



HAL
open science

Transportation-information inequalities for Markov processes

Arnaud Guillin, Christian Léonard, Liming Wu, Nian Yao

► **To cite this version:**

Arnaud Guillin, Christian Léonard, Liming Wu, Nian Yao. Transportation-information inequalities for Markov processes. *Probability Theory and Related Fields*, 2009, 144 (3-4), pp.669-695. hal-00158258

HAL Id: hal-00158258

<https://hal.science/hal-00158258>

Submitted on 28 Jun 2007

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.

TRANSPORTATION-INFORMATION INEQUALITIES FOR MARKOV PROCESSES

ARNAUD GUILLIN, CHRISTIAN LÉONARD, LIMING WU, AND NIAN YAO

ABSTRACT. In this paper, one investigates the following type of transportation-information T_cI inequalities: $\alpha(T_c(\nu, \mu)) \leq I(\nu|\mu)$ for all probability measures ν on some metric space (\mathcal{X}, d) , where μ is a given probability measure, $T_c(\nu, \mu)$ is the transportation cost from ν to μ with respect to some cost function $c(x, y)$ on \mathcal{X}^2 , $I(\nu|\mu)$ is the Fisher-Donsker-Varadhan information of ν with respect to μ and $\alpha : [0, \infty) \rightarrow [0, \infty]$ is some left continuous increasing function. Using large deviation techniques, it is shown that T_cI is equivalent to some concentration inequality for the occupation measure of a μ -reversible ergodic Markov process related to $I(\cdot|\mu)$, a counterpart of the characterizations of transportation-entropy inequalities, recently obtained by Gozlan and Léonard in the i.i.d. case [24]. Tensorization properties of T_cI are also derived.

Let d be a metric. One denotes $W_1I := T_dI$ and $W_2I := T_{d^2}I$ the transportation-information inequalities associated with the metric cost $c = d$ and the quadratic cost $c = d^2$.

It is proved that W_2I is stronger than Poincaré inequality, weaker than log-Sobolev inequality, and equivalent to it when Bakry-Emery's curvature is bounded from below.

For the trivial metric cost d , one establishes the sharp transportation-information inequality W_1I in terms of the spectral gap. In particular, a Hoeffding type concentration inequality for Markov processes is derived and one shows that W_1I implies a Poincaré inequality.

For a general metric cost d , it is established that the spectral gap in the space of Lipschitz functions of the Markov diffusion process implies W_1I . A sharp estimate of the constant is obtained for general one-dimensional diffusion processes. Finally, a Lyapunov function condition for W_1I is proposed. It may be applied to a wide class of examples; some examples are worked out.

CONTENTS

1. Introduction	2
2. General results on T_cI	4
3. Poincaré inequality implies Hoeffding's deviation inequality	16
4. Spectral gap in the space of Lipschitz functions implies W_1I for diffusion processes	24
5. Lyapunov function conditions	30
References	36

1. INTRODUCTION

Let $M_1(\mathcal{X})$ be the space of all probability measures on some complete separable metric space (\mathcal{X}, d) and consider the *cost function* $c(x, y) : \mathcal{X}^2 \rightarrow [0, +\infty]$ with $c(x, x) = 0$ (for all $x \in \mathcal{X}$), which is lower semicontinuous on \mathcal{X}^2 . Given $\mu, \nu \in M_1(\mathcal{X})$, the *transportation cost* $T_c(\nu, \mu)$ from ν to μ with respect to the cost function c is defined by

$$T_c(\nu, \mu) = \inf_{\pi \in M_1(\mathcal{X}^2): \pi_0 = \nu, \pi_1 = \mu} \iint_{\mathcal{X}^2} c(x, y) \pi(dx, dy) \quad (1.1)$$

where $\pi_0(dx) = \pi(dx \times \mathcal{X})$, $\pi_1(dy) = \pi(\mathcal{X} \times dy)$ are the marginal distributions of π . When $c(x, y) = d^p(x, y)$ where $p \geq 1$, $(T_c(\nu, \mu))^{1/p} = W_p(\nu, \mu)$ is the L^p - Wasserstein distance between ν and μ .

The relative entropy (or Kullback information) of ν with respect to μ is given by

$$H(\nu|\mu) := \begin{cases} \int_{\mathcal{X}} f \log f d\mu, & \text{if } \nu \ll \mu \text{ and } f := \frac{d\nu}{d\mu} \\ +\infty, & \text{otherwise.} \end{cases} \quad (1.2)$$

The usual transportation inequalities for a given $\mu \in M_1(\mathcal{X})$, introduced by K. Marton [31, 32] and M. Talagrand [37], compare the Wasserstein metric $W_p(\nu, \mu)$ with the relative entropy $H(\nu|\mu)$. The following extension of these inequalities:

$$\alpha(T_c(\nu, \mu)) \leq H(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}), \quad (T_cH)$$

has recently been proposed and developed by Gozlan and Léonard [24]. Here $\alpha : [0, \infty) \rightarrow [0, +\infty]$ is some left continuous and increasing function with $\alpha(0) = 0$.

Let us denote

$$\alpha^{\circledast}(\lambda) := \sup_{r \geq 0} (\lambda r - \alpha(r)) \quad (1.3)$$

the monotone conjugate of α . With α as above, one sees that α^{\circledast} is the restriction to $[0, \infty)$ of the usual convex conjugate $\tilde{\alpha}^*(\lambda) = \sup_{r \in \mathbb{R}} (\lambda r - \tilde{\alpha}(r))$ of $\tilde{\alpha}(r) = \mathbf{1}_{r \geq 0} \alpha(r)$, $r \in \mathbb{R}$. We also denote $\mu(v) := \int_{\mathcal{X}} v d\mu$.

Notation. In the special case $c(x, y) = d^p(x, y)$ where $p \geq 1$ and d is a metric, $T_c(\nu, \mu) = W_p(\nu, \mu)^p$. We shall use the notation $W_p H$ instead of $T_{d^p} H$. In particular, $W_1 H$ stands for $T_d H$.

As an extension of the Bobkov-Götze criterion [5], we have

Theorem 1.1 (Gozlan-Léonard [24]). *Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of \mathcal{X} -valued i.i.d. random variables with common law μ and α be moreover convex. Then the following properties are equivalent:*

- (a) *The transportation inequality $T_c H$ holds;*
- (b) *For any couple of bounded and measurable functions $u, v : \mathcal{X} \rightarrow \mathbb{R}$ such that $u(x) - v(y) \leq c(x, y)$ over \mathcal{X}^2 ,*

$$\log \int_{\mathcal{X}} e^{\lambda u} d\mu \leq \lambda \mu(v) + \alpha^{\circledast}(\lambda), \quad \forall \lambda \geq 0;$$

- (c) *For all $n \geq 1$ and $r > 0$ and for any couple of bounded and measurable functions $u, v : \mathcal{X} \rightarrow \mathbb{R}$ such that $u(x) - v(y) \leq c(x, y)$ over \mathcal{X}^2 , the following concentration inequality holds*

$$\mathbb{P} \left(\frac{1}{n} \sum_{k=1}^n u(X_k) \geq \mu(v) + r \right) \leq e^{-n\alpha(r)};$$

(c') The following large deviation upper bound holds for any couple of bounded and measurable functions $u, v : \mathcal{X} \rightarrow \mathbb{R}$ such that $u(x) - v(y) \leq c(x, y)$ over \mathcal{X}^2 ,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left(\frac{1}{n} \sum_{k=1}^n u(X_k) \geq \mu(v) + r \right) \leq -\alpha(r), \quad \forall r \geq 0.$$

The main purpose of this paper. In this paper, instead of the transportation-entropy inequality $T_c H$, one investigates the following transportation-information inequality

$$\alpha(T_c(\nu, \mu)) \leq I(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}) \quad (T_c I)$$

for some given probability measure μ . Here $I(\nu|\mu)$ is the Fisher-Donsker-Varadhan information of ν with respect to μ

$$I(\nu|\mu) = \begin{cases} \mathcal{E}(\sqrt{f}, \sqrt{f}) & \text{if } \nu = f\mu, \sqrt{f} \in \mathbb{D}(\mathcal{E}) \\ +\infty & \text{otherwise} \end{cases} \quad (1.4)$$

associated with the Dirichlet form \mathcal{E} on $L^2(\mu)$ with domain $\mathbb{D}(\mathcal{E})$.

Notation. In the special case where $c(x, y) = d^p(x, y)$, we use the notation $W_p I$ instead of $T_{d^p} I$. In particular, $W_1 I$ stands for $T_d I$.

Organization of the paper. This paper is organized as follows. In the next section we characterize $T_c I$ by means of concentration inequalities for the empirical means $L_t(u) = \frac{1}{t} \int_0^t u(X_s) ds$ of observables u along the symmetric Markov process (X_t) associated with the Dirichlet form \mathcal{E} , extending Theorem 1.1 from i.i.d. sequences to time-continuous Markov processes. The method of proof is borrowed from Gozlan and Léonard [24] who proved Theorem 1.1 by means of large deviations of the empirical measure of an i.i.d. sequence. In the present paper, it relies on the large deviations of the occupation measure of (X_t) . The tensorization of $T_c I$ is proved, and the relations between $W_2 I$, Poincaré and log-Sobolev are exhibited with the help of Otto-Villani [34].

In Section 3, we prove $W_1 I$ for the trivial metric $d(x, y) = \mathbf{1}_{x \neq y}$ with the sharp constant in terms of the spectral gap and derive a sharp Hoeffding concentration inequality for Markov processes. Furthermore, we also prove that $W_1 I$ implies the existence of a positive spectral gap in the symmetric and uniform positive improving case by means of a result of L. Wu [46].

For a general metric, using Lyons-Meyer-Zheng forward-backward martingale decomposition, we obtain in Section 4 sharp $W_1 I$ under the spectral gap existence of the Markov diffusion process in the space of Lipschitz functions, and an explicit and sharp constant is provided for one-dimensional diffusions.

Finally in Section 5 we propose a practical Lyapunov condition for $W_1 I$ (or a more general $T_\Phi I$), which, although not providing the sharp constant, provides a good order.

About the literature. Let us give some historical notes on the usual transportation inequality $W_p H$. K. Marton [31] first noticed that $W_1 H$ implies the concentration inequality for μ by a very elementary and neat argument, and she established $W_1 H$ for the law of a Dobrushin-contractive Markov chain in [32]. M. Talagrand [37] established $W_2 H$ for the Gaussian measure μ with $\alpha(r) = r/2C$ and provided the sharp constant C (this particular case of $T_c H$ is often called Talagrand's transportation inequality). Bobkov and Götze [5] obtained the characterization of $W_p H$ in Theorem 1.1 with [$p = 1$, α quadratic] and [$p = 2$, α linear]. Otto and Villani [34] proved that the log-Sobolev inequality is stronger than Talagrand's transportation inequality and presented a differential geometrical point of view on $M_1(\mathcal{X})$ equipped with the W_2 -metric. Bobkov, Gentil and Ledoux [4] shed light on a profound relation between Talagrand's transportation inequality, log-Sobolev inequality, inf-convolution and some Hamilton-Jacobi equation. Djellout, Guillin and Wu [13] obtained a necessary and sufficient condition for $W_1 H$ with a quadratic α by means of the Gaussian integrability of $d(x, x_0)$ under μ , and gave a direct proof of Talagrand's transportation inequality for the law of a diffusion process by means of Girsanov's formula, without appealing to log-Sobolev inequality. Bolley and Villani [6] and later Gozlan and Léonard [24] refined the result of [13] under a Gaussian integrability condition. Cattiaux and Guillin [9] constructed the first example for which Talagrand's transportation inequality holds but not log-Sobolev inequality, and Gozlan [23] found a necessary and sufficient condition for Talagrand's transportation inequality with $\mu(dx) = e^{-V(x)} dx$ on \mathbb{R} when the Bakry-Emery curvature V'' is lower bounded. Otto-Villani's differential geometrical point of view on $M_1(\mathcal{X})$ equipped with the W_2 -metric is very fruitful, as developed by the recent works of Sturm [35, 36] and Lott and Villani [30]. The reader is referred to the textbooks by Ledoux [27] and Villani [39, 38] for further references pertaining to this very active field.

Convention and notation. Throughout this paper either (\mathcal{X}, d) is a complete separable metric space with the associated Borel σ -field \mathcal{B} .

- The space of all real bounded and \mathcal{B} -measurable functions is denoted by $b\mathcal{B}$.
- The functions to be considered later are assumed to be measurable without warning.
- For $\mu, \nu \in M_1(\mathcal{X})$, $\|\nu - \mu\|_{TV} := \sup_{u: |u| \leq 1} \int u d(\nu - \mu)$ is the total variation norm.
- Throughout this paper a cost function c is a nonnegative lower semicontinuous function on \mathcal{X}^2 such that $c(x, x) = 0$ for all $x \in \mathcal{X}$.

2. GENERAL RESULTS ON $T_c I$

2.1. Markov processes, Fisher-Donsker-Varadhan information and Feynman-Kac semigroup. The main probabilistic object to be considered in this paper is an \mathcal{X} -valued time-continuous Markov process $(\Omega, \mathcal{F}, (X_t)_{t \geq 0}, (\mathbb{P}_x)_{x \in \mathcal{X}})$ with an invariant probability measure μ . The transition semigroup is denoted $(P_t)_{t \geq 0}$.

Assumption: Ergodicity. It is assumed that the invariant probability measure μ is ergodic: if $f \in b\mathcal{B}$ satisfies $P_t f = f$, μ -a.e. for all $t \geq 0$, then f is constant μ -a.e. Denoting $\mathbb{P}_\beta(\cdot) := \int_{\mathcal{X}} \mathbb{P}_x(\cdot) \beta(dx)$ for any initial probability measure β , the previous condition on μ amounts to stating that $((X_t)_{t \geq 0}, \mathbb{P}_\mu)$ is a stationary ergodic process.

Assumption: Closability of the Dirichlet form. It is assumed that (P_t) is strongly continuous on $L^2(\mu) := L^2(\mathcal{X}, \mathcal{B}, \mu)$. Let \mathcal{L} be its generator with domain $\mathbb{D}_2(\mathcal{L})$ on $L^2(\mu)$. It is also assumed that

$$\mathcal{E}(g, g) := \langle -\mathcal{L}g, g \rangle_\mu, \quad g \in \mathbb{D}_2(\mathcal{L})$$

is closable in $L^2(\mu)$. Its closure which is denoted by $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$ is a Dirichlet form: the symmetrized Dirichlet form associated with the Markov process (X_t) (or (P_t)). Notice that $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$ corresponds to a self-adjoint generator \mathcal{L}^σ (formally $\mathcal{L}^\sigma = (\mathcal{L} + \mathcal{L}^*)/2$), and $P_t^\sigma = e^{t\mathcal{L}^\sigma}$ is the symmetrized Markov semigroup of (P_t) . When P_t is symmetric on $L^2(\mu)$, the above closability assumption is always satisfied and the domain $\mathbb{D}(\mathcal{E})$ of the Dirichlet form coincides with the domain $\mathbb{D}_2(\sqrt{-\mathcal{L}})$ in $L^2(\mu)$.

These above assumptions of ergodicity and closability of the Dirichlet form prevail for the whole paper.

Fisher-Donsker-Varadhan information. The following definition is motivated by standard large deviation results.

Definition 2.1. *Given the Dirichlet form \mathcal{E} with domain $\mathbb{D}(\mathcal{E})$ on $L^2(\mu)$, the Fisher-Donsker-Varadhan information of ν with respect to μ is defined by*

$$I(\nu|\mu) := \begin{cases} \mathcal{E}(\sqrt{f}, \sqrt{f}), & \text{if } \nu = f\mu, \sqrt{f} \in \mathbb{D}(\mathcal{E}) \\ +\infty, & \text{otherwise.} \end{cases} \quad (2.1)$$

Convention. We adopt the following convention for the Fisher-Donsker-Varadhan information on a Riemannian manifold \mathcal{X} : if $\nu = f\mu$ with $f > 0$ smooth,

$$I(\nu|\mu) = \frac{1}{4} \int_{\mathcal{X}} \frac{|\nabla f|^2}{f} d\mu.$$

This means that $I = I_F/4$ where I_F is the standard Fisher information. This will lead to transportation-information inequalities with a natural interpretation in terms of large deviations, see (2.2) below.

When (P_t) is μ -symmetric, $\nu \mapsto I(\nu|\mu)$ is exactly the Donsker-Varadhan entropy i.e. the rate function governing the large deviation principle of the empirical measure

$$L_t := \frac{1}{t} \int_0^t \delta_{X_s} ds$$

for large time t . This was proved by Donsker and Varadhan [15, 16, 17] under some conditions of absolute continuity and regularity of $P_t(x, dy)$, and established in full generality by L. Wu [44, Corollary B.11]. When $\mu = e^{-V(x)} dx / Z$ (Z is the normalization constant) with $V \in C^1$ on a complete connected Riemannian manifold $\mathcal{X} = M$, the diffusion (X_t) generated by $\mathcal{L} = \Delta - \nabla V \cdot \nabla$ (Δ, ∇ are respectively the Laplacian and the gradient on M) is μ -reversible and the corresponding Dirichlet form is given by

$$\mathcal{E}_\mu(g, g) = \int_M |\nabla g|^2 d\mu, \quad g \in \mathbb{D}(\mathcal{E}_\mu) = H^1(\mathcal{X}, \mu)$$

where $H^1(\mathcal{X}, \mu)$ is the closure of $C_b^\infty(M)$ (the space of infinitely differentiable functions f on M with $|\nabla^n f|$ bounded for all n) with respect to the norm

$$\sqrt{\mu(|g|^2 + |\nabla g|^2)}.$$

It also matches with the space of these $g \in L^2(M)$ such that $\nabla g \in L^2(M \rightarrow TM; \mu)$ in distribution. In this case, if $\nu = f\mu$ with $0 < f \in C^1(M)$, then

$$I(\nu|\mu) = \int_{\mathcal{X}} |\nabla \sqrt{f}|^2 d\mu = \frac{1}{4} \int_{\mathcal{X}} \frac{|\nabla f|^2}{f} d\mu. \quad (2.2)$$

Feynman-Kac semigroup. The derivation of the large deviation results for L_t as t tends to infinity is intimately related to the Feynman-Kac semigroup

$$P_t^u g(x) := \mathbb{E}^x g(X_t) \exp\left(\int_0^t u(X_s) ds\right). \quad (2.3)$$

When u is bounded, (P_t^u) is a strongly continuous semigroup of bounded operators on $L^2(\mu)$ whose generator is given by $\mathcal{L}^u g = \mathcal{L}g + ug$, for all $g \in \mathbb{D}_2(\mathcal{L}^u) = \mathbb{D}_2(\mathcal{L})$.

