Probab. Theory Relat. Fields (2009) 144:669–695
DOI 10.1007/s00440-008-0159-5

Transportation-information inequalities for Markov processes

Arnaud Guillin · Christian Léonard ·
Liming Wu · Nian Yao

Received: 22 July 2007 / Revised: 28 April 2008 / Published online: 15 July 2008

Abstract In this paper, one investigates the transportation-information $T_c I$ inequalities: $\alpha(T_c(\nu, \mu)) \leq I(\nu|\mu)$ for all probability measures ν on a 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 the 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 $\alpha : [0, \infty) \to [0, \infty]$ is a left continuous increasing function. Using large deviation techniques, it is shown that $T_c I$ is equivalent to some concentration inequality for the occupation measure of a μ-reversible ergodic Markov process related to $I(\cdot|\mu)$. The tensorization property of $T_c I$ and comparisons of $T_c I$ with Poincaré and log-Sobolev inequalities are investigated. Several easy-to-check sufficient conditions are provided for special important cases of $T_c I$ and several examples are worked out.

Mathematics Subject Classification (2000) 26D10 · 47D07 · 60E15 · 60F10 · 60G99 · 60J25 · 60J60

A. Guillin (✉)
Ecole Centrale de Marseille et LATP, Centre de Mathématiques et Informatique,
Technopôle de Château-Gombert, 13453 Marseille, France
e-mail: guillin@cmi.univ-mrs.fr

C. Léonard
Modal-X, Université Paris Ouest, Bât. G, 200 av. de la République,
92001 Nanterre, France
e-mail: Christian.Leonard@u-paris10.fr

L. Wu
Laboratoire de Mathématiques Appliquées, CNRS-UMR 6620,
Université Blaise Pascal, 63177 Aubière, France
e-mail: Li-Ming.Wu@math.univ-bpclermont.fr

L. Wu · N. Yao
Department of Mathematics, Wuhan University, 430072 Hubei, China

1 Introduction

Let $M_1(\mathcal{X})$ be the space of all probability measures on a complete separable metric space $(\mathcal{X}, d)$ and consider the *cost function* $c(x, y) : \mathcal{X}^2 \to [0, +\infty]$ 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

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

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

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

The usual transportation inequalities for a given $\mu \in M_1(\mathcal{X})$, introduced by Marton [32] and Talagrand [38], compare the Wasserstein metric $W_p(\nu, \mu)$ with the relative entropy $H(\nu|\mu)$. The following extension of these inequalities:

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

has recently been proposed and developed by Gozlan and Léonard [22]. Here $\alpha : [0, \infty) \to [0, +\infty]$ is some left continuous and increasing function with $\alpha(0) = 0$.

Let us denote

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

the monotone conjugate of α. With α as above, one sees that $\alpha^{\circledast}$ 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 $\mu(v) := \int_{\mathcal{X}} v d\mu$.

As an extension of the Bobkov and Götze criterion [4], we have

Theorem 1.1 (Gozlan and Léonard [22]) *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_cH holds*;
(b) *For any couple of bounded and measurable functions $u, v : \mathcal{X} \to \mathbb{R}$ such that $u(x) - v(y) \leq c(x, y)$ over $\mathcal{X}^2$, $\log \int_{\mathcal{X}} e^{\lambda u} d\mu \leq \lambda\mu(v) + \alpha^{\circledast}(\lambda)$, $\forall \lambda \geq 0$;*
(c) *For all $n \geq 1$ and $r > 0$ and for any couple of bounded and measurable functions $u, v : \mathcal{X} \to \mathbb{R}$ such that $u(x) - v(y) \leq c(x, y)$ over $\mathcal{X}^2$, the following concentration inequality holds $\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n u(X_k) \geq \mu(v) + r\right) \leq e^{-n\alpha(r)}$.*

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

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

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

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

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_pI instead of $T_{d^p}I$. In particular, W_1I stands for T_dI.

Organization of the paper. This paper is organized as follows. In the next section we characterize T_cI 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, extending Theorem 1.1 from i.i.d. sequences to time-continuous Markov processes. The method of proof is borrowed from [22] who proved Theorem 1.1 by means of large deviations of the empirical measure of an i.i.d. sequence. Here, it relies on the large deviations of the occupation measure of (X_t). The tensorization of T_cI is proved, and the relations between W_2I, Poincaré and log-Sobolev are exhibited with the help of [34].

In Sect. 3, W_1I is proved for the trivial metric $d(x,y) = \mathbf{1}_{x\neq y}$ with the sharp constant in terms of the spectral gap as well as a sharp Hoeffding concentration inequality for Markov processes.

For a general metric, using Lyons–Meyer–Zheng forward–backward martingale decomposition, we obtain in Sect. 4 a sharp W_1I inequality under the spectral gap existence of the Markov diffusion process in the space of Lipschitz functions.

Finally in Sect. 5 we propose a practical Lyapunov condition for W_1I (or a more general $T_\Phi I$) which, although not providing the sharp constant, yields a good order.

About the literature. Let us give some historical notes on the usual transportation inequality W_pH. Marton [31] first noticed that W_1H implies the concentration inequality for μ by a very elementary and neat argument, and she established W_1H for the law of a Dobrushin-contractive Markov chain in [32]. Talagrand [38] established W_2H for the Gaussian measure μ with $\alpha(r) = r/2C$ and provided the sharp constant C (this particular case of T_cH is often called Talagrand's transportation inequality). Bobkov and Götze [4] obtained the characterization of W_pH 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 et al. [3] shed light on a profound relation between Talagrand's transportation inequality, log-Sobolev inequality, inf-convolution and some Hamilton–Jacobi equation. Djellout et al. [13] obtained a necessary and sufficient condition for W_1H 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. See Fernique [17], Feyel and Ustunel [18], Bogachev and Kolesnikov [5] for the approach of Girsanov's transform to optimal mass transportation and T_cH.

Bolley and Villani [6] and later Gozlan and Léonard [22] 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 [21] 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 [30,36,37]. The reader is referred to the textbooks by Ledoux [27] and Villani [39,40] for further references pertaining to this very active field.

The transportation-information inequalities T_cI are new objects.

Convention and notation. Throughout this paper $(\mathcal{X}, d)$ 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 $\mu, \nu \in M_1(\mathcal{X})$, $\|\nu - \mu\|_{\mathrm{TV}} := \sup_{u:|u|\le 1} \int u\, d(\nu - \mu)$ is the total variation norm.
- A cost function c is a non-negative lower semicontinuous function on $\mathcal{X}^2$ such that $c(x, x) = 0$ for all $x \in \mathcal{X}$.

2 General results on T_cI

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 $(\Omega, \mathcal{F}, (X_t)_{t\ge 0}, (\mathbb{P}_x)_{x\in\mathcal{X}})$ with an invariant probability measure μ. The transition semigroup is denoted by $(P_t)_{t\ge 0}$.

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 $t \ge 0$, 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 $((X_t)_{t\ge 0}, \mathbb{P}_\mu)$ is a stationary ergodic process.

Assumption: *Closability of the symmetrized 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

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

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 $\mathcal{L}^\sigma$ (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

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

Remarks 2.2 (*I as rate function*) 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] under some conditions of absolute continuity and regularity of $P_t(x, dy)$, and established in full generality by Wu [43, Corollary B.11].

Remarks 2.3 (*Framework of Riemannian manifold*) 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 $\nabla g \in L^2(M \to TM; \mu)$ in distribution. In this case, if $\nu = f\mu$ with $0 < f \in C^1(M)$, then

$$I(\nu|\mu) = \int_{\mathcal{X}} |\nabla\sqrt{f}|^2 \, d\mu = \frac{1}{4}\int_{\mathcal{X}} \frac{|\nabla f|^2}{f} \, d\mu = \frac{1}{4} I_{\mathrm{F}}(f|\mu) \tag{2.2}$$

where $I_{\mathrm{F}}(f|\mu)$ is the classical Fisher information of the probability density f.

