The Annals of Applied Probability
2020, Vol. 30, No. 3, 1276–1320
https://doi.org/10.1214/19-AAP1530

TRANSPORT-INFORMATION INEQUALITIES FOR MARKOV CHAINS

BY NENG-YI WANG[1] AND LIMING WU[2]

[1] *School of Mathematics and Statistics, Huazhong University of Science and Technology, wangnengyi@hust.edu.cn*
[2] *Laboratoire de Mathématiques Blaise Pascal, CNRS-UMR 6620, Université Clermont Auvergne, Li-Ming.Wu@math.univ-bpclermont.fr*

This paper is the discrete time counterpart of the previous work in the continuous time case by Guillin, Léonard, the second named author and Yao [*Probab. Theory Related Fields* **144** (2009), 669–695]. We investigate the following transport-information $T_{\mathcal{V}}I$ inequality: $\alpha(T_{\mathcal{V}}(\nu,\mu)) \le I(\nu|P,\mu)$ for all probability measures ν on some metric space $(\mathcal{X}, d)$, where μ is an invariant and ergodic probability measure of some given transition kernel $P(x, dy)$, $T_{\mathcal{V}}(\nu,\mu)$ is some transportation cost from ν to μ, $I(\nu|P,\mu)$ is the Donsker–Varadhan information of ν with respect to (P,μ) and $\alpha : [0,\infty) \to [0,\infty]$ is some left continuous increasing function. Using large deviation techniques, we show that $T_{\mathcal{V}}I$ is equivalent to some concentration inequality for the occupation measure of the μ-reversible Markov chain $(X_n)_{n\ge 0}$ with transition probability $P(x, dy)$. Its relationships with the transport-entropy inequalities are discussed. Several easy-to-check sufficient conditions are provided for $T_{\mathcal{V}}I$. We show the usefulness and sharpness of our general results by a number of applications and examples. The main difficulty resides in the fact that the information $I(\nu|P,\mu)$ has no closed expression, contrary to the continuous time or independent and identically distributed case.

1. Introduction. Let $M_1(\mathcal{X})$ be the space of all probability measures on the Polish space $\mathcal{X}$ and consider a *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 (w.r.t. in short) 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 d is a metric on $\mathcal{X}$ which is lower semicontinuous on $\mathcal{X}^2$ (not necessarily compatible with the topology of $\mathcal{X}$) and $p \ge 1$, $(T_c(\nu,\mu))^{1/p} = W_{p,d}(\nu,\mu)$ is the L^p-Wasserstein distance between ν and μ. Throughout the paper we write in short $W_p(\nu,\mu)$ for $W_{p,d}(\nu,\mu)$ for the given metric d, and keep the notation $W_{p,\rho}(\nu,\mu)$ for the other metric ρ.

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

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

The usual transport inequality for a given $\mu \in M_1(\mathcal{X})$, called *transport-entropy inequality* in this paper, compares the Wasserstein metric $W_p(\nu,\mu)$ with the relative entropy $H(\nu|\mu)$. It was introduced by Marton [36, 37] and Talagrand [51] as a powerful tool for showing the concentration of measure for the deviation of a Lipschitzian observable from its median. Bobkov

Received October 2017; revised July 2019.
MSC2010 subject classifications. 60E15, 60F10, 60J05.
Key words and phrases. Transport-information inequality, concentration inequality, Donsker–Varadhan information.

and Götze [2] characterizes the transport inequality $W_1H(C)$ (see Corollary 2.6 below) by the Gaussian bound of the Laplace transform of Lipschitzian function, which implies the Gaussian concentration inequality of a Lipschitzian observable from its expectation. Gozlan and Léonard [22] completed the circle by proving that the Gaussian concentration inequality implies $W_1H(C)$.

The following extension of these inequalities,

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

has been proposed and developed by Gozlan and Léonard [22]. Here $\alpha : [0,\infty) \to [0,+\infty]$ is some left continuous and nondecreasing function with $\alpha(0) = 0$. It is proved in [22] that (T_cH) is equivalent to some concentration inequality for the empirical measure of a sequence of independent and identically distributed (i.i.d. in short) random variables $(X_n)_{n\ge 0}$, with common law μ. Their main idea comes from large deviations of i.i.d. sequences. The reader is referred to the books by Ledoux [28], Villani [52, 53] and the survey by Gozlan and Léonard [23] for literature on this topic.

Guillin et al. [24–26] proposed the following transport information inequality:

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

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

$$(1.3) \qquad I(\nu|\mu) = \begin{cases} \mathcal{E}(\sqrt{f},\sqrt{f}) & \text{if } \nu = f\mu, \sqrt{f} \in \mathbb{D}(\mathcal{E}), \\ +\infty & \text{otherwise} \end{cases}$$

associated with the Dirichlet form $\mathcal{E}$ on $L^2(\mu)$ with domain $\mathbb{D}(\mathcal{E})$. It is proved that (T_cI) is equivalent to some concentration inequality for the empirical measure of the continuous-time μ-reversible Markov process $(X_t)_{t\ge 0}$, associated with the Dirichlet form $\mathcal{E}$. Furthermore several useful sufficient conditions for (T_cI) and its relationships with functional inequalities are established.

The main common point between (T_cH) and (T_cI) is: $\nu \to H(\nu|\mu)$ is the rate function governing the large deviation principle of the empirical measure $L_n := \frac{1}{n}\sum_{k=0}^{n-1}\delta_{X_k}$ of an i.i.d. sequence $(X_n)_{n\ge 0}$ of common law μ (Sanov's theorem); whereas $\nu \to I(\nu|\mu)$ is the rate function governing the large deviation principle of the empirical measure $L_t := \frac{1}{t}\int_0^t \delta_{X_s}\,ds$ (Donsker–Varadhan's theorem) of the continuous-time μ-reversible Markov process $(X_t)_{t\ge 0}$ associated with $\mathcal{E}$. Here $\delta_\cdot$ is the Dirac measure at point $\cdot$.

The main purpose of this paper. This paper is a sequel to [24–26]. Our main purpose is to generalize the results in the continuous time in [24–26] to discrete time Markov chains. Let $(X_n)_{n\ge 0}$ be a Markov chain valued in $\mathcal{X}$ with transition probability kernel $P(x,dy)$, and μ be an invariant and ergodic probability measure of P. Let $I(\nu|P,\mu)$ be the Donsker–Varadhan information (or entropy) of ν w.r.t. (P,μ) (see Section 2 for definition). The counterpart of (T_cI) above becomes

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

The objective of this paper is twofold:

(i) to characterize (T_cI) above by the concentration inequality on the trapeze-type empirical measure

$$(1.4) \qquad \tilde{L}_n := \frac{1}{n}\left(\frac{1}{2}(\delta_{X_0} + \delta_{X_n}) + \sum_{k=1}^{n-1}\delta_{X_k}\right), \quad n \ge 1;$$

(ii) to present several easy-to-check sufficient conditions for ($T_c I$):

(a) Poincaré inequality;
(b) Lipschitzian spectral gap;
(c) hypercontractivity or hyperboundedness;
(d) Lyapunov function condition.

Related literature for concentration of Markov chains. We mention two lines of developments. The first one for obtaining concentration of Markov chains is to establish transport-entropy inequalities in the process level. This was begun again by Marton [36], who proved the transport-entropy inequality in process level for contracting (in the Dobrushin sense) Markov chains by means of coupling (known today as Marton's coupling). Her result was generalized to ϕ-mixing sequences by Rio [48] and Samson [50]. Djellout, Guillin and the second named author [12] further extended those results w.r.t. a general metric rather than the discrete one. Ollivier [42] introduced and studied the Ricci curvature of Markov chains, and Joulin and Ollivier [27] proved the concentration inequalities for Markov chains of positive Ricci curvature, with the constant more close to the best possible one—the variance (in the central limit theorem) for non large deviation, rather than that in [12]. Wintenberger [58] obtained the Bernstein-type concentration inequalities by introducing weak transport inequalities.

Another line is to use the spectral gap for concentration of Markov chains. Gillman [21] used the perturbation theory of operators to obtain deviation inequalities of symmetric Markov chains. The successive refinements of techniques allowed Dinwoodie [11] and Lezaud [31] to improve this bound. In the reversible case, the definite result is due to Léon and Perron [30]: they obtained a sharp Chernoff (or Hoeffding) concentration inequality by comparison with a two-states Markov chain. More recently Paulin [44, 45] improved the Bernstein inequality of Lezaud [31] in the nonsymmetric case (another aspect of his work generalized the results of Samson [50] by coupling method of Marton), and Miasojedow [39] and Fan et al. [47] extended the result of Léon and Perron [30] to the nonsymmetric case. This line is particularly fruitful in the continuous time case because of functional inequalities: see the second named author [60], Lezaud [32], Guillin et al. [26] etc, where the reader could find sharp concentration inequalities of different types. There are less recent works on concentration inequalities of discrete time Markov chains in this line, just because of lack of functional inequalities (except spectral gap) in the discrete time.

Our study of the transport-information inequality will benefit from those two lines of developments.

Organization of the paper. This paper is organized as follows. We present the main results in the next section. At first we characterize ($T_c I$) in the symmetric case, by means of concentration inequalities for the trapeze-type empirical mean $\tilde{L}_n(u)$ of observable u, extending Gozlan–Léonard's result from i.i.d. sequences to discrete time Markov chains. Some transport-information inequality in the nonsymmetric case is also proposed for obtaining concentration inequalities of the empirical mean $L_n(u)$. Next we propose several easy-to-check sufficient conditions for the transport-information inequality. In Section 3 we apply our general results to a series of examples, showing the usefulness and sharpness of those theoretical results.

The last four sections (from Section 4 to Section 7) are devoted to the proofs of the results in Section 2.

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 $b\mathcal{B}$. For $\mu \in M_1(\mathcal{X})$, $L^p(\mu) := L^p(\mathcal{X}, \mathcal{B}, \mu)$.
– The functions to be considered later are assumed to be measurable without warning.
– For a measurable function f on $\mathcal{X}$, $\mu(f) := \int f\, d\mu = \int_{\mathcal{X}} f(x)\, d\mu(x)$.
– 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.
– 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 $x \in \mathcal{X}$.
– For a function f, $\|f\|_p$ denotes its norm in $L^p(\mu)$; and for a bounded operator A on $L^p(\mu)$, $\|A\|_p$ is the operator norm on $L^p(\mu)$.

2. Main results.

2.1. *Markov chains, Donsker–Varadhan information, Feynman–Kac semigroup.*

Markov chain. Our probabilistic object is a discrete time Markov chain $(X_n)_{n\in\mathbb{N}}$ taking values on $\mathcal{X}$, defined on $(\Omega, \mathcal{F}, (\mathbb{P}_x)_{x\in\mathcal{X}})$, with the probability transition kernel $P(x, dy)$ and with an invariant and ergodic probability measure μ, where $\mathbb{P}_x(X_0 = x) = 1$, $\forall x \in \mathcal{X}$. Here the invariance of μ means $\mu P = \mu$, and the ergodicity of μ means that if a real bounded and measurable function f on $\mathcal{X}$ (say $f \in b\mathcal{B}$) satisfies $Pf = f$, μ-a.s., then f is constant μ-a.s.

For an initial probability measure β, $\mathbb{P}_\beta(\cdot) := \int \mathbb{P}_x(\cdot)\, d\beta(x)$ and $\mathbb{E}_\beta(\cdot) = \int \cdot\, d\mathbb{P}_\beta$, $\mathbb{E}_x(\cdot) := \mathbb{E}_{\delta_x}(\cdot)$.

Donsker–Varadhan information. The following definition is motivated by standard large deviation results ([16–18]).

DEFINITION 2.1. The Donsker–Varadhan information of ν with respect to (P, μ) is defined by

$$I(\nu|P, \mu) := \begin{cases} \sup_{1\le u\in b\mathcal{B}} \int_{\mathcal{X}} \log \dfrac{u}{Pu}\, d\nu & \text{if } \nu \ll \mu; \\ +\infty & \text{otherwise.} \end{cases} \tag{2.1}$$

When the context is clear (i.e., P, μ are fixed), we write in short $I(\nu)$ in place of $I(\nu|P, \mu)$.

REMARK 2.2 (I as rate function). $\nu \mapsto I(\nu)$ is exactly the modified Donsker–Varadhan information, that is, the rate function governing the local large deviation principle of the empirical measure

$$L_n := \frac{1}{n} \sum_{k=0}^{n-1} \delta_{X_k} \tag{2.2}$$

under $\mathbb{P}_\beta$ with initial measure $\beta \ll \mu$, in the τ-topology for large time n. This was proved by Donsker and Varadhan [16–18] for the weak convergence topology under some conditions of absolute continuity and regularity of $P(x, dy)$, and established in full generality by the second named author [61].

