^\sharp F| e^{\lambda F} d\sigma(x) \right], \end{aligned}$$

where we used the elementary inequality $e^x - 1 \leq xe^x$, $x \geq 0$. \square

Actually, such a modified log-Sobolev inequality has been established in [17], through the inequality (3.4), with the slight difference that the gradient is replaced by its square, i.e. a L^2 -type gradient is considered. However, we cannot expect a L^2 -behavior in the log-Sobolev inequality (12) since our method relies on the L^1 -inequality (8). The same remark holds for the concentration results below.

As a by-product of the so-called Herbst's method applied to the modified log-Sobolev inequality above, see e.g. Propositions 10 and 11 in [4] or Proposition 3.1 in [17], we obtain the following L^1 -type concentration result:

Corollary 1. *Given a Lipschitz functional $F \in \rho_1 - \text{Lip}_1$, we have the following concentration inequality: for any deviation level $x \geq 0$,*

$$(13) \quad \mu_\sigma(F - \mathbb{E}_{\mu_\sigma}[F] \geq x) \leq \exp \left\{ -\frac{x}{2} \log \left(1 + \frac{x}{4 \|\nabla^\sharp F\|_{1,\infty}} \right) \right\}.$$

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

Proof. By translation invariance, it is sufficient to prove the result for centered functional. Letting $H(\lambda) = \frac{1}{\lambda} \log \mathbb{E}_{\mu_\sigma} [e^{\lambda F}]$ with $H(0) = \mathbb{E}_{\mu_\sigma} [F] = 0$, we have

$$H'(\lambda) = \frac{\text{Ent}_{\mu_\sigma}(e^{\lambda F})}{\lambda^2 \mathbb{E}_{\mu_\sigma} [F]}.$$

Denoting $c = \|\nabla F\|_{1,\infty}$, we obtain by Theorem 8 the inequality $H'(\lambda) \leq ce^\lambda$, so that $H(\lambda) \leq c(e^\lambda - 1)$. In other words, we get the following bound on the Laplace transform:

$$\mathbb{E}_{\mu_\sigma} [e^{\lambda F}] \leq \exp \left\{ c\lambda(e^\lambda - 1) \right\}.$$

Hence by Chebychev's inequality,

$$\mu_\sigma (F > x) \leq \inf_{\lambda > 0} \exp \left\{ -\lambda x + c\lambda(e^\lambda - 1) \right\}.$$

On the one hand, if $x \in (0, 4c)$ then choose $\lambda = \frac{x}{4c}$ so that

$$\begin{aligned} \mu_\sigma (F > x) &\leq \exp \left\{ -\lambda x + c\lambda(e^\lambda - 1) \right\} \\ &\leq \exp \left\{ -\lambda x + 2c\lambda^2 \right\} \\ &= \exp \left\{ -\frac{x^2}{8c} \right\} \\ &\leq \exp \left\{ -\frac{x}{2} \log \left(1 + \frac{x}{4c} \right) \right\}. \end{aligned}$$

On the other hand, if $x \geq 4c$, then choose $\lambda = \log \left(\frac{x}{2c} \right)$ so that

$$\begin{aligned} \mu_\sigma (F > x) &\leq \exp \left\{ -\lambda x + c\lambda(e^\lambda - 1) \right\} \\ &\leq \exp \left\{ -\frac{x}{2} \log \left(\frac{x}{2c} \right) \right\} \\ &\leq \exp \left\{ -\frac{x}{2} \log \left(1 + \frac{x}{4c} \right) \right\}. \end{aligned}$$

The proof is thus complete. \square

The constants in the concentration inequality (13) have no reason to be sharp since the choices of λ in the proof above are far from optimal. For instance, to obtain a slightly better behavior, we might proceed as follows. Let $F \in \rho_1 - \text{Lip}_1$ be centered and let $\lambda > 0$. Denote $Z_\lambda = \mathbb{E}_{\mu_\sigma} [e^{\lambda F}]$ 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 the proof of Theorem 8,

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

where in the last inequality we used the fact that the function $x \mapsto (e^x - 1)/x$ is non-decreasing on $(0, +\infty)$. Hence we obtain the following bound on the Laplace transform:

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

Finally using Chebychev's inequality, we get the improved concentration inequality available for any $x \geq 0$:

$$\begin{aligned} \mu_\sigma (F - \mathbb{E}_{\mu_\sigma} [F] \geq x) &\leq \exp \left\{ x - (x + \|\nabla F\|_{1,\infty}) \log \left(1 + \frac{x}{\|\nabla F\|_{1,\infty}} \right) \right\} \\ &\leq \exp \left\{ -\frac{x}{2} \log \left(1 + \frac{x}{\|\nabla F\|_{1,\infty}} \right) \right\}. \end{aligned}$$

See also [9] for a somewhat similar concentration inequality, but however not comparable.

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 L^1 -gradient is well-defined on $\text{Dom } \nabla^\#$.