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

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

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 [40]) states that for any $\nu, \mu \in M_1(\mathcal{X})$ so that $T_c(\nu, \mu) < +\infty$,

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

where $\Phi_c := \{(u, v) \in (b\mathcal{B})^2 : u(x) - v(y) \le c(x, y),\ \forall (x, y) \in \mathcal{X}^2\}$. This motivates us to introduce as in [22]

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

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

(A1) $u \le v$ for all $(u, v) \in \Phi$;
(A2) For all $\nu_1, \nu_2 \in M_1(\mathcal{X})$, there exists $(u, v) \in \Phi$ such that $\int u\, d\nu_1 - \int v\, d\nu_2 \ge 0$.

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

(a) *The following transportation inequality holds*

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

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

$$\|P_t^{\lambda u}\|_{L^2(\mu)} \le e^{t[\lambda\mu(v) + \alpha^{\circledast}(\lambda)]} \tag{2.6}$$

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

(b′) *For all* $(u,v)\in\Phi$ *and all* $\lambda\ge 0$

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

(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* $r,t>0$,

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

(c′) *For all* $(u,v)\in\Phi$ *and for any* $r\ge 0$, *there exists* $\beta\in M_1(\mathcal{X})$ *such that* $\beta\ll\mu$, $d\beta/d\mu\in L^2(\mu)$ *and* $\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}_\beta\left(\frac{1}{t}\int_0^t u(X_s)\,ds\ge\mu(v)+r\right)\le-\alpha(r)$

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.*

Its proof is postponed to the end of the paper. From Theorem 2.4 we derive easily

Corollary 2.5 (The inequalities $W_1I(c)$ and $W_2I(c)$) *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* $W_1I(c)$ *inequality holds true:*

$$W_1^2(\nu,\mu)\le 4c^2\,I(\nu|\mu),\quad\forall\nu\in M_1(\mathcal{X});\tag{$W_1I(c)$}$$

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

$$\|P_t^{\lambda u}\|_{L^2(\mu)}\le\exp\left(t[\lambda\mu(u)+c^2\lambda^2]\right);$$

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

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

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

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

2. *The statements below are equivalent*:
   (a) *The following $W_2 I(c)$ inequality holds true*:

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

   (b) *For any* $v \in b\mathcal{B}$, $\|P_t^{\frac{1}{4c^2} Qv}\|_{L^2(\mu)} \le e^{\frac{t}{4c^2}\mu(v)}$, $\forall t \ge 0$ *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* $u \in b\mathcal{B}$, $\|P_t^{\frac{1}{4c^2} u}\|_{L^2(\mu)} \le e^{\frac{t}{4c^2}\mu(Su)}$, $\forall t \ge 0$ *where* $Su(y) = \sup_{x \in \mathcal{X}} \{u(y) - d^2(x, y)\}$ *is the so-called* "*sup-convolution*" *of* u.

Notation. The best constants $c > 0$ in $W_1 I(c)$ and $W_2 I(c)$ will be denoted, respectively, by $c_{W_1 I}(\mu)$ and $c_{\mathrm{W_2 I}}(\mu)$.

Remarks 2.6 (i) The best constants $c_{W_1 I}(\mu)$ and $c_{\mathrm{W_2 I}}(\mu)$ depend on the metric d and the Dirichlet form $\mathcal{E}$. Of course $c_{W_1 I}(\mu) \le c_{\mathrm{W_2 I}}(\mu)$.
(ii) The above corollary may be seen as the counterpart of Bobkov–Götze's characterizations of $W_p H$ $(p = 1, 2)$ for Markov processes.

The following simple example illustrates difference between $W_1 I$ and $W_2 I$.

Example 2.7 (Bernoulli distribution). Let μ be the Bernoulli distribution on $\mathcal{X} = \{0, 1\}$ with $\mu(\{1\}) = p \in (0, 1)$. Consider the Dirichlet form $\mathcal{E}(g, g) = (g(1) - g(0))^2$. By Theorem 3.1-(a) in Sect. 3, we see that

$$W_1(\nu, \mu)^2 \le pq I(\nu|\mu)$$

where W_1 is built with the trivial metric. The constant pq is sharp. However μ does not satisfy any $W_2 I(c)$ as is easily seen with $\nu = \mu_\varepsilon = (1 + \varepsilon g)\mu$.

Remarks 2.8
In all concentration inequalities of $L_t(u)$ in this paper, $\|d\beta/d\mu\|_2$ is used (which may become very big if $\beta = \delta_x$ and $\mathcal{X}$ is discrete); however in the study of the relaxation time $T_r = \inf\{t; \|\beta P_t - \mu\|_{\mathrm{TV}} \le e^{-1}\}$, one can control it in terms of $H(\beta|\mu)$ by assuming the log-Sobolev inequality (see the course of [35] and references therein). We believe that the main reason for this difference resides in the nature of our problem: the empirical measure $L_t(u) = \frac{1}{t}\int_0^t u(X_s)ds$ contains the whole sample path $X_s, 0 \le s \le t$ (unlike the relaxation problem). To avoid the term $\|d\beta/d\mu\|_2$ in our estimates, one can combine the relaxation time idea with ours in the following ways. Consider instead of $L_t(u)$, $L_t(u) \circ \theta_T = \frac{1}{t}\int_T^{T+t} u(X_s)ds$ (where θ_T is the shift such that $X_t \circ \theta_T = X_{T+t}$). We have

$$\mathbb{P}_\beta(L_t(u) \circ \theta_T \in \cdot) = \mathbb{P}_{\beta P_T}(L_t(u) \in \cdot) \le \frac{1}{2}\|\nu P_T - \mu\|_{\mathrm{TV}} + \mathbb{P}_\mu(L_t(u) \in \cdot)$$

where the first term at the r.h.s. can be controlled as in [35] and the last term by those in this work. A slightly different way is to use the following relaxation time $T := \inf\{t \geq 0;\ \|d(\beta P_t)/d\mu\|_2 \leq e\}$, which can be estimated by the hypercontractity and the entropy $H(\beta|\mu)$, and then apply our result to $\mathbb{P}_\beta(L_t(u) \circ \theta_T \in \cdot) = \mathbb{P}_{\beta P_T}\left(\frac{1}{t}\int_0^t u(X_s)ds \in \cdot\right)$. Considering $L_t(u)\circ\theta_T$ is very natural: before T, the process is too far from the equilibrium measure μ.

2.3 Relations between $W_2 I$, Poincaré and log-Sobolev inequalities

In the rest of the paper we are interested in two particular cases of $T_c I$: $W_1 I(c)$ and $W_2 I(c)$ introduced at Corollary 2.5.

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

$$\mathrm{Var}_\mu(g) \leq c\,\mathcal{E}(g, g), \quad \forall g \in \mathbb{D}_2(\mathcal{L}) \tag{2.8}$$

for some *finite* $c \geq 0$. We denote $c_{\mathrm{P}}(\mu)$ the best constant c in the above Poincaré inequality. It is the inverse of the *spectral gap* of $\mathcal{L}$.

The following result is just a reformulation of the work of [34].

Proposition 2.9 *In the framework of Riemannian manifold in Remarks 2.3, the followings hold.*

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

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

is satisfied, then μ satisfies $W_2 I(c)$.

(b) *$W_2 I(c)$ implies the Poincaré inequality with constant c, i.e., $c_{\mathrm{W_2I}}(\mu) \geq c_{\mathrm{P}}(\mu)$.*

(c) *Assume that the Bakry–Emery curvature* Ric + HessV *is bounded from below by $K \in \mathbb{R}$, where* Ric *is the Ricci curvature and* 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 in the case $K > 0$), then we have the following log-Sobolev inequality*

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

Proof Before the proof, let us remind the reader that $I = I_{\mathrm{F}}/4$ where I_{F} is I in Otto–Villani's paper [34].