REMARK 2.3 (Relation with relative entropy). The following identity is well known ([9, 10, 16, 17]): $\forall \nu \ll \mu$,

$$I(\nu|P, \mu) = \inf_{Q\in M_1(\mathcal{X}^2), Q_0=Q_1=\nu} H(Q|\nu \otimes P), \tag{2.3}$$

where Q_0, Q_1 are respectively the marginal law of the first coordinate and the second of Q on $\mathcal{X}^2$, and $\nu \otimes P(dx, dy) = \nu(dx)P(x, dy)$. In the i.i.d. case $P(x, dy) = \mu(dy)$,

$$I(\nu|P, \mu) = H(\nu|\mu).$$

Feynman–Kac semigroups. We consider three kinds of *Feynman–Kac* semigroups for $u \in b\mathcal{B}$. The first one is $g \to e^u Pg$ for which we have for all $n \geq 1$,

$$(e^u P)^n g(x) = \mathbb{E}_x g(X_n) \exp\left(\sum_{k=0}^{n-1} u(X_k)\right) = \mathbb{E}_x g(X_n) \exp(nL_n(u)), \tag{2.4}$$

where L_n is given in (2.2); and the second one is $g \to P(e^u g)$ for which we have for all $n \geq 1$,

$$(Pe^u)^n g(x) = \mathbb{E}_x g(X_n) \exp\left(\sum_{k=1}^{n} u(X_k)\right) = \mathbb{E}_x g(X_n) \exp(nL_n(u) \circ \theta), \tag{2.5}$$

where θ is the shift on Ω such that $X_n \circ \theta = X_{n+1}$. Though those two are often used in large deviations ([9, 10]), whereas in the symmetric case, we will use the third one: $g \to e^{u/2} P(e^{u/2} g) =: P^u g$ for which we have for all $n \geq 1$,

$$\begin{aligned}(P^u)^n g(x) &= \mathbb{E}_x g(X_n) \exp\left(\frac{u(X_0) + u(X_n)}{2} + \sum_{k=1}^{n-1} u(X_k)\right) \\ &= \mathbb{E}_x g(X_n) \exp(n\tilde{L}_n(u)),\end{aligned} \tag{2.6}$$

where $\tilde{L}_n$ is the trapeze-type empirical measure given by (1.4).

2.2. *Transportation cost $T_\mathcal{V}$ and transport-information inequality.* The Kantorovich–Rubinstein duality theorem (see [52]) states that for any $\nu, \mu \in M_1(\mathcal{X})$ so that $T_c(\nu, \mu) < +\infty$,

$$T_c(\nu, \mu) = \sup_{(u,v) \in \mathcal{V}_c} \int u\, d\nu - \int v\, d\mu, \tag{2.7}$$

where

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

In particular for $c(x, y) = d(x, y)$, we have

$$W_1(\nu, \mu) = \sup_{g : \|g\|_{\mathrm{Lip}} = 1} \int g\, d(\nu - \mu). \tag{2.8}$$

This motivates the introduction of the following general transportation cost (slightly more general than that in [22]):

$$T_\mathcal{V}(\nu, \mu) = \sup_{(u,v) \in \mathcal{V}} \int u\, d\nu - \int v\, d\mu, \tag{2.9}$$

where $\mathcal{V}$ is some given family of couples $(u, v) \in (b\mathcal{B})^2$ such that $u \leq v$. It may be negative.

The main objective of this paper is to investigate the following transport-information inequality:

$$\alpha(T_\mathcal{V}(\nu, \mu)) \leq I(\nu|P, \mu) \quad \forall \nu \in M_1(\mathcal{X}), \tag{α-$T_\mathcal{V} I$}$$

where $\alpha : \mathbb{R} \to [0, \infty]$ *is some nondecreasing left-continuous convex function with* $\alpha(r) = 0$ *for* $r \leq 0$, *that will be assumed throughout the paper except explicit contrary statement.*

When (α-$T_{\mathcal{V}}I$) holds, we say that α is a *$T_{\mathcal{V}}I$-deviation function*. Sometimes we write in short $T_{\mathcal{V}}I$ in place of (α-$T_{\mathcal{V}}I$). In the i.i.d. case this inequality becomes the transport-entropy inequality (α-$T_{\mathcal{V}}H$) introduced by Gozlan–Léonard [22].

Let us consider the convex conjugate

$$\alpha^*(\lambda) := \sup_{r\in\mathbb{R}}\big(\lambda r - \alpha(r)\big). \tag{2.10}$$

We have $\alpha^*(\lambda) = +\infty$ for $\lambda < 0$ and

$$\alpha^*(\lambda) = \sup_{r\ge 0}\big(\lambda r - \alpha(r)\big), \quad \lambda \ge 0. \tag{2.11}$$

2.3. *Characterization of* $(\alpha - T_{\mathcal{V}}I)$ *in the symmetric case.* The objective of this subsection is to illustrate the probabilistic meaning of the transport-information inequalities. We begin with the symmetric case.

THEOREM 2.4. *Assume that P is symmetric on $L^2(\mu)$, that is, the Markov chain $(X_n)_{n\ge 0}$ is μ-reversible. Let $\alpha : \mathbb{R} \to [0,\infty]$ be a left continuous nondecreasing convex function such that $\alpha(r) = 0$ for $r \le 0$, $\mathcal{V}$ as above. Then the following properties are equivalent*:

(a) (α-$T_{\mathcal{V}}I$) *holds*.

(b) *For all $(u,v) \in \mathcal{V}$ and all $\lambda \ge 0$, $n \ge 1$*

$$\|(P^{\lambda u})^n\|_2 \le e^{n[\lambda\mu(v)+\alpha^*(\lambda)]}, \tag{2.12}$$

where $P^u(x,dy) = e^{(u(x)+u(y))/2}P(x,dy)$ and α^* is defined by (2.10).

(b′) *For all $(u,v) \in \mathcal{V}$ and all $\lambda \ge 0$*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}_\mu \exp\big(\lambda n \tilde{L}_n(u)\big) \le \lambda\mu(v) + \alpha^*(\lambda).$$

(c) *For any initial measure $\beta \ll \mu$ with $d\beta/d\mu \in L^2(\mu)$ and for all $(u,v) \in \mathcal{V}$, the following concentration inequality holds*:

$$\mathbb{P}_\beta\big(\tilde{L}_n(u) > \mu(v) + r\big) \le \left\|\frac{d\beta}{d\mu}\right\|_2 e^{-n\alpha(r)} \quad \forall r > 0, n \ge 1. \tag{2.13}$$

(c′) *For all $(u,v) \in \mathcal{V}$ and for any $r > 0$, there exists $\beta \in M_1(\mathcal{X})$ such that $\beta \ll \mu$, $d\beta/d\mu \in L^2(\mu)$ and*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\beta\big(\tilde{L}_n(u) > \mu(v) + r\big) \le -\alpha(r).$$

In the i.i.d. case, as $I(\nu|P,\mu) = H(\nu|\mu)$, the result above is Gozlan–Léonard's characterization of (α-$T_{\mathcal{V}}H$) in [22].

The characterization (2.13) of the transport-information inequality ($T_{\mathcal{V}}I$) gives a robust concentration inequality for the trapeze-type empirical mean $\tilde{L}_n(u)$. In practice, if $\|d\beta/d\mu\|_2$ is too big, one may run the Markov chain till step N such that $\|d(\beta P^N)/d\mu\|_2$ is not so big (possible if the Poincaré inequality holds), and consider $\tilde{L}_n \circ \theta^N$ (see [26, 27] for explanation).

We explain now why the equivalences above work for the trapeze empirical mean $\tilde{L}_n(u)$, but not for $L_n(u)$. One crucial reason is: as $P^u = e^{u/2}Pe^{u/2}$ is again symmetric on $L^2(\mu)$, we have the remarkable equality $\|(P^u)^n\|_2 = \|P^u\|_2^n$ for all $n \in \mathbb{N}^*$ (by the spectral decomposition of a bounded symmetric operator). The other reason will be given after Lemma 4.3 (due to [29] under the uniform integrability of P on $L^2(\mu)$). However e^uP and Pe^u are no longer symmetric, the previous equality is lost.

It seems that Lei [29] was the first to use $\tilde{L}_n(u)$ instead of $L_n(u)$.

REMARK 2.5. As will be seen from the proof of Theorem 2.4, we have

$$(\mathrm{a}) \Longrightarrow (\mathrm{b}) \Longrightarrow (\mathrm{b}'), \qquad (\mathrm{a}) \Longrightarrow (\mathrm{c}) \Longrightarrow (\mathrm{c}')$$

without the convexity of α; and (c$'$) $\Longrightarrow$ (a) holds true *even in the nonsymmetric case and without the convexity of* α.

Furthermore since $|L_n(u) - \tilde{L}_n(u)| \le \sup_x |u(x)|/n$, (c$'$) is equivalent to

(c$''$) *For all* $(u, v) \in \mathcal{V}$ *and for any* $r > 0$, *there exists* $\beta \in M_1(\mathcal{X})$ *such that* $\beta \ll \mu$, $d\beta/d\mu \in L^2(\mu)$ *and*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\beta\big(L_n(u) > \mu(v) + r\big) \le -\alpha(r).$$

From Theorem 2.4 we derive easily

COROLLARY 2.6 (The inequalities ($W_1I(C)$) and ($W_2I(C)$)). *Let* $C > 0$ *and assume the symmetry of* P *on* $L^2(\mu)$.

(1) *The statements below are equivalent*:

(a) *the following* ($W_1I(C)$) *inequality holds true*:

$$(W_1I(C)) \qquad W_1^2(\nu, \mu) \le 2CI(\nu|P, \mu) \quad \forall \nu \in M_1(\mathcal{X});$$

(b) *for all bounded Lipschitzian function* u *on* $\mathcal{X}$ *with* $\|u\|_{\mathrm{Lip}} \le 1$ *and all* $\lambda \ge 0, n \ge 1$,

$$\|(P^{\lambda u})^n\|_2 \le \exp(n[\lambda\mu(u) + C\lambda^2/2]);$$

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

$$\limsup_{n\to+\infty} \frac{1}{n} \log \mathbb{E}_\mu \exp\big(\lambda n \tilde{L}_n(u)\big) \le C\lambda^2/2;$$

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

$$\mathbb{P}_\beta\big(\tilde{L}_n(u) > \mu(u) + r\big) \le \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-\frac{nr^2}{2C\|u\|_{\mathrm{Lip}}^2}\right).$$

(2) *Assume that the metric* d *generates the topology of* $\mathcal{X}$. *The statements below are equivalent*:

(a) *the following* ($W_2I(C)$) *inequality holds true*:

$$(W_2I(C)) \qquad W_2(\nu, \mu)^2 \le 2CI(\nu|P, \mu) \quad \forall \nu \in M_1(\mathcal{X});$$

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

$$\|(P^{\frac{1}{2C}Qv})^n\|_2 \le e^{\frac{n}{2C}\mu(v)} \quad \forall n \ge 1,$$

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* $u \in b\mathcal{B}$,

$$\|(P^{\frac{1}{2C}u})^n\|_2 \le e^{\frac{n}{2C}\mu(Su)} \quad \forall n \ge 1,$$

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

In the i.i.d. case, $W_pI(C)$ becomes $W_pH(C)$ ($p=1,2$), and the equivalence between (a) and (b) both in part (1) and in part (2) of this corollary is the well-known characterization of Bobkov and Götze [2]. $W_2H(C)$ is known often as Talagrand's T_2-transport inequality.

REMARK 2.7. From part (1.d), we see that the best constant C in ($W_1I(C)$) is just the best sub-Gaussian constant in the concentration inequality of $\tilde{L}_n(u)$ for Lipschitzian observersable u with $\|u\|_{\mathrm{Lip}} \le 1$. In the i.i.d. case, the best constant C_G in ($W_1I(C)$) coincides with the best constant $C_H(\mu)$ in $W_1H(C)$, called *sub-Gaussian constant* of μ in Bobkov, Houdré and Tetali [3]. Following them,

DEFINITION 2.8. We call the best constant C in ($W_1I(C)$), *sub-Gaussian constant* of (P,μ), denoted by $C_G(P,\mu)$, or C_G if without ambiguity.

Let us give immediately a lower bound of C_G. Let $\sigma_n^2(u) := n\,\mathrm{Var}_{\mathbb{P}_\mu}(\tilde{L}_n(u))$ and

$$\sigma^2(u) := \lim_{n\to\infty} \sigma_n^2(u) = \mathrm{Var}_\mu(u) + 2\sum_{k=1}^{\infty}\langle u-\mu(u), P^k u\rangle_\mu \tag{2.14}$$

if the last series is convergent, where $\mathrm{Var}_\mu(u)$ is the variance of u under μ. $\sigma^2(u)$ is the asymptotic variance in the central limit theorem. If ($W_1I(C)$) holds, using part (1.b) above and looking at the second order Taylor expansion, we have

$$C_G \ge \sup_{n\ge 1}\sup_{\|u\|_{\mathrm{Lip}}=1} \sigma_n^2(u) \ge \sup_{\|u\|_{\mathrm{Lip}}=1} \sigma^2(u) =: V(P,\mu). \tag{2.15}$$

2.4. *Nonsymmetric case.* We now turn to the nonsymmetric case.

In the continuous time case the unique choice of symmetrization of a Markov generator $\mathcal{L}$ is $(\mathcal{L}+\mathcal{L}^*)/2$. In the discrete time case, one has three choices: $P^\sigma = (P+P^*)/2$, P^*P, PP^*. For instance P^σ turns out to be the best one for the Poincaré inequality and for the central limit theorem (ref. the second named author [59]), whereas PP^* (or P^*P) is the best tool for convergence of P^n to μ in $L^2(\mu)$.