It is no surprise that this semigroup also plays a role in the present investigation.

2.2. Characterizations of $T_c I$. Recall that Kantorovich's duality theorem (see [39]) states that for any $\nu, \mu \in M_1(\mathcal{X})$ so that $T_c(\nu, \mu) < +\infty$,

$$T_c(\nu, \mu) = \sup_{(u,v) \in \Phi_c} \int u d\nu - \int v d\mu \quad (2.4)$$

where

$$\Phi_c := \{(u, v) \in (b\mathcal{B})^2 : u(x) - v(y) \leq c(x, y), \forall (x, y) \in \mathcal{X}^2\}.$$

This motivates us to introduce as in [24]

$$T_\Phi(\nu, \mu) = \sup_{(u,v) \in \Phi} \int u d\nu - \int v d\mu \quad (2.5)$$

where $\Phi \subset (b\mathcal{B})^2$ (non-empty) satisfies

- (A1) $u \leq v$ for all $(u, v) \in \Phi$ (or equivalently $T_\Phi(\nu, \nu) = 0$ for all $\nu \in M_1(\mathcal{X})$);
- (A2) For all $\nu_1, \nu_2 \in M_1(\mathcal{X})$, there exists $(u, v) \in \Phi$ such that $\int u d\nu_1 - \int v d\nu_2 \geq 0$ (or equivalently $T_\Phi(\nu_1, \nu_2) \geq 0$ for all $\nu_1, \nu_2 \in M_1(\mathcal{X})$).

Note that for (A1) and (A2) to be satisfied when $\Phi = \Phi_c$, it is enough that $c(x, x) = 0$ for all x . The main result of this section is the following generalization of Theorem 1.1.

Theorem 2.2. *Let $((X_t)_{t \geq 0}, \mathbb{P}_\mu)$ be a stationary ergodic Markov process with the symmetrized Dirichlet form $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$, Φ be as above and $\alpha : [0, \infty) \rightarrow [0, \infty]$ be a left continuous increasing function such that $\alpha(0) = 0$. Consider the following properties:*

- (a) *The following transportation inequality holds*

$$\alpha(T_\Phi(\nu, \mu)) \leq I(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}) \quad (T_\Phi I)$$

- (b) *For all $(u, v) \in \Phi$ and all $\lambda, t \geq 0$*

$$\|P_t^{\lambda u}\|_{L^2(\mu)} \leq e^{t[\lambda\mu(v) + \alpha^*(\lambda)]} \quad (2.6)$$

where $P_t^{\lambda u}$ is the Feynman-Kac semigroup (2.3) and α^* is defined at (1.3).

- (b') *For all $(u, v) \in \Phi$ and all $\lambda \geq 0$*

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}_\mu \exp\left(\lambda \int_0^t u(X_s) ds\right) \leq \lambda\mu(v) + \alpha^*(\lambda)$$

(c) For any initial measure $\beta \ll \mu$ with $d\beta/d\mu \in L^2(\mu)$ and for all $(u, v) \in \Phi$ and $r, t > 0$,

$$\mathbb{P}_\beta \left(\frac{1}{t} \int_0^t u(X_s) ds \geq \mu(v) + r \right) \leq \left\| \frac{d\beta}{d\mu} \right\|_2 e^{-t\alpha(r)} \quad (2.7)$$

(c') For all $(u, v) \in \Phi$ and for any $r \geq 0$, there exists $\beta \in M_1(E)$ such that $\beta \ll \mu$, $d\beta/d\mu \in L^2(\mu)$ and

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{P}_\beta \left(\frac{1}{t} \int_0^t u(X_s) ds \geq \mu(v) + r \right) \leq -\alpha(r)$$

We have

(1) (a) \Rightarrow (b) \Rightarrow (b') and (a) \Rightarrow (c) \Rightarrow (c').

(2) If α is convex, then (a) \Leftrightarrow (b).

(3) If (P_t) is symmetric on $L^2(\mu)$, then (a) \Leftrightarrow (c) \Leftrightarrow (c').

If furthermore α is convex, (a), (b), (b'), (c) and (c') are equivalent.

The proof of this result is similar to [24, Theorems 2 and 15]'s ones. It takes advantage of large deviation results previously obtained by L. Wu. Namely,

- the identification of the rate function in the symmetric case and the large deviation lower bound are taken from [44] and
- the non-asymptotic Cramér's upper bounds which are used in [24] are replaced by the following result.

Lemma 2.3 (L. Wu [43]). *For any $u \in b\mathcal{B}$ with $\mu(|u|) < +\infty$ and any $t > 0$, the following statements hold true.*

(1) Denoting

$$\Lambda(u) := \sup \left\{ \int ug^2 d\mu - \mathcal{E}(g, g); g \in \mathbb{D}(\mathcal{E}), \mu(g^2) = 1, \mu(g^2|u|) < +\infty \right\}, \quad (2.8)$$

one has

$$\|P_t^u\|_{L^2(\mu)} \leq e^{t\Lambda(u)} \quad (2.9)$$

and the equality holds in the symmetric case;

(2) For all $r > 0$,

$$\mathbb{P}_\beta \left(\frac{1}{t} \int_0^t u(X_s) ds - \mu(u) \geq r \right) \leq \left\| \frac{d\beta}{d\mu} \right\|_2 \exp \left(-t \lim_{\delta \downarrow 0} I_u(\mu(u) + r - \delta) \right) \quad (2.10)$$

where

$$I_u(r) := \inf \{ I(\nu|\mu); \nu \in M_1(\mathcal{X}), \nu(u) = r \}, \quad r \in \mathbb{R}.$$

It is proved in [43, 44] that in the symmetric case, $I_u(r)$ is exactly the rate function governing the large deviation principle of $\frac{1}{t} \int_0^t u(X_s) ds$ for bounded u . In these papers no mixing assumptions are required, this is in contrast with the usual assumptions for the large deviation principle as discovered by Donsker and Varadhan [15, 16, 17] and reconsidered by Deuschel and Stroock [12]. This relaxation of the usual assumptions is allowed by the assumed restriction that the initial law is absolutely continuous with respect to the ergodic measure μ .

Proof of Theorem 2.2. Part (1). As $\nu \rightarrow I(\nu|\mu)$ is convex on $M_1(\mathcal{X})$, so is $I_u : \mathbb{R} \rightarrow [0, +\infty]$. Since $I_u(\mu(u)) = 0$, I_u is increasing on $[\mu(u), +\infty)$. For all $(u, v) \in \Phi$ and all $\lambda \geq 0$, we have

$$\Lambda(\lambda u) = I_u^*(\lambda) \quad (2.11)$$

where I_u^* is the convex conjugate of I_u . Indeed for $\lambda \geq 0$, by (2.8)

$$\begin{aligned} \Lambda(\lambda u) &= \sup\{\lambda \int u g^2 d\mu - \mathcal{E}(g, g); g \in \mathbb{D}(\mathcal{E}), \mu(g^2) = 1\} \\ &= \sup\{\lambda \int u g^2 d\mu - \mathcal{E}(g, g); 0 \leq g \in \mathbb{D}(\mathcal{E}), \mu(g^2) = 1\} \\ &= \sup\{\lambda \int u d\nu - I(\nu|\mu); \nu \in M_1(\mathcal{X})\} \\ &= \sup_{a \in \mathbb{R}}\{\lambda a - I_u(a)\} \end{aligned}$$

where the second equality follows from the fact that $\mathcal{E}(|g|, |g|) \leq \mathcal{E}(g, g)$ for all $g \in \mathbb{D}(\mathcal{E})$.

Note also that $T_\Phi I$ implies that for any $(u, v) \in \Phi$,

$$I_u(\mu(v) + r) \geq \tilde{\alpha}(r), \quad \forall r \in \mathbb{R} \quad (2.12)$$

where $\tilde{\alpha}(r) = \alpha(r)$ for $r \geq 0$ and $= 0$ for $r \leq 0$. Indeed it is trivial for $r \leq 0$ and for any $r \geq 0$ and $\nu \in M_1(\mathcal{X})$ such that $\nu(u) = \mu(v) + r$, $T_\Phi I$ implies that

$$I(\nu|\mu) \geq \alpha(T_\Phi(\nu, \mu)) \geq \alpha(\nu(u) - \mu(v)) = \alpha(r).$$

• (a) \Rightarrow (b): Putting together (2.11) and (2.12) leads us to

$$\Lambda(\lambda u) = \sup_{a \in \mathbb{R}} [\lambda a - I_u(a)] \leq \sup_{r \in \mathbb{R}} [\lambda(\mu(v) + r) - \tilde{\alpha}(r)] = \lambda\mu(v) + \alpha^\otimes(\lambda)$$

for all $\lambda \geq 0$. Statement (b) now follows from inequality (2.9).

• (a) \Rightarrow (c): This follows from (2.10) and (2.12), noting that by (A1), $\mu(u) \leq \mu(v)$ for all $(u, v) \in \Phi$.

• (b) \Rightarrow (b') and (c) \Rightarrow (c'): These implications are trivial.

Part (2). (b) \Rightarrow (a) in the case where α is convex. By (2.6), we have for $(u, v) \in \Phi$ fixed and for any $g \in \mathbb{D}_2(\mathcal{L})$,

$$\langle P_t^{\lambda u} g, P_t^{\lambda u} g \rangle_\mu \leq e^{2t(\lambda\mu(v) + \alpha^\otimes(\lambda))} \langle g, g \rangle_\mu.$$

Differentiating at time zero we obtain

$$2\langle g, \mathcal{L}g + \lambda u g \rangle_\mu = 2(\lambda\mu(g^2 u) - \mathcal{E}(g, g)) \leq 2(\lambda\mu(v) + \alpha^\otimes(\lambda))\mu(g^2).$$

Then for all $g \in \mathbb{D}_2(\mathcal{L})$,

$$\lambda[\mu(g^2 u) - \mu(v)\mu(g^2)] - \alpha^\otimes(\lambda)\mu(g^2) \leq \mathcal{E}(g, g).$$

It can be extended to $g \in \mathcal{D}(\mathcal{E})$. Now for any $\nu \in M_1(\mathcal{X})$ such that $I(\nu|\mu) < +\infty$, applying the above inequality to $g = \sqrt{\frac{d\nu}{d\mu}}$, we get

$$\lambda[\nu(u) - \mu(v)] - \alpha^\otimes(\lambda) \leq I(\nu|\mu).$$

Taking the supremum over all $\lambda \in \mathbb{R}$, as α assumed to be convex and $\alpha^{\otimes} = \tilde{\alpha}^*$ on $[0, \infty)$ (see the remark below (1.3)), we get

$$\tilde{\alpha}(\nu(u) - \mu(v)) \leq I(\nu|\mu)$$

and taking the supremum over all $(u, v) \in \Phi$ leads to the desired result.

Part (3). Let us assume from now on that the semigroup (P_t) is symmetric in $L^2(\mu)$.

• $(c') \Rightarrow (a)$: By the large deviation lower bound in [44, Theorem B.1] and the identification of the rate function in the symmetric case in [44, Corollary B.11], we have for any initial probability measure $\beta \ll \mu$,

$$\liminf_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{P}_\beta \left(\frac{1}{t} \int_0^t u(X_s) ds \geq \mu(v) + r \right) \geq -\inf\{I(\nu|\mu); \nu(u) > \mu(v) + r\}.$$

This together with (c') implies that for any $r \geq 0$,

$$\inf\{I(\nu|\mu); \nu(u) > \mu(v) + r\} \geq \alpha(r).$$

Fix now ν such that $r_0 = T_\Phi(\nu, \mu) > 0$ (otherwise $T_\Phi I$ is obviously true.) Choosing a sequence $(u_n, v_n) \in \Phi$ so that $\nu(u_n) - \mu(v_n) > r_0 - 1/n$, for all large enough n ,

$$\alpha(r_0 - 1/n) \leq I(\nu|\mu)$$

where $T_\Phi I$ follows by letting $n \rightarrow \infty$ and by the left-continuity of α .

• α is convex and (P_t) is symmetric. $(b') \Rightarrow (c')$ with $\beta = \mu$: The proof is standard and consists in optimizing exponential upper bounds. So doing, one obtains by means of (b') the asymptotic upper bound (c') with the convex envelope of $\tilde{\alpha}$ instead of $\tilde{\alpha}$. As α is assumed to be convex, (c') is proved.

This completes the proof of the theorem. \square

We now investigate two particular cases of Theorem 2.2.

Corollary 2.4 (The inequalities $W_1 I(c)$ and $W_2 I(c)$). *Let $c > 0$ and let (X_t) be a μ -reversible and ergodic Markov process such that $\int d^2(x, x_0) d\mu(x) < +\infty$.*

(1) *The statements below are equivalent:*

(a) *The following $W_1 I(c)$ inequality holds true:*

$$W_1^2(\nu, \mu) \leq 4c^2 I(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}); \quad (W_1 I(c))$$

(b) *For all Lipschitz function u on \mathcal{X} with $\|u\|_{\text{Lip}} \leq 1$ and all $\lambda, t \geq 0$,*

$$\|P_t^{\lambda u}\|_{L^2(\mu)} \leq \exp(\lambda \mu(u) + c^2 \lambda^2);$$

(c) *For all Lipschitz function u on \mathcal{X} with $\|u\|_{\text{Lip}} \leq 1$, $\mu(u) = 0$ and all $\lambda \geq 0$,*

$$\limsup_{t \rightarrow +\infty} \frac{1}{t} \log \mathbb{E}_\mu \exp \left(\lambda \int_0^t u(X_s) ds \right) \leq c^2 \lambda^2;$$

(d) *For all Lipschitz function u on \mathcal{X} , $r > 0$ and $\beta \in M_1(\mathcal{X})$ such that $d\beta/d\mu \in L^2(\mu)$,*

$$\mathbb{P}_\beta \left(\frac{1}{t} \int_0^t u(X_s) ds \geq \mu(u) + r \right) \leq \left\| \frac{d\beta}{d\mu} \right\|_2 \exp \left(-\frac{r^2}{4c^2 \|u\|_{\text{Lip}}^2} \right).$$

(2) *The statements below are equivalent:*

(a) The following $W_2I(c)$ inequality holds true:

$$W_2^2(\nu, \mu) \leq 4c^2 I(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}); \quad (W_2I(c))$$

(b) For any $v \in b\mathcal{B}$,

$$\|P_t^{\frac{1}{4c^2}Qv}\|_{L^2(\mu)} \leq e^{\frac{t}{4c^2}\mu(v)}, \quad \forall t \geq 0$$

where $Qv(x) = \inf_{y \in \mathcal{X}} \{v(y) + d^2(x, y)\}$ is the so-called “inf-convolution” of v ;

(c) For any $u \in b\mathcal{B}$,

$$\|P_t^{\frac{1}{4c^2}Su}\|_{L^2(\mu)} \leq e^{\frac{t}{4c^2}\mu(Su)}, \quad \forall t \geq 0$$

where $Su(y) = \sup_{x \in \mathcal{X}} \{u(x) - d^2(x, y)\}$ is the so-called “sup-convolution” of u .

Notation. The best constants $c > 0$ in $W_1I(c)$ and $W_2I(c)$ will be denoted respectively by $c_{W_1I}(\mu)$ and $c_{W_2I}(\mu)$.

Remarks 2.5.

- (i) The best constants $c_{W_1I}(\mu)$ and $c_{W_2I}(\mu)$ depend on the metric d and the Dirichlet form \mathcal{E} . Of course $c_{W_1I}(\mu) \leq c_{W_2I}(\mu)$.
- (ii) The above corollary may be seen as the counterpart of Bobkov-Götze’s characterizations of W_pH ($p = 1, 2$) for Markov processes.
- (iii) For a justification of the choice of the constant $4c^2$ in $WI(c)$, see Example 2.12 below, where c_{W_1I} and c_{W_2I} are identified as standard error in a Gaussian model.

Proof. Part (1). Notice that $W_1(\nu, \mu) = T_\Phi(\nu, \mu)$, where $\Phi := \{(u, u); \|u\|_{\text{Lip}} \leq 1, u \in b\mathcal{B}\}$. The result is a direct consequence of Theorem 2.2 in case u is bounded. The transition from a bounded to an unbounded u follows from an elementary monotone convergence argument.

Part (2). The equivalence of (b) and (c) is direct. Though (a) \Leftrightarrow (b) follows easily from Theorem 2.2, we nevertheless present a simple proof. In the present symmetric case, by Lemma 2.3, for any $u \in L^1(\mu)$ the equality is achieved in inequality (2.9):

$$\|P_t^u\|_{L^2(\mu)} = e^{t\Lambda(u)}, \quad \forall t \geq 0$$

(possibly infinite) where $\Lambda(u)$ is given at (2.8). Notice that Qv is upper semicontinuous and Su is lower semicontinuous.

- (a) \Rightarrow (b). Since for any $\nu \in M_1(E)$,

$$\int Qv d\nu - \int v d\mu \leq W_2^2(\nu, \mu) \leq 4c^2 I(\nu|\mu)$$

then

$$\frac{1}{4c^2} \int Qv d\nu - I(\nu|\mu) \leq \frac{1}{4c^2} \mu(v)$$

Taking the supremum over all ν yields (b).