- (a). The proof is direct, as by Otto and Villani [34] or Bobkov et al. [3] a logarithmic Sobolev inequality implies the $W_2 H$ (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.
- (b). The proof follows from the usual linearization procedure. Set $\mu_\varepsilon = (1+\varepsilon g)\mu$ for some smooth and compactly supported g with $\int g\, d\mu = 0$, we easily get $\lim_{\varepsilon\to 0} I(\mu_\varepsilon|\mu)/\varepsilon^2 = \frac{1}{4}\mathcal{E}_\nabla(g, g)$ and by Otto and 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 $W_2 I(c)$ we

get

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

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

- (c) is a direct application of the HWI inequality [34, Theorem 3] in the Euclidean case and [3] for a general Riemannian manifold: $H(\nu|\mu) \le 2W_2(\nu,\mu)\sqrt{I(\nu|\mu)} - \frac{K}{2}W_2^2(\nu,\mu)$. □

2.4 Tensorization of $T_c I$

Assume that $\mu_i \in M_1(\mathcal{X}_i)$ satisfies

$$\alpha_i(T_{c_i}(\nu,\mu_i)) \le I_i(\nu|\mu_i), \quad \forall \nu \in M_1(\mathcal{X}_i) \tag{2.9}$$

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 $\mu := \otimes_{i=1}^n \mu_i$, consider the sum-cost function

$$\oplus_i c_i(x,y) := \sum_{i=1}^n c(x_i,y_i), \quad \forall x,y \in \mathcal{X}^{(n)} \tag{2.10}$$

and the *inf-convolution* of (α_i)

$$\alpha_1 \Box \cdots \Box \alpha_n(r) := \inf\left\{\sum_{i=1}^n \alpha(r_i);\ r_i \ge 0, \sum_{i=1}^n r_i = r\right\}. \tag{2.11}$$

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 [22]). Define the sum-Dirichlet form of $\oplus_i \mathcal{E}_i$ by

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

$$\oplus_i \mathcal{E}_i(g,g) := \int_{\mathcal{X}^{(n)}} \sum_{i=1}^n \mathcal{E}_i(g_i,g_i)\,d\mu, \quad g \in \mathbb{D}(\mathcal{E}) \tag{2.12}$$

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 2.10 *Assume* (2.9) *for each $i = 1,\dots,n$ with α_i moreover convex. Define $c, \alpha, \mathcal{E}$, respectively, by* (2.10)–(2.12). *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*

$$\alpha_1 \Box \cdots \Box \alpha_n (T_{\oplus c_i}(\nu, \mu)) \le I_{\oplus_i \mathcal{E}_i}(\nu|\mu), \quad \forall \nu \in M_1\left(\mathcal{X}^{(n)}\right). \tag{2.13}$$

This result is similar to [22, 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 result [31] where an ordering of sites is required.

Lemma 2.11 *Let* $\mu = \otimes_{i=1}^n \mu_i$. *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* (2.10), $T_{\oplus c_i}(\mu, \nu) \le \int \sum_{i=1}^n T_{c_i}(\mu_i, \nu_i) d\nu$.

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 $\mu = \otimes_{i=1}^n \mu_i$, 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 $(Z_j)_{j \ne i}$ is a coupling of $(\mu_i(dx_i), \nu_i(dx_i|Y_j, j \ne i)$ and $\mathbb{P}$-a.s.,

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

Obviously $(X_j, j \ne i; \tilde{X}_i)$ and $(Y_j, j \ne i; \tilde{Y}_i)$ (more precisely their joint law) constitute a coupling of (μ, ν). Thus $\mathbb{E} \sum_j c_j(X_j, Y_j) \le \mathbb{E}[\sum_{j \ne i} c_j(X_j, Y_j) + c_i(\tilde{X}_i, \tilde{Y}_i)]$ or $\mathbb{E} c_i(X_i, Y_i) \le \mathbb{E} c_i(\tilde{X}_i, \tilde{Y}_i) = \mathbb{E} T_{c_i}(\mu_i, \nu_i(\cdot|Y_j, j \ne i))$. Consequently

$$\begin{aligned} T_{\oplus c_i}(\mu, \nu) &= \mathbb{E} \sum_{i=1}^n T_{c_i}(X_i, Y_i) \le \mathbb{E} \sum_{i=1}^n T_{c_i}(\mu_i, \nu_i(\cdot|Y_j, j \ne i)) \\ &= \int \sum_{i=1}^n T_{c_i}(\mu_i, \nu_i)\, d\nu. \end{aligned}$$

□

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

Lemma 2.12 *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*

$$I_{\oplus_i \mathcal{E}_i}(\nu|\mu) = \mathbb{E}^\nu \sum_i I_i(\nu_i|\mu_i). \tag{2.14}$$

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. □

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

Proof of Theorem 2.10 Without loss of generality we may assume that $I(\nu|\mu) < +\infty$. For simplicity write $\alpha = \alpha_1 \Box \cdots \Box \alpha_n$. By Lemma 2.11, Jensen's inequality and the definition of α,

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

The last quantity is equal to $I_{\oplus \mathcal{E}_i}(\nu|\mu)$, by Lemma 2.12. □

As an example of application, let $(X_t^i)_{t\ge 0}, i = 1, \ldots, n$ 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 $(X_0^i)_{i=1,\ldots,n}$ is fixed. Then $X_t := (X_t^1, \ldots, X_t^n)$ 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 2.13 *Assume that μ satisfies $T_c I$ on $\mathcal{X}$ with α convex. Then $\mu^{\otimes n}$ satisfies*

$$n\alpha\left(\frac{T_{\oplus_n c}(\nu, \mu^{\otimes n})}{n}\right) \le I_{\oplus_n \mathcal{E}}(\nu|\mu^{\otimes n}), \quad \forall \nu \in M_1(\mathcal{X}^n). \tag{2.15}$$

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 $t, r > 0$,

$$\mathbb{P}_\beta\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{t}\int_0^t u(X_s^i)\, ds \ge \mu(v) + r\right) \le \left\|\frac{d\beta}{d\mu^{\otimes n}}\right\|_2 e^{-nt\alpha(r)}. \tag{2.16}$$

Proof As $\alpha^{\Box n}(r) = n\alpha(r/n)$, (2.15) follows from Theorem 2.10. 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 (2.16) by Theorem 2.4. □

The tensorization of $W_p I$ in the dependent Gibbs measure case is carried out in [19].

3 Poincaré inequality implies Hoeffding's deviation inequality

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*

$$\mathrm{Var}_\mu(g) \le c_{\mathrm{P}}\, \mathcal{E}(g, g), \quad \forall g \in \mathbb{D}_2(\mathcal{L}) \tag{3.1}$$

implies

$$\|\nu - \mu\|_{\mathrm{TV}}^2 \le 4c_{\mathrm{P}}\, I(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}) \tag{3.2}$$

and for $u \in b\mathcal{B}$ so that $\|u\|_\infty \le 1$, $\mu(u) = 0$,

$$\int u d(\nu - \mu) \le \sqrt{4c_{\mathrm{P}} I(\nu|\mu) \left(\mathrm{Var}_\mu(u) + \sqrt{c_{\mathrm{P}} I(\nu|\mu)/2}\right)}. \tag{3.3}$$

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

$$\mathbb{P}_\beta\left(\frac{1}{t}\int_0^t u(X_s)\, ds \ge r\right) \tag{3.4}$$

$$\le \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-\frac{t}{c_{\mathrm{P}}} \max\left[\frac{r^2}{\delta(u)^2}, 4\varepsilon(\varepsilon+\sigma^2)\left(\sqrt{1+\frac{r^2}{2\varepsilon(\varepsilon+\sigma^2)^2\|u\|_\infty^2}}-1\right)\right]\right) \tag{3.5}$$

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

(b) *Conversely, assume that (P_t) is symmetric on $L^2(\mu)$. If there is some left-continuous and increasing $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ with $\alpha(1) > 0$ such that*

$$\alpha\left(\|\nu - \mu\|_{\mathrm{TV}}\right) \le I(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}), \tag{3.6}$$

then the Poincaré inequality (3.1) holds with $c_{\mathrm{P}} \le 1/\alpha(1)$.