DEFINITION 2.9. The symmetrized Donsker–Varadhan information of ν w.r.t. the invariant and ergodic probability measure μ of P is defined as

$$I^\sigma(\nu) := \frac{1}{2} I\big(\nu | PP^*, \mu\big),$$

where P^* is the adjoint operator of P on $L^2(\mu)$.

We begin with exhibiting a relationship between H, I^σ and I.

PROPOSITION 2.10.

(a) *It holds always that*

$$I^\sigma(\nu) \le I(\nu) \quad \forall \nu \in M_1(\mathcal{X}). \tag{2.16}$$

(b) *If P is symmetric and nonnegative definite on $L^2(\mu)$, then*

$$I(\nu|P,\mu) \le 2H(\nu|\mu) \quad \forall \nu \in M_1(\mathcal{X}). \tag{2.17}$$

In particular if $(\alpha - T_{\mathcal{V}}I)$ holds, then $\alpha(T_{\mathcal{V}}(\nu,\mu)) \le 2H(\nu|\mu)$.

(c) *Without the symmetry of P, the following inequality always holds*:

$$I^{\sigma}(\nu) \le H(\nu|\mu) \quad \forall \nu \in M_1(\mathcal{X}). \tag{2.18}$$

In particular if (α-$T_{\mathcal{V}}I^{\sigma}$) *below holds*

$$\alpha\big(T_{\mathcal{V}}(\nu,\mu)\big) \le I^{\sigma}(\nu|\mu) \quad \forall \nu \in M_1(\mathcal{X}) \tag{α-$T_{\mathcal{V}}I^{\sigma}$}$$

then the transport-entropy inequality $\alpha(T_{\mathcal{V}}(\nu,\mu)) \le H(\nu|\mu)$, $\forall \nu \in \mathcal{M}_1(\mathcal{X})$ *holds*.

REMARK 2.11. In Proposition 2.10(b), without the nonnegative definiteness of P, (2.17) does not hold. A counter example is as follows: let $\mathcal{X} = \{0, 1\}$,

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

then $\mu(0) = \mu(1) = 1/2$. By (2.3), $I(\nu|P,\mu) = +\infty$ for all $\nu \ne \mu$. Hence (2.17) does not hold.

From Proposition 2.10, we see that under some mild condition our transport information inequality is stronger than transport entropy $T_{\mathcal{V}}H$ inequality. This means roughly that if a Markov chain with the invariant and ergodic measure μ is concentrated, then the i.i.d. sequence of common law μ is concentrated: a quite natural idea.

Below we explain the probabilistic meaning of $(\alpha - T_{\mathcal{V}}I^{\sigma})$.

THEOREM 2.12. *Without the symmetry of P on $L^2(\mu)$ nor convexity of α, assume $(\alpha - T_{\mathcal{V}}I^{\sigma})$. Then for any initial measure $\beta \ll \mu$ with $d\beta/d\mu \in L^2(\mu)$ and for all $(u,v) \in \mathcal{V}$ and $r > 0$, $n \ge 1$*,

$$\mathbb{P}_{\beta}\big(L_n(u) > \mu(v) + r\big) \le \left\| \frac{d\beta}{d\mu} \right\|_2 e^{-n\alpha(r)}. \tag{2.19}$$

More generally if for some $N \ge 1$,

$$\alpha\big(T_{\mathcal{V}}(\nu,\mu)\big) \le \frac{1}{2N} I\big(\nu|P^N(P^*)^N, \mu\big) \quad \forall \nu \in M_1(\mathcal{X}), \tag{2.20}$$

the concentration inequality (2.19) *still holds for all $n \ge N$*.

REMARK 2.13. As seen in its proof, when α is moreover convex, $(\alpha - T_{\mathcal{V}}I^{\sigma})$ is in fact equivalent to

$$\|e^{\lambda u}P\|_2 \le \exp\big(\lambda\mu(v) + \alpha^*(\lambda)\big) \quad \forall (u,v) \in \mathcal{V}, \lambda \ge 0.$$

Having explained the probabilistic meaning of the transport-information inequality $T_{\mathcal{V}}I$ or $T_{\mathcal{V}}I^{\sigma}$ above, let us now present several easy-to-check sufficient conditions for $T_{\mathcal{V}}I$ or $T_{\mathcal{V}}I^{\sigma}$.

2.5. *Poincaré inequality is equivalent to Hoeffding's deviation inequality.* Let $\sigma(P|_{L^2(\mu)})$ be the spectrum of P on $L^2(\mu)$. The largest element $\lambda_0(P)$ of $\sigma(P|_{L^2(\mu)})$ is 1. If P is symmetric, the second one denoted by $\lambda_1(P)$, is the supremum of $\sigma(P|_{L^2(\mu)}) \setminus \{1\}$. Its relation with the Poincaré inequality

$$\mathrm{Var}_{\mu}(g) \le c_{\mathrm{P}}\langle g, (I-P)g\rangle_{\mu} \quad \forall g \in L^2(\mu) \tag{2.21}$$

is: the best constant c_P coincides with $(1 - \lambda_1(P))^{-1}$.

We now state the sharp Hoeffding's inequality of León and Perron [30] under the spectral gap condition in the transport-information inequality form, extend it to the nonsymmetric case, and show especially that the converse is also true.

THEOREM 2.14. *Let $((X_n)_{n\geq 0}, \mathbb{P}_\mu)$ be a stationary ergodic Markov chain.*

(a) (a variant of León and Perron [30], Theorem 1) *Assume that P is symmetric on $L^2(\mu)$ and $\lambda_1^+ := \max\{0, \lambda_1(P)\} < 1$ (the spectral gap condition). Then*

$$\|\nu - \mu\|_{TV}^2 \leq 2\frac{1+\lambda_1^+}{1-\lambda_1^+} I(\nu) \leq 4c_{\mathrm{P}} I(\nu) \quad \forall \nu \in M_1(\mathcal{X}). \tag{2.22}$$

In particular for $u \in b\mathcal{B}$, for any initial probability measure $\beta \ll \mu$ with $d\beta/d\mu \in L^2(\mu)$ and for all $r > 0$ and $n \geq 1$,

$$\mathbb{P}_\beta(\tilde{L}_n(u) > \mu(u) + r) \leq \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-n\frac{2r^2(1-\lambda_1^+)}{\|u\|_{\mathrm{osc}}^2(1+\lambda_1^+)}\right), \tag{2.23}$$

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

(b) *Without the symmetry of P, if for some $N \geq 1$, some $\delta \in [0, 1)$,*

$$\mathrm{Var}_\mu(P^N u) \leq \delta^2 \mathrm{Var}_\mu(u) \quad \forall u \in L^2(\mu) \tag{2.24}$$

then

$$\|\nu - \mu\|_{TV}^2 \leq 2\frac{1+\delta^2}{1-\delta^2} I(\nu | P^N (P^*)^N, \mu) \quad \forall \nu \in M_1(\mathcal{X}) \tag{2.25}$$

in particular for $u \in b\mathcal{B}$, and all $n \geq N$, $r > 0$

$$\mathbb{P}_\beta(L_n(u) > \mu(u) + r) \leq \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-n\frac{(1-\delta^2)r^2}{N(1+\delta^2)\|u\|_{\mathrm{osc}}^2}\right). \tag{2.26}$$

(c) *Conversely in the symmetric case, if $\alpha(\|\nu - \mu\|_{\mathrm{TV}}) \leq I(\nu|P, \mu)$, $\forall \nu \in M_1(\mathcal{X})$, for some nonnegative nondecreasing left-continuous function $\alpha : \mathbb{R}^+ \to [0, +\infty]$ with $\alpha(1) > 0$, then the Poincaré inequality (2.21) holds with*

$$c_P \leq \frac{1}{1 - e^{-\alpha(1)}} \quad \textit{or equivalently} \quad \lambda_1(P) \leq e^{-\alpha(1)}. \tag{2.27}$$

REMARK 2.15.

(i) For Markov chains with finite states, Lezaud [32] proved a Bernstein-type (Gaussian-exponential for small and large deviations respectively) concentration inequality for $L_n(u)$. For reversible Markov chains with finite states, León and Perron [30], Theorem 1, using comparison with a two-states Markov chain, proved an optimal concentration inequality which implies in particular the following sharp Hoeffding's inequality:

$$\mathbb{P}_\mu(L_n(u) > \mu(u) + r) \leq \exp\left(-2n\frac{r^2}{\|u\|_{\mathrm{osc}}^2}\frac{1-\lambda_1^+}{1+\lambda_1^+}\right), \quad r > 0, n \geq 1, \tag{2.28}$$

which is generalized to general state space reversible Markov chains by Miasojedow [39] and Fan et al. [47]; please refer to them for the detailed proof. That implies the transport-information inequality (2.22) by Theorem 2.4[$(c') \Longrightarrow (a)$] (taking the discrete metric $d(x, y) = 1_{x\neq y}$). In other words part (a) in Theorem 2.14 is a reformulation of their result. Let $\mathcal{V} = \{(u, u) \in (b\mathcal{B})^2; \|u\|_{\mathrm{osc}} \leq 1\}$ and $d(x, y) = \mathbf{1}_{x\neq y}$ (the discrete metric). Then

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

Hence (2.22) is exactly the inequality ($W_1 I(C)$) with $C = \frac{1+\lambda_1^+}{4(1-\lambda_1^+)}$. The concentration inequality (2.23) is a direct consequence of (2.22) by Corollary 2.6.

The part (b) above generalizes the result of León and Perron [30], Theorem 1, to the nonsymmetric case. Miasojedow [39] and Fan et al. [47] already generalized the result of León-Perron to the nonsymmetric case, but only for $N = 1$. We mention that Paulin [44, 45] generalized the Bernstein-type inequality of Lezaud in the nonsymmetric case. Our proof, based on part (a), will be very short and completely different from those. The converse part (c) is new and completes this long history on Hoeffding's inequality of Markov chains under the spectral gap condition.

(ii) In the i.i.d. case, as $\lambda_1(P) = 0$, (2.22) above yields

$$\|\nu - \mu\|_{\mathrm{TV}}^2 \le 2H(\nu|\mu)$$

which is the famous CKP (Csiszär–Kullback–Pinsker) inequality.

(iii) If one imposes the Dobrushin contraction condition as in Marton [36]

$$\frac{1}{2}\|P(x,\cdot) - P(y,\cdot)\|_{\mathrm{TV}} \le r_D < 1 \quad \forall x, y \in \mathcal{X}$$

by [36] we have for any u with $\|u\|_{\mathrm{osc}} \le 1$,

$$\mathbb{P}_\beta\big(L_n(u) > \mathbb{E}_\beta L_n(u) + r\big) \le \exp\big(-2n(1 - r_D)^2 r^2\big)$$

which, with $\beta = \mu$, implies by Theorem 2.4.

$$\frac{1}{2}(1 - r_D)^2\|\nu - \mu\|_{\mathrm{TV}}^2 \le I(\nu). \tag{2.29}$$

Let us compare (2.22) and (2.29) in the symmetric case. Since $\|P^n u\|_{\mathrm{osc}} \le r_D^n \|u\|_{\mathrm{osc}}$, by [62], Lemma 5.4, the spectrum of P in $L^2(\mu)$ is contained in $[-r_D, r_D]$. Thus $\lambda_1^+ \le r_D$ and $\frac{1+\lambda_1^+}{1-\lambda_1^+} \le \frac{1+r_D}{1-r_D} \le \frac{1}{(1-r_D)^2}$. Hence the inequality (2.22) is better than (2.29), especially when r_D is close to 1.

The condition "$r_D < 1$" implies the L^2-contraction condition (2.24) with $N = 1$, $\delta^2 = r_D$ in the nonsymmetric case, by applying Del Moral et al. [8] to P. Thus for any u with $\|u\|_{\mathrm{osc}} \le 1$,

$$\mathbb{P}_\beta\big(L_n(u) > \mu(u) + r\big) \le \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-n\frac{1 - r_D}{1 + r_D} r^2\right), \tag{2.30}$$

where the constant in the exponential is better than the one in Marton's concentration inequality above when $r_D > 1/\sqrt{2}$, but worse otherwise.

The condition $r_D < 1$ is much stronger than the L^2-contraction condition (2.24). For example, let (P_t) be the standard *Ornstein–Uhlenbeck* semigroup generated by $\Delta - x \cdot \nabla$, for $P = P_1$, (2.24) is satisfied with $\delta = e^{-1}$ but $r_D = 1$. The advantage of Marton's result is that her transport-entropy inequality is in process level, which implies the concentration inequality for all nonlinear Lipschitzian functionals (rather than the only additive functionals, here).