- (b) \Rightarrow (a). Reverse the above proof. □

Remarks 2.6. We have seen that, by Theorem 2.2, $T_\Phi I$ inequalities lead to exponential deviation inequalities when the initial measure β is such that $d\beta/d\mu \in L^2(\mu)$. It is of course a limitation for the applications. Let us see that in the diffusion case we may overcome this limitation. As remarked by Wu [43, p.441-442], this assumption can be replaced by $d\beta/d\mu \in L^q(\mu)$ for $1 \leq q < 2$, provided that one replaces (2.9) in Lemma 2.3 by $\|P_t^u\|_p \leq e^{t\Lambda_p(u)}$, with

$$\Lambda_p(u) := \sup \left\{ \int u|f|^p d\mu + \langle \text{sgn}(f)|f|^{p-1}, \mathcal{L}f \rangle_\mu; f \in \mathbb{D}_p(\mathcal{L}) \text{ and } \int |f|^p d\mu = 1 \right\},$$

where p and q are conjugate numbers. Now, suppose that \mathcal{L} admits a carré du champ Γ . One can integrate by parts and get

$$\langle \text{sgn}(f)|f|^{p-1}, \mathcal{L}f \rangle_\mu = - (4(p-1)/p^2) \int \Gamma(|f|^{p/2}) d\mu.$$

Taking $g = |f|^{p/2}$ in the definition of Λ , we obtain that

$$\Lambda_p(u) = (4(p-1)/p^2) \Lambda((p^2/4(p-1))u).$$

Once again a deviation inequality is obtained, however with worse constants.

2.3. Tensorization of $T_c I$. Assume that $\mu_i \in M_1(\mathcal{X}_i)$ satisfies

$$\alpha_i(T_{c_i}(\nu, \mu_i)) \leq I_i(\nu|\mu_i), \quad \forall \nu \in M_1(\mathcal{X}_i) \quad (2.13)$$

where $I_i(\nu|\mu_i)$ is the Fisher-Donsker-Varadhan information related to the Dirichlet form $(\mathcal{E}_i, \mathbb{D}(\mathcal{E}_i))$, and α_i is moreover convex. On the product space $\mathcal{X}^{(n)} := \prod_{i=1}^n \mathcal{X}_i$ equipped with the product measure $\mu := \prod_{i=1}^n \mu_i$, consider the sum-cost function

$$\oplus_i c_i(x, y) := \sum_{i=1}^n c(x_i, y_i), \quad \forall x, y \in \mathcal{X}^{(n)} \quad (2.14)$$

and the *inf-convolution* of (α_i)

$$\alpha_1 \square \cdots \square \alpha_n(r) := \inf \left\{ \sum_{i=1}^n \alpha(r_i); r_i \geq 0, \sum_{i=1}^n r_i = r \right\}. \quad (2.15)$$

It also shares the following properties of every α_i : it is increasing, left continuous and convex on \mathbb{R}^+ with $\alpha(0) = 0$ (see [24]). Define the sum-Dirichlet form of $\oplus_i \mathcal{E}_i$ by

$$\begin{aligned} \mathbb{D}(\oplus_i \mathcal{E}_i) &:= \left\{ g \in L^2(\mu) : g_i \in \mathbb{D}(\mathcal{E}_i), \text{ for } \mu\text{-a.e. } \hat{x}_i \text{ and } \int_{\mathcal{X}^{(n)}} \sum_{i=1}^n \mathcal{E}_i(g_i, g_i) d\mu < +\infty \right\} \\ \oplus_i \mathcal{E}_i(g, g) &:= \int_{\mathcal{X}^{(n)}} \sum_{i=1}^n \mathcal{E}_i(g_i, g_i) d\mu, \quad g \in \mathbb{D}(\mathcal{E}) \end{aligned} \quad (2.16)$$

where $g_i(x_i) := g(x_1, \dots, x_i, \dots, x_n)$ with $\hat{x}_i := (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ fixed.

Theorem 2.7. Assume (2.13) for each $i = 1, \dots, n$ with α_i moreover convex. Define c, α, \mathcal{E} respectively by (2.14), (2.15) and (2.16). Let $I_{\oplus_i \mathcal{E}_i}(\nu|\mu)$ be the Fisher-Donsker-Varadhan information associated with $(\oplus_i \mathcal{E}_i, \mathbb{D}(\oplus_i \mathcal{E}_i))$. Then

$$\alpha_1 \square \cdots \square \alpha_n(r)(T_{\oplus c_i}(\nu, \mu)) \leq I_{\oplus_i \mathcal{E}_i}(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}^{(n)}). \quad (2.17)$$

This result is similar to [24, Corollary 5], but the proof will be different. It is based on the following sub-additivity result for the transportation cost of a product measure which is different from Marton's original proof [31] where an ordering of sites is required.

Lemma 2.8. *Given a probability measure ν on $\prod_{i=1}^n \mathcal{X}_i$, let ν_i be the regular conditional distribution of x_i knowing \hat{x}_i . Then with the cost function c given at (2.14),*

$$T_{\oplus c_i}(\mu, \nu) \leq \mathbb{E}^\nu \sum_{i=1}^n T_{c_i}(\mu_i, \nu_i).$$

Notation. The expectation \mathbb{E}^ν simply means integration with respect to ν .

Proof. Let $(Z_i = (X_i, Y_i))_{i=1, \dots, n}$ be a sequence of random variables valued in $\prod_{i=1}^n \mathcal{X}_i^2$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, realizing $T_{\oplus c_i}(\mu, \nu)$, i.e., the law of $X = (X_i)_{i=1, \dots, n}$ is $\mu = \prod_{i=1}^n \mu_i$, the law of $Y = (Y_i)_{i=1, \dots, n}$ is ν , and

$$\mathbb{E} \sum_i c_i(X_i, Y_i) = T_{\oplus c_i}(\mu, \nu).$$

For each i fixed, construct a couple of r.v. $(\tilde{X}_i, \tilde{Y}_i)$ so that its conditional law given $(Z_j)_{j \neq i}$ is a coupling of $(\mu_i(dx_i), \nu_i(dx_i|Y_j, j \neq i))$ and \mathbb{P} -a.s.,

$$\mathbb{E}[c_i(\tilde{X}_i, \tilde{Y}_i) | Z_j, j \neq i] = T_{c_i}(\mu_i, \nu_i(\cdot | Y_j, j \neq i)).$$

Obviously $(X_j, j \neq i; \tilde{X}_i)$ and $(Y_j, j \neq i; \tilde{Y}_i)$ (more precisely their joint law) constitute a coupling of (μ, ν) . Thus $\mathbb{E} \sum_j c_j(X_j, Y_j) \leq \mathbb{E}[\sum_{j \neq i} c_j(X_j, Y_j) + c_i(\tilde{X}_i, \tilde{Y}_i)]$ or

$$\mathbb{E} c_i(X_i, Y_i) \leq \mathbb{E} c_i(\tilde{X}_i, \tilde{Y}_i) = \mathbb{E} T_{c_i}(\mu_i, \nu_i(\cdot | Y_j, j \neq i)).$$

Consequently

$$T_{\oplus c_i}(\mu, \nu) = \mathbb{E} \sum_{i=1}^n T_{c_i}(X_i, Y_i) \leq \mathbb{E} \sum_{i=1}^n T_{c_i}(\mu_i, \nu_i(\cdot | Y_j, j \neq i)) = \int \sum_{i=1}^n T_{c_i}(\mu_i, \nu_i) d\nu.$$

□

The following additivity property of the Fisher information will be needed. It holds even in the dependent case.

Lemma 2.9. *Let ν, μ be probability measures on $\prod_{i=1}^n \mathcal{X}_i$ such that $I(\nu|\mu) < +\infty$, let μ_i, ν_i be the regular conditional distributions of x_i knowing \hat{x}_i under μ, ν . Then*

$$I_{\oplus \mathcal{E}_i}(\nu|\mu) = \mathbb{E}^\nu \sum_i I_i(\nu_i|\mu_i). \quad (2.18)$$

Proof. Let $f = d\nu/d\mu$. Then $d\nu_i/d\mu_i = f/\mu_i(f) = f_i/\mu_i(f_i)$, ν -a.s. (recalling that f_i is the function f of x_i with \hat{x}_i fixed). For ν -a.e. \hat{x}_i fixed,

$$I_i(\nu_i|\mu_i) = \mathcal{E}_i \left(\sqrt{\frac{f_i}{\mu_i(f_i)}}, \sqrt{\frac{f_i}{\mu_i(f_i)}} \right) = \frac{1}{\mu_i(f_i)} \mathcal{E}_i(\sqrt{f_i}, \sqrt{f_i})$$

(for $\mu_i(f_i)$ is constant with \hat{x}_i fixed). We obtain

$$\begin{aligned} \mathbb{E}^\nu \sum_{i=1}^n I_i(\nu_i|\mu_i) &= \mathbb{E}^\mu f \sum_{i=1}^n \frac{1}{\mu_i(f_i)} \mathcal{E}_i(\sqrt{f_i}, \sqrt{f_i}) \\ &= \mathbb{E}^\mu \sum_{i=1}^n \mathcal{E}_i(\sqrt{f_i}, \sqrt{f_i}) \\ &= \oplus_i \mathcal{E}_i(\sqrt{f}, \sqrt{f}) = I_{\oplus_i \mathcal{E}_i}(\nu|\mu), \end{aligned}$$

which completes the proof. \square

The above additivity is different from the super-additivity of the Fisher information for product measure obtained by E. Carlen [7].

Proof of Theorem 2.7. Without loss of generality we may assume that $I(\nu|\mu) < +\infty$. For simplicity write $\alpha = \alpha_1 \square \cdots \square \alpha_n$. By Lemma 2.8, Jensen's inequality and the definition of α ,

$$\begin{aligned} \alpha(T_{\oplus c_i}(\nu, \mu)) &\leq \alpha \left(\mathbb{E}^\nu \sum_{i=1}^n T_{c_i}(\nu_i, \mu_i) \right) \\ &\leq \mathbb{E}^\nu \alpha \left(\sum_{i=1}^n T_{c_i}(\nu_i, \mu_i) \right) \\ &\leq \mathbb{E}^\nu \sum_{i=1}^n \alpha_i(T_{c_i}(\nu_i, \mu_i)) \\ &\leq \mathbb{E}^\nu \sum_{i=1}^n I_i(\nu_i|\mu_i). \end{aligned}$$

The last quantity is equal to $I_{\oplus_i \mathcal{E}_i}(\nu|\mu)$, by Lemma 2.9. \square

As an example of application, let $(X_t^i)_{t \geq 0}, i = 1, \dots, n$ be n Markov processes with the same transition semigroup (P_t) and the same symmetrized Dirichlet form \mathcal{E} on $L^2(\mu)$, and conditionally independent once $(X_0^i)_{i=1, \dots, n}$ is fixed. Then $X_t := (X_t^1, \dots, X_t^n)$ is a Markov process with the symmetrized Dirichlet form given by

$$\oplus_n \mathcal{E}(g, g) = \int \sum_{i=1}^n \mathcal{E}(g_i, g_i) \mu(dx_1) \cdots \mu(dx_n)$$

which is the n -fold sum-Dirichlet form of \mathcal{E} .

Corollary 2.10. *Assume that μ satisfies $T_c I$ on \mathcal{X} with α convex. Then $\mu^{\otimes n}$ satisfies*

$$n\alpha \left(\frac{T_{\oplus_n c}(\nu, \mu^{\otimes n})}{n} \right) \leq I_{\oplus_n \mathcal{E}}(\nu|\mu^{\otimes n}), \quad \forall \nu \in M_1(\mathcal{X}^n). \quad (2.19)$$

In particular for all $(u, v) \in \Phi_c$, for all initial measure β on \mathcal{X}^n with $d\beta/d\mu^{\otimes n} \in L^2(\mu^{\otimes n})$ and for any $t, r > 0$,

$$\mathbb{P}_\beta \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{t} \int_0^t u(X_s^i) ds \geq \mu(v) + r \right) \leq \left\| \frac{d\beta}{d\mu^{\otimes n}} \right\|_2 e^{-nt\alpha(r)}. \quad (2.20)$$

Proof. As $\alpha^{\square n}(r) = n\alpha(r/n)$, (2.19) follows from Theorem 2.7. Noting that for $u, v \in \Phi_c$, $(\sum_{i=1}^n u(x_i), \sum_{i=1}^n v(x_i))$ as a couple of functions on \mathcal{X}^n belongs to $\Phi_{\oplus_n c}$, we obtain (2.20) by Theorem 2.2. \square

The tensorization of $W_p I$ in the dependent Gibbs measure case is carried out in Gao and Wu [20].

2.4. Relations between $W_2 I$, Poincaré and log-Sobolev inequalities. In the rest of the paper we are interested in two particular cases of $T_c I$: $W_1 I(\kappa)$ and $W_2 I(\kappa)$ introduced at Corollary 2.4.

Notation (Spectral gap). As usual, one says that μ satisfies a Poincaré inequality if

$$\mathrm{Var}_\mu(g) \leq c \mathcal{E}(g, g), \quad \forall g \in \mathbb{D}_2(\mathcal{L})$$

for some *finite* $c \geq 0$ and a Dirichlet form \mathcal{E} which is closable in $L^2(\mu)$. We denote $c_P(\mu)$ the best constant c in the above Poincaré inequality. It is the inverse of the *spectral gap* of \mathcal{L} .

From the work of Otto-Villani [34], we have the following observations.

Proposition 2.11. *Let \mathcal{X} be a complete connected Riemannian manifold and $\mu = e^{-V(x)} dx / Z$ where dx is the Riemannian volume measure, $V \in C^2(\mathcal{X})$ and $Z = \int_{\mathcal{X}} e^{-V} dx < +\infty$. Let $\mathbb{D}(\mathcal{E})$ be the space $H^1(\mathcal{X}, \mu)$ of those functions $g \in L^2(\mathcal{X}, \mu)$ such that $\nabla g \in L^2(TM, \mu)$ in the sense of distribution and consider the Dirichlet form,*

$$\mathcal{E}_\nabla(g, g) := \int_{\mathcal{X}} |\nabla g|^2 d\mu, \quad g \in \mathbb{D}(\mathcal{E})$$

and the associated Fisher-Donsker-Varadhan information $I(\nu|\mu)$, see (2.2).

(a) *If the log-Sobolev inequality below*

$$H(\nu|\mu) \leq 2c I(\nu|\mu), \quad \forall \nu$$

is satisfied, then μ satisfies $W_2 I(c)$. In other words the best constant $c_{\mathrm{LS}}(\mu)$ in the log-Sobolev inequality above satisfies

$$c_{\mathrm{LS}}(\mu) \geq c_{W_2 I}(\mu).$$

(b) *If $W_2 I(c)$ holds, then the Poincaré inequality holds with constant c . In other words the inverse spectral gap $c_P(\mu)$ of \mathcal{E}_∇ satisfies*

$$c_{W_2 I}(\mu) \geq c_P(\mu).$$

(c) *Assume that the Bakry-Emery curvature $\mathrm{Ric} + \mathrm{Hess}V$ is bounded from below by $K \in \mathbb{R}$, where Ric is the Ricci curvature and $\mathrm{Hess}V$ is the Hessian of V . If $W_2 I(c)$ holds with $cK \leq 1$ (this is possible by Part (a) and Bakry-Emery's criterion), then the log-Sobolev inequality*

$$H(\nu|\mu) \leq 2(2c - c^2 K) I(\nu|\mu), \quad \forall \nu$$

is also satisfied or equivalently

$$2c_{W_2 I}(\mu) - Kc_{W_2 I}(\mu)^2 \geq c_{\mathrm{LS}}(\mu).$$

Proof. Before the proof, let us remind the reader that $I = I_F/4$ where I_F is I in Otto-Villani's paper [34].

• Statement (a). The proof is direct, as by [34] or [4] a logarithmic Sobolev inequality implies the W_2H (sometimes called T_2) inequality so that

$$W_2(\nu, \mu) \leq \sqrt{2cH(\nu|\mu)} \leq 2c\sqrt{I(\nu|\mu)}$$

which is the announced conclusion.

• Statement (b). The proof follows from the usual linearization procedure. Set $\mu_\varepsilon = (1 + \varepsilon g)\mu$ for some smooth and bounded g with $\int g d\mu = 0$, we easily get: as $\varepsilon \rightarrow 0$,

$$I(\mu_\varepsilon|\mu)/\varepsilon^2 \rightarrow \frac{1}{4}\mathcal{E}_\nabla(g, g)$$

and by Otto-Villani [34, p.394], there exists r such that

$$\int g^2 d\mu \leq \sqrt{\mathcal{E}_\nabla(g, g)} \frac{W_2(\mu_\varepsilon, \mu)}{\varepsilon} + \frac{r}{\varepsilon} W_2^2(\mu_\varepsilon, \mu).$$

Using now $W_2I(c)$ we get

$$\int g^2 d\mu \leq 2c\sqrt{\mathcal{E}_\nabla(g, g)}\sqrt{\frac{I(\mu_\varepsilon|\mu)}{\varepsilon^2}} + \frac{4rc^2}{\varepsilon}I(\mu_\varepsilon|\mu).$$

Letting $\varepsilon \rightarrow 0$ gives the result.