Remarks 3.2 (i) Let $d(x,y)=\mathbf{1}_{x\neq y}$ (the trivial metric) and $\Phi=\{(u,u);\,\delta(u)\le 1\}$. Then $\frac{1}{2}\|\nu-\mu\|_{\mathrm{TV}}=W_1(\nu,\mu)=T_\Phi(\nu,\mu)$. Hence (3.2) is exactly the inequality $T_\Phi I$ or $W_1 I(c)$ with $4c^2=c_{\mathrm{P}}$. In the symmetric case, $\sqrt{c_{\mathrm{P}}}/4\le c_{W_1 I}(\mu)\le \sqrt{c_{\mathrm{P}}}/2$ by (b).

(ii) Lezaud [28] proved a better deviation inequality using the asymptotic variance $V(u)=\lim_{t\to\infty}\frac{1}{t}\mathrm{Var}_{\mathbb{P}_\mu}\left(\int_0^t u(X_s)ds\right)$ in the central limit theorem instead of $\mathrm{Var}_\mu(u)$, which is sharp in the moderate deviation scale (r very small), nevertheless his proof involves Kato's theory of perturbation of operators and difficult combinatorial techniques. Our result (3.3) is similar to [8].

Proof of Theorem 3.1 • (a). Equation (3.4) follows by (3.2) and (3.3) by Theorem 2.4. We first prove the transportation inequality (3.2). To this end recall the following known inequality in statistics (see [20]) $\frac{1}{4}\|\nu-\mu\|_{\mathrm{TV}}^2\le d_H^2(\nu,\mu)[2-d_H^2(\nu,\mu)]$ where $d_H^2(\nu,\mu)=\frac{1}{2}\int(1-\sqrt{f})^2d\mu=1-\mu(\sqrt{f})$ is the square of the Hellinger distance between $\nu=f\mu$ and μ. The above right-hand side is exactly $1-[\mu(\sqrt{f})]^2=\mathrm{Var}_\mu(\sqrt{f})$. In other words we have for every μ-probability density f, i.e. $f\ge 0$ and $\mu(f)=1$,

$$\|f\mu-\mu\|_{\mathrm{TV}}^2\le 4\mathrm{Var}_\mu(\sqrt{f}). \tag{3.7}$$

Now for every probability density f so that $\sqrt{f}\in\mathbb{D}(\mathcal{E})$, we have by (3.7) and the assumed Poincaré inequality, $\frac{1}{4}\|\nu-\mu\|_{\mathrm{TV}}^2\le \mathrm{Var}_\mu(\sqrt{f})\le c_{\mathrm{P}}\mathcal{E}(\sqrt{f},\sqrt{f})=c_{\mathrm{P}}I(\nu|\mu)$ which is (3.2).

For (3.3) it is enough to prove that if $\|u\|_\infty\le 1$, $\mu(u^2)\le\sigma^2(\le 1)$, then for every probability density f

$$\int u(fd\mu-d\mu)\le\sqrt{4\mathrm{Var}_\mu(\sqrt{f})\left(\sigma^2+\sqrt{\mathrm{Var}_\mu(\sqrt{f})/2}\right)} \tag{3.8}$$

Indeed by Cauchy–Schwarz inequality

$$\begin{aligned}\int u(fd\mu-d\mu)&\le\sqrt{\int(\sqrt{f}-1)^2d\mu\int u^2(\sqrt{f}+1)^2d\mu}\\&=\sqrt{4\mathrm{Var}_\mu(\sqrt{f})\int u^2(\sqrt{f}+1)^2d\mu/\mu[(\sqrt{f}+1)^2]}\end{aligned}$$

which gives us (3.7) (again). By the fact that $\delta(u^2)\le 1$, $\int u^2d(\nu-\mu)\le\|\nu-\mu\|_{\mathrm{TV}}/2$, we get by (3.7)

$$\begin{aligned}\int u^2(\sqrt{f}+1)^2d\mu/\mu[(\sqrt{f}+1)^2]&\le\mu(u^2)+\sqrt{\mathrm{Var}_\mu\left(\sqrt{(\sqrt{f}+1)^2/\mu[(\sqrt{f}+1)^2]}\right)}\\&\le\sigma^2+\sqrt{\mathrm{Var}_\mu(\sqrt{f})/2}\end{aligned}$$

which yields (3.8).

- (b). This converse part is based on the following well known fact ([11]): $1/c_P \ge \inf_{D\in\mathcal{B}:\mu(D)\le 1/2}\lambda_0(D)$, where $\lambda_0(D) = \inf \mathcal{E}(g,g)$ with the *inf* taken for all $g \in \mathbb{D}(\mathcal{E})$ such that $\|g\|_2 = 1$ and $g = 0$ a.e. on D^c, is the Dirichlet eigenvalue restricted in D. For any such domain D and g, letting $\nu := g^2\mu$, by the assumed $T_{\mathcal{V}}I$ and the fact that $\|\nu-\mu\|_{\mathrm{TV}} \ge |\nu(D)-\mu(D)|+|\nu(D^c)-\mu(D^c)| = 2(\mu(D^c)-\nu(D^c)) \ge 1$ we have

$$\mathcal{E}(g,g) \ge \mathcal{E}(|g|,|g|) = I(\nu|\mu) \ge \alpha(\|\nu-\mu\|_{\mathrm{TV}}) \ge \alpha(1)$$

  where it follows $\lambda_0(D) \ge \alpha(1)$ and then $c_P \le 1/\alpha(1)$. □

Note that $W_1 I$ for the trivial metric implies $W_1 I$ for any bounded metric. So our next purpose is to obtain $W_1 I$ for unbounded metrics. Our study is naturally separated into two sections. Next Sect. 4 is concerned with estimating sharply $c_{W_1 I}$ under Lyapunov spectral gap condition. In Sect. 5, Lyapunov function conditions for $W_1 I$ or more general $T_\Phi I$ are taken into consideration.

4 Spectral gap in the space of Lipschitz functions implies $W_1 I$ for diffusion processes

Let $((X_t), \mathbb{P}_\mu)$ be a reversible ergodic Markov process with generator $\mathcal{L}$, Dirichlet form $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$ and 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 $\Gamma : \mathbb{D}(\mathcal{E}) \times \mathbb{D}(\mathcal{E}) \to L^1(\mu)$ (symmetric, bilinear definite non-negative form):

$$\mathcal{E}(h,h) = \int_{\mathcal{X}} \Gamma(h,h)\, d\mu, \quad \forall h \in \mathbb{D}(\mathcal{E}). \tag{4.1}$$

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

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

Theorem 4.1 *Assume that $\int_{\mathcal{X}} d^2(x,x_0)\, d\mu(x) < +\infty$ and for any $g \in C_{\mathrm{Lip}}(\mathcal{X}, d)$ bounded with $\mu(g) = 0$, then $g \in \mathbb{D}(\mathcal{E})$ and*

$$\sqrt{\Gamma(g,g)} \le \sigma \|g\|_{\mathrm{Lip}}, \quad \mu\text{-a.s.} \tag{4.2}$$