It may be amusing to give:

A DIRECT PROOF OF (2.29). For any $\nu \in M_1(\mathcal{X})$, we have

$$\|\nu - \mu\|_{\mathrm{TV}} \le \|\nu - \nu P\|_{\mathrm{TV}} + \|\nu P - \mu P\|_{\mathrm{TV}} \le \|\nu - \nu P\|_{\mathrm{TV}} + r_D\|\nu - \mu\|_{\mathrm{TV}}$$

then by the CKP inequality,

$$(1 - r_D)\|\nu - \mu\|_{\mathrm{TV}} \le \|\nu - \nu P\|_{\mathrm{TV}} \le \sqrt{2H(\nu|\nu P)} \le \sqrt{2I(\nu|P, \mu)},$$

where the last inequality follows by (2.3). $\square$

2.6. *Lipschitzian spectral gap criteria.* The carré-du-champs operator associated with the generator $\mathcal{L} = P - I$ is

$$\Gamma(G)(x) = \frac{1}{2}\int \big(G(y) - G(x)\big)^2 P(x, dy)$$
$$= \frac{1}{2}\{\mathrm{Var}_{P(x,\cdot)}(G) + \big(PG(x) - G(x)\big)^2\}.$$

Besides the given metric d, consider another metric ρ such that:

$$\sup_{G:\|G\|_{\mathrm{Lip}(\rho)}=1} \big\|\sqrt{\Gamma(G)}\big\|_\infty \le M. \tag{2.31}$$

So the constant M bounds the spread rate of the Markov chain viewed from ρ-Lipschitzian functions G with $\|G\|_{\mathrm{Lip}(\rho)} := \sup_{x\neq y} \frac{|G(x)-G(y)|}{\rho(x,y)} = 1$ (while the Lipschitzian norm $\|G\|_{\mathrm{Lip}}$ is taken always w.r.t. the metric d). We denote by $C_{\mathrm{Lip}}(\mathcal{X}, \rho)$ (resp. $C_{\mathrm{Lip},0}(\mathcal{X}, \rho)$) the space of all ρ-Lipschitzian functions G (resp. with moreover $\mu(G) = 0$). When $\rho = d$, we write simply $C_{\mathrm{Lip}}(\mathcal{X}) := C_{\mathrm{Lip}}(\mathcal{X}, d)$, $C_{\mathrm{Lip},0}(\mathcal{X}) := C_{\mathrm{Lip},0}(\mathcal{X}, d)$.

Notice that (2.31) is satisfied once if

$$\frac{1}{2}\int \rho^2(x, y)P(x, dy) \le M^2 \quad \forall x \in \mathcal{X}. \tag{2.32}$$

Introduce

$$I_c(\nu) := \begin{cases} \langle\sqrt{f}, (I - P)\sqrt{f}\rangle_\mu = \langle\sqrt{f}, \big(I - P^\sigma\big)\sqrt{f}\rangle_\mu & \text{if } \nu \ll \mu,\ f := \dfrac{d\nu}{d\mu}, \\ +\infty & \text{otherwise} \end{cases} \tag{2.33}$$

which is the Donsker–Varadhan–Fisher information of the continuous time symmetric Markov process generated by $P^\sigma - I := \frac{P+P^*}{2} - I$ ([26]).

THEOREM 2.16. *Assume that P is μ-symmetric, and ρ is a metric on $\mathcal{X}$ satisfying* (2.31) *and $\int [d^2(x, x_0) + \rho^2(x, x_0)]\, d\mu(x) < +\infty$, where $x_0 \in \mathcal{X}$ is some fixed point. Suppose that the Poisson operator $(I - P)^{-1}$ is bounded from $C_{\mathrm{Lip},0}(\mathcal{X}, d)$ to $C_{\mathrm{Lip},0}(\mathcal{X}, \rho)$, that is, there is some best constant $c_{\mathrm{Lip}}(d, \rho) < +\infty$ such that for any bounded $g \in C_{\mathrm{Lip},0}(\mathcal{X}, d)$, there is $G \in C_{\mathrm{Lip},0}(\mathcal{X}, \rho)$ solving the Poisson equation $(I - P)G = g$, μ-a.s. and satisfying*

$$\|G\|_{\mathrm{Lip}(\rho)} \le c_{\mathrm{Lip}}(d, \rho)\|g\|_{\mathrm{Lip}(d)}. \tag{2.34}$$

Then

$$W_1(\nu, \mu)^2 \le 4\big(Mc_{\mathrm{Lip}}(d, \rho)\big)^2 I_c(\nu) \le 4\big(Mc_{\mathrm{Lip}}(d, \rho)\big)^2 I(\nu) \quad \forall \nu \in \mathcal{M}_1(\mathcal{X}). \tag{2.35}$$

In particular for any bounded Lipschitzian function g, and $n, r > 0$,

$$\mathbb{P}_\beta\big(\tilde{L}_n(g) > \mu(g) + r\big) \le \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-\frac{nr^2}{4(Mc_{\mathrm{Lip}}(d, \rho)\|g\|_{\mathrm{Lip}})^2}\right). \tag{2.36}$$

The inequality in the result above may be sharp, as seen from the discrete cube model in Section 3. We present at first a geometric application (a counterpart of Bonnet–Myers diameter theorem).

COROLLARY 2.17. *Assume the μ-symmetry of P, the condition* (2.31) *and $c_{\mathrm{Lip}}(d, \rho) < +\infty$. Assume also that μ charges all nonempty open subsets of $\mathcal{X}$. Then the diameter* $\mathrm{Diam}(\mathcal{X}) := \sup_{x,y\in\mathcal{X}} d(x, y)$ *of $\mathcal{X}$ is bounded and*

$$\mathrm{Diam}(\mathcal{X}) \le 4Mc_{\mathrm{Lip}}(d, \rho). \tag{2.37}$$

As a consequence of this Corollary, any reversible ergodic Markov chain of bounded range (i.e., the diameter of the topological support of $P(x, dy)$ is bounded by some constant R for all x) on a unbounded space $(\mathcal{X}, d)$ has *no Lipschitzian spectral gap, that is*,

$$c_{\mathrm{Lip}} := c_{\mathrm{Lip}}(d, d) = +\infty.$$

PROOF. For any $l < \mathrm{Diam}(\mathcal{X})$, we can find $\delta > 0$ and $x_1, x_2 \in \mathcal{X}$ such that $d(x_1, x_2) > l + 2\delta$. Since μ charges all nonempty open subsets, then consider the balls $B(x_i, \delta) = \{x \in \mathcal{X}; d(x, x_i) < \delta\}$, $i = 1, 2$ and the probability measures

$$\nu_i = \frac{\mathbf{1}_{B(x_i,\delta)}}{\mu(B(x_i, \delta))}\mu, \quad i = 1, 2.$$

We see clearly that $W_1(\nu_1, \nu_2) \ge l$. On the other hand for any μ-probability density f,

$$I_c(f\mu) = \langle\sqrt{f}, (I - P)\sqrt{f}\rangle_\mu \le \mu(f) = 1,$$

hence by (2.35) in Theorem 2.16, we have

$$l \le W_1(\nu_1, \nu_2) \le W_1(\nu_1, \mu) + W_1(\nu_2, \mu) \le 4Mc_{\mathrm{Lip}}(d, \rho).$$

As $l < \mathrm{Diam}(\mathcal{X})$ is arbitrary, the desired result follows. □

Another application of Theorem 2.16 is:

COROLLARY 2.18 (Bernstein-type inequality). *In the context of Theorem 2.16, assume moreover that $c_P < +\infty$ (i.e., the spectral gap exists). Then for any bounded d-Lipschitzian function g with $\|g\|_{\mathrm{Lip}(d)} = 1$,*

$$\begin{aligned} \nu(g) - \mu(g) &\le \sqrt{I_c(\nu)\big[2V_c(g) + 4\big[Mc_{\mathrm{Lip}}(d, \rho)\big]^2\sqrt{c_P I_c(\nu)}\big]} \\ &\le \sqrt{I(\nu)\big[2V_c(g) + 4\big[Mc_{\mathrm{Lip}}(d, \rho)\big]^2\sqrt{c_P I(\nu)}\big]} \quad \forall \nu \in M_1(\mathcal{X}), \end{aligned} \tag{2.38}$$

where

$$V_c(g) = 2\langle g - \mu(g), G\rangle_\mu = \sigma^2(g) + \mathrm{Var}_\mu(g), \tag{2.39}$$

where $\sigma^2(g)$ is the asymptotic variance in the central limit theorem, given by (2.14). Equivalently (by Theorem 2.4 with $\mathcal{V} = \{(g, g)\}$) for all $n \ge 1$, $r > 0$,

$$\mathbb{P}_\beta\big(\tilde{L}_n(g) > \mu(g) + \sqrt{r\big[2V_c(g) + 4\big[Mc_{\mathrm{Lip}}(d, \rho)\big]^2\sqrt{c_P r}\big]}\big) \le \left\|\frac{d\beta}{d\mu}\right\|_2 e^{-nr}. \tag{2.40}$$

The following result is inspired by Djellout et al. [12], Theorem 2.11, and based on the Lyons–Meyer–Zheng forward-backward martingale decomposition developed in [59].

PROPOSITION 2.19. *Assume the μ-symmetry of P. Suppose that $(I - P)^{-1}$ is bounded from $C_{\mathrm{Lip},0}(X, d)$ to $C_{\mathrm{Lip},0}(X, \rho)$ with the norm denoted by $c_{\mathrm{Lip}}(d, \rho)$, and $P(x, \cdot)$ satisfies the transport-entropy inequality w.r.t. the metric ρ uniformly over x, that is, there is a positive constant $c_H(P, \rho)$ such that for all $x \in \mathcal{X}$, $\nu \in M_1(\mathcal{X})$,*

$$W_{1,\rho}^2\big(\nu, P(x, \cdot)\big) \le 2c_H(P, \rho)H\big(\nu|P(x, \cdot)\big). \tag{2.41}$$

Then

$$W_1^2(\nu, \mu) \le 2\big(c_{\mathrm{Lip}}(d, \rho)\big)^2 c_H(P, \rho)I(\nu) \quad \forall \nu \in M_1(\mathcal{X}) \tag{2.42}$$

or equivalently for any bounded d-Lipschitzian function g with $\|g\|_{\mathrm{Lip}(d)} = 1$ and all $n \ge 1$, $r > 0$,

$$\mathbb{P}_\beta\big(\tilde{L}_n(g) > \mu(g) + r\big) \le \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-n\frac{r^2}{2(c_{\mathrm{Lip}}(d, \rho))^2 c_H(P, \rho)}\right). \tag{2.43}$$

Let us make some comments on the results above and related works.

REMARK 2.20. Let us compare Proposition 2.19 with Djellout et al. [12], Theorem 2.11, for dependent tensorization of transport-entropy inequality. The Lipschitzian spectral gap condition here for P is weaker than the coupling condition there in the symmetric case (just look at periodic Markov chains); but the Hoeffding's concentration inequality in Proposition 2.19 holds only for additive functionals.

REMARK 2.21. The W_1H condition (2.41) on P is imposed in almost all known works for the Gaussian concentration inequality of Markov chains since Marton [37] and Djellout et al. [12]. It is equivalent to the exponential integrability of $\rho^2(y_1, y_2)$ under $P(x, dy_1)P(x, dy_2)$ by [12]. But for the Gaussian concentration inequality (2.36) in Theorem 2.16, only the uniform square integrability condition (2.32) of $\rho(x, y)$ under $P(x, dy)$ is demanded. That is a little bit curious even in the i.i.d. case. Take $\rho = d$, below. In the i.i.d. case, $c_{\mathrm{Lip}}(d, d) = 1$, and when $\mathcal{X}$ is a closed subset of $\mathbb{R}^N$ equipped with the Euclidean metric $d(x, y) = |x - y|$, since $\int |x - y|^2\, d\mu(y) = |x - \mu(y)|^2 + \int |y - \mu(y)|^2\, d\mu(y)$, the uniform square integrability condition (2.32) is equivalent to the boundedness of $\mathcal{X}$. In other words the uniform square integrability condition (2.32), so weak in appearance, is in fact a very strong condition in the i.i.d. case. This condition is not well adapted to the i.i.d. case, but well adapted for Markov chains: it means that starting from $X_0 = x$, $\mathbb{E}_x d^2(X_0, X_1)^2 \le 2M^2$, that is, X_1 should be not very far from x.

Therefore Theorem 2.16 is a complement to [12], Theorem 2.11, in the symmetric Markov chain case.

REMARK 2.22. Ollivier [42] called

$$\kappa(x, y) = 1 - \frac{W_1(P(x, \cdot), P(y, \cdot))}{d(x, y)}$$

Ricci curvature of the Markov chain (w.r.t. the metric d). If the Ricci curvature lower bound $\kappa := \inf_{x \ne y} \kappa(x, y) > 0$, then $\|P\|_{C_{\mathrm{Lip}}(X,d)} \le 1 - \kappa$. Using

$$(I - P)^{-1} = \sum_{n=0}^{\infty} P^n \quad \text{on } C_{\mathrm{Lip},0}(\mathcal{X}, d)$$

we see that $c_{\mathrm{Lip}}(d, d) \le \sum_{n=0}^{\infty} \|P^n\|_{C_{\mathrm{Lip}}(X,d)} \le \sum_{n=0}^{\infty} (1 - \kappa)^n \le 1/\kappa$, which can be also derived from the results of Paulin [46]. This is a simple way for estimating $c_{\mathrm{Lip}}(d, d)$. But there are many situations where the Ricci curvature lower bound $\kappa \le 0$, whereas $c_{\mathrm{Lip}}(d, d)$ is finite. For example if $\mathcal{X}$ is finite, $c_{\mathrm{Lip}}(d, \rho) < +\infty$ (by the theory of matrices); moreover any nearest-neighbor random walk (i.e., $P(x, y) > 0$ iff x, y are neighbors) on a finite connected sub-graph of $\mathbb{Z}^m$ (m here denotes the dimension) or on the discrete circle $\mathbb{Z}/(N\mathbb{Z})$ ($N \ge 4$), we see that $\kappa = 0$ (it is enough to look at $W_1(P(x, \cdot), P(y, \cdot))$ for two neighbors x, y).