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

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

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 $F \in \text{Dom } \nabla^\#$:

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

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

$$(14) \quad h \mathbb{E}_{\mu_\sigma} [|F - \mathbb{E}_{\mu_\sigma} [F]|] \leq 2 \mathbb{E}_{\mu_\sigma} \left[\int_\Lambda |\nabla_x^\# F| \, d\sigma(x) \right], \quad F \in \text{Dom } \nabla^\#.$$

We have the following theorem.

Theorem 9. *Assume that the measure σ is finite. Then we have*

$$(15) \quad 1 \leq h_{\mu_\sigma} \leq \frac{\sigma(\Lambda)}{1 - e^{-\sigma(\Lambda)}}.$$

In particular, we have the asymptotic $h_{\mu_\sigma} \approx 1$ as $\sigma(\Lambda)$ is close to 0.

Note that Houdré and Privault established first the inequality $h_{\mu_\sigma} \geq 1$ by using Poincaré inequality, cf. Proposition 6.4 in [10]. Hence we recover their result via another approach. On the other hand, our estimate in the right-hand-side of (15) 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 $h_{\mu_\sigma} \geq 1$, let us establish the inequality (14) with $h = 1$. By homogeneity, it is sufficient to prove the result for functionals $F \in \text{Dom } \nabla^\sharp$ 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,

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

Hence using Theorem 2 with ρ_0 (recall that the pair (∇^\sharp, ρ_0) satisfies the Rademacher property by Lemma 2), we get the inequality (14) with $h = 1$, thus obtaining the desired inequality $h_{\mu_\sigma} \geq 1$.

On the other hand, to provide the upper bound in (15), 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. \square

REFERENCES

- [1] S. Albeverio, Y. G. Kondratiev, and M. Rockner, Analysis and geometry on configuration spaces, *J. Funct. Anal.*, 154(2):444-500, 1998.
- [2] A. D. Barbour, T. C. Brown, and A. Xia, Point processes in time and Stein's method, *Stochastics Stochastics Rep.*, 65(1-2):127-151, 1998.
- [3] A. D. Barbour, L. Holst, and S. Janson, *Poisson approximation*, Oxford Studies in Probability (2), The Clarendon Press Oxford University Press, New York, 1992.
- [4] S. G. Bobkov and M. Ledoux, On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures, *J. Funct. Anal.*, 156(2):347-365, 1998.
- [5] L. Decreusefond, Perturbation Analysis and Malliavin Calculus, *Ann. Appl. Probab.*, 8:496-523, 1998.
- [6] L. Decreusefond, Wasserstein distance on configurations space, *Potential Anal.*, 28(3):283-300, 2008.
- [7] A. Dermoune, P. Krée, and L. Wu, Calcul stochastique non adapté par rapport à la mesure aléatoire de Poisson, *Séminaire de Probabilités, XXII*, Lecture Notes in Math., 1321:477-484, Springer, Berlin, 1988.
- [8] D. Feyel and A. S. Üstünel, Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space, *Probab. Theory Related Fields*, 128(3):347-385, 2004.
- [9] C. Houdré and N. Privault, Concentration and deviation inequalities in infinite dimensions via covariance representations, *Bernoulli*, 8(6):697-720, 2002.
- [10] C. Houdré and N. Privault, Isoperimetric and related bounds on configuration spaces, *Statist. Probab. Lett.*, 78(14):2154-2164, 2008.
- [11] D. Nualart and J. Vives, Anticipative calculus for the Poisson process based on the Fock space, *Séminaire de probabilités, XXIV*, Lecture Notes in Math., 1426:154-165, Springer, Berlin, 1988.
- [12] J. Ren, M. Röckner, and X. Zhang, Kusuoka-Stroock formula on configuration space and regularities of local times with jumps, *Potential Anal.*, 26(4):363-396, 2007.
- [13] M. Röckner and A. Schied, Rademacher's theorem on configuration spaces and applications, *J. Funct. Anal.*, 169(2):325-356, 1999.

- [14] J. Ruiz de Chavez, Espaces de Fock pour les processus de Wiener et de Poisson, *Séminaire de probabilités*, XIX, Lecture Notes in Math., 1123:230-241, Springer, Berlin, 1985.
- [15] D. Schuhmacher, Estimation of distances between point process distributions, PhD Thesis, Universität Zürich, 2005.
- [16] C. Villani, *Optimal transport: old and new*, Grundlehren der mathematischen Wissenschaften, Springer, Berlin, 2009.
- [17] L. Wu, A new modified logarithmic Sobolev inequality for Poisson point processes and several applications, *Probab. Theory Related Fields*, 118(3):427-438, 2000.

INSTITUT TELECOM, TELECOM PARISTECH, CNRS LTCI, PARIS, FRANCE
E-mail address: `Laurent.Decreusefond@telecom-paristech.fr`

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`

UNIVERSITÉ DE TOULOUSE, UNIVERSITÉ PAUL SABATIER, INSTITUT DE MATHÉMATIQUES DE TOULOUSE, F-31062 TOULOUSE, FRANCE
E-mail address: `nicolas.savy@math.univ-toulouse.fr`