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

$$\|\tilde{h}\|_{\mathrm{Lip}} \le C\|g\|_{\mathrm{Lip}} \tag{4.3}$$

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

$$\mathbb{E}_\mu \phi \left(\int_0^t g(X_s)\, ds \right) \le \mathbb{E}_\mu \phi (B_{2\sigma^2 C^2 \|g\|_{\mathrm{Lip}}^2 t}) \tag{4.4}$$

where B *is a standard Brownian Motion. In particular*

$$\mathbb{E}_\beta \exp \left(\lambda \int_0^t g(X_s)\, ds \right) \le \| \frac{d\beta}{d\mu} \|_2 e^{\lambda^2 (\sigma C)^2 \|g\|_{\mathrm{Lip}}^2 t}, \quad \forall \lambda \in \mathbb{R}, \quad t > 0. \tag{4.5}$$

and μ *satisfies* $W_1 I(\sigma C)$ *on* $(\mathcal{X}, d)$. *Furthermore let* $V(g) = \lim_{t\to\infty} \frac{1}{t} \mathrm{Var}_{\mathbb{P}_\mu} \left(\int_0^t g(X_s) ds \right)$ *and* $c_{\mathrm{P}} = c_{\mathrm{P}}(\mu) \le C$, *we have for all* $\lambda \in \mathbb{R}, t > 0$, $p > 1$ (*and* $1/p + 1/q = 1$),

$$\mathbb{E}_\beta \exp \left(\lambda \int_0^t g(X_s)\, ds \right) \le \| \frac{d\beta}{d\mu} \|_2 \exp \left(t \left[\frac{\lambda^2}{2} p V(g) + \frac{\lambda^4}{4} p^2 q c_{\mathrm{P}} (\sigma C)^4 \|g\|_{\mathrm{Lip}}^4 \right] \right). \tag{4.6}$$

Remarks 4.2 (i) Let C_{Lip}^0 be the Banach space of those $g \in C_{\mathrm{Lip}}$ with $\mu(g) = 0$, equipped with $\| \cdot \|_{\mathrm{Lip}}$. Hence the best constant C in (4.3) is exactly

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

By the spectral decomposition we always have (cf. [45, Lemma 5.4])

$$C = \|(-\mathcal{L})^{-1}\|_{C_{\mathrm{Lip}}^0} \ge \|(-\mathcal{L})^{-1}\|_{L^2(\mu) \bigcap \{g \in L^2(\mu);\, \mu(g)=0\}} = c_{\mathrm{P}}(\mu).$$

But the converse is false: the symmetric exponential measure $\mu = \frac{1}{2} e^{-|x|} dx$ on $\mathbb{R}$ satisfies the Poincaré inequality but the associated Dirichlet form $\mathcal{E}(g, g) = \int_{\mathbb{R}} g'^2 d\mu$ does not have spectral gap in C_{Lip}^0 w.r.t. the Euclidean metric.

(ii) The concentration inequality (4.5) and its equivalent $W_1 I(\sigma C)$ are sharp, as seen for one-dimensional Ornstein–Uhlenbeck process. The estimate (4.6) for p close to 1 is sharp in the moderate deviation scale (λ very small) and extends the result of [28] to unbounded g.

(iii) Klein et al. [26] developed *convex concentration inequality* (4.4) for semimartingales instead of $S_t(g)$, by means of a forward–backward martingale calculus, but their result cannot be applied directly here.

(iv) In [25], Joulin obtained a similar result for Markov processes with jumps.

Proof of Theorem 4.1 Let $\Phi = \{(g, g);\ \|g\|_{\mathrm{Lip}} \le 1, g \text{ bounded}\}$. Then $W_1(\nu, \mu) = T_\Phi(\nu, \mu)$ by Kantorovich–Rubinstein's theorem. Let us verify that (b') of Theorem 2.4 holds.

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

$$M_t(h) := h(X_t) - h(X_0) + \int_0^t g(X_s)\,ds \quad \text{and}$$

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

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

$$S_t(g) := \int_0^t g(X_s)\,ds = \frac{1}{2}(M_t(h) + M_t^*(h)), \tag{4.7}$$

it follows that for any convex function ϕ on $\mathbb{R}$, $\mathbb{E}_\mu\phi(S_t(g)) \le \frac{1}{2}\mathbb{E}_\mu[\phi(M_t(h) + \phi(M_t^*(h))] = \mathbb{E}_\mu\phi(M_t(h))$. 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 by our conditions (4.2) and (4.3), $\langle M(h)\rangle_t = 2\int_0^t \Gamma(h,h)(X_s)\,ds \le 2(\sigma C)^2 t$, by Jensen's inequality we obtain for all convex function ϕ on $\mathbb{R}$ that $\mathbb{E}_\mu\phi(S_t(g)) \le \mathbb{E}\phi(B_{\tau_t}) = \mathbb{E}\phi(\mathbb{E}[B_{2(\sigma C)^2 t}|\hat{\mathcal{F}}_{\tau_t}) \le \mathbb{E}\phi(B_{2(\sigma C)^2 t})$, which is (4.4). Applying this to $\phi(x) = e^{\lambda x}$, we get (4.5) for $\beta = \mu$. Hence Theorem 2.4-(b') holds with $\Phi = \{(g,g); \|g\|_{\text{Lip}} \le 1, g \text{ bounded}\}$ and $\alpha(r) = r^2/(4(\sigma C)^2)$, and Theorem 2.4 also gives us (4.5) for β.

For (4.6), it is enough to show it for $\beta = \mu$ by Theorem 2.4 and for g with $\|g\|_{\text{Lip}} = 1$. By Hölder's inequality,

$$\mathbb{E}_\mu e^{\lambda S_t(g)} \le \mathbb{E}_\mu e^{\lambda M_t} \le \left(\mathbb{E}_\mu \exp\left(\lambda p M_t - \frac{\lambda^2 p^2}{2}\langle M\rangle_t\right)\right)^{1/p} \left(\mathbb{E}_\mu \exp\left(\frac{\lambda^2 pq}{2}\langle M\rangle_t\right)\right)^{1/q}$$

where $M_t = M_t(h)$. Since $\exp\left(\lambda p M_t - \frac{\lambda^2 p^2}{2}\langle M\rangle_t\right)$ is an exponential martingale, its expectation is 1. To estimate the last term above we use the analogue of (3.2) at Theorem 3.1, given by Theorem 2.4(b). Noting $\delta(\Gamma(h,h)) \le \|\Gamma(h,h)\|_\infty \le (\sigma C)^2$ and $2\mu(\Gamma(h,h)) = V(g)$, this provides us with

$$\begin{aligned}\mathbb{E}_\mu \exp\left(\frac{\lambda^2 pq}{2}\langle M\rangle_t\right) &= \mathbb{E}_\mu \exp\left(\lambda^2 pq \int_0^t \Gamma(h,h)(X_s)ds\right) \\ &\le \exp\left(t\left[\frac{\lambda^2}{2} pq V(g) + \frac{\lambda^4 p^2 q^2}{4} \cdot c_P(\sigma C)^4\right]\right)\end{aligned}$$

where (4.6) follows. □

Remarks Using results from [14], one can be even more precise in the one dimensional case. Using the metric induced by the carré-du-champ operator of the diffusion, and conditions on Feller's scale and speed functions, we get sharp constant on the spectral gap for Lipschitz functions. It will be developed in [23].