• Statement (c). This result is a direct application of the HWI inequality [34, Th.3] in the Euclidean case and [4] for a general Riemannian manifold:

$$H(\nu|\mu) \leq 2W_2(\nu, \mu)\sqrt{I(\nu|\mu)} - \frac{K}{2}W_2^2(\nu, \mu). \quad (2.21)$$

□

Example 2.12. Let $\mu = \mathcal{N}(0, \Sigma)$ be the centered Gaussian measure on \mathbb{R}^n , with positive definite covariance matrix Σ . We claim that with respect to the Euclidean metric $|\cdot|$,

$$c_{W_1I}(\mu) = c_{W_2I}(\mu) = \lambda_{\min}^{-1}(\Sigma^{-1})$$

where λ_{\min} denotes the minimal eigenvalue. Indeed it is well known that with respect to the usual gradient ∇ on \mathbb{R}^n (cf. Ledoux [27]),

$$c_P(\mu) = c_{LS}(\mu) = \lambda_{\min}^{-1}(\Sigma^{-1}).$$

Then by Proposition 2.11(a) and (b),

$$c_{W_1I}(\mu) \leq c_{W_2I}(\mu) = \lambda_{\min}^{-1}(\Sigma^{-1}).$$

On the other hand, let $\nu = \mathcal{N}(m, \Sigma)$ where $m \in \mathbb{R}^n$. Then $W_1(\nu, \mu) = |m|$ (indeed $W_p(\nu, \mu) = |m|$ for all $p \geq 1$), and $I(\nu|\mu) = \frac{1}{4}|\Sigma^{-1}m|^2$. Thus

$$c_{W_1I}(\mu) \geq \left(\inf_{m \in \mathbb{R}^n} \frac{|\Sigma^{-1}m|}{|m|} \right)^{-1} = \lambda_{\min}^{-1}(\Sigma^{-1})$$

completing the proof of the claim.

As is seen from this proposition, in the bounded below curvature case, $W_2I(\kappa)$ is (qualitatively as constants are lost) equivalent to a logarithmic Sobolev inequality. It is an interesting question to know whether or not it is the case in full generality. We do not believe this to be true, and the hint for this conjecture comes from the fact that even for the usual W_2H inequality, the only cases of measures satisfying W_2H but not log-Sobolev inequality which are known so far have an infinite curvature [9]. That is why in the rest of the paper we are mainly interested in $W_1I(\kappa)$. Nevertheless, note that W_2I may easily be applied to obtain tensorization results in some situations where dependence occurs and the log-Sobolev inequality seems to be unfruitful, see [20].

3. POINCARÉ INEQUALITY IMPLIES Hoeffding's Deviation Inequality

3.1. Relations between Poincaré and W_1I inequalities. The purpose of this section is to establish

Theorem 3.1. *Let $((X_t), \mathbb{P}_\mu)$ be a stationary ergodic Markov process.*

(a) *The Poincaré inequality*

$$\mathrm{Var}_\mu(g) \leq c_P \mathcal{E}(g, g), \quad \forall g \in \mathbb{D}_2(\mathcal{L}) \quad (3.1)$$

implies

$$\|\nu - \mu\|_{TV}^2 \leq 4c_P I(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}). \quad (3.2)$$

In particular for every initial probability measure $\beta \ll \mu$ with $d\beta/d\mu \in L^2(\mu)$ and for all $u \in b\mathcal{B}$, $t, r > 0$,

$$\mathbb{P}_\beta \left(\frac{1}{t} \int_0^t u(X_s) ds - \mu(u) \geq r \right) \leq \left\| \frac{d\beta}{d\mu} \right\|_2 \exp \left(-\frac{tr^2}{c_P \delta(u)^2} \right) \quad (3.3)$$

where $\delta(u) := \sup_{x, y \in \mathcal{X}} |u(x) - u(y)|$ is the oscillation of u .

(b) *Conversely, under the additional assumption that (X_t, \mathbb{P}_μ) is reversible and $R_1 := \int_0^\infty e^{-t} P_t dt$ is μ -uniformly positive improving, if there is some left-continuous and increasing α such that $\alpha(r) > 0$ for all $r > 0$ and*

$$\alpha(\|\nu - \mu\|_{TV}) \leq I(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}),$$

then the Poincaré inequality (3.1) holds.

(c) *In other words when d is the trivial metric, W_1I is equivalent to Poincaré's inequality in the symmetric and uniformly positive improving case.*

Here the kernel R_1 is said μ -uniformly positive improving, if for any $\varepsilon > 0$,

$$\inf_{A, B: \mu(A), \mu(B) \geq \varepsilon} \mu(1_A R_1 1_B) > 0.$$

Note that if the symmetric semigroup (P_t) is irreducible, i.e. $R_1(x, A) = \int_0^\infty e^{-t} P_t(x, A) dt > 0$, $\forall x \in \mathcal{X}$ for any $A \in \mathcal{B}$ charged by μ , then R_1 is μ -uniformly positive improving for every $t > 0$ (cf. [46, 22]). A typical example of this situation is when $P_t(x, dy) = p_t(x, y)\mu(dy)$ and $p_t(x, y) > 0$, $\mu \otimes \mu$ -a.s. for some $t > 0$. See [46, 22] for related results and references on this subject.

Remarks 3.2.

(i) Let $d(x, y) = \mathbf{1}_{x \neq y}$ (the trivial metric) and $\Phi = \{(u, u); \delta(u) \leq 1\}$. Then

$$\frac{1}{2} \|\nu - \mu\|_{\text{TV}} = W_1(\nu, \mu) = T_\Phi(\nu, \mu).$$

Hence (3.2) is exactly the inequality $T_\Phi I$ or $W_1 I(c)$ with $4c^2 = c_P$, and (3.3) is a direct consequence of (3.2) and Corollary 2.4-(1-d).

- (ii) Hoeffding type inequality (3.3) improves a similar result by Cattiaux and Guillin [8]. Lezaud [28] proved a deviation inequality, which is better than (3.3) in the moderate deviation scale (r very small), nevertheless his proof involves a difficult argument based on Kato's theory of perturbation of operators.
- (iii) Inequality (3.3) is only meaningful for r small enough, since its left-hand side vanishes as soon as $r > \delta(u)$.
- (iv) About the deviation inequality in Theorem 2.2(c), several variants are already known. K. Marton [32] proved a Gaussian deviation inequality for Doeblin recurrent Markov chains $(X_n)_{n \in \mathbb{N}}$ by means of L^1 -transportation inequality for the law of the chain. Her result is next generalized in Djellout and al. [13].
- (v) Does $W_1 I(c)$ imply the exponential decay: $W_1(\nu P_t, \mu) \leq C e^{-\delta t} W_1(\nu, \mu)$? In the trivial metric case this decay means Doeblin recurrence and Theorem 3.1 gives a negative answer. Indeed, there exist reversible Markov processes having a positive spectral gap which are not Doeblin recurrent, for instance the Ornstein-Uhlenbeck processes.

Proof of Theorem 3.1. • Statement (a). As noticed at Remark 3.2-(i), all we have to prove is the transportation inequality (3.2). To this end we may assume that $f = d\nu/d\mu$ satisfies $\sqrt{f} \in \mathbb{D}(\mathcal{E})$. By the inequality (3.5) in Theorem 3.3 below and the assumed Poincaré inequality,

$$\frac{1}{4} \|\nu - \mu\|_{\text{TV}}^2 \leq \text{Var}_\mu(\sqrt{f}) \leq c_P \mathcal{E}(\sqrt{f}, \sqrt{f}) = c_P I(\nu|\mu).$$

• Statement (b). This converse part is based on the third author's paper [46]. Indeed by Theorem 2.2 and our assumption,

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{P}_\mu \left(\left| \frac{1}{t} \int_0^t u(X_s) ds - \mu(u) \right| > r \right) < 0$$

for all $u : \mathcal{X} \rightarrow \mathbb{R}^d$ bounded and measurable ($d \geq 1$) and any $r > 0$. This implies by [46, Theorem 3.9] that (P_t) satisfies the Resolvent Tail-Norm Condition (named in [46]). This last property together with the uniform positive improving property implies the existence of a spectral gap by [46, Theorem 4.1] or [22, Theorem 4.4].

• Finally, (c) is a direct consequence of (a) and (b). □

3.2. A CKP type inequality. During the proof of (a), we have used inequality (3.5) which is part of the Theorem 3.3 below. The usual CKP (Csiszár-Kullback-Pinsker) inequality is

$$\frac{1}{2} \|\nu - \mu\|_{\text{TV}}^2 \leq H(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}).$$

We shall see during the proof of Lemma 3.7 at (3.12) that (3.5) is of the similar form

$$\frac{1}{4} \|\nu - \mu\|_{\text{TV}}^2 \leq I(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X})$$

for some well chosen I . This is the reason why it is called a CKP type inequality.

Let $((X_t)_{t \geq 0}, \mathbb{P})$ be the pure jump Markov process on the state space \mathcal{X} with generator $\mathcal{L}g(x) = \int_{\mathcal{X}} [g(y) - g(x)] \mu(dy)$, $g \in b\mathcal{B}$ and initial law $\mu \in M_1(\mathcal{X})$. A representation of X is given by

$$X_t = Y_{N_t}, \quad t \geq 0 \quad (3.4)$$

where N is a Poisson process with parameter one which is independent of the sequence $(Y_n)_{n \geq 0}$ of independent identically μ -distributed \mathcal{X} -valued random variables.

Theorem 3.3. *Let μ be any probability measure on \mathcal{X} and X be its associated process defined at (3.4).*

(1) *The following two equivalent families of inequalities hold true:*

(a) *For every μ -probability density f , i.e. $f \geq 0$ and $\mu(f) = 1$,*

$$\|f\mu - \mu\|_{TV}^2 \leq 4\text{Var}_\mu(\sqrt{f}); \quad (3.5)$$

(b) *For all $\lambda \in \mathbb{R}$ and $u \in b\mathcal{B}$ such that $\mu(u) = 0$ and $\delta(u) \leq 2$, we have*

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E} \exp \left(\lambda \int_0^t u(X_s) ds \right) \leq \rho(\lambda) \quad (3.6)$$

where $\rho(\lambda) = \mathbf{1}_{|\lambda| \leq 1} \lambda^2 + \mathbf{1}_{|\lambda| > 1} (2|\lambda| - 1)$.

(2) *The constant 4 in (3.5) is sharp and the equality holds if and only if*

$$\mu \circ f^{-1} = p\delta_{\frac{1-p}{p}} + (1-p)\delta_{\frac{p}{1-p}} \quad (3.7)$$

for some $0 < p < 1$.

(3) *The function ρ is the best right-hand side for the inequality (3.6) and the equality is achieved for some $\lambda \in \mathbb{R}$ and some $u \in b\mathcal{B}$ such that $\mu(u) = 0$ and $\delta(u) \leq 2$, if and only if there exists $0 < p < 1$ such that*

$$\lambda = 1 - 2p := \lambda(p)$$

and

$$\mu \circ u^{-1} = p\delta_{2-2p} + (1-p)\delta_{-2p}. \quad (3.8)$$

Proof. The statement of Theorem 3.3 is simply a gathering of Lemmas 3.5 and 3.7 below. These lemmas provide two distinct proofs of inequality (3.5). \square

Remarks 3.4.

- (i) Note the symmetry $p \leftrightarrow 1 - p$ in (3.7) and the antisymmetry in (3.8) : $(\lambda, u) \leftrightarrow (-\lambda, -u)$ gives $p \leftrightarrow 1 - p$.
- (ii) Let us recall some well-known facts about optimal transportation [39, 38]. The total variation $\|\mu - \nu\|_{TV}$ is the minimal transportation cost $T_c(\nu, \mu)$ associated with the cost function $c(x, y) = 2 \cdot \mathbf{1}_{x \neq y}$ (see Remarks 3.2) :

$$\|\mu - \nu\|_{TV} = 2 \inf_{\pi \in P(\nu, \mu)} \pi(\{(x, y); x \neq y\})$$

where $P(\nu, \mu) = \{\pi \in M_1(\mathcal{X}^2) : \pi_0 = \nu, \pi_1 = \mu\}$. The infimum is attained on $P(\nu, \mu)$, these minimizers are often called optimal couplings of ν and μ . One has the following characterization: $\pi \in P(\nu, \mu)$ is optimal if and only if there exists some measurable function u on \mathcal{X} such that

$$\pi(\{(x, y) \in \mathcal{X}^2; u(x) - u(y) = 2 \cdot \mathbf{1}_{x \neq y}\}) = 1. \quad (3.9)$$

Such a u is often called an optimal Kantorovich potential.

Let $0 < p < 1$ and f satisfy (3.7). Any optimal coupling π of $f\mu$ and μ satisfies

$$\pi(\{(x, y); x \neq y\}) = |1 - 2p| = |\lambda(p)|$$

and it admits an optimal Kantorovich potential u (see (3.9)) satisfying (3.8). More precisely,

- (a) $\{u = 2 - 2p\} = \{f = (1 - p)/p\}$ if $0 < p < 1/2$;
- (b) $\{u = 2 - 2p\} = \{f = p/(1 - p)\}$ if $1/2 < p < 1$;
- (c) When $p = 1/2$, (3.7) is equivalent to $f = 1$ μ -a.e., that is $\nu = \mu$. On the other hand, any u satisfying (3.8) with $p = 1/2$ is an optimal potential for the trivial optimal coupling $\pi(dxdy) = \mu(dx)\delta_x(dy)$.

The equalities in (a) and (b) are satisfied up to μ -negligible sets.

- (iii) Inequality (3.5) is already known in statistics. Indeed for $\nu = f\mu$, the Hellinger distance between ν and μ is given by

$$d_H^2(\nu, \mu) = \frac{1}{2} \int (1 - \sqrt{f})^2 d\mu = 1 - \mu(\sqrt{f}).$$

The known inequality (see Gibbs and Su [21]) is

$$\frac{1}{4} \|\nu - \mu\|_{\text{TV}}^2 \leq d_H^2(\nu, \mu) [2 - d_H^2(\nu, \mu)]$$

and the above right-hand side is exactly $1 - [\mu(\sqrt{f})]^2 = \text{Var}_\mu(\sqrt{f})$.

We are going to give two different new proofs of (3.5). The first one, at Lemma 3.5, is elementary and provides a characterization of these f 's which achieve the equality in (3.5). The second one, at Lemma 3.7, is in the spirit of this paper since it is a corollary of Theorem 2.2. It also provides a characterization of the real parameters λ which achieve the equality in the ‘‘dual’’ inequality (3.6).

A first proof of the CKP type inequality (3.5) and more. It may be seen as an amusing exercise in graduate courses.

Lemma 3.5. *The inequality (3.5) holds for every μ -probability density f . Moreover, equality is achieved in (3.5) if and only (3.7) is satisfied for some $0 < p < 1$.*

Proof of Lemma 3.5. Assume that $\mu(f = 1) < 1$ (trivial otherwise) in the following. Then $0 < \mu(f < 1), \mu(f > 1) < 1$.

• *Step 1. Reduction to the two-values case.* This step might be the most difficult. Let $A = \{f < 1\}, B = \{f \geq 1\}$ and $\bar{f} := \mathbb{E}^\mu(f | \sigma(A)) = \alpha \mathbf{1}_A + \beta \mathbf{1}_B$ with $\alpha = \mu(\mathbf{1}_A f) / \mu(A)$ and $\beta = \mu(\mathbf{1}_B f) / \mu(B)$. As $\mu(f < 1) > 0$ and $\mu(f > 1) > 0$, one sees that $0 \leq \alpha < 1 < \beta$. Therefore, $f < 1$ if and only if $\bar{f} < 1$ and

$$\frac{1}{2} \mu(|1 - f|) = \int_{f < 1} (1 - f) d\mu = \int_{f < 1} (1 - \bar{f}) d\mu = \int_{\bar{f} < 1} (1 - \bar{f}) d\mu = \frac{1}{2} \mu(|1 - \bar{f}|)$$

On the other hand, by Jensen's inequality

$$\text{Var}_\mu \left(\sqrt{\bar{f}} \right) = 1 - \left[\mu \left(\sqrt{\bar{f}} \right) \right]^2 \leq 1 - [\mu(\sqrt{f})]^2$$

and the equality holds if and only if $f = \bar{f}$. It follows that for f to satisfy (3.5) it is enough that (3.5) also holds for \bar{f} . Without loss of generality, we may assume from now on that $f = \bar{f}$, i.e. there are two numbers $0 \leq a < 1 < b$ such that

$$p = \mu(f = a^2) \in (0, 1), \quad q = \mu(f = b^2) \in (0, 1), \quad p + q = 1.$$

• *Step 2.* Let $\xi \in (0, \pi/2)$ such that $\sqrt{p} = \cos \xi$, $\sqrt{q} = \sin \xi$. Since

$$1 = \mu(f) = pa^2 + qb^2$$

we may choose $\theta \in [0, \pi/2]$ such that

$$a = \frac{1}{\sqrt{p}} \cos \theta, \quad b = \frac{1}{\sqrt{q}} \sin \theta.$$

As $a \in [0, 1)$, $\theta > \xi$. Now noting that

$$\text{Var}_\mu(\sqrt{f}) = (b-a)^2 pq, \quad \frac{1}{2} \mu(|1-f|) = \mu(1_{f < 1}(1-f)) = (1-a^2)p$$

the inequality (3.5) amounts to saying that

$$\begin{aligned} (b-a)\sqrt{pq} - (1-a^2)p &= \sin \theta \cos \xi - \cos \theta \sin \xi - (\cos^2 \xi - \cos^2 \theta) \\ &= \sin(\theta - \xi) - (\cos^2 \xi - \cos^2 \theta) =: g(\xi, \theta) \geq 0 \end{aligned} \quad (3.10)$$

for all $0 < \xi < \theta \leq \pi/2$. Fix ξ and put $h(\theta) = g(\xi, \theta)$. We have

$$h'(\theta) = \cos(\theta - \xi) - \sin(2\theta) = \sin\left(\frac{\pi}{2} - (\theta - \xi)\right) - \sin(2\theta)$$

and then $h'(\theta) = 0$ if and only if $\frac{\pi}{2} - (\theta - \xi) = 2\theta$ (if $2\theta \leq \pi/2$) or $\frac{\pi}{2} - (\theta - \xi) = \pi - 2\theta$ (if $2\theta \geq \pi/2$), i.e. $\theta = \theta_1 = \frac{\pi}{6} + \frac{\xi}{3}$ or $\theta = \theta_2 = \frac{\pi}{2} - \xi$.