Joulin and Ollivier [27] established some concentration inequalities for general Markov chains of positive Ricci curvature without the symmetry, improving the constant estimate in [12], Theorem 2.11, for non-large deviations. As our constant M uses only the Lipschitzian functions (without the bounded range condition in [27]), Theorem 2.16 generalizes theirs, but only in the symmetric case.

Notice that if $d(x, y) = 1_{x \ne y}$, $\|P\|_{C_{\mathrm{Lip},0}(X,d)}$ coincides with the Dobrushin contraction coefficient r_D in Remark 2.15(iii), and (2.43) yields (2.29).

REMARK 2.23. About the *Bonnet–Myers* diameter theorem, Ollivier proved in [42], Prop. 23, with an argument of one line

$$\mathrm{Diam}(\mathcal{X}) \le \frac{2\sup_x W_1(\delta_x, P(x,\cdot))}{\kappa}.$$

See also Lin, Lu and Yau [33], Theorem 4.1, for a similar bound. We compare this result with ours in Corollary 2.17 on a connected and at most countable graph $\mathcal{X}$ equipped with the graph distance $d = d_G$. If (X_n) is a nearest-neighbor random walk, that is, $P(x, y) > 0$ iff $d_G(x, y) = 1$, we see that the condition (2.31) is satisfied for $\rho = d_G$ with $M = 1/\sqrt{2}$. Thus by Corollary 2.17,

$$\mathrm{Diam}(\mathcal{X}) \le 2\sqrt{2} c_{\mathrm{Lip}}(d_G, d_G)$$

which is not greater than $2\sqrt{2}/\kappa$ if $\kappa > 0$. So in the positive curvature case the result of Ollivier [42], Prop. 23, above is better. Our result in terms of $c_{\mathrm{Lip}}(d_G, d_G)$ generalizes their result to the case where κ may be not positive, but only in the symmetric case.

REMARK 2.24. For symmetric P, we have always $c_P \le c_{\mathrm{Lip}}(d, d)$ (by [62], Lemma 5.4). This is an important tool for estimating the Poincaré constant c_P, see the second named author [64] for diffusions on Riemannian manifolds, Liu and Ma [34] for birth-death processes, Djellout and the second named author [13] for one-dimensional diffusions.

REMARK 2.25. In the range of moderate deviations, that is,

$$1/\sqrt{n} \ll t = \sqrt{r\big[2V_c(g) + 4\big[M c_{\mathrm{Lip}}(d, \rho)\big]^2 \sqrt{c_P r}\big]} \ll 1$$

our Bernstein-type inequality (2.40) in Corollary 2.18 gives a Gaussian upper bound $\exp(-nt^2/2[V_c(g) + \varepsilon])$. But as $V_c(g) = \sigma^2(g) + \mathrm{Var}_\mu(g) > \sigma^2(g)$ if g is not constant, it is not sharp in moderate deviations. It is however better than the Gaussian concentration inequality in Theorem 2.16 for small $r > 0$, because $V_c(g) \le 2(M c_{\mathrm{Lip}}(d, \rho))^2$, by Guillin et al. [26]. We mention that recently Paulin [45] proved the Bernstein-type inequality with $\sigma^2(g) + 0.8\,\mathrm{Var}_\mu(g) < V_c(g)$ in place of $V_c(g)$, in the discrete metric $d(x, y) = 1_{x\neq y}$ case. However for large deviation $t > 0$, the upper bound in (2.40) becomes $O(\exp(-nct^{4/3}))$, better than the exponential-type bound e^{-nct} in the usual Bernstein inequality. This curious phenomenon, pointed out by a referee, is not strange: for large deviation t, even the Gaussian bound $\exp(-nct^2)$ holds by Theorem 2.16.

How to obtain the sharp Bernstein inequality in the discrete time Markov chains (as in the i.i.d. or continuous time reversible Markov processes [19]) remains an open question.

2.7. *Transport-information inequalities $T_\mathcal{V} I$ under hyperboundedness or ultra-boundedness.* What is the counterpart of the log-Sobolev inequality in the discrete time Markov chain case? By Gross' Theorem the reader can naturally guess that it is the hypercontractivity, that is, for some $p > 2$, $\|P\|_{2,p} = \sup_{f:\|f\|_2\le 1} \|Pf\|_p = 1$. The question is: in many cases, one can check easily that $\|P\|_{2,p} < +\infty$ (i.e., P is hyperbounded on $L^2(\mu)$), but the hypercontractivity is much more difficult to verify. A practical criterion for bounding $\|P\|_{2,p}$ is: if $P(x, dy) = p(x, y)\mu(dy)$ is an absolutely continuous transition kernel, then

$$\|P\|_{2,p} \le \left[\int \left(\int p(x, y)^2\, d\mu(y)\right)^{p/2} d\mu(x)\right]^{1/p}, \quad p \ge 2$$

which is a direct consequence of Hölder's inequality. The finiteness of the right hand side above implies that P is of Hilbert–Schmidt type.

Note that in the actual absolute continuous case, the spectral gap exists (i.e., $c_P < +\infty$) once P is hyperbounded, by the uniform integrability criterion in [61].

Recall that in the continuous time case, the hyperboundedness of transition semigroup is equivalent to the defective log-Sobolev inequality, which can be tightened in terms of the Poincaré inequality by Rothaus lemma.

THEOREM 2.26. *Without the symmetry of P, assume that P is hyperbounded, that is, $\|P\|_{2,p} := \|P\|_{L^2(\mu)\to L^p(\mu)} < \infty$ for some $p \in (2, +\infty)$. Then the following defective log-Sobolev inequality holds*:

$$H(\nu|\mu) \le \frac{2p}{p-2}\big[I^\sigma(\nu) + \log\|P\|_{2,p}\big] \quad \forall \nu \in M_1(\mathcal{X}). \tag{2.44}$$

In particular

(a) (Hypercontractivity) *If $\|P\|_{2,p} = 1$ and $\alpha(T_{\mathcal{V}}(\nu,\mu)) \le H(\nu|\mu)$ for all $\nu \in M_1(\mathcal{X})$ (the (α-$T_{\mathcal{V}}H$) inequality holds), then*

$$\alpha\big(T_{\mathcal{V}}(\nu,\mu)\big) \le \frac{2p}{p-2} I^\sigma(\nu) = \frac{p}{p-2} I\big(\nu|PP^*,\mu\big) \quad \forall \nu \in M_1(\mathcal{X}).$$

In particular for any $u : \mathcal{X} \to \mathbb{R}$ such that $\mu(u) = 0$, $|u| \le 1$, the following Bennett-type inequality holds: for all $N, n \ge 1$, $r > 0$,

$$\mathbb{P}_\beta\big(L_n(u)\circ\theta^N > r\big) \le \left\|\frac{d\beta}{d\mu}\right\|_{q_N} \exp\left(-n\frac{(p-2)r}{4p}\log\left(1+\frac{r}{\mu(u^2)}\right)\right), \tag{2.45}$$

where $q_N := \frac{1}{1-(2^{N-1}/p^N)}$ is the conjugate number of $p_N = p^N/2^{N-1}$.

(b) (Hyperboundedness) *Assume moreover that for some $\gamma \in (0,1)$,*

$$\|Pu\|_2 \le \gamma\|u\|_2 \quad \forall u \in L_0^2(\mu) := \{f \in L^2(\mu); \mu(f) = 0\} \tag{2.46}$$

that is, P is contractive in $L_0^2(\mu)$. If $g \in L^1(\mu)$ satisfies the Gaussian integrability condition:

$$b(\delta) := \frac{1}{\delta}\log\int \exp\{\delta\big(g-\mu(g)\big)^2\}\,d\mu < +\infty$$

for some $\delta > 0$, then for $\mathcal{V} = \{(g,g)\}$,

$$T_{\mathcal{V}}(\nu,\mu) = \nu(g) - \mu(g) \le \sqrt{2C_G(g)I^\sigma(\nu)} = \sqrt{C_G(g)I\big(\nu|PP^*\big)}, \tag{2.47}$$

where the sub-Gaussian constant $C_G(g)$ is bounded by

$$C_G(g) \le \frac{2p}{\delta(p-2)} + \frac{2}{1-\gamma^2}\left\{\mathrm{Var}_\mu(g) + b(\delta) + \frac{2p}{\delta(p-2)}\log\|P\|_{2,p}\right\}.$$

In particular for all $N, n \ge 1$, $r > 0$,

$$\mathbb{P}_\beta\big(L_n(g)\circ\theta^N > \mu(g) + r\big) \le \left\|\frac{d\beta}{d\mu}\right\|_{q_N} \|P\|_{2,p}^{\frac{p}{p-2}} \exp\left(-n\frac{r^2}{2C_G(g)}\right),$$

where q_N is given in Part (a).

Furthermore, if for some $\delta > 0$,

$$L(\delta) := \frac{1}{\delta}\log\int\exp\left\{\delta\left(\int d(x,y)\,d\mu(y)\right)^2\right\}d\mu(x) < +\infty \tag{2.48}$$

then

$$W_1(\nu,\mu)^2 \le 2C_G^\sigma I^\sigma(\nu) = C_G^\sigma I\big(\nu|PP^*,\mu\big) \quad \forall \nu \in M_1(\mathcal{X}), \tag{2.49}$$

where the (symmetrized) sub-Gaussian constant C_G^σ satisfies

$$
\begin{aligned}
C_G^\sigma \le{} & \frac{2p}{\delta(p-2)} \\
& + \frac{2}{1-\gamma^2}\left\{\frac{1}{2}\iint d^2(x,y)\,d\mu(x)\,d\mu(y) + L(\delta) + \frac{2p}{\delta(p-2)}\log\|P\|_{2,p}\right\}.
\end{aligned}
\tag{2.50}
$$

In the theorem above, as q_N approaches to 1 exponentially fast in N, $\|d\beta/d\mu\|_{q_N}$ approaches to $\|d\beta/d\mu\|_1 = 1$ also exponentially fast as N goes to infinity if $d\beta/d\mu \in L^{1+\delta}(\mu)$ for some $\delta > 0$. That improves the factor constant $\|d\beta/d\mu\|_2$ when this latter term is too big in practice.

When $\|P\|_{2,\infty} < +\infty$, P is said to be *ultrabounded*. It is related to the inequalities of Sobolev, Nash, Gagliardo–Nirenberg, etc., in the continuous time case; see Saloff–Coste [49]. If $\|P\|_{2,\infty} < +\infty$, then $P(x,dy) = p(x,y)\mu(dy)$ for μ-a.e. x (absolutely continuous) and

$$
\|P\|_{2,\infty} = \operatorname*{esssup}_{x\in\mathcal{X}}\left(\int p(x,y)^2\,d\mu(y)\right)^{1/2}
$$

(see Wang [56]). By Jensen's inequality if (X_n) is not i.i.d. under $\mathbb{P}_\mu$, then $\|P\|_{2,\infty} > 1$.

THEOREM 2.27 (Ultraboundedness). *Assume that $\|P\|_{2,\infty} < +\infty$, P is contractive on $L_0^2(\mu)$, that is, verifies (2.46), and the metric d satisfies the Gaussian integrability condition (2.48). Then for any d-Lipschitzian function g with $\|g\|_{\mathrm{Lip}} = 1$, any initial probability measure $\beta \ll \mu$ and for all $n, r > 0$,*

$$
\mathbb{P}_\beta\big(L_n(g)\circ\theta > \mu(g) + r\big) \le \|P\|_{2,\infty}\exp\left(-n\frac{r^2}{2C_G^\sigma}\right),
\tag{2.51}
$$

where C_G^σ is some positive constant bounded by the limit of the right hand side of (2.50) as $p \to \infty$.

REMARK 2.28. The inequality (2.44) is exactly the counterpart of the usual defective log-Sobolev inequality in the discrete time case. For applications of log-Sobolev inequality in concentration, Otto and Villani [43] proved that the log-Sobolev inequality implies Talagrand's W_2H (or T_2) inequality (for generalizations, see Bobkov, Gentil and Ledoux [1] for an ingenious approach basing on Hamilton–Jacobi's equation and Wang [54]). The second named author [60] showed that log-Sobolev inequality implies the concentration inequality of the empirical mean as in the i.i.d. case, and Gao et al. [19] established the sharp Bernstein inequality, both for continuous time Markov processes.

REMARK 2.29. We have seen in Proposition 2.10 that the transport-information inequality ($\alpha - T_\mathcal{V} I^\sigma$) is stronger than the transport-entropy inequality (α-$T_\mathcal{V} H$). Hence the Gaussian integrability condition (2.48) of the metric d is *necessary* to the $W_1 I^\sigma$ inequality (2.49), by Djellout et al. [12], Theorem 2.3, for the transport-entropy inequality ($W_1H(C)$).

Part (a) of Theorem 2.26 shows that the converse of Proposition 2.10 is true under the hypercontractivity, and it says that the Markov chain has the same type concentration inequalities as in the i.i.d. case, extending the same result of [60] from the continuous time to discrete time. For applying (a) above, the reader is referred to Djellout et al. [12], Bolley and Villani [4] and Gozlan and Léonard [22, 23] for numerous known results on W_pH and $T_\mathcal{V} H$.