5 Lyapunov function conditions

5.1 Main result

We will use in this section general conditions on the generator of the process, known as Lyapunov function conditions, for deriving $W_1 I$ or more generally $T_\Phi I$ where $\Phi = \{(u, u); |u| \le \phi\}$ with ϕ unbounded. 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 $\mathbb{P}_\mu$-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 $\int_0^t |g|(X_s)\,ds < +\infty$, $\mathbb{P}_\mu$-a.s. and

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

is a local $\mathbb{P}_\mu$-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 $U : \mathcal{X} \to [1, +\infty)$ in $\mathbb{D}_e(\mathcal{L})$, a non-negative function ϕ and a constant $b > 0$ such that

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

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 [16], large or moderate deviations and essential spectral radii [44,45]. 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 $c_P < \infty$ 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 $p \in [1, +\infty)$, we have*

$$\alpha_{p,R}\left(\|\phi^{1/p}(\nu - \mu)\|_{\mathrm{TV}}\right) \le I(\nu|\mu), \quad \forall \nu \tag{5.1}$$

where $R = 2^{1/q}\left(\frac{3}{c_P} + 6b + 2\sqrt{2}\|\phi\|_2\right)^{1/p}$, $1/p + 1/q = 1$ *and*

$$\alpha_{p,R}(r) = \begin{cases} r^2/(R^2 c_P), & \text{if } 0 \le r \le R; \\ r^p/(R^p c_P), & \text{if } r > R. \end{cases} \tag{5.2}$$

In particular, if moreover $d(x, x_0) \le C\phi^{1/p}$, $\forall x \in \mathcal{X}$ *for some* $x_0 \in \mathcal{X}$, *then*

$$\alpha_{p,R}(W_1(\nu, \mu)/C) \le I(\nu|\mu). \tag{5.3}$$

Remarks 5.2 Since $\|\phi(\nu - \mu)\|_{TV} = \sup_{u:|u|\le\phi} \int u\, d(\nu - \mu)$, the inequality (5.1) in this theorem may be regarded as $T_\Phi I$ in Theorem 2.4 with $\Phi = \{(u, u); u \in b\mathcal{B}, |u| \le \phi^{1/p}\}$. Since

$$W_1(\nu, \mu) = \sup_{f:\|f\|_{\mathrm{Lip}}\le 1} \int f\, d[\nu - \mu] \le \inf_{x_0\in\mathcal{X}} \|d(\cdot, x_0)(\nu - \mu)\|_{TV}, \tag{5.4}$$

one sees that the $W_1 I$ inequality (5.3) is a direct consequence of (5.1).

An important feature of this result is: $(1/t)\int_0^t u(X_s)ds$ has different concentration behaviors according to different p so that $|u| \le C\phi^{1/p}$, by (5.1).

The explicit constants in the inequalities of this theorem, produced by the Lyapunov function condition (H), are in general far from being optimal, but are sharp in order, as will be seen for the Ornstein–Uhlenbeck process at Example 5.5.

Example 5.3 (*M/M/∞* 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

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

(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 with $\phi(n) := (n+1)(1 - e^{-c})$ and $b = e^c + e^{-c} - 2$, and it is well known that $c_P(\mu) = \lambda$. Noting $\|\phi\|_2 = (1 - e^{-c})\sqrt{\lambda^2 + 3\lambda + 1}$, we have by Theorem 5.1 that for the distance $d(m, n) := |\sqrt{m} - \sqrt{n}|$,

$$W_1(\nu, \mu)^2 \le K(c) I(\nu|\mu), \quad \forall \nu \in M_1(\mathbb{N}), c > 0$$

where $K(c) = \frac{2}{1-e^{-c}}\left[3\left(1+2(e^c+e^{-c}-2)\lambda\right)+2\sqrt{2}(1-e^{-c})\sqrt{\lambda^2+3\lambda+1}\lambda\right]$. By Theorem 2.4, this gives the Gaussian deviation inequality for any observable u so that $|u(n+1)-u(n)| \le C|\sqrt{n+2}-\sqrt{n+1}|$ or $|u(n)| \le C\sqrt{1+n}$. Furthermore if $|u(n)| \le C[(1-e^{-c})(1+n)]^{1/p}$, we have by (5.2) and Theorem 2.4

$$\mathbb{P}_\beta\left(\frac{1}{t}\int_0^t u(X_s)ds > \mu(u)+r\right) \le \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-t\alpha_{p,R}(Cr)\right), \quad \forall t, r > 0.$$

See Joulin [24] and Liu and Ma [29] for previous studies on deviation inequalities of this model. Note that they only obtain Poisson tail for observable u so that $|u(n+1)-u(n)| \le C|\sqrt{n+2}-\sqrt{n+1}|$.

Corollary 5.4 *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*

$$\gamma := \limsup_{d(x,x_0)\to\infty} \frac{\Delta V(x)}{|\nabla V|^2(x)} < 1 \tag{5.5}$$

and for some $p \ge 1$, $d(x,x_0) \le C(1+|\nabla V|^2(x))^{1/p}$, $\forall x \in \mathcal{X}$. Then for every $\delta \in (0,(1-\gamma)^2/4)$, the Lyapunov function condition (H) is satisfied with $\phi = \delta(1+|\nabla V|^2(x))$ and some $b = b(\delta) > 0$, and $c_P = c_P(\mu) < +\infty$. If moreover ϕ is in $L^2(\mu)$, then w.r.t. the Riemannian metric d,

$$\alpha_{p,R}(W_1(\nu,\mu)/C) \le I(\nu|\mu), \quad \forall \nu \in M_1(\mathcal{X}). \tag{5.6}$$

where $\alpha_{p,R}$ is given in (5.2). In particular for every Lipschitz function u with $\|u\|_{\mathrm{Lip}} \le 1$ and any initial law β with $d\beta/d\mu \in L^2(\mu)$,

$$\mathbb{P}_\beta\left(\frac{1}{t}\int_0^t u(X_s)ds > \mu(u)+r\right) \le \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-t\alpha_{p,R}(Cr)\right), \quad \forall t, r > 0. \tag{5.7}$$

Proof For any $0 < \delta < (1-\gamma)^2/4$ let $\varepsilon \in (0, 1-\gamma)$ so that $\delta = (1-\gamma-\varepsilon)^2/4$ (or $\gamma+\varepsilon = 1-2\sqrt{\delta}$). Choose $\lambda = \sqrt{\delta}$ we have $\lambda - \lambda^2 = (\gamma+\varepsilon)\lambda + \delta$. For $U = e^{\lambda V}$, we have

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

where $b := \delta + \sqrt{\delta}\sup_{\mathcal{X}}\left(\Delta V - (1-2\sqrt{\delta})|\nabla V|^2\right)$ is finite under our assumption (5.5). Thus (H) is satisfied with $\phi = \delta(1+|\nabla V|^2)$ which is in $L^2(\mu)$. 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 $c_P(\mu) < \infty$. The statement now follows directly from Theorem 5.1. □

In the example below the constant C may change from one place to other.

Example 5.5 Let $\mathcal{X} = \mathbb{R}^n$, $V(x) = a|x|^\beta$ where $\beta > 1, a > 0$. Then (5.5) is verified with $\gamma = 0$ and then (H) is satisfied with $\phi = \delta(1 + |\nabla V|^2) \sim C|x|^{2(\beta-1)}$ when $|x|$ large (though V is not C^2 at $x = 0$, one can choose $U = e^{\lambda\tilde{V}}$ where $\tilde{V} \in C^2(\mathbb{R}^n)$ and $\tilde{V}(x) = V(x)$ for $|x| > 1$ in the proof of Corollary 5.4), where $0 < \delta < 1/4$.

(i) If $\beta \geq 3/2$, then the condition in Corollary 5.4 is verified with $p = 2(\beta-1) \geq 1$, so we have (5.7) for Lipschitz observable u with $p = 2(\beta - 1)$ [we can prove that (5.7) is false once $p > 2(\beta - 1)$]. 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(5.1), $\|\psi(\nu - \mu)\|_{\mathrm{TV}}^2 \leq CI(\nu|\mu)$. Then the Gaussian deviation inequality holds true for the observable u satisfying $|u| \leq C(1 + |x|)^{\beta-1}$.