Case 1. $\xi \in [\pi/4, \pi/2)$: Since $\theta_1, \theta_2 \leq \xi$ and $h'(\pi/2) > 0$, we have $h'(\theta) > 0$ for all $\theta \in (\xi, \pi/2]$. Hence for all $\theta \in (\xi, \pi/2]$, $g(\xi, \theta) = h(\theta) > h(\xi) = 0$.

Case 2. $\xi \in (0, \pi/4)$: In this case $\xi < \theta_1 < \theta_2 < \pi/2$. Since $h'(\xi) > 0$, $h'(\pi/2) > 0$, $h(\xi) = h(\theta_2) = 0$, h' is positive, negative and positive respectively on (ξ, θ_1) , (θ_1, θ_2) and $(\theta_2, \pi/2)$. Consequently for all $\theta \in (\xi, \pi/2]$, $g(\xi, \theta) = h(\theta) \geq 0$ and the equality holds if and only if $\theta = \theta_2 = \frac{\pi}{2} - \xi$.

• *Step 3. Equality in (3.5).* If $f = 1$, μ -a.s., then the equality in (3.5) holds. Now, assume $\mu(f = 1) < 1$ and the equality in (3.5). By Step 1, $f = \bar{f}$, i.e. f takes only two values $\sqrt{a} < 1 < \sqrt{b}$. By Step 2, this is possible if and only if $\xi < \pi/4$ and $\theta = \pi/2 - \xi$, i.e. $p = \cos^2 \xi > \frac{1}{2}$, $\cos^2 \theta + \cos^2 \xi = a^2 p + p = 1$. That is $p(1 + a^2) = 1$. Therefore, either there exist two numbers $0 < a < 1 < b$ such that $\mu(f \in \{a^2, b^2\}) = 1$ and

$$\begin{cases} pa^2 + (1-p)b^2 &= 1 \\ p(1+a^2) &= 1 \end{cases}$$

or $f = 1$, μ -a.e. This proves the desired parametrization (3.7) for $1/2 \leq p < 1$, and hence for all $0 < p < 1$ because of the symmetry in (3.7). \square

Example 3.6. (Bernoulli distribution). Let μ be the Bernoulli distribution on $\mathcal{X} = \{0, 1\}$ with $\mu(\{1\}) = p \in (0, 1)$. Consider the Dirichlet form $\mathcal{E}(g, g) = (g(1) - g(0))^2$. Then $c_P(\mu) = pq$. By Theorem 3.1-(a) and Remark 3.2-(i), we see that

$$pqW_1(\nu, \mu)^2 \leq I(\nu|\mu)$$

where W_1 is built with the trivial metric. The constant pq is sharp. However μ does not satisfy any $W_2I(\kappa)$ as is easily seen with $\nu = \mu_\varepsilon = (1 + \varepsilon g)\mu$.

A second proof of the CKP type inequality (3.5) and more. Recall that $((X_t)_{t \geq 0}, \mathbb{P})$ is the pure jump Markov process defined at (3.4).

Lemma 3.7.

- (1) *The inequality (3.5) holds for all probability density f if and only if for all $\lambda \in \mathbb{R}$ and $u \in b\mathcal{B}$ such that $\mu(u) = 0$ and $\delta(u) \leq 2$, we have*

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E} \exp \left(\lambda \int_0^t u(X_s) ds \right) \leq \lambda^2. \quad (3.11)$$

- (2) *For all $\lambda \in \mathbb{R}$ and $u \in b\mathcal{B}$ such that $\mu(u) = 0$ and $\delta(u) \leq 2$, the inequality (3.11) holds true.*
(3) *The equality is achieved in (3.11) for some $\lambda \in \mathbb{R}$ and some $u \in b\mathcal{B}$ such that $\mu(u) = 0$ and $\delta(u) \leq 2$, if and only if $\lambda = 1 - 2p$ and (3.8) hold for some $0 < p < 1$.*
(4) *The function ρ is sharp in inequality (3.6), that is*

$$\rho(\lambda) = \sup_u \left\{ \limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E} \exp \left(\lambda \int_0^t u(X_s) ds \right) \right\}$$

for all real λ , where the supremum is taken over all $u \in b\mathcal{B}$ such that $\mu(u) = 0$ and $\delta(u) \leq 2$.

- (5) *For all $\lambda \in \mathbb{R}$ and $u \in b\mathcal{B}$ such that $\mu(u) = 0$ and $\delta(u) \leq 2$, the inequalities (3.5), (3.6) and (3.11) are equivalent. The equality in (3.6) is never achieved whenever $|\lambda| \geq 1$.*

Proof. • Statement (1): Clearly, μ is a reversing measure for the process X and the associated Dirichlet form is

$$\mathcal{E}(g, g) = \text{Var}_\mu(g), \quad g \in \mathbb{D}_2(\mathcal{L}) = L^2(\mu). \quad (3.12)$$

Therefore, statement (1) is a direct consequence of Theorem 2.2-(b') applied with $\Phi = \Phi_c$, $c(x, y) = 2 \cdot \mathbf{1}_{x \neq y}$ and $\alpha(a) = a^2/4$. Note that one passes from $\lambda \geq 0$ to $\lambda \in \mathbb{R}$, by considering $-u$ instead of u . This is possible since $\delta(-u) = \delta(u) \leq 2$ and $\mu(-u) = \mu(u) = 0$.

- Statement (2): Let us introduce the notations $\psi(t) = \log \varphi(t)$ with $\varphi(t) = \mathbb{E} \exp \left(\int_0^t v(X_s) ds \right)$ and $v = \lambda u$. We wish to get an upper bound for $\psi(t)$. For all $t \geq 0$,

$$\begin{aligned} \psi'(t) &= \frac{\varphi'(t)}{\varphi(t)} = \mathbb{E}^{v,t}[v(X_t)] \\ \psi''(t) &= \left[\frac{\varphi''}{\varphi} - \left(\frac{\varphi'}{\varphi} \right)^2 \right] (t) \end{aligned}$$

where $\mathbb{E}^{v,t}$ is the expectation with respect to

$$\mathbb{P}^{v,t} = \exp\left(\int_0^t v(X_s) - \psi(t)\right) \cdot \mathbb{P}.$$

In order to compute $\psi''(t)$, let us apply Itô's formula to $Y_t = v(X_t) \exp \int_0^t v(X_s) ds$. This gives

$$dY_t = \exp\left(\int_0^t v(X_s) ds\right) \left[(\mathcal{L}v(X_t) + v(X_t)^2) dt + dM_t \right]$$

where M is some martingale. It follows that

$$\begin{aligned} \varphi'(t) &= \mathbb{E}Y_t \\ &= \mathbb{E}\left(v(X_0) + \int_0^t dY_s\right) \\ &= \mathbb{E}v(X_0) + \int_0^t \mathbb{E}\left(\exp\left(\int_0^s v(X_r) dr\right) [\mathcal{L}v(X_s) + v(X_s)^2]\right) ds \end{aligned}$$

Hence $\varphi''(t) = \mathbb{E}\left(\exp\left(\int_0^t v(X_s) ds\right) [\mathcal{L}v(X_t) + v(X_t)^2]\right)$ and we obtain

$$\psi''(t) - \mathbb{E}^{\lambda u,t}(\mathcal{L}[\lambda u](X_t)) = \lambda^2 \mathbb{V}^{\lambda u,t}[u(X_t)]$$

where $\mathbb{V}^{\lambda u,t}$ is the variance with respect to $\mathbb{P}^{\lambda u,t}$.

Since $\mu(u) = 0$, $\mathcal{L}u(x) = \mu(u) - u(x) = -u(x)$, one sees that

$$\psi''(t) + \psi'(t) = \lambda^2 \mathbb{V}^{\lambda u,t}[u(X_t)], \quad t \geq 0. \quad (3.13)$$

As it is assumed that $\delta(u) \leq 2$, we have $\mathbb{V}^{\lambda u,t}[u(X_t)] \leq 1$ and

$$\psi''(t) + \psi'(t) \leq \lambda^2, \quad t \geq 0. \quad (3.14)$$

Clearly, $\psi(0) = 0$ and $\psi'(0) = \mathbb{E}v(X_0) = \mu(v) = 0$. Let θ be the solution of

$$\theta''(t) + \theta'(t) = \lambda^2, \quad t \geq 0 \quad (3.15)$$

with the same initial conditions as ψ : $\theta(0) = \theta'(0) = 0$. Denote $\gamma(t) = \theta(t) - \psi(t)$. As $\gamma''(t) + \gamma'(t) \geq 0$, for all $t \geq 0$, we see that $\gamma'(t) + \gamma(t) \geq \gamma'(0) + \gamma(0) = 0$. By Gronwall's inequality, it follows that $\gamma(t) \geq 0$ for all $t \geq 0$. This means that

$$\psi(t) \leq \lambda^2(e^{-t} + t - 1), \quad \forall t \geq 0$$

where the right hand side is $\theta(t)$ which is obtained by elementary differential calculus. The upper bound (3.11) follows immediately.

• Statement (3): Now, let's investigate the equality in (3.11). Inspecting the proof of statement (2), one sees that this equality holds if and only if the equality holds in (3.14) asymptotically as t tends to infinity, i.e.

$$\lim_{t \rightarrow \infty} \mathbb{V}^{\lambda u,t}[u(X_t)] = 1 \quad (3.16)$$

To see this, remark that the solution to (3.15) with general initial conditions $\theta(0) = \theta_o$, $\theta'(0) = \theta'_o$ is $\theta(t) = \lambda^2(e^{-t} + t - 1) + \theta_o + \theta'_o(1 - e^{-t})$, $t \geq 0$ so that $\lim_{t \rightarrow \infty} \theta(t)/t = \lambda^2$, for any initial conditions. Using Gronwall's inequality as above, one sees now with (3.13) that the desired equality holds if and only if $\liminf_{t \rightarrow \infty} \mathbb{V}^{\lambda u,t}[u(X_t)] = 1$. But this is equivalent to (3.16) since $\mathbb{V}^{\lambda u,t}[u(X_t)] \leq 1$ as $\delta(u) \leq 2$.

It is an easy exercise to show that for any random variable Z such that $-1 \leq Z \leq 1$ almost surely, we have $\text{Var}(Z) \leq 1$ and $\text{Var}(Z) = 1$ if and only if the law of Z is $\frac{1}{2}(\delta_{-1} + \delta_{+1})$. It follows that (3.16) holds if and only if u only takes two values $a + 1$ and $a - 1$ where a is some real such that $\mu(u) = 0$ and

$$\lim_{t \rightarrow \infty} \mathbb{P}^{\lambda u, t}(u(X_t) = a + 1) = \lim_{t \rightarrow \infty} \mathbb{P}^{\lambda u, t}(u(X_t) = a - 1) = 1/2. \quad (3.17)$$

One immediately sees that with $p := \mu(u = a + 1)$, $\mu(u) = 0$ implies that $a = a(p) = 1 - 2p$. Therefore, the image law $\mu \circ u^{-1}$ of μ by u must satisfy (3.8) for some $0 < p < 1$.

Fix $0 < p < 1$ and take u as in (3.8). Let us introduce the $\{+, -\}$ -valued process defined by $Z_t^{(p)} = u(X_t) - a(p)$, $t \geq 0$ and denote χ the identity on $\{+, -\}$: $\chi(+)=+1$, $\chi(-)=-1$. Rewriting (3.17) as a ratio, one obtains

$$\begin{aligned} 1 &= \lim_{t \rightarrow \infty} \frac{\mathbb{E} \left[\mathbf{1}_{\{u(X_t)=a(p)+1\}} \left\{ \exp \left(\lambda \int_0^t u(X_s) ds \right) - \psi(t) \right\} \right]}{\mathbb{E} \left[\mathbf{1}_{\{u(X_t)=a(p)-1\}} \left\{ \exp \left(\lambda \int_0^t u(X_s) ds \right) - \psi(t) \right\} \right]} \\ &= \lim_{t \rightarrow \infty} \frac{\mathbb{E} \left[\mathbf{1}_+(Z_t^{(p)}) \exp \left(\lambda \int_0^t \chi(Z_s^{(p)}) ds \right) \right]}{\mathbb{E} \left[\mathbf{1}_-(Z_t^{(p)}) \exp \left(\lambda \int_0^t \chi(Z_s^{(p)}) ds \right) \right]} \end{aligned}$$

The process $Z^{(p)}$ is still Markov and its generator is given for all $\sigma \in \{+, -\}$ and $g \in \mathbb{R}^{\{+, -\}}$ by

$$A_p g(\sigma) = m_p(g) - g(\sigma)$$

where $m_p = p\delta_+ + (1-p)\delta_-$ is the image of μ by $x \in \mathcal{X} \mapsto u(x) - a(p)$.

The Feynman-Kac semigroup associated with $Z^{(p)}$ and the potential $\lambda\chi$ is defined for all $\sigma \in \{+, -\}$ and $g \in \mathbb{R}^{\{+, -\}}$ by

$$P_t^\lambda g(\sigma) := \mathbb{E} \left[g(Z_t^{(p)}) \exp \left(\lambda \int_0^t \chi(Z_s^{(p)}) ds \right) \middle| Z_0^{(p)} = \sigma \right].$$

This allows to rewrite

$$\begin{aligned} \frac{\mathbb{E} \left[\mathbf{1}_+(Z_t^{(p)}) \exp \left(\lambda \int_0^t \chi(Z_s^{(p)}) ds \right) \right]}{\mathbb{E} \left[\mathbf{1}_-(Z_t^{(p)}) \exp \left(\lambda \int_0^t \chi(Z_s^{(p)}) ds \right) \right]} &= \frac{\langle m_p, P_t^\lambda \mathbf{1}_+ \rangle}{\langle m_p, P_t^\lambda \mathbf{1}_- \rangle} \\ &= \frac{\langle m_p, \exp[t(A_p + \lambda\chi)] \mathbf{1}_+ \rangle}{\langle m_p, \exp[t(A_p + \lambda\chi)] \mathbf{1}_- \rangle} \end{aligned}$$

which can be computed explicitly by means of elementary linear algebra in \mathbb{R}^2 . Indeed, seeing the functions as vectors: $g = \begin{pmatrix} g(+), \\ g(-) \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} = x\mathbf{1}_+ + y\mathbf{1}_-$ where $(\mathbf{1}_+, \mathbf{1}_-)$ is the canonical base of $\mathbb{R}^{\{+, -\}}$, one immediately identifies the operator $A_p + \lambda\chi$ with the matrix $M = \begin{pmatrix} -1 + p + \lambda & 1 - p \\ p & -p - \lambda \end{pmatrix}$. It is a simple exercise to show that M has two real distinct eigenvalues $s_1 > s_2$ given by $s_1 = -1/2 + \sqrt{\Delta}/2$ and $s_2 = -1/2 - \sqrt{\Delta}/2$ with $\Delta = 1 + 4\lambda(\lambda + 2p - 1)$ and $\Delta > 0$ since $0 < p < 1$. One can check that $\mathbf{1}_+ = (v_1 - v_2)/\sqrt{\Delta}$

and $\mathbf{1}_- = (-\frac{p+\lambda+s_2}{p}v_1 + \frac{p+\lambda+s_1}{p}v_2)/\sqrt{\Delta}$ where $v_1 = \begin{pmatrix} p+\lambda+s_1 \\ p \end{pmatrix}$ and $v_2 = \begin{pmatrix} p+\lambda+s_2 \\ p \end{pmatrix}$ are eigenvectors associated respectively with s_1 and s_2 . Using the elementary remark that $s_1 > s_2$ implies that $\lim_{t \rightarrow \infty} e^{-ts_1} e^{tM}(av_1 + bv_2) = av_1$ for all $a, b \in \mathbb{R}$, one obtains

$$\lim_{t \rightarrow \infty} \frac{\langle m_p, P_t^\lambda \mathbf{1}_+ \rangle}{\langle m_p, P_t^\lambda \mathbf{1}_- \rangle} = -\frac{p}{p+\lambda+s_2}$$

for all $0 < p < 1$ and $\lambda \in \mathbb{R}$. Putting everything together, we conclude that the equality in (3.11) holds if and only if $-p/(p+\lambda+s_2) = 1$ which in turn is also equivalent to $\lambda = 1 - 2p$.

• Statement (4): Let us denote

$$R(\lambda) := \sup_u \left\{ \limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E} \exp \left(\lambda \int_0^t u(X_s) ds \right) \right\}, \quad \lambda \in \mathbb{R}$$

where the supremum is taken over all $u \in b\mathcal{B}$ such that $\mu(u) = 0$ and $\delta(u) \leq 2$. We have to show that $R = \rho$.

Let us first prove that R is a convex function. As a log-Laplace transform, for each u and t , $\log \mathbb{E} \exp \left(\lambda \int_0^t u(X_s) ds \right)$ is a convex function of λ . Since X has stationary independent increments, by a standard sub-additivity argument, one shows that $\limsup_{t \rightarrow \infty}$ is a genuine limit: $\lim_{t \rightarrow \infty}$. It follows that for each u , $\limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E} \exp \left(\lambda \int_0^t u(X_s) ds \right)$ is a convex function of λ . Finally, R is convex as it is the supremum of convex functions.

Because of statements (2) and (3), it is already seen that $R(\lambda) = \rho(\lambda) = \lambda^2$ for all $-1 < \lambda < 1$. On the other hand, since $\sup |u| \leq 2$ for each u such that $\mu(u) = 0$ and $\delta(u) \leq 2$, it is clear that $\limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E} \exp \left(\lambda \int_0^t u(X_s) ds \right) \leq 2|\lambda|$ for all real λ . Therefore, for all λ , $\rho(\lambda) \leq R(\lambda) \leq \text{cvr}(\lambda)$ where cvr is the convex envelope of $r(\lambda) = \min(\lambda^2, 2|\lambda|)$. Indeed, the first inequality holds since ρ is the lowest convex function which matches with λ^2 on $\lambda \in (-1, 1)$ while the second one follows from the inequality $R(\lambda) \leq \min(\lambda^2, 2|\lambda|)$ for all λ and the convexity of R . One concludes that $R = \rho$, remarking that $\text{cvr} = \rho$.