REMARK 2.30. There are few criteria for the hypercontractivity of a single kernel P (see however Remark below), but there are many for that of semigroups $(P_t)_{t\geq 0}$ by means of the log-Sobolev inequality. In the latter continuous time situation, one approaches $\mu(u)$ by the empirical mean $(1/T)\int_0^T u(X_t)\,dt$. This last object is not what is used numerically: one may use $(1/n)\sum_{k=0}^{n-1} u(X_{kh})$ where $h > 0$ is small and $nh = T$. To this last process $(X_{nh})_{n\geq 0}$ our result above applies and yields sharp concentration inequalities.

REMARK 2.31. If $\|P^m\|_{2,p} = 1$ for some $m \geq 1$, then we can apply part (a) above to P^m for obtaining the concentration inequality by Theorem 2.12. F. Y. Wang [56], Theorem 5.2.1, showed that $\lambda_1(P^*P) < 1$ and $\|P^N\|_{2,4} = 1$ for some $N \geq 1$ once if $\|P\|_{2,4}^4 < 2$ without the absolute continuity condition.

Furthermore, if $\|P^k\|_{2,p} < +\infty$ only for some $k \geq 1$ (i.e., the case for the Markov chain $(Y_n = (X_n, X_{n+1}))$ even if P is hyperbounded), the spectral gap of P^k (i.e., 1 is an isolated eigenvalue in the spectrum of P^k on $L^2(\mu)$) exists by L. Miclo [40] and F. Y. Wang [55], so for P. If P is moreover aperiodic, we will get $\|P^m - \mu\|_2 \leq Ce^{-\delta m}$ for some positive constants C, δ and for all $m \geq 1$. Then for some $m \geq k$, $\|P^m\|_{L_0^2(\mu)} < 1$ and $\|P^m\|_{2,p} \leq \|P^k\|_{2,p} < +\infty$. One can apply part (b) above to P^m instead of P to obtain the transport-information inequality of type (2.20) with quadratic α and then derive the Gaussian concentration inequality by Theorem 2.12.

REMARK 2.32. It is well known that the log-Sobolev inequality implies the Poincaré inequality. In the discrete time case, if $\|P\|_{2,p} = 1$ for some $p > 2$, by the CKP inequality and (2.44), we have

$$\|\nu - \mu\|_{\mathrm{TV}}^2 \leq 2H(\nu|\mu) \leq \frac{2p}{p-2} I(\nu|PP^*, \mu)$$

which implies by Theorem 2.14(c)

$$\lambda_1(PP^*) \leq \exp\left(-\frac{p-2}{2p}\right).$$

2.8. *Lyapunov function criterion*. Djellout et al. [12] proved that μ satisfies the W_1H-inequality $W_1(\nu, \mu)^2 \leq 2CH(\nu|\mu)$, $\forall \nu \in M_1(\mathcal{X})$ if there is some constant $\delta > 0$ such that

$$\int e^{\delta d(x,x_0)^2}\, d\mu(x) < +\infty \tag{2.52}$$

and later Bolley and Villani [4] obtained the following better estimate of W_1H-constant C under (2.52):

$$C = \inf_{\delta>0, x_0\in\mathcal{X}} \frac{1}{\delta}\left(1 + \log\int e^{\delta d(x,x_0)^2}\, d\mu(x)\right). \tag{2.53}$$

For progresses on this subject, see E. Milman [41] for the corresponding dimension free estimate.

At first glance this criterion is quite far from the well-known Lyapunov function method for the ergodicity of Markov chains (see Meyn and Tweedie [38]). But as H is just the Donsker–Varadhan's entropy for the i.i.d. sequence, we will show that this is indeed a special case of a general Lyapunov function method.

We state the Lyapunov condition for geometric ergodicity as follows:

(H) There exist a measurable function $U : \mathcal{X} \to [1, +\infty)$, a nonnegative function ϕ and a constant $b > 0$ such that $PU(x) < +\infty$, μ-a.s. and

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

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 ϕ such that $\inf_{\mathcal{X}\setminus C}\phi > 0$ in (H) is equivalent to the geometric ergodicity of the Markov chain ([38]), or the Poincaré inequality (2.21) in the symmetric case ([62]).

THEOREM 2.33. *Assume that P is μ-symmetric and satisfies Poincaré inequality* (2.21) *with best constant $c_{\mathrm{P}} < \infty$ and the Lyapunov condition (H) holds. Suppose moreover that $\phi \in L^1(\mu)$. Then*

$$\|\sqrt{\phi}(\nu-\mu)\|_{\mathrm{TV}}^2 \le 2CI(\nu) \quad \forall \nu \in M_1(\mathcal{X}); \qquad C = 2[1 + c_P(\mu(\phi)+b)]. \tag{2.54}$$

In particular, if there are $x_0 \in \mathcal{X}$ and $\delta > 0$ such that $\delta d(x,x_0)^2 \le \phi$, $\forall x \in \mathcal{X}$, then

$$W_1(\nu,\mu)^2 \le 2\tilde{C}I(\nu) \quad \forall \nu \in M_1(\mathcal{X}); \qquad \tilde{C} := \frac{2}{\delta}[1 + c_P(\mu(\phi)+b)]. \tag{2.55}$$

REMARK 2.34. In the i.i.d. case, under condition (2.52), choose $U := \exp(\delta d^2(x,x_0))$, we have

$$\log\frac{U}{PU} = \delta d^2(x,x_0) - \log\int e^{\delta d^2(x,x_0)}\,d\mu(x).$$

In other words (H) holds with $b := \log\int e^{\delta d^2(x,x_0)}\,d\mu(x)$ and $\phi = \delta d^2(x,x_0)$. This condition (H) is exactly the necessary and sufficient condition in Djellout et al. [12] for W_1H-inequality. Moreover as $c_P = 1$, (2.55) is read as $W_1(\nu,\mu)^2 \le 2\tilde{C}H(\nu|\mu)$, $\forall \nu \in M_1(\mathcal{X})$ where

$$\tilde{C} := \frac{2}{\delta}\left[1 + \delta\mu(d^2(\cdot,x_0)) + \log\int e^{\delta d^2(x,x_0)}\,d\mu(x)\right].$$

This constant is slightly worse than that of Bolley–Villani's (2.53).

We did not manage to generalize another transport inequality in Bolley and Villani [4], although we tried our best. In the actual Makov chains setting, it may be read as: in the framework of Theorem 2.33,

$$\|\phi(\nu-\mu)\|_{\mathrm{TV}} \le C_1\sqrt{I(\nu)} + C_2 I(\nu) \quad \forall \nu \in M_1(\mathcal{X}).$$

That was successfully proved in [26] for continuous time symmetric Markov processes. The main difficulty in extending such results outside of the cases of i.i.d. or continuous time symmetric Markov processes resides in the fact that $I(\nu) = I(\nu|P,\mu)$ has no longer closed expression.

The explicit constants in the inequalities of this theorem, produced by the Lyapunov function condition (H), are in general far from being optimal, but may be correct in order, as seen in the actual i.i.d. case.

REMARK 2.35. Since $\|\psi(\nu-\mu)\|_{\mathrm{TV}} = \sup_{u:|u|\le\psi}\int u\,d(\nu-\mu)$ for any $\psi \ge 0$, the inequality (2.54) in this theorem may be regarded as $T_{\mathcal{V}}I$ in Theorem 2.4 with $\mathcal{V} = \{(u,u); u \in b\mathcal{B}, |u| \le \phi^{1/2}\}$. Furthermore as noticed by Gozlan and Léonard [22], $\|\psi(\nu-\mu)\|_{\mathrm{TV}} = W_{1,d_\psi}(\nu,\mu)$ where $d_\psi(x,y) = 1_{x\neq y}(\psi(x)+\psi(y))$, we can also interpret (2.54) as a $(W_1I(C))$-inequality.

REMARK 2.36. As

$$W_1(\nu,\mu) \le \inf_{x_0\in\mathcal{X}}\|d(\cdot,x_0)(\nu-\mu)\|_{\mathrm{TV}} \tag{2.56}$$

(see [52]), we see that the W_1I-inequality (2.55) is a direct consequence of (2.54).

2.9. *Comparison with the continuous case.* As the reader sees, the results of this paper in the discrete time case are quite similar to those in the continuous time case in [26]. In fact the main idea of this paper is to show that many approaches in continuous time work again in the discrete time.

But apart from the similarity in appearance, there are several important differences. At first the empirical mean $\tilde{L}_n$ of trapeze-type used here is different. The most important difference is that the information $I(\nu|P,\mu)$ has now no longer explicit expression: this creates many new difficulties. The third one is: functional inequalities are powerful tools to prove the transport-information inequalities in the continuous time case, but there are no counterparts of usual functional inequalities in the discrete time case (except the Poincaré inequality).

In the discrete time case, we did not manage to prove the tensorization of the $T_\mathcal{V} I$ inequality, although we tried our best. This important dimension-free property remains one main open question in our investigation.

The reader will see more clearly those differences between discrete and continuous time cases, in the examples and proofs below.

3. Several examples. Given (P,μ), in this section we apply the general results of Section 2 to a series of examples and our main purpose is to give estimates of the sub-Gaussian constant—the best (or least) constant $C_G = C_G(P,\mu)$ in the ($W_1 I(C)$) inequality

$$W_1^2(\nu,\mu) \le 2C_G I(\nu) \quad \forall \nu \in M_1(\mathcal{X}).$$

Assume that P is symmetric on $L^2(\mu)$, we note (by the inequality (2.15) and Proposition 2.10),

$$C_G \ge V(P,\mu), \quad \text{and} \quad C_G \ge C_H(\mu)/2 \quad \text{if } P \text{ is nonnegative definite,} \tag{3.1}$$

where $C_H(\mu)$ is the best constant in $W_1(\nu,\mu)^2 \le 2C_H(\mu)H(\nu|\mu)$ (transport-entropy inequality), and $V(P,\mu)$ is the maximal asymptotic variance of 1-Lipschitzian function u given in (2.15).

3.1. *Two points model.* We begin with the simplest Markov chain on $\mathcal{X} = \{0,1\}$ equipped with the discrete metric d (then $d(0,1)=1$), with transition matrix $P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$, where $a,b \in (0,1]$. Notice that P is symmetric w.r.t. μ given by

$$\mu(0) = \frac{b}{a+b} =: q, \qquad \mu(1) = \frac{a}{a+b} =: p.$$

Though this model is simple, but its study is abundant: for the Dirichlet form

$$\mathcal{E}_P(g,g) = \langle g,(I-P)g\rangle_\mu = \frac{ab}{a+b}\big(g(1)-g(0)\big)^2$$

associated with the continuous time Markov process generated by $\mathcal{L} = P - I$,

(1) the best log-Sobolev constant is known, see [49];

(2) the best constant $C_H(p)$ in $W_1(\nu,\mu)^2 \le 2C_H(p)H(\nu|\mu)$ is obtained by Bobkov, Houdré and Tetali [3]:

$$C_H(p) = \frac{p-q}{2(\log p - \log q)} \quad (:= 1/4 \text{ if } p = q); \tag{3.2}$$

(3) the best rate $\kappa > 0$ in the exponential entropy convergence $H(\nu P_t|\mu) \le e^{-\kappa t}H(\nu|\mu)$ is unknown, and only some accurate estimates are known, see Chen [6].

(4) the best constant C in $W_1(\nu|\mu)^2 \le 2CI_c(\nu)$ (where $I_c(\nu) = \langle (I-P)\sqrt{\frac{d\nu}{d\mu}}, \sqrt{\frac{d\nu}{d\mu}}\rangle_\mu$) is known: $C = 1/[2(a+b)]$ by the work [26].

In those works, it is enough to work with the Dirichlet form $\mathcal{E}^0(g,g) := (g(1)-g(0))^2$ and then essentially only $p = \mu(1)$ is important.

Since the second largest eigenvalue $\lambda_1(P)$ is equal to $1-(a+b)$, by the sharp Hoeffding's inequality of León and Perron [30], Proposition 1, for two states model and Corollary 2.6, the best sub-Gaussian constant C_G in $W_1(\nu|\mu)^2 \le 2C_G I(\nu)$ is given by

$$C_G \le \frac{2-(a+b)}{4(a+b)}, \quad \text{if } a+b \le 1; \qquad C_G \le \frac{1}{4}, \quad \text{if } a+b > 1. \tag{3.3}$$

Let us see what give the Lipschitzian spectral gap criteria. If $(I-P)G = g$ where $\mu(g) = 0$, then $G(1) - G(0) = (a+b)^{-1}(g(1)-g(0))$, so

$$c_{\mathrm{Lip}}(d,d) = c_P = (a+b)^{-1},$$

$$\sigma^2(g) = 2\langle G, g\rangle_\mu - \mathrm{Var}_\mu(g) = \left(\frac{2}{a+b} - 1\right)\frac{ab}{(a+b)^2}[g(1)-g(0)]^2$$

(however its Ricci curvature $\kappa = 1 - |1-(a+b)|$). The constant M appearing in the condition (2.32) equals to $\sqrt{(a\vee b)/2}$. Hence Theorem 2.16 gives us

$$C_G \le \frac{a\vee b}{(a+b)^2},$$

which is slightly worse. A sharp estimate of the $W_1 I$-constant is given by Proposition 2.19: since the best constant $c_H(P)$ appearing in the condition (2.41) is (by (3.2))

$$\begin{aligned} c_H(P) &= \max\{C_H(a), C_H(b)\} \\ &= \max\left\{\frac{a-(1-a)}{2(\log a - \log(1-a))}, \frac{b-(1-b)}{2(\log b - \log(1-b))}\right\}, \end{aligned}$$

and then

$$C_G \le \frac{\max\{C_H(a), C_H(b)\}}{(a+b)^2} \tag{3.4}$$

which becomes *equality* if $a+b=1$ (i.e., i.i.d. case, for $C_H(a) = C_H(b) = C_H(\mu)$).