(iii) Let $\beta = 2$ (Ornstein–Uhlenbeck process). (5.1) holds with $\phi(x) = \delta(1 + |x|^2)$ and so does (5.3) with $p = 2$. They are both correct in order. Indeed for any $p \in [1, +\infty)$ fixed, if $\psi(x) \gg |x|^{2/p} \sim C\phi^{1/p}$ at infinity with $\mu(\psi) < +\infty$, one cannot hope that

$$\alpha_{p,R}(\|\psi(\nu - \mu)\|_{\mathrm{TV}}) \leq I(\nu|\mu), \quad \forall \nu$$

for some $R > 0$, since by Theorem 2.4, this would imply that

$$\mathbb{P}_\mu\left(\int_0^1 \psi(X_s)ds > \mu(\psi) + r\right) \leq e^{-\alpha_{p,R}(r)}, \quad \forall r > 0$$

which is impossible for large r.

For this example and for the empirical mean with unbounded u, only Gaussian concentration inequality is known and that is only in the case $\beta = 2$ (cf. Djellout et al. [13]).

5.2 Proof of Theorem 5.1

The starting point is the following large deviation result.

Lemma 5.6 *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,*

$$\int -\frac{\mathcal{L}U}{U} g^2\, d\mu \leq \mathcal{E}(g, g), \quad \forall g \in \mathbb{D}(\mathcal{E}). \tag{5.8}$$

When U is bounded, this is contained in [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 $\mathbb{P}_\beta$-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 $\mathbb{P}_\beta$-martingale. As (N_t) is non-negative, it is also a $\mathbb{P}_\beta$-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

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

On the other hand by the lower bound of large deviation in [43, Theorem B.1, Corollary B.11] and Laplace–Varadhan principle, as in the proof of $(c') \Rightarrow (a)$ in Theorem 2.4, $F(u_n) \geq \sup\{\nu(u_n) - I(\nu|\mu);\ \nu \in M_1(E)\}$. Thus $\int u_n d\nu \leq I(\nu|\mu)$, which yields (by letting $n \to \infty$ and monotone convergence)

$$\int -\frac{\mathcal{L}U}{U}\,d\nu \leq I(\nu|\mu), \quad \forall \nu \in M_1(E). \tag{5.9}$$

This is equivalent to (5.8) 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 [43, Theorem B.1] under $\mathbb{P}_\beta$ since β is absolutely continuous with respect to μ. In addition, in the symmetric case, [43, Corollary B.11] states that the large deviation rate function is $I(\cdot|\mu)$; it does not depend on β under the underlying assumption that $\mathbb{P}_\mu$ 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 $\nu \mapsto \nu(u_n)$. □

Lemma 5.7 *In the framework of Theorem 5.1, for all $a \geq 2$ and $\nu \in M_1(\mathcal{X})$,*

$$\|\phi(\nu - \mu)\|_{\mathrm{TV}} \leq (1 + 2bc_{\mathrm{P}})\,\frac{a+1}{a-1} I(\nu|\mu) + a\sqrt{2}\|\phi\|_2\sqrt{c_{\mathrm{P}}\, I(\nu|\mu)}. \tag{5.10}$$

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

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

First consider the last term. Since $t^2 - 1 - h^2(t) \leq a(t-1)$ for $t \in [1, a]$, and $= 0$ for $t \geq a \geq 2$, $\int \phi[|f-1| - h^2(\sqrt{f})]d\mu = \int \phi[1_{\{f \leq 1\}}(1-f) + 1_{\{1 \leq f \leq a^2\}} a(\sqrt{f} - 1)]d\mu \leq a \int \phi|1 - \sqrt{f}|d\mu$ which is not greater than $a\|\phi\|_2\|1 - \sqrt{f}\|_2 = a\|\phi\|_2\sqrt{2}\sqrt{1 - \mu(\sqrt{f})} \leq a\|\phi\|_2\sqrt{2\text{Var}_\mu(\sqrt{f})} \leq a\sqrt{2c_{\text{P}}}\|\phi\|_2\sqrt{I(\nu|\mu)}$.

We turn now to bound the crucial first term by means of (5.8):

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

Substituting these two estimates into our previous decomposition, we obtain (5.10). □

Proof of Theorem 5.1 As noticed in Remarks 5.2, (5.3) follows directly from (5.1). Note that if $p = 1$, (5.1) is a direct consequence of Lemma 5.7 (5.10) with $a = 2$. It remains to show (5.1) in the case $p > 1$. Noting that $\int |f - 1|d\mu \leq 2\min\{1, \sqrt{c_{\text{P}} I(\nu|\mu)}\}$ by Theorem 3.1, we have by Hölder's inequality and (5.10) with $a = 2$,

$$\begin{aligned}\|\phi^{1/p}(\nu - \mu)\|_{\text{TV}} &\leq \left(\int |f - 1|d\mu\right)^{1/q} \left(\int \phi|f - 1|d\mu\right)^{1/p} \\ &\leq 2^{1/q} \min\left\{1, \sqrt{c_{\text{P}} I(\nu|\mu)}^{1/q}\right\} \\ &\quad \times \left(3\,(1 + 2bc_{\text{P}})\, I(\nu|\mu) + 2\sqrt{2}\|\phi\|_2\sqrt{c_{\text{P}} I(\nu|\mu)}\right)^{1/p}\end{aligned}$$

which entails immediately the desired (5.1). □

6 Proof of Theorem 2.4

The proof of this result is similar to [22, Theorems 2 and 15]'s ones. It takes advantage of large deviation results previously obtained by Wu. Namely,

- the identification of the rate function in the symmetric case and the large deviation lower bound are taken from [43] and
- the non-asymptotic Cramér's upper bounds which are used in [22] are replaced by the following result.

Lemma 6.1 (Wu [42]) *For any $u \in b\mathcal{B}$ and any $t > 0$, the following statements hold true.*

(1) *Denoting*

$$\Lambda(u) := \sup\left\{\int ug^2\,d\mu - \mathcal{E}(g,g);\, g \in \mathbb{D}(\mathcal{E}),\, \mu(g^2) = 1,\, \mu(g^2|u|) < +\infty\right\}, \tag{6.1}$$

one has

$$\|P_t^u\|_{L^2(\mu)} \leq e^{t\Lambda(u)} \tag{6.2}$$

and the equality holds in the symmetric case;

(2) *For all $r > 0$,*

$$\mathbb{P}_\beta\left(\frac{1}{t}\int_0^t u(X_s)\,ds - \mu(u) \geq r\right) \leq \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-t\lim_{\delta\downarrow 0} I_u(\mu(u)+r-\delta)\right) \tag{6.3}$$

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

It is proved in [42,43] 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] 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 2.4 Part (1). As $\nu \to I(\nu|\mu)$ is convex on $M_1(\mathcal{X})$, so is $I_u : \mathbb{R} \to [0, +\infty]$. 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

$$\Lambda(\lambda u) = I_u^*(\lambda) \tag{6.4}$$

where I_u^* is the convex conjugate of I_u. Indeed for $\lambda \geq 0$, by (6.1)

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

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 $(u, v) \in \Phi$,

$$I_u(\mu(v) + r) \geq \widetilde{\alpha}(r), \quad \forall r \in \mathbb{R} \tag{6.5}$$

where $\widetilde{\alpha}(r) = \alpha(r)$ 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 (6.4) and (6.5) leads us to