• Statement (5) is a direct consequence of statement (1) and the proof of statement (4). \square

Note that W_1I for the trivial metric implies W_1I for any bounded metric. So our next purpose is to obtain W_1I for unbounded metrics. Our study is naturally separated into two sections. Next Section 4 is concerned with estimating sharply c_{W_1I} under strong dissipative conditions. In Section 5, Lyapunov function conditions for W_1I or more general $T_\Phi I$ are taken into consideration.

4. SPECTRAL GAP IN THE SPACE OF LIPSCHITZ FUNCTIONS IMPLIES W_1I FOR DIFFUSION PROCESSES

4.1. General observations. We begin with the particular case where $\mu(dx) = e^{-V(x)} dx / Z$ (Z is the normalization constant) with $V \in C^2(\mathcal{X})$ on a connected and complete Riemannian manifold \mathcal{X} , the diffusion (X_t) generated by $\mathcal{L} = \Delta - \nabla V \cdot \nabla$ (Δ, ∇ are respectively

the Laplacian and gradient on M) is reversible with respect to μ , and the corresponding Dirichlet form is given by

$$\mathcal{E}(h, h) = \int_{\mathcal{X}} |\nabla h|^2 d\mu, \quad \forall h \in \mathbb{D}(\mathcal{E}) = H^1(\mathcal{X}, \mu).$$

Theorem 4.1. *Assume that $\int_{\mathcal{X}} d^2(x, x_0) d\mu(x) < +\infty$ and \mathcal{L} has a spectral gap on the space $C_{\text{Lip}}(\mathcal{X})$ of Lipschitz functions with respect to the Riemannian metric d , i.e. there is a best finite constant $C > 0$ such that for any $g \in C_{\text{Lip}} \cap b\mathcal{B}$ with $\mu(g) = 0$, there is $h \in \mathbb{D}_2(\mathcal{L})$ with $\mu(h) = 0$ solving the Poisson equation*

$$-\mathcal{L}h = -\Delta h + \nabla V \cdot \nabla h = g, \quad \mu\text{-a.s.}$$

such that one μ -version \tilde{h} of h verifies

$$\|\tilde{h}\|_{\text{Lip}} \leq C \|g\|_{\text{Lip}}. \quad (4.1)$$

Then μ satisfies $W_1 I(C)$:

$$W_1(\nu, \mu)^2 \leq 4C^2 I(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}) \quad (4.2)$$

or equivalently for any Lipschitz function u on \mathcal{X} and any initial probability measure β with $d\beta/d\mu \in L^2(\mu)$

$$\mathbb{P}_\beta \left(\frac{1}{t} \int_0^t u(X_s) ds \geq \mu(u) + r \right) \leq \left\| \frac{d\beta}{d\mu} \right\|_2 \exp \left(-t \frac{r^2}{4C^2 \|u\|_{\text{Lip}}^2} \right), \quad \forall r, t > 0. \quad (4.3)$$

Remarks 4.2.

- (i) Let C_{Lip}^0 be the Banach space of those $g \in C_{\text{Lip}}$ with $\mu(g) = 0$, equipped with $\|\cdot\|_{\text{Lip}}$. Hence the best constant C in (4.1) is exactly

$$\|(-\mathcal{L})^{-1}\|_{C_{\text{Lip}}^0}.$$

By the spectral decomposition we always have (cf. [47, Proof of Lemma 4.3])

$$C = \|(-\mathcal{L})^{-1}\|_{C_{\text{Lip}}^0} \geq \|(-\mathcal{L})^{-1}\|_{L^2(\mu) \cap \{g \in L^2(\mu); \mu(g)=0\}} = c_P(\mu).$$

- (ii) The constant in the concentration inequality (4.3) is sharp. Indeed let $dX_t = \sqrt{2} dB_t - X_t dt$, which is reversible with respect to $\mu = \mathcal{N}(0, 1)$ on \mathbb{R} . For this model we have $\nabla P_t = e^{-t} P_t \nabla$ and then

$$\nabla(-\mathcal{L})^{-1}g = \nabla \int_0^\infty P_t g dt = (1 - \mathcal{L})^{-1} \nabla g, \quad \forall g \in C_b^\infty(\mathbb{R}), \mu(g) = 0.$$

This implies $C = 1$ in (4.1). On the other hand for $u(x) = x$, under \mathbb{P}_μ , the law of

$$\frac{1}{t} \int_0^t u(X_s) ds = \frac{1}{t} \int_0^t X_s ds$$

is $\mathcal{N}(0, \sigma^2(t))$ where the variance $\sigma^2(t)$ is given by

$$\sigma^2(t) = \frac{2}{t^2} \iint_{0 \leq a \leq b \leq t} \mathbb{E}_\mu X_a X_b da db = \frac{2}{t^2} \iint_{0 \leq a \leq b \leq t} e^{-(b-a)} da db = \frac{2}{t} - \frac{2}{t^2} (1 - e^{-t})$$

from which we get

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{P}_\mu \left(\frac{1}{t} \int_0^t X_s ds > r \right) = -\frac{r^2}{4}.$$

This coincides with the upper bound $-r^2/(4C^2)$ derived from (4.3), showing the sharpness of (4.3).

Proof of Theorem 4.1. Let $\Phi = \{(g, g); \|g\|_{\text{Lip}} \leq 1, g \text{ bounded}\}$. Then $W_1(\nu, \mu) = T_\Phi(\nu, \mu)$ by Kantorovich-Rubinstein's theorem. Let us verify that (b') of Theorem 2.2 holds.

For any $g \in C_{\text{Lip}}$ with $\|g\|_{\text{Lip}} \leq 1$, let $h \in C_{\text{Lip}} \cap \mathbb{D}_2(\mathcal{L})$ such that $-\mathcal{L}h = g$. Hence

$$M_t(h) := h(X_t) - h(X_0) + \int_0^t g(X_s) ds$$

and

$$M_t^*(h) := h(X_0) - h(X_t) + \int_0^t g(X_s) ds$$

have the same law under \mathbb{P}_μ by the reversibility of $((X_t), \mathbb{P}_\mu)$. Consequently from Lyons-Meyer-Zheng's forward-backward martingale decomposition

$$S_t(g) := \int_0^t g(X_s) ds = \frac{1}{2}(M_t(h) + M_t^*(h)), \quad (4.4)$$

it follows that for any convex function ϕ on \mathbb{R} ,

$$\mathbb{E}_\mu \phi(S_t(g)) \leq \frac{1}{2} \mathbb{E}_\mu [\phi(M_t(h) + \phi(M_t^*(h)))] = \mathbb{E}_\mu \phi(M_t(h))$$

As $M_t(h)$ is a (forward) continuous martingale, $M_t(h) = B_{\tau_t}$ where (B_t) is some Brownian motion with respect to another time-changed filtration $(\hat{\mathcal{F}}_t)$, and $\tau_t = \langle M(h) \rangle_t$ is a $(\hat{\mathcal{F}}_t)$ -stopping time (a well known result). Since

$$\langle M(h) \rangle_t = 2 \int_0^t |\nabla h|^2(X_s) ds \leq 2C^2t$$

By Jensen's inequality, we obtain for all convex function ϕ on \mathbb{R} that

$$\mathbb{E}_\mu \phi(S_t(g)) \leq \mathbb{E} \phi(B_{\tau_t}) = \mathbb{E} \phi(\mathbb{E}[B_{2C^2t} | \hat{\mathcal{F}}_{\tau_t}]) \leq \mathbb{E} \phi(B_{2C^2t}) \quad (4.5)$$

Applying this to $\phi(x) = e^{\lambda x}$, we get

$$\mathbb{E}_\mu \exp \left(\lambda \int_0^t g(X_s) ds \right) \leq \mathbb{E} e^{\lambda B_{2C^2t}} = e^{\lambda^2 C^2 t}, \quad \forall \lambda \in \mathbb{R}. \quad (4.6)$$

Hence Theorem 2.2-(b') holds with $\Phi = \{(g, g); \|g\|_{\text{Lip}} \leq 1, g \text{ bounded}\}$ and $\alpha(r) = r^2/(4C^2)$. Therefore (4.2) and (4.3) follow from Theorem 2.2. \square

Klein-Ma-Privault [26] developed *convex concentration inequality* (4.5) for semimartingales instead of $S_t(g)$, by means of a forward-backward martingale calculus, but their result cannot be applied directly here.

Before estimating the constant C in condition (4.1), we extend the above result to general symmetric Markov diffusions by following Bakry [1].

Let $((X_t), \mathbb{P}_\mu)$ be a reversible ergodic Markov process with the Dirichlet form $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$, with continuous sample paths valued in some separable complete metric space (\mathcal{X}, d) (called Markov diffusion). We assume that $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$ is given by the carré-du-champs $\Gamma : \mathbb{D}(\mathcal{E}) \times \mathbb{D}(\mathcal{E}) \rightarrow L^1(\mu)$ (symmetric, bilinear definite nonnegative form):

$$\mathcal{E}(h, h) = \int_{\mathcal{X}} \Gamma(h, h) d\mu, \quad \forall h \in \mathbb{D}(\mathcal{E}). \quad (4.7)$$

The continuity of sample paths of (X_t) implies that Γ is a differentiation, that is: for all $(h_k)_{1 \leq k \leq n} \subset \mathbb{D}(\mathcal{E})$, $g \in \mathbb{D}(\mathcal{E})$ and $F \in C_b^1(\mathbb{R}^n)$,

$$\Gamma(F(h_1, \dots, h_n), g) = \sum_{i=1}^n \partial_i F(h_1, \dots, h_n) \Gamma(h_i, g).$$

With exactly the same proof as that of Theorem 4.1 we have

Theorem 4.3. *Assume that $\int_{\mathcal{X}} d^2(x, x_0) d\mu(x) < +\infty$ and for any $g \in C_{\text{Lip}}(\mathcal{X}, d)$ bounded with $\mu(g) = 0$, then $g \in \mathbb{D}(\mathcal{E})$ and*

$$\sqrt{\Gamma(g, g)} \leq \sigma \|g\|_{\text{Lip}}, \quad \mu\text{-a.s.} \quad (4.8)$$

and there is some $h \in \mathbb{D}_2(\mathcal{L})$ such that $-\mathcal{L}h = g$ (μ -a.e.) and a μ -continuous version \tilde{h} of h satisfying

$$\|\tilde{h}\|_{\text{Lip}} \leq C \|g\|_{\text{Lip}} \quad (4.9)$$

where $\sigma, C > 0$ are fixed constants. Then for any $u \in C_{\text{Lip}}(\mathcal{X}, d)$ and any convex function ϕ on \mathbb{R} ,

$$\mathbb{E}_\mu \phi(S_t(g)) \leq \mathbb{E} \phi(B_{2\sigma^2 C^2 t})$$

where B is a standard Brownian Motion. In particular

$$\mathbb{E}_\mu \exp\left(\lambda \int_0^t g(X_s) ds\right) \leq e^{\lambda^2 (\sigma C)^2 \|g\|_{\text{Lip}}^2 t}, \quad \forall \lambda \in \mathbb{R}, t > 0.$$

and μ satisfies $W_1 I(\sigma C)$ on (\mathcal{X}, d) .

4.2. Multi-dimensional diffusions. Let us show now how to estimate the constant C in (4.1) or (4.9) by means of some examples.

A first example. At first in the framework of Theorem 4.1, if the Bakry-Emery curvature is positive

$$\text{Ric} + \nabla^2 V \geq K > 0$$

then it is well known that for $g \in C_b^1(\mathcal{X})$ with $\mu(g) = 0$,

$$|\nabla P_t g| \leq e^{-Kt} P_t |\nabla g|$$

and then $h := \int_0^\infty P_t g dt$ is absolutely convergent in $C_{\text{Lip}} \subset L^2(\mu)$ (for $\int_{\mathcal{X}} d^2(x, x_0) d\mu < +\infty$). Hence $h \in \mathbb{D}_2(\mathcal{L})$, $-\mathcal{L}h = g$ and

$$\|h\|_{\text{Lip}} \leq \|\nabla g\|_\infty \int_0^\infty e^{-Kt} dt = \frac{1}{K} \|g\|_{\text{Lip}}.$$

In other words condition (4.1) holds with $C = K^{-1}$ and μ satisfies $W_1 I(K^{-1})$. Of course one can also derive this sharp transportation inequality from the log-Sobolev inequality of Bakry-Emery [3] and Proposition 2.11-(b).

A second example. Now we turn to another situation where the log-Sobolev inequality is unknown as in Djellout et al. [13]. Consider the stochastic differential equation

$$dX_t = \sqrt{2}\sigma(X_t) dB_t + b(X_t) ds \quad (4.10)$$

where $\sigma : \mathbb{R}^d \rightarrow \mathcal{M}_{d \times n}$ (the space of real $d \times n$ -matrices) and $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ are locally Lipschitz, and (B_t) is a standard Brownian motion in \mathbb{R}^n . Assume that for some $\delta > 0$,

$$\text{tr}[(\sigma(y) - \sigma(x))(\sigma(y) - \sigma(x))^T] + \langle y - x, b(y) - b(x) \rangle \leq -\delta|y - x|^2, \quad \forall x, y \in \mathbb{R}^d \quad (4.11)$$

Here $\text{tr}(\cdot)$ denotes the trace and A^T the transposition of matrix A . With Itô's formula one easily obtains

$$\mathbb{E}|X_t(y) - X_t(x)|^2 \leq e^{-2\delta t}|x - y|^2, \quad \forall x, y \in \mathbb{R}^d, t \geq 0$$

where $X_t(x)$ is the solution of (4.10) with $X_0 = x$. This implies that (X_t) has a unique invariant probability measure μ . Hence for any $g \in C_{\text{Lip}}(\mathbb{R}^d)$ (with respect to the Euclidean norm $|\cdot|$) with $\mu(g) = 0$,

$$\|P_t g\|_{\text{Lip}} = \sup_{x \neq y} \frac{|\mathbb{E}g(X_t(y)) - \mathbb{E}g(X_t(x))|}{|y - x|} \leq e^{-\delta t} \|g\|_{\text{Lip}}. \quad (4.12)$$

Then $h := \int_0^\infty P_t g dt$ is absolutely convergent in $C_{\text{Lip}}(\mathbb{R}^d)$ and $\|h\|_{\text{Lip}} \leq \delta^{-1} \|g\|_{\text{Lip}}$. In other words (4.9) holds with $C = \delta^{-1}$. Finally as the carré-du-champ of (X_t) is given by

$$\Gamma(h, h)(x) = \langle \sigma \sigma^T \nabla h(x), \nabla h(x) \rangle, \quad h \in C_0^2(\mathbb{R}^d)$$

the constant σ in (4.8) can be identified as $\|\sigma(\cdot)\|_\infty := \sup_x \|\sigma(x)\|_{\mathbb{R}^n \rightarrow \mathbb{R}^d}$, at least for $h \in C_0^2(\mathbb{R}^d)$.

Corollary 4.4. *Assume that σ, b are locally Lipschitz such that $\|\sigma(\cdot)\|_\infty < +\infty$ and satisfy the dissipativity condition (4.11). Suppose moreover that μ is absolutely continuous and its transition semigroup (P_t) is symmetric with respect to the unique invariant measure μ . Then the Dirichlet form $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$ on $L^2(\mu)$ is given by the closure of*

$$\mathcal{E}(h, h) = \int_{\mathbb{R}^d} \langle \sigma \sigma^T \nabla h, \nabla h \rangle d\mu, \quad h \in C_0^\infty(\mathbb{R}^d)$$

and μ satisfies $W_1 I(c)$ on \mathbb{R}^d with respect to the Euclidean metric with

$$c = \|\sigma(\cdot)\|_\infty / \delta.$$

Proof. As $C_{\text{Lip}} \cap \mathbb{D}_2(\mathcal{L})$ is stable by (P_t) and contains $C_0^\infty(\mathbb{R}^d)$, $C_{\text{Lip}} \cap \mathbb{D}_2(\mathcal{L})$ is an operator core for $(\mathcal{L}, \mathbb{D}_2(\mathcal{L}))$, hence a form core for $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$. Since any $h \in C_{\text{Lip}}$ can be approached by a sequence h_n in $C_0^\infty(\mathbb{R}^d)$ with respect to the norm

$$\sqrt{\|h\|_2^2 + \int_{\mathbb{R}^d} \langle \sigma \sigma^T \nabla h, \nabla h \rangle d\mu}$$

one sees that $C_{\text{Lip}} \subset \mathbb{D}(\mathcal{E})$,

$$\mathcal{E}(h, h) = \int_{\mathbb{R}^d} \langle \sigma \sigma^T \nabla h, \nabla h \rangle d\mu, \quad \forall h \in C_{\text{Lip}}$$

and $C_0^\infty(\mathbb{R}^d)$ is a form core for $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$. This proves the first claim. It also follows that the conditions in Theorem 4.3 are verified. The remaining part follows from Theorem 4.3. \square

Remark. The crucial formula (4.12) for estimating our condition (4.9) is equivalent to

$$W_1(P_t(x, \cdot), P_t(y, \cdot)) \leq e^{-\delta t} d(x, y) \quad (4.13)$$

which can be obtained by means of numerous coupling techniques, see M.F. Chen [11]. Note that this condition is the one introduced by Joulin [25] under the name of Wasserstein's curvature, with which he obtains Poisson type deviation inequality.