By the calculation of $\sigma^2(g)$ and the lower bound in (3.1), we have

$$C_G \ge \left(\frac{2}{a+b} - 1\right)\frac{ab}{(a+b)^2}.$$

Notice that if $a = b = 1$, $c_H(P) = 0$ and then $C_G = 0$ by (3.4).

We do not know the exact expression of C_G.

Before moving further, let us notice the following consequence of the estimate (3.2) of Bobkov, Houdré and Tetali [3]:

LEMMA 3.1. *On an arbitrary probability space $(\mathcal{X}, \mathcal{B}, \mu)$ equipped with the discrete metric d, $C_H(\mu) = \sup_{A\in\mathcal{B}} C_H(\mu(A)) = C_H(p)$, where $p = \sup\{\mu(A); A \in \mathcal{B}, \mu(A) \le 1/2\}$.*

PROOF. At first for any $A \in \mathcal{B}$, restricting μ to $\sigma\{A\}$ (the σ-algebra generated by A), we have $C_H(\mu(A)) \le C_H(\mu)$. Conversely for any $\nu = f\mu$, letting $A := \{f > 1\}$, $\bar{f} := \mathbb{E}^\mu(f|\sigma\{A\})$ and $\bar{\nu} = \bar{f}\mu$, we have

$$W_1(\nu,\mu) = \frac{1}{2}\int |f-1|\,d\mu = \int_{f>1}(f-1)\,d\mu = \int_{\bar{f}>1}(\bar{f}-1)\,d\mu = W_1(\bar{\nu},\mu)$$

and now apply (3.2): the right-hand side (r.h.s. in short) above is not larger than

$$\sqrt{2C_H\big(\mu(A)\big)H(\bar{\nu}|\mu)} \le \sqrt{2C_H\big(\mu(A)\big)H(\nu|\mu)}$$

by the convexity of H. The desired result follows for $C_H(\mu(A)) = C_H(1-\mu(A))$. □

3.2. *Complete graph.* Let $\mathcal{X}$ be the complete graph of N (≥ 3) vertices, that is, any two vertices are connected by an edge and then the graph distance is given by $d(x,y) = 1_{x\neq y}$ (the discrete metric). The probability transition matrix is given by $P(x,y) = \frac{1}{N-1}$ for all $y \neq x$. It is symmetric w.r.t. the uniform measure $\mu(x) = 1/N$ (for each $x \in \mathcal{X}$). It is easy to see that

$$(I-P)G = \frac{N}{N-1}\big(G - \mu(G)\big).$$

Thus $c_P = \frac{N-1}{N} = c_{\mathrm{Lip}}(d,d)$. By the spectral gap criterion in Theorem 2.14, we have

$$C_G \le \frac{N-1}{2N}.$$

By Lemma 3.1, the W_1H constant $c_H(P)$ of $P(x,\cdot)$ in condition (2.41) is given as

$$c_H(P) = 1/4, \quad \text{if } N \text{ is odd;}$$

$$c_H(P) = \frac{1}{2(N-1)[\log N - \log(N-2)]} \le \frac{1}{4}, \quad \text{if } N \text{ is even.}$$

Proposition 2.19 yields the better

$$C_G \le \frac{1}{4}\cdot\left(\frac{N-1}{N}\right)^2. \tag{3.5}$$

Let us compare it with the best Gaussian constant $C_H(\mu)$ in the i.i.d. case. By Lemma 3.1, if N is even, $C_H(\mu) = 1/4$; and if N is odd,

$$C_H(\mu) = \frac{1}{2N(\log(N+1) - \log(N-1))} \ge \frac{N-1}{4N}.$$

Thus in both cases, $C_H(\mu) \ge C_G$. In other words on the complete graph, the Markov chain is more concentrated than the i.i.d. sequence.

The estimate (3.5) is sharp for large N. Indeed it is very easy to calculate the asymptotic variance of g from the expression of $(I-P)G$ above:

$$\sigma^2(g) = \frac{N-2}{N}\mathrm{Var}_\mu(g).$$

Now for even N, $\sup_{\|g\|_{\mathrm{Lip}}=1}\mathrm{Var}_\mu(g) = 1/4$; and for odd N,

$$\sup_{\|g\|_{\mathrm{Lip}}=1}\mathrm{Var}_\mu(g) = \big(1 - N^{-2}\big)/4,$$

we get

$$C_G \ge V(P,\mu) = \begin{cases} \dfrac{N-2}{4N} & \text{if } N \text{ is even;} \\ \dfrac{N-2}{4N}\left(1 - \dfrac{1}{N^2}\right) & \text{if } N \text{ is odd.} \end{cases}$$

3.3. *Discrete cube.* Let $\mathcal{X} = \{0,1\}^N$ ($N \ge 2$) equipped with the Hamming metric: $d(x,y) := \sum_{k=1}^N 1_{x_k \neq y_k}$, and $\mu = \alpha^{\otimes N}$ (product measure) where $\alpha(1) = p$, $\alpha(0) = q = 1 - p$ (Bernoulli law). The Markov transition kernel is described at follows: starting from $x = (x_1, \dots, x_N) \in \mathcal{X}$, next pick at random $1 \le k \le N$ and change x_k according to the distribution α with the other coordinates the same as those of x; or equivalently

$$Pg(x) = \frac{1}{N}\sum_{k=1}^N \mu_k(g),$$

where $\mu_k(g)(x) = \int_{\{0,1\}} g(x)\, d\alpha(x_k)$. We claim that

$$(2N-1)Npq \le C_G \le \max\{p,q\}N^2. \tag{3.6}$$

We begin with the calculus of the Lipschitzian spectral gap constant $c_{\mathrm{Lip}}(d,d)$. Since $(P - I)g = \frac{1}{N}\mathcal{L}g$, where $\mathcal{L}g := \sum_{k=1}^N (\mu_k(g) - g)$, by [63], Proposition 2.5, we get $c_{\mathrm{Lip}}(d,d) \le N$. But for $g(x) := S_N - Np$ where $S_N = \sum_{k=1}^N x_k$, it is easy to check that $(I - P)Ng = g$ and then $N \le c_P \le c_{\mathrm{Lip}}(d,d) \le N$.

Since $M^2 = \max\{p,q\}/2$ (see (2.32) for the definition of M^2), Theorem 2.16 entails $C_G \le \max\{p,q\}N^2$, the upper bound in (3.6). For the lower bound take $g(x) = S_N - Np$ whose Lipschitzian coefficient w.r.t. the Hamming metric d is 1. Since $(I-P)^{-1}g = Ng$ (in $C_{\mathrm{Lip},0}$) as noted above, we have

$$\sigma^2(g) = 2\langle (I-P)^{-1}g, g\rangle_\mu - \langle g, g\rangle_\mu = (2N-1)\operatorname{Var}_\mu(S_N) = (2N-1)Npq.$$

Thus $C_G \ge \sigma^2(g) = (2N-1)Npq$, the lower bound in (3.6).

When $p = q = 1/2$ and N is large, (3.6) shows that the general result (i.e., Theorem 2.16) is sharp. Recall that by tensorization and (3.2), $C_H(\mu) = NC_H(p)$ which is only of order N, in contrast with (3.6). That is quite natural: in this Monte Carlo Markov Chain (MCMC in short), one requires at least N steps so that the law of X_N is close to μ, instead of one only step in the i.i.d. case.

3.4. *Line random walks on* $\mathbb{N} \cap [0,N]$. Consider the Markovian random walk on $\mathcal{X} := \mathbb{N} \cap [0,N]$ where $3 \le N \in \mathbb{N} \cup \{+\infty\}$, whose transition probabilities are given by for $\forall k \in \mathcal{X}$,

$$P(k,k-1) = a_k > 0, \qquad P(k,k+1) = b_k > 0, \qquad P(k,k) = c_k \ge 0, \quad a_k + b_k + c_k = 1,$$

where -1 is identified with 0 and $N+1$ is identified with N if N is finite. When $c_k > 0$, the walker is lazy. Then $(P-I)g(k) = b_k(g(k+1) - g(k)) + a_k(g(k-1) - g(k))$, that is, $P - I$ is the generator of a birth-death process with birth rate b_k and death rate a_k. So we can profit from the rich theory of birth-death processes [5, 7]. At first P is symmetric w.r.t. the measure m given by

$$m(0) = 1, \qquad m(n) = \frac{b_0 b_1 \cdots b_{n-1}}{a_1 a_2 \cdots a_n}, \quad 1 \le n \in \mathcal{X}. \tag{3.7}$$

Of course we assume that $m(\mathbb{N}) = \sum_{n\ge 0} m(n) < +\infty$ if $N = +\infty$ (or equivalently the process is positively recurrent) and denote the unique invariant probability measure by $\mu(n) = m(n)/m(\mathbb{N} \cap [0,N])$. Given now an increasing function $h : \mathcal{X} \to \mathbb{R}$ in $L^2(\mu)$, consider the metric $d(i,j) := d_h(i,j) := |h(j) - h(i)|$. Liu and Ma [34] have given the exact Lipschitzian spectral gap constant w.r.t. the metric d_h:

$$c_{\mathrm{Lip}}(d_h, d_h) = \sup_{i \ge 1} \frac{\sum_{k=i}^N \mu(k)(h(k) - \mu(h))}{\mu(i)a_i(h(i) - h(i-1))} =: c_{\mathrm{Lip}}(h) \tag{3.8}$$

(though they have studied only the infinite case, but their proof works for finite N). Notice that if $N = +\infty$ and $h(n) = n$, $c_{\mathrm{Lip}}(h) = +\infty$ by Corollary 2.17..

REMARK 3.2.

(1) Note that the r.h.s. of (3.8) is exactly the Lipschitzian coefficient of the solution G to the Poisson equation $(I-P)G = h-\mu(h)$ (see Liu and Ma [34]).

(2) Chen–Wang's variational formula for spectral gap in L^2 ([5, 34]) can be read as

$$c_P = \inf_h c_{\mathrm{Lip}}(h).$$

From Theorem 2.16, Proposition 2.19 and Lemma 3.1 we get immediately

COROLLARY 3.3. *Let $d = d_h$. Assume that $c_{\mathrm{Lip}}(h) < +\infty$.*

(a) *If*

$$M^2 := \frac{1}{2}\sup_{k\in\mathbb{N}}\{a_k[h(k)-h(k-1)]^2 + b_k[h(k+1)-h(k)]^2\} < +\infty,$$

then

$$W_1(\nu,\mu)^2 \le 4[c_{\mathrm{Lip}}(h)M]^2 I(\nu) \quad \forall \nu \in M_1(\mathcal{X}). \tag{3.9}$$

(b) *If $c_H := c_H(P) < +\infty$ w.r.t. the metric d_h is finite, then*

$$W_1(\nu,\mu)^2 \le 2[c_{\mathrm{Lip}}(h)]^2 c_H I(\nu) \quad \forall \nu \in M_1(\mathcal{X}). \tag{3.10}$$

EXAMPLE 3.4 (Uniform distribution). Let $a_k = b_k = p$, $c_k = 1-2p$ $(0 < p \le 1/2)$ and $N \ge 3$ finite. Consider the Euclidean metric $d(i,j) = |i-j|$ which corresponds to $h(i) = i$. Then $\mu(n) = 1/(N+1)$ for all $0 \le n \le N$. We claim that

$$\frac{N^4}{60p}(1-\varepsilon(N)) \le C_G \le \frac{1}{2p}\{([N/2]+1)(N-[N/2])\}^2 \le \frac{1}{32p}(N+2)^4, \tag{3.11}$$

where $[x]$ denotes the integer part of $x \in \mathbb{R}$ and $\lim_{N\to\infty}\varepsilon(N) = 0$. For the upper bound we have at first $M^2 = p$. From (3.8) it is easy to see that

$$c_{\mathrm{Lip}}(d,d) = \frac{1}{2p}([N/2]+1)(N-[N/2]).$$

Hence from (3.9) we get the upper bound in (3.11). For the lower bound let $g(k) = k - N/2$. It is easy to check (see [34]) that the solution G of the Poisson equation $(I-P)G = g$ verifies

$$G(i+1)-G(i) = -\frac{\sum_{j=0}^{i}\mu(j)g(j)}{\mu(i+1)a_{i+1}} = -\frac{1}{p}\sum_{j=0}^{i}\left(j-\frac{N}{2}\right) = -\frac{1}{2p}(i+1)(i-N).$$

Now using $\sigma^2(g) = 2\langle G,g\rangle_\mu - \langle g,g\rangle_\mu$ and the fact that $\langle g,g\rangle_\mu = \frac{N^2+2N}{12}$ and

$$\begin{aligned}\langle G,g\rangle_\mu = \langle G,(I-P)G\rangle_\mu &= \sum_{0\le i<j\le N}(G(j)-G(i))^2\mu(i)P(i,j)\\ &= \frac{p}{N+1}\sum_{i=0}^{N-1}(G(i+1)-G(i))^2 = \frac{1}{4p(N+1)}\sum_{i=0}^{N-1}((i+1)(N-i))^2\\ &\ge \frac{1}{4p(N+1)}\int_0^N x^2(N-x)^2\,dx = \frac{N^5}{120p(N+1)}\end{aligned}$$

we obtain the lower bound in (3.11).