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

  for all $\lambda \geq 0$. Statement (b) now follows from inequality (6.2).
- $(a) \Rightarrow (c)$: This follows from (6.3) and (6.5), 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 $g \in \mathbb{D}_2(\mathcal{L})$, $\langle P_t^{\lambda u} g, P_t^{\lambda u} g\rangle_\mu \leq e^{2t(\lambda\mu(v)+\alpha^{\circledast}(\lambda))}\langle g, g\rangle_\mu$. Differentiating at time zero we obtain $2\langle g, \mathcal{L}g + \lambda ug\rangle_\mu = 2(\lambda\mu(g^2u) - \mathcal{E}(g, g)) \leq 2(\lambda\mu(v) + \alpha^{\circledast}(\lambda))\mu(g^2)$. Then for all $g \in \mathbb{D}_2(\mathcal{L})$, $\lambda[\mu(g^2u) - \mu(v)\mu(g^2)] - \alpha^{\circledast}(\lambda)\mu(g^2) \leq \mathcal{E}(g, g)$. 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 $\lambda[\nu(u) - \mu(v)] - \alpha^{\circledast}(\lambda) \leq I(\nu|\mu)$. Taking the supremum over all $\lambda \in \mathbb{R}$, as α assumed to be convex and $\alpha^{\circledast} = \tilde{\alpha}^*$ 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 [43, Theorem B.1] and the identification of the rate function in the symmetric case in [43, Corollary B.11], we have for any initial probability measure $\beta \ll \mu$,

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

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 $n \to \infty$ 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. □

References

1. Bakry, D.: L'hypercontractivité et son utilisation en théorie des semigroupes. In: Ecole d'Eté de Probabilités de Saint-Flour (1992). Lecture Notes in Mathematics, vol. 1581. Springer, New York (1994)
2. Bakry, D., Cattiaux, P., Guillin, A.: Rates of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. J. Funct. Anal. **254**(3), 727–759 (2008)
3. Bobkov, S.G., Gentil, I., Ledoux, M.: Hypercontractivity of Hamilton-Jacobi equations. J. Math. Pures Appl. **80**(7), 669–696 (2001)
4. Bobkov, S.G., Götze, F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. **163**, 1–28 (1999)
5. Bogachev, V., Kolesnikov, A.: Integrability of absolutely continuous transformations of measures and applications to optimal mass transportation. Probab. Theory Appl. **50**(3), 3–25 (2005)
6. Bolley, F., Villani, C.: Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. Ann. Fac. Sci. Toulouse **14**, 331–352 (2005)
7. Carlen, E.A.: Superadditivity of Fisher's information and logarithmic Sobolev inequalities. J. Funct. Anal. **101**(1), 194–211 (1991)
8. Cattiaux, P., Guillin, A.: Deviation bounds for additive functionals of Markov process. ESAIM P S **12**, 12–29 (2008)
9. Cattiaux, P., Guillin, A.: On quadratic transportation cost inequalities. J. Math. Pures Appl. **86**(9), 341–361 (2006)
10. Cattiaux, P., Guillin, A., Wu, L., Wang, F.Y.: Preprint, available on Arxiv (2007)
11. Chen, M.F.: Eigenvalues, inequalities, and ergodic theory. In: Probability and its Applications. Springer, New York (2005)
12. Deuschel, J.-D., Stroock, D.W.: Large Deviations, vol. 137 of Pure and Applied Mathematics. Academic Press, London (1989)
13. Djellout, H., Guillin, A., Wu, L.: Transportation cost-information inequalities for random dynamical systems and diffusions. Ann. Probab. **32**(3B), 2702–2732 (2004)
14. Djellout, H., Wu, L.: Lipschitzian spectral gap for one dimensional diffusions. In preparation (2008)
15. Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluations of certain Markov process expectations for large time, I, III, IV. Commun. Pure Appl. Math. **28**:1–47 (1975), **29**:389–461 (1976), **36**:183–212 (1983)
16. Douc, R., Fort, G., Guillin, A.: Subgeometric rates of convergence of f-ergodic strong Markov processes. Preprint, available on Arxiv (2006)

17. Fernique, X.: Extension du théorème de Cameron-Martin aux translations aléatoires. II. Intégrabilité des densités. In: High Dimensional Probability III (Sandjberg 2002), Progresses in Probability, vol. 55, 95-102. Birkhäuser, Basel (2003)
18. Feyel, D., Ustunel, A.S.: The Monge-Kantorovitch problem and Monge-Ampère equation on Wiener space. Probab. Theor. Relat. Fields **128**(3), 347–385 (2004)
19. Gao, F.Q., Wu, L.: Transportation-information inequalities for Gibbs measures. Preprint (2007)
20. Gibbs, A., Su, F.: On choosing and bounding probability metrics. Int. Stat. Rev. **70**(3), 419–435 (2002)
21. Gozlan, N.: Characterization of Talagrand's like transportation cost inequalities on the real line. To appear in J. Funct. Anal. (2006)
22. Gozlan, N., Léonard, C.: A large deviation approach to some transportation cost inequalities. To appear in Probab. Theory Relat. Fields (2008)
23. Guillin, A., Léonard, C., Wu, L.: Transportation cost inequalities. In preparation (2008)
24. Joulin, A.: Concentration et fluctuations de processus stochastiques avec sauts. Ph.D. thesis, Université La Rochelle (2006)
25. Joulin, A.: A new Poisson-type deviation inequality for Markov jump process with positive Wasserstein curvature. Preprint (2007)
26. Klein, T., Ma, Y., Privault, N.: Convex concentration inequalities and forward-backward stochastic calculus. Electron. J. Probab. **11**, 486–512 (2006)
27. Ledoux, M.: The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs, vol. 89. American Mathematical Society, Providence, RI (2001)
28. Lezaud, P.: Chernoff and Berry-Esséen inequalities for Markov processes. ESAIM Probab. Stat. **5**: 183–201 (2001) (electronic)
29. Liu, W., Ma, Y.: Spectral gap and deviation inequalities for birth-death processes. Preprint 2006, contained in the Ph.D. thesis of Y. Ma at Université La Rochelle 2007. To appear in Ann. IHP Probab. Stat. (2008)
30. Lott, J., Villani, C.: Ricci curvature for metric-measure spaces via optimal transport. To appear in Ann. Math. (2008)
31. Marton, K.: Bounding $\bar{d}$-distance by informational divergence: a way to prove measure concentration. Ann. Probab. **24**, 857–866 (1996)
32. Marton, K.: A measure concentration inequality for contracting Markov chains. Geom. Funct. Anal. **6**, 556–571 (1997)
33. Meyn, S.P., Tweedie, R.L.: Markov chains and stochastic stability. Communications and Control Engineering Series. Springer, New York (1993)
34. Otto, F., Villani, C.: Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. J. Funct. Anal. **173**, 361–400 (2000)
35. Saloff-Coste, L.: Lectures on finite Markov chains. École d'été de Probabilités de Saint-Flour 1996, LNM, vol. 1685, pp. 301–413. Springer, New York (1997)
36. Sturm, K.-T.: On the geometry of metric measure spaces, I. Acta Math. **196**, 65–131 (2006)
37. Sturm, K.-T.: On the geometry of metric measure spaces, II. Acta Math. **196**, 133–177 (2006)
38. Talagrand, M.: Transportation cost for gaussian and other product measures. Geom. Funct. Anal. **6**, 587–600 (1996)
39. Villani, C.: Saint-Flour Lecture Notes. Optimal transport, old and new. http://www.umpa.ens-lyon.fr/~cvillani/ (2005)
40. Villani, C.: Topics in Optimal Transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence, RI (2003)
41. Wang, F.Y.: Functional Inequalities, Markov Semigroup and Spectral Theory. Chinese Sciences Press, Beijing (2005)
42. Wu, L.: A deviation inequality for non-reversible Markov processes. Ann. Inst. Henri Poincaré (Sér. Probab. Stat.) **36**, 435–445 (2000)
43. Wu, L.: Uniformly integrable operators and large deviations for Markov processes. J. Funct. Anal. **172**, 301–376 (2000)
44. Wu, L.: Large and moderate deviations for stochastic damping Hamiltonian systems. Stoch. Proc. Appl. **91**, 205–238 (2001)
45. Wu, L.: Essential spectral radius for Markov semigroups (I) : discrete time case. Probab. Theory Relat. Fields **128**, 255–321 (2004)