4.3. One-dimensional diffusions. Now let us consider one-dimensional diffusion processes with values in the interval (x_0, y_0) and generated by

$$\mathcal{L}h = ah'' + bh', \quad h \in C_0^\infty(x_0, y_0)$$

where a, b are continuous with $a > 0$. Let $((X_t)_{0 \leq t < \tau}, \mathbb{P}_x)$ be the martingale solution associated with \mathcal{L} and initial position x , where τ is the explosion time. With a fixed $c \in (x_0, y_0)$,

$$s'(x) := \exp\left(-\int_c^x \frac{b(z)}{a(z)} dz\right), \quad m'(x) := \frac{1}{a(x)} \exp\left(\int_c^x \frac{b(z)}{a(z)} dz\right)$$

are respectively the derivatives of Feller's scale and speed functions. Assume that

$$Z := \int_{x_0}^{y_0} m'(x) dx < +\infty \quad (4.14)$$

and let $\mu(dx) = m'(x)dx/Z$. It is well known that $(\mathcal{L}, C_0^\infty(x_0, y_0))$ is symmetric on $L^2(\mu)$. Assume also that

$$\int_c^{y_0} s'(x) dx \int_c^x m'(x) dx = \int_{x_0}^c s'(x) \int_x^c m'(x) dx = +\infty \quad (4.15)$$

which, in Feller's classification, means that x_0 and y_0 are not accessible or equivalently $\tau = \infty$, \mathbb{P}_x -a.s. In this case by the L^1 -uniqueness in [42, 19], the Dirichlet form

$$\begin{aligned} \mathbb{D}(\mathcal{E}) &= \left\{ h \in \mathcal{AC}(x_0, y_0) \cap L^2(\mu); \int_{x_0}^{y_0} (h')^2 d\mu < +\infty \right\}, \\ \mathcal{E}(h, h) &= \int_{x_0}^{y_0} (h')^2 d\mu, \quad h \in \mathbb{D}(\mathcal{E}) \end{aligned}$$

is associated with (X_t) , where $\mathcal{AC}(x_0, y_0)$ is the space of the absolutely continuous functions on (x_0, y_0) .

Fix some $\rho \in C^1(x_0, y_0)$ such that $\rho \in L^2(\mu)$ and $\rho'(x) > 0$ everywhere, consider the metric $d_\rho(x, y) = |\rho(x) - \rho(y)|$. A function h on (x_0, y_0) is Lipschitz with respect to d_ρ (one writes $h \in C_{\text{Lip}(\rho)}$) if and only if $h \in \mathcal{AC}(x_0, y_0)$ and

$$\|h\|_{\text{Lip}(\rho)} = \sup_{x_0 < x < y < y_0} \frac{|h(y) - h(x)|}{\rho(y) - \rho(x)} = \|h'/\rho'\|_\infty < \infty$$

The argument below is borrowed from [14]. Assume that

$$C(\rho) := \sup_{x \in (x_0, y_0)} \frac{1}{\rho'(x)} \int_x^{y_0} [\rho(z) - \mu(\rho)] m'(z) dz < +\infty. \quad (4.16)$$

For every $g \in C_{\text{Lip}(\rho)}$ with $\mu(g) = 0$, then the C^2 function $h(x) = \int_c^x dy \int_y^{y_0} g(z)m'(z) dz - A$ solves

$$-(ah'' + bh') = g. \quad (4.17)$$

It is obvious that $\|h\|_{\text{Lip}(\rho)} = \sup_{x \in (x_0, y_0)} \frac{1}{\rho'(x)} \int_x^{y_0} g(z)m'(z) dz$. An elementary exercise (see [14]) shows that this quantity is not greater than $C(\rho)\|g\|_{\text{Lip}(\rho)}$. Thus h belongs to $L^2(\mu)$ whenever ρ is in $L^2(\mu)$. By Itô's formula, $h \in \mathbb{D}_2(\mathcal{L})$. With the constant A so that $\mu(h) = 0$, because of the ergodicity of (X_t) , h is the unique solution of (4.17) in $L^2(\mu)$ with zero mean. One also sees that $C(\rho)$ is the best constant by taking $g = \rho - \mu(\rho)$. In other words, condition (4.9) is satisfied with $C = C(\rho)$. Hence, with Theorem 4.3 one obtains

Corollary 4.5. *Let $a, b : (x_0, y_0) \rightarrow \mathbb{R}$ be continuous with $a > 0$ and let conditions (4.14) and (4.15) be satisfied. Assume (4.16) and $\sigma := \sup_{x \in (x_0, y_0)} \sqrt{a(x)}\rho'(x) < +\infty$. Then μ satisfies $W_1I(\kappa)$ on $((x_0, y_0), d_\rho)$ with $\kappa = (\sigma C(\rho))^{-1}$. In particular for*

$$\rho_a(x) = \int_c^x \frac{dz}{\sqrt{a(z)}}$$

if $C(\rho_a) < +\infty$, then μ satisfies $W_1I(c)$ on $((x_0, y_0), d_{\rho_a})$ with $c = C(\rho_a)$.

Remarks 4.6.

- (1) d_{ρ_a} is the metric associated with the carré-du-champ operator of the diffusion.
- (2) The quantity $C(\rho)$ in (4.16) is not innocent: Chen-Wang's variational formula for the spectral gap tells us that ([11, 41]): $c_P(\mu) = \inf_\rho C(\rho)$.

5. LYAPUNOV FUNCTION CONDITIONS

We will use in this section general conditions on the generator of the process, known as Lyapunov function conditions, for deriving W_1I or more generally $T_\Phi I$ where $\Phi = \{(u, u); |u| \leq \phi\}$ with ϕ unbounded, and log-Sobolev inequality. To state properly the Lyapunov function condition, it is necessary to enlarge the domain of the generator. In this section, the Markov process $((X_t), \mathbb{P}_\mu)$ is reversible and its sample paths are \mathbb{P}_μ -càdlàg (possibly with jumps).

A continuous function h is said to be in the μ -extended domain $\mathbb{D}_e(\mathcal{L})$ of the generator of the Markov process $((X_t), \mathbb{P}_\mu)$ if there is some measurable function g such that $\int_0^t |g|(X_s) ds < +\infty$, \mathbb{P}_μ -a.s. and

$$M_t(h) := h(X_t) - h(X_0) - \int_0^t g(X_s) ds$$

is a local \mathbb{P}_μ -martingale. It is obvious that g is uniquely determined up to μ -equivalence. In such case one writes $h \in \mathbb{D}_e(\mathcal{L})$ and $\mathcal{L}h = g$.

The Lyapunov condition can now be stated:

- (H) There exist a continuous function $U : \mathcal{X} \rightarrow [1, +\infty)$ in $\mathbb{D}_e(\mathcal{L})$, a nonnegative function ϕ and a constant $b > 0$ such that

$$-\frac{\mathcal{L}U}{U} \geq \phi - b, \quad \mu\text{-a.s.}$$

When the process is irreducible and the constant b is replaced by $b1_C$ for some “small set” C , then it is well-known that the existence of a positive bounded ϕ such that $\inf_{\mathcal{X} \setminus C} \phi > 0$ in (H) is equivalent to Poincaré inequality (see [2], for instance).

Lyapunov conditions are widely used to study the speed of convergence of Markov chains [33] or Markov processes [18], large or moderate deviations and essential spectral radii [45, 47]. More recently, they have been used to study functional inequalities as weak Poincaré inequality [2] or super-Poincaré inequality [10]. See Wang [41] on weak and super Poincaré inequalities.

Theorem 5.1. *Assume that μ satisfies a Poincaré inequality with best constant $c_P(\mu) < \infty$ and that the Lyapunov condition (H) holds. Suppose moreover that $\phi \in L^2(\mu)$, that is $\|\phi\|_2 := (\int \phi^2 d\mu)^{1/2} < \infty$. Then, for any $a \geq 2$ and for every probability measure ν ,*

$$\|\phi(\nu - \mu)\|_{TV} \leq (1 + 2bc_P(\mu)) \frac{a+1}{a-1} I(\nu|\mu) + a\sqrt{2}\|\phi\|_2 \sqrt{c_P(\mu) I(\nu|\mu)} \quad (5.1)$$

and

$$\|\sqrt{\phi}(\nu - \mu)\|_{TV}^2 \leq 2 \left[3(1 + 2bc_P(\mu)) + 2\sqrt{2}\|\phi\|_2 c_P(\mu) \right] I(\nu|\mu) \quad (5.2)$$

Remarks 5.2. Since $\|\phi(\nu - \mu)\|_{TV} = \sup_{u: |u| \leq \phi} \int u d(\nu - \mu)$, the inequalities in this theorem may be regarded as $T_\Phi I$ in Theorem 2.2 with $\Phi = \{(u, u); u \in b\mathcal{B}, |u| \leq \phi\}$. Since

$$W_1(\nu, \mu) = \sup_{f: \|f\|_{Lip} \leq 1} \int f d[\nu - \mu] \leq \inf_{x_0 \in \mathcal{X}} \|d(\cdot, x_0)(\nu - \mu)\|_{TV},$$

one sees that (5.2) implies $W_1 I(c)$ in Corollary 2.4 as soon as $d(\cdot, x_0) \leq C\sqrt{\phi}$ for some $x_0 \in \mathcal{X}$ and $C > 0$. As will be seen with the Ornstein-Uhlenbeck process at Example 5.9, the order of this inequality is sharp.

Theorem 5.3. *In the framework of Proposition 2.11, assume that the Bakry-Emery’s curvature of $\mu = e^{-V} dx/Z$ is bounded from below by some constant $K \leq 0$. Assume that μ satisfies a Poincaré inequality with best constant $c_P(\mu) < \infty$.*

If the Lyapunov condition (H) holds with $\phi(x) = cd(x, x_0)^2$ where $c > 0$ and $x_0 \in \mathcal{X}$ is some fixed point, then μ satisfies the log-Sobolev inequality on the Riemannian manifold \mathcal{X} .

Their proofs are based on the following large deviation result.

Lemma 5.4. *For every continuous function $U \geq 1$ in $\mathbb{D}_e(\mathcal{L})$ such that $-\mathcal{L}U/U$ is μ -a.e. lower bounded,*

$$\int -\frac{\mathcal{L}U}{U} g^2 d\mu \leq \mathcal{E}(g, g), \quad \forall g \in \mathbb{D}(\mathcal{E}). \quad (5.3)$$

When U is bounded, this is contained in Deuschel-Stroock [12, Lemme 4.2.35].

Proof. For any initial law β ,

$$N_t = U(X_t) \exp \left(- \int_0^t \frac{\mathcal{L}U}{U}(X_s) ds \right)$$

is a local \mathbb{P}_β -martingale. Indeed, denoting $A_t := \exp\left(-\int_0^t \frac{\mathcal{L}U}{U}(X_s)ds\right)$, Itô's formula is $dN_t = A_t [dM_t(U) + \mathcal{L}U(X_t) dt] - \frac{\mathcal{L}U}{U}(X_t)A_t U(X_t) dt = A_t dM_t(U)$ where $M(U)$ is a local \mathbb{P}_β -martingale. As (N_t) is nonnegative, it is also a \mathbb{P}_β -supermartingale. Choosing $\beta := U^{-1}\mu/Z$ with $0 < Z = \mu(U^{-1}) \leq 1$, one sees that for all $t \geq 0$

$$\mathbb{E}_\beta \exp\left(-\int_0^t \frac{\mathcal{L}U}{U}(X_s)ds\right) \leq \mathbb{E}_\beta N_t \leq \beta(U) = 1/Z < +\infty.$$

Let $u_n := \min\{-\mathcal{L}U/U, n\}$. The previous estimation implies that

$$F(u_n) := \limsup_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E}_\beta \exp\left(\int_0^t u_n(X_s)ds\right) \leq 0.$$

On the other hand by the lower bound of large deviation in [44, Theorem B.1, Corollary B.11] and Laplace-Varadhan principle, as in the proof of (c') \Rightarrow (a) in Theorem 2.2,

$$F(u_n) \geq \sup\{\nu(u_n) - I(\nu|\mu); \nu \in M_1(E)\}.$$

Thus $\int u_n d\nu \leq I(\nu|\mu)$, which yields to (by letting $n \rightarrow \infty$ and monotone convergence)

$$\int -\frac{\mathcal{L}U}{U} d\nu \leq I(\nu|\mu), \quad \forall \nu \in M_1(E). \quad (5.4)$$

This is equivalent to (5.3) by the fact that $\mathcal{E}(|h|, |h|) \leq \mathcal{E}(h, h)$ for all $h \in \mathbb{D}(\mathcal{E})$.

Note that one was allowed to apply the large deviation lower bound [44, Theorem B.1] under \mathbb{P}_β since β is absolutely continuous with respect to μ . In addition, in the symmetric case, [44, Corollary B.11] states that the large deviation rate function is $I(\cdot|\mu)$; it doesn't depend on β under the underlying assumption that \mathbb{P}_μ is ergodic. As this lower bound holds for the topology of probability measures weakened by all bounded measurable test functions (sometimes called τ -topology), one can apply the Laplace-Varadhan principle to the continuous bounded function $\nu \mapsto \nu(u_n)$. \square

Proof of Theorem 5.3. It is a combination of the Lyapunov function condition and the HWI inequality of Otto-Villani.

We begin with the following fact ([39, Proposition 7.10]):

$$W_2^2(\nu, \mu) \leq 2\|d(\cdot, x_0)^2(\nu - \mu)\|_{TV}.$$

Now for every function g with $|g| \leq \phi(x) := cd(x, x_0)^2$, we have by (H),

$$\begin{aligned} \int g d(\nu - \mu) &\leq \nu(\phi) + \mu(\phi) \\ &\leq \int \left(-\frac{\mathcal{L}U}{U} + b\right) d\nu + \mu(\phi) \\ &\leq I(\nu|\mu) + b + \mu(\phi) \end{aligned}$$

where the last inequality follows by Lemma 5.4. Taking the supremum over all such g , we get

$$\frac{c}{2}W_2^2(\nu, \mu) \leq c\|d(\cdot, x_0)^2(\nu - \mu)\|_{TV} \leq I(\nu|\mu) + b + \mu(\phi),$$

which yields to (by the inequality at the beginning)

$$W_2^2(\nu, \mu) \leq \frac{2}{c}I(\nu|\mu) + \frac{2}{c}[b + \mu(\phi)].$$

Substituting it into the HWI inequality of Otto-Villani (2.21), we obtain (using $2ab \leq a^2 + b^2$)

$$\begin{aligned} H(\nu|\mu) &\leq 2\sqrt{\frac{2}{c}I(\nu|\mu) + \frac{2}{c}[b + \mu(\phi)]}\sqrt{I(\nu|\mu)} - \frac{K}{2}\left(\frac{2}{c}I(\nu|\mu) + \frac{2}{c}[b + \mu(\phi)]\right) \\ &\leq AI(\nu|\mu) + B \end{aligned} \quad (5.5)$$

where

$$A = \left(1 - \frac{K}{2}\right)\frac{2}{c} + 1, \quad B = \frac{2}{c}[b + \mu(\phi)]\left(1 - \frac{K}{2}\right).$$

Finally by Rothaus' lemma the non-tight log-Sobolev inequality (5.5) together with the spectral gap implies the tight log-Sobolev inequality

$$H(\nu|\mu) \leq [A + (B + 2)c_{\mathbb{P}}(\mu)]I(\nu|\mu).$$

□

Remarks 5.5. In the case that the Bakry-Emery's curvature is bounded from below by a negative constant K , Wang's criterion [40] says that the log-Sobolev inequality holds if $\int e^{\lambda d^2(x, x_0)} d\mu(x) < +\infty$ for some $\lambda > |K|$. Our Lyapunov condition (H) above is complementary to that result and sharp in order (as seen for $V(x) = |x|^\alpha$ on $\mathcal{X} = \mathbb{R}^d$). Furthermore our proof here is completely different and gives an explicit estimate of the log-Sobolev constant.

Proof of Theorem 5.1. We may assume that $\nu = f\mu$ with $\sqrt{f} \in \mathbb{D}(\mathcal{E})$ (trivial otherwise). For any $a \geq 2$, define $h : \mathbb{R} \rightarrow \mathbb{R}^+$ by

$$h(t) = \begin{cases} 0 & \text{if } t \leq 1; \\ \sqrt{\frac{a+1}{a-1}}(t-1) & \text{if } t \in [1, a]; \\ \sqrt{t^2-1} & \text{if } t \geq a. \end{cases}$$

It is easy to see that $\|h\|_{\text{Lip}} \leq \sqrt{\frac{a+1}{a-1}}$. Decompose

$$\|\phi(\nu - \mu)\|_{TV} = \int \phi|f-1|d\mu = \int \phi h^2(\sqrt{f})d\mu + \int \phi[|f-1| - h^2(\sqrt{f})]d\mu.$$

First consider the last term. Since $t^2 - 1 - h^2(t) \leq a(t-1)$ for $t \in [1, a]$, and $= 0$ for $t \geq a \geq 2$,

$$\begin{aligned} \int \phi[|f-1| - h^2(\sqrt{f})]d\mu &= \int \phi[1_{\{f \leq 1\}}(1-f) + 1_{\{1 \leq f \leq a^2\}}a(\sqrt{f}-1)]d\mu \\ &\leq a \int \phi|1 - \sqrt{f}|d\mu \end{aligned}$$

which is not greater than

$$\begin{aligned} a\|\phi\|_2\|1 - \sqrt{f}\|_2 &= a\|\phi\|_2\sqrt{2}\sqrt{1 - \mu(\sqrt{f})} \leq a\|\phi\|_2\sqrt{2\text{Var}_\mu(\sqrt{f})} \\ &\leq a\sqrt{2c_P(\mu)}\|\phi\|_2\sqrt{I(\nu|\mu)}. \end{aligned}$$

We turn now to bound the crucial first term by means of (5.3):

$$\begin{aligned} \int \phi h^2(\sqrt{f})d\mu &\leq \int \left(-\frac{\mathcal{L}U}{U} + b \right) h^2(\sqrt{f})d\mu \\ &\leq \mathcal{E}(h(\sqrt{f}), h(\sqrt{f})) + b\|h\|_{\text{Lip}}^2 \int (\sqrt{f} - 1)^2 d\mu \\ &\leq \|h\|_{\text{Lip}}^2 \mathcal{E}(\sqrt{f}, \sqrt{f}) + 2b\|h\|_{\text{Lip}}^2 \text{Var}_\mu(\sqrt{f}) \\ &\leq (1 + 2bc_P(\mu)) \frac{a+1}{a-1} I(\nu|\mu). \end{aligned}$$

Substituting these two estimates into our previous decomposition, we obtain (5.1).

For (5.2), noting that with Theorem 3.1: $\int |f - 1|d\mu \leq 2 \min\{1, \sqrt{c_P(\mu)I(\nu|\mu)}\}$, we have by Cauchy-Schwarz inequality and (5.1)

$$\begin{aligned} &\|\sqrt{\phi}(\nu - \mu)\|_{TV}^2 \\ &\leq \int |f - 1|d\mu \int \phi |f - 1|d\mu \\ &\leq 2 \min\left(1, \sqrt{c_P(\mu)I(\nu|\mu)}\right) \left[(1 + 2bc_P(\mu)) \frac{a+1}{a-1} I(\nu|\mu) + a\sqrt{2}\|\phi\|_2\sqrt{c_P(\mu)I(\nu|\mu)} \right] \\ &\leq 2 \left[(1 + 2bc_P(\mu)) \frac{a+1}{a-1} + a\sqrt{2}\|\phi\|_2 c_P(\mu) \right] I(\nu|\mu) \end{aligned}$$

which gives (5.2) with $a = 2$. □

Remarks 5.6. When $\sqrt{2}(c_P(\mu)^{-1} + 2b) \geq \|\phi\|_2$, optimizing $a \geq 2$ in the proof of (5.2) above, we get the slightly better inequality:

$$\|\sqrt{\phi}(\nu - \mu)\|_{TV}^2 \leq 2 \left(\sqrt{2}\|\phi\|_2 c_P(\mu) + (1 + 2bc_P(\mu)) + 2^{5/4} \sqrt{\|\phi\|_2(1 + 2bc_P(\mu))c_P(\mu)} \right) I(\nu|\mu).$$

Notice that by Lemma 5.4 and condition (H), $b \geq \mu(\phi)$ (in practice b is much bigger).

From now on the positive constant C may change from one place to another. One can do some variation of the proof of (5.2) above. For every $p > 1$ and its conjugate number $q = p/(p - 1)$, instead of Cauchy-Schwarz we apply Hölder inequality to get for $\nu = f\mu$,

$$\begin{aligned} \|\phi^{1/p}(\nu - \mu)\|_{TV} &\leq \left(\int |f - 1|d\mu \right)^{1/q} \left(\int \phi |f - 1|d\mu \right)^{1/p} \\ &\leq 2^{1/q} \min\{1, \sqrt{c_P(\mu)I(\nu|\mu)}^{1/q}\} (C_1 I(\nu|\mu) + C_2 \sqrt{I(\nu|\mu)})^{1/p} \\ &\leq C[(1 + I(\nu|\mu))^{2/p} - 1]^{1/2}. \end{aligned}$$

In other words, we have proved

Corollary 5.7. *Under the conditions of Theorem 5.1, for any $p > 1$, there exists some constant $\kappa > 0$ such that for $\alpha(r) = \kappa[(1 + r^2)^{p/2} - 1]$,*

$$\alpha(\|\phi^{1/p}(\nu - \mu)\|_{TV}) \leq I(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}). \quad (5.6)$$

Corollary 5.8. *Let $\mu = e^{-V} dx/Z$ be a probability measure where $V \in C^\infty(\mathcal{X})$ is bounded from below and $|\nabla V|^2 \in L^2(\mu)$. Let $\mathcal{L} = \Delta - \nabla V \cdot \nabla$ be the generator of the diffusion (X_t) on the non-compact connected complete Riemannian manifold \mathcal{X} . Assume that for some $p > 1$,*

$$d(x, x_0) \leq C(1 + |\nabla V|^2(x))^{1/p}, \quad \forall x \in \mathcal{X}$$

and

$$\gamma := \limsup_{d(x, x_0) \rightarrow \infty} \frac{\Delta V(x)}{|\nabla V|^2(x)} < 1.$$

Then w.r.t. the Riemannian metric d , there exists $\kappa > 0$ such that with $\alpha(r) = \kappa[(1 + r^2)^{p/2} - 1]$,

$$\alpha(W_1(\nu, \mu)) \leq I(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}). \quad (5.7)$$

In particular for every Lipschitz function u with $\|u\|_{\text{Lip}} \leq 1$ and any initial law β with $d\beta/d\mu \in L^2(\mu)$,

$$\mathbb{P}_\beta \left(\frac{1}{t} \int_0^t u(X_s) ds > \mu(u) + r \right) \leq \left\| \frac{d\beta}{d\mu} \right\|_2 \exp(-t\kappa[(1 + r^2)^{p/2} - 1]), \quad \forall t, r > 0. \quad (5.8)$$

Proof. Let $\gamma' \in (\gamma, 1)$ and $\lambda, \delta \in (0, 1)$ sufficiently small so that $\lambda - \lambda^2 > \gamma'\lambda + \delta$. For $U = e^{\lambda V}$, we have

$$-\frac{\mathcal{L}U}{U} = -\lambda\mathcal{L}V - \lambda^2|\nabla V|^2 = (\lambda - \lambda^2)|\nabla V|^2 - \lambda\Delta V \geq \delta(1 + |\nabla V|^2) - b$$

where $b := \delta + \sup_{\mathcal{X}}(\lambda\Delta V - \gamma'\lambda|\nabla V|^2)$ is finite under our assumptions. Thus (H) is satisfied with $\phi = \delta(1 + |\nabla V|^2)$ which is in $L^2(\mu)$ under our assumptions. On the other hand our assumptions imply that ϕ tends to infinity at infinity. Hence $(1 - \mathcal{L})^{-1}$ is compact on $L^2(\mu)$ and $c_P(\mu) < \infty$. Noting that

$$W_1(\nu, \mu) \leq \inf_{x_0 \in \mathcal{X}} \|d(\cdot, x_0)(\nu - \mu)\|_{TV},$$

the statement now follows directly from Theorem 5.1 and Corollary 5.7. \square

Example 5.9. Let $\mathcal{X} = \mathbb{R}^n$, $V(x) = C|x|^\beta$ for $|x| > 1$ where $\beta \geq 1, C > 0$. Then (H) is satisfied for $\phi = \delta(1 + |\nabla V|^2) \sim C|x|^{2(\beta-1)}$ (when $|x|$ large).

- (i) If $\beta > 3/2$, then the condition in Corollary 5.8 is verified with $p = 2(\beta - 1) > 1$, so we have (5.8) for Lipschitz observable u with $p = 2(\beta - 1)$. Then we have Gaussian behavior for small r , and even a super-Gaussian tail for large r whenever $\beta > 2$.
- (ii) Let $\beta \in [1, 3/2]$. Then for $\psi = (1 + |x|)^{\beta-1}$, we have by Theorem 5.1(5.2),

$$\|\psi(\nu - \mu)\|_{TV}^2 \leq CI(\nu|\mu).$$

Then the Gaussian deviation inequality holds true for the observable u satisfying $|u| \leq C(1 + |x|)^{\beta-1}$.

- (iii) If $\beta = 2$ (Ornstein-Uhlenbeck process), the inequality (5.2) for $\sqrt{\phi} \sim C|x|$ (proved in (i)) becomes the correct one in order: indeed if $\psi(x) \gg |x|$ at infinity with $\mu(\psi) < +\infty$, one cannot hope that

$$\|\psi(\nu - \mu)\|_{TV}^2 \leq CI(\nu|\mu), \quad \forall \nu$$

since by Theorem 2.2, this would imply that

$$\mathbb{E}_\mu \exp \left(\lambda \int_0^1 \psi(X_s) ds \right) \leq e^{\lambda\mu(\psi) + C\lambda^2/2}, \quad \forall \lambda \in \mathbb{R}$$

which yields, integrating w.r.t. $N(0, \sigma^2)(d\lambda)$ with variance $\sigma^2 < 1/C$,

$$\mathbb{E}_\mu \exp \left(\delta \left(\int_0^1 \psi(X_s) ds \right)^2 \right) < +\infty$$

for some $\delta > 0$. But this is impossible.

We conclude by an example of jump process.

Example 5.10. ($M/M/\infty$ queue). In this example $\mathcal{X} = \mathbb{N}$, μ is the Poisson measure with mean $\lambda > 0$ and the Dirichlet form is

$$\mathcal{E}(h, h) = \sum_{n \in \mathbb{N}} (h(n+1) - h(n))^2 \mu(n)$$

The associated generator is

$$\mathcal{L}h(n) = \lambda(h(n+1) - h(n)) + n(h(n-1) - h(n)), \quad \forall n \geq 0$$

(with the convention that $h(-1) = h(0)$). Let $U(n) = e^{cn}$ where $c > 0$. We have

$$-\frac{\mathcal{L}U}{U}(n) = n(1 - e^{-c}) - (e^c - 1).$$

Thus condition (H) is satisfied, and we have by Theorem 5.1 that for $\psi(n) := \sqrt{1+n}$,

$$\|\psi(\nu - \mu)\|_{TV}^2 \leq CI(\nu|\mu), \quad \forall \nu \in M_1(\mathbb{N}).$$

By Theorem 2.2, this gives the Gaussian deviation inequality for any observable u so that $|u(n)| \leq C\sqrt{1+n}$. See Joulin [25] and Liu-Ma [29] for previous studies on deviation inequalities of this model. Note that they only obtain Poisson tail by their approach for the same test function. Remark also that our result provides exponential tail for $u(n) = n$, which is close of the conjectured Poisson behavior.

REFERENCES

- [1] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. In *Ecole d'Eté de Probabilités de Saint-Flour (1992)*, number 1581 in Lecture Notes in Mathematics. Springer-Verlag, 1994.
- [2] D. Bakry, P. Cattiaux, and A. Guillin. Rates of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. Preprint, 2007.
- [3] D. Bakry and M. Emery. Diffusions hypercontractives. In *Séminaire de Probabilités*, number 1123 in Lecture Notes in Mathematics. Springer-Verlag, 1985.
- [4] S.G. Bobkov, I. Gentil, and M. Ledoux. Hypercontractivity of Hamilton-Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 80(7):669–696, 2001.
- [5] S.G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *Journal of Functional Analysis.*, 163:1–28, 1999.

- [6] F. Bolley and C. Villani. Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Annales de la Faculté des Sciences de Toulouse.*, 14:331–352, 2005.
- [7] E.A. Carlen. Superadditivity of Fisher’s information and logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 101(1):194–211, 1991.
- [8] P. Cattiaux and A. Guillin. Deviation bounds for additive functionals of Markov process. To appear in *ESAIM P&S*.
- [9] P. Cattiaux and A. Guillin. On quadratic transportation cost inequalities. To appear in *Journal de Mathématiques Pures et Appliquées*.
- [10] P. Cattiaux, A. Guillin, L. Wu, and F.Y. Wang. In preparation.
- [11] M.F. Chen. *Eigenvalues, inequalities, and ergodic theory*. Probability and its Applications. Springer-Verlag, 2005.
- [12] J-D. Deuschel and D.W. Stroock. *Large Deviations*, volume 137 of *Pure and Applied Mathematics*. Academic Press, 1989.
- [13] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities for random dynamical systems and diffusions. *Annals of Probability*, 32(3B):2702–2732, 2004.
- [14] H. Djellout, Y. Miao, and L. Wu. Transportation and log-Sobolev inequalities for Gibbs measures. In preparation.
- [15] M.D. Donker and S.R.S. Varadhan. Asymptotic evaluations of certain Markov process expectations for large time, I. *Comm. Pure Appl. Math.*, 28:1–47, 1975.
- [16] M.D. Donker and S.R.S. Varadhan. Asymptotic evaluations of certain Markov process expectations for large time, III. *Comm. Pure Appl. Math.*, 29:389–461, 1976.
- [17] M.D. Donker and S.R.S. Varadhan. Asymptotic evaluations of certain Markov process expectations for large time, IV. *Comm. Pure Appl. Math.*, 36:183–212, 1983.
- [18] R. Douc, G. Fort, and A. Guillin. Subgeometric rates of convergence of f -ergodic strong Markov processes. Preprint, 2006.
- [19] A. Eberle. *Uniqueness and Non-Uniqueness of Semigroups Generated by Singular Diffusion Operators*, volume 1718 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1999.
- [20] F.Q. Gao and L. Wu. Transportation-information inequalities for Gibbs measures. Preprint, 2007.
- [21] A. Gibbs and F. Su. On choosing and bounding probability metrics. *Internat. Statist. Rev.*, 70(3):419–435, 2002.
- [22] F.Z. Gong and L. Wu. Spectral gap of positive operators and applications. *J. Math. Pures Appl*, 85:151–191, 2006.
- [23] N. Gozlan. Characterization of Talagrand’s like transportation cost inequalities on the real line. Preprint, 2006.
- [24] N. Gozlan and C. Léonard. A large deviation approach to some transportation cost inequalities. To appear in *Probability Theory and Related Fields*.
- [25] A. Joulin. *Concentration et fluctuations de processus stochastiques avec sauts*. PhD thesis, Université La Rochelle, 2006.
- [26] T. Klein, Y. Ma, and N. Privault. Convex concentration inequalities and forward-backward stochastic calculus. *Electronic Journal of Probability*, 11:486–512, 2006.
- [27] M. Ledoux. *The Concentration of Measure Phenomenon*. Mathematical Surveys and Monographs 89. American Mathematical Society, Providence RI, 2001.
- [28] P. Lezaud. Chernoff and Berry-Esséen inequalities for Markov processes. *ESAIM, Probab. Statist.*, pages 183–201, 2001. (electronic).
- [29] W. Liu and Y. Ma. Spectral gap and deviation inequalities for birth-death processes. Preprint, contained in the Ph.D thesis of Y. Ma at Université La Rochelle 2007, 2006.
- [30] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. Preprint.
- [31] K. Marton. Bounding \bar{d} -distance by informational divergence: a way to prove measure concentration. *Annals of Probability*, 24:857–866, 1996.
- [32] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6:556–571, 1997.
- [33] S.P. Meyn and R.L. Tweedie. *Markov chains and stochastic stability*. Communications and Control Engineering Series. Springer-Verlag, 1993.

- [34] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis*, 173:361–400, 2000.
- [35] K-T. Sturm. On the geometry of metric measure spaces, I. *Acta Math*, 196:65–131, 2006.
- [36] K-T. Sturm. On the geometry of metric measure spaces, II. *Acta Math*, 196:133–177, 2006.
- [37] M. Talagrand. Transportation cost for gaussian and other product measures. *Geometric and Functional Analysis*, 6:587–600, 1996.
- [38] C. Villani. Saint-Flour Lecture Notes, 2005. Optimal transport, old and new. Available online via <http://www.umpa.ens-lyon.fr/~cvillani/>.
- [39] C. Villani. *Topics in Optimal Transportation*. Graduate Studies in Mathematics 58. American Mathematical Society, Providence RI, 2003.
- [40] F.Y. Wang. Logarithmic Sobolev inequalities on noncompact Riemannian manifolds. *Probab. Theory Related Fields* 109 (1997), no. 3, 417–424.
- [41] F.Y. Wang. *Functional inequalities, Markov Semigroup and Spectral Theory*. Chinese Sciences Press, Beijing, New York, 2005.
- [42] L. Wu. Uniqueness of Nelson’s diffusions. *Probability Theory Rel. Fields*, 114:549–585, 1999.
- [43] L. Wu. A deviation inequality for non-reversible Markov processes. *Ann. Inst. Henri Poincaré (série Probab. Stat.)*, 36:435–445, 2000.
- [44] L. Wu. Uniformly integrable operators and large deviations for Markov processes. *J. Funct. Anal*, 172:301–376, 2000.
- [45] L. Wu. Large and moderate deviations for stochastic damping Hamiltonian systems. *Stoch. Proc. Appl.*, 91:205–238, 2001.
- [46] L. Wu. Uniform positive improvingness, tail norm condition and spectral gap. Preprint, 2001.
- [47] L. Wu. Essential spectral radius for Markov semigroups (I) : discrete time case. *Probab. Th. Rel. Fields*, 128:255–321, 2004.

ARNAUD GUILLIN. ECOLE CENTRALE DE MARSEILLE ET LATP, CENTRE DE MATHÉMATIQUES ET INFORMATIQUE. TECHNOPOLE DE CHÂTEAU-GOMBERT. 13453 MARSEILLE, FRANCE
E-mail address: guillin@cmi.univ-mrs.fr

CHRISTIAN LÉONARD. MODAL-X, UNIVERSITÉ PARIS 10. BÂTIMENT G, 200 AVENUE DE LA RÉPUBLIQUE.92001 NANTERRE, FRANCE

CHRISTIAN LÉONARD. CMAP, ÉCOLE POLYTECHNIQUE. 91128 PALAISEAU, FRANCE
E-mail address: christian.leonard@polytechnique.fr

LIMING WU. LABORATOIRE DE MATHÉMATIQUES APPLIQUÉES, CNRS-UMR 6620, UNIVERSITÉ BLAISE PASCAL, 63177 AUBIÈRE, FRANCE. AND DEPARTMENT OF MATHEMATICS, WUHAN UNIVERSITY, 430072 HUBEI, CHINA
E-mail address: Li-Ming.Wu@math.univ-bpclermont.fr

NIAN YAO. DEPARTMENT OF MATHEMATICS, WUHAN UNIVERSITY, 430072 HUBEI, CHINA