Since $c_H(P(k,\cdot)) \ge \mathrm{Var}_{P(k,\cdot)}(h) = 2p = 2M^2$ for $1 \le k \le N-1$, Proposition 2.19 does not furnish better upper bound of C_G. The above bound in p is quite natural: the case $p = 1/2$ is the most mixing and C_G should increase if p close to 0 (i.e., when the walker is very lazy).

EXAMPLE 3.5 (Binomial distribution). Let $\mathcal{X} = \mathbb{N} \cap [0, N]$ (with $N \geq 3$ finite) be equipped with the Euclidean metric, $a_k = kq/N$ $(1 \leq k \leq N)$, $b_k = (N-k)p/N$ $(0 \leq k < N)$ and $a_0 = q$, $b_N = p$ and $c_k = 1 - (a_k + b_k)$, where $p \in [1/2, 1)$ and $q = 1 - p$. The corresponding matrix P is symmetric w.r.t. the binomial distribution with success probability p: $\mu(k) = C_N^k p^k q^{N-k}$, $0 \leq k \leq N$, by the formula (3.7).

Let us calculate $c_{\mathrm{Lip}}(h)$ with $h(k) = k$. Instead of using (3.8), let us look at directly the Poisson equation $(I - P)G = h - \mu(h)$ and as recalled in Remark 3.2, $c_{\mathrm{Lip}}(h) = \|G\|_{\mathrm{Lip}(d_h)} = \max_{0 \leq k < N} |G(k+1) - G(k)|$.

The equation $(I - P)G = h - \mu(h)$ says that for each $0 \leq k < N$,

$$b_k\big(G(k+1) - G(k)\big) + a_k\big(G(k-1) - G(k)\big) = Np - k.$$

Hence $G(1) - G(0) = Np/b_0 = N$. And if $G(k) - G(k-1) = N$, then

$$b_k\big[G(k+1) - G(k)\big] = Np - k + a_k N = Np - kp = b_k N,$$

where it follows that $G(k+1) - G(k) = N$. By recurrence $G(k+1) - G(k) = N$ for all $k = 0, \ldots, N-1$. Thus $c_{\mathrm{Lip}}(h) = N$ and $h - \mu(h)$ is an eigenfunction of $I - P$ associated with the eigenvalue $1/N$. Hence $N \leq c_P \leq c_{\mathrm{Lip}}(h) = N$, that is, $c_P = N$ too.

$$\sigma^2(h) = 2\langle G, h - \mu(h)\rangle_\mu - \mathrm{Var}_\mu(h) = (2N-1)\,\mathrm{Var}_\mu(h) = (2N-1)Npq.$$

On the other hand the quantity M^2 given in Corollary 3.3(a) equals to $p/2$. Thus Corollary 3.3(a) yields (recalling that $p \in [1/2, 1)$)

$$(2N-1)Npq = \sigma^2(h) \leq C_G \leq 2\big[c_{\mathrm{Lip}}(h)M\big]^2 = N^2 p. \tag{3.12}$$

This example shows the sharpness of Corollary 3.3(a) (and of Theorem 2.16 again).

EXAMPLE 3.6 (Geometric distribution). Let $\mathcal{X} = \mathbb{N}$, $a_k = a$, $b_k = b = 1 - a$ for all $k \geq 0$ such that $a > 1/2$. The invariant measure μ is the geometric distribution: $\mu(n) = q^n p$, $\forall n \in \mathbb{N}$, $q = b/a$, $p = 1 - q$.

(1) *Euclidean metric*. Consider the Euclidean metric $d(i, j) = |i - j|$. In this metric $C_G = +\infty$, because otherwise we would have for $h(k) = k$ and $\lambda > 0$ (by Corollary 2.6)

$$\int e^{\lambda h}\, d\mu \leq \|e^{\lambda h} P e^{\lambda h}\|_2 \leq e^{2\lambda\mu(h) + 2\lambda^2 C_G}$$

which implies $\mu(h > 2\mu(h) + r) \leq e^{-r^2/(8C_G)}$ for all $r > 0$, the Gaussian tail. That is impossible for the geometric distribution μ. It is worth mentioning that $c_{\mathrm{Lip}}(h) = +\infty$ for the Euclidean metric by Corollary 2.17.

(2) *Discrete metric*. As it is well known that $c_P = (\sqrt{b} - \sqrt{a})^2$ ([7], Example 9.22), we have $C_G \leq c_P/2 = (\sqrt{b} - \sqrt{a})^2/2$ for the discrete metric $d(x, y) = 1_{x \neq y}$, by Theorem 2.14.

3.5. *A random scan Gibbs sampler under the Dobrushin's uniqueness condition*. Throughout the subsection $(E, \mathscr{B})$ is a Polish space equipped with the Borel field $\mathscr{B}$, and d is a metric which is lower semicontinuous on E^2 (not necessarily compatible with the topology of E). Let μ be a Gibbs probability measure on $\mathcal{X} = E^N$ equipped with the product topology and the sum-metric

$$d_{l^1}(x, y) := \sum_{i=1}^N d(x^i, y^i), \quad x = (x^1, \ldots, x^N), y = (y^1, \ldots, y^N) \in E^N.$$

We always assume that for some $x_0 \in E^N$, $\int_{E^N} d_{l^1}^2(x, x_0)\, d\mu(x) < \infty$. Consider the local specification $\{\mu_i(\cdot|x); x \in E^N, i = 1, \ldots, N\}$ where each μ_i is a regular conditional distribution of x^i knowing $(x^j, j \neq i)$ under μ. It is very difficult to calculate the high-dimensional integral $\mu(f)$. Our purpose is to propose a random scan Gibbs sampler (a Markov chain Monte Carlo algorithm) for approximating $\mu(f)$, and to establish its concentration inequality (i.e., the estimate of C_G).

The corresponding Markov chain $(X_n)_{n\geq 0}$ can be described as follows: if $X_n = x \in E^N$, then choose randomly an index i among $\{1, \dots, N\}$ (uniformly), and once i is chosen, $X_{n+1} = (x^1, \dots, x^{i-1}, Z^i, x^{i+1}, \dots, x^N)$ where the random variable Z^i is generated according to the conditional distribution $\mu_i(dz^i|x)$ (knowing $X_n = x$ and the index i).

The d-Dobrushin interdependence matrix $C := (c_{ij})_{i,j=1,\dots,N}$ is defined by

$$c_{ij} := \sup_{x=y \text{ off } j} \frac{W_{1,d}(\mu_i(\cdot|x), \mu_i(\cdot|y))}{d(x^j, y^j)}, \quad i, j = 1, \dots, N. \tag{3.13}$$

Obviously $c_{ii} = 0$. Then the Dobrushin uniqueness condition (see [14, 15]) is

$$(H1) \quad \|C\|_1 := \max_{1\leq j\leq N} \sum_{i=1}^N c_{ij} < 1.$$

Let $Pg(x) = \frac{1}{N}\sum_{k=1}^N \mu_k g(x)$, $\mu_k g(x) = \int_E \mu_k(dz^k|x) g(x(z^k))$, $k = 1, \dots, N$, where

$$\left(x(z^k)\right)^i := \begin{cases} z^k & \text{if } i = k, \\ x^i & \text{if } i \neq k. \end{cases} \tag{3.14}$$

Let $\mathcal{L}g = \sum_{k=1}^N (\mu_k(g) - g)$, then $(P - I)g = \frac{1}{N}\mathcal{L}g$. By [63], Propsition 2.5, we have the following:

LEMMA 3.7. *Under* $(H1)$, $c_{\mathrm{Lip}}(d_{l^1}, d_{l^1}) \leq \frac{N}{1-\|C\|_1}$.

PROOF.

$$\begin{aligned} c_{\mathrm{Lip}}(d_{l^1}, d_{l^1}) &= \|(I - P)^{-1}\|_{\mathrm{Lip}} = \left\|\left(-\frac{1}{N}\mathcal{L}\right)^{-1}\right\|_{\mathrm{Lip}} \\ &= N\|(-\mathcal{L})^{-1}\|_{\mathrm{Lip}} \\ &\leq N\int_0^\infty e^{-t(1-\|C\|_1)}\, dt = \frac{N}{1 - \|C\|_1}, \end{aligned}$$

where the third inequality holds by [63], Proposition 2.5. □

COROLLARY 3.8. *Under* $(H1)$, *if*

$$M^2 := \frac{1}{2N} \sup_{x\in E^N} \sum_{k=1}^N \int_E \mu_k(dz^k|x)\, d^2(x^k, z^k) < \infty,$$

then

$$C_G \leq 2\big(M c_{\mathrm{Lip}}(d_{l^1}, d_{l^1})\big)^2 \leq 2\left(\frac{MN}{1 - \|C\|_1}\right)^2.$$

PROOF. Given $x \in E^N$,

$$\begin{aligned} \frac{1}{2}\int_{E^N} d_{l^1}^2(x, y) P(x, dy) &= \frac{1}{2N}\sum_{k=1}^N \int_E \mu_k(dz^k|x)\, d_{l^1}^2(x, x(z_k)) \\ &= \frac{1}{2N}\sum_{k=1}^N \int_E \mu_k(dz^k|x)\, d^2(x^k, z^k) \leq M^2. \end{aligned}$$

Thus (2.32) holds with M^2 given in the corollary. By Lemma 3.7 and Theorem 2.16, we complete the proof. $\square$

REMARK 3.9. For the discrete metric on E, we have $M^2 \le \frac{1}{2}$, and so

$$C_G \le \left(\frac{N}{1-\|C\|_1}\right)^2.$$

On the other hand the transportation-entropy inequality is known (due to the second named author [63], Theorem 4.3):

$$C_H(\mu) \le \frac{N}{2(1-\|C\|_1)^2}.$$

REMARK 3.10. When (E,d) is unbounded, then $M = +\infty$ in the free case (i.e., when μ is a product measure). In that case how this MCMC is concentrated remains to be studied.

We show at first that N^2 is the correct order of the sub-Gaussian concentration constant C_G.

EXAMPLE 3.11 (No interaction). Let $E = \{0,1\}$, d is the discrete metric on E, and $\mu = \alpha^{\otimes N}$ where $\alpha(1) = p$, $\alpha(0) = 1-p$, $0 < p < 1$. For this model (discrete cube), the Dobrushin matrix $C = 0$ and since $M^2 \le \max\{p, 1-p\}/2$, we have $C_G \le \max\{p, 1-p\}N^2$ by Corollary 3.8. This upper bound coincides with that in (3.6), and $C_G \ge (2N-1)Np(1-p)$. In other words N^2 is the correct order of C_G.

EXAMPLE 3.12 (Interaction). Let $E = \{1,-1\}$, d is the trivial metric on E, and μ is the probability measure on $\{1,-1\}^N$ given by

$$\mu(x) = \frac{\exp(-\sum_{S\subset\{1,\dots,N\}} J(S)x^S)}{\sum_{y\in\{1,-1\}^N}\exp(-\sum_{S\subset\{1,\dots,N\}} J(S)y^S)}, \quad x \in \{1,-1\}^N,$$

where $J(S)$ is a function in S, and $x^S := \prod_{i\in S} x_i$.

For this model, by [20, 63], the Dobrushin matrix C satisfies

$$\|C\|_1 \le \sup_{i\in\{1,\dots,N\}} \sum_{S:S\ni i} \bigl(|S|-1\bigr)\tanh\bigl|J(S)\bigr|. \tag{3.15}$$

Hence if the last quantity is less than one, we can apply the corollary above.

It is well known that the Dobrushin uniqueness condition is sharp for the phase transition of the mean field models (ref. [20]).

4. Some preparations. Though the Donsker–Varadhan information $I(\nu) := I(\nu|P,\mu)$ has no closed expression in the actual discrete time case, we establish several properties necessary to the results of this paper. Throughout this section, μ is a fixed invariant and ergodic probability measure of P.

4.1. *Cramer functionals.* Let

$$P^V g := e^{V/2} P\big(e^{V/2} g\big) \quad \text{or} \quad P^V(x, dy) = e^{[V(x)+V(y)]/2} P(x, dy).$$

Consider the Cramer functionals on $V \in b\mathcal{B}$,

$$\Lambda_\beta(V|P) := \limsup_{n\to\infty} \frac{1}{n} \log \int \big(P^V\big)^n 1\, d\beta = \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}_\beta e^{n\tilde{L}_n(V)}; \tag{4.1}$$

$$\Lambda_p(V|P) := \lim_{n\to\infty} \frac{1}{n} \log \big\| \big(P^V\big)^n \big\|_p. \tag{4.2}$$

Notice that for $(e^V P)(x, dy) := e^{V(x)} P(x, dy)$, $e^V P f = e^{V/2} P^V e^{-V/2} f$, and then the spectrum of $e^V P$ is the same as that of P^V on $L^p(\mu)$ (for $V \in b\mathcal{B}$) and

$$\big(e^V P\big)^n f = e^{V/2} \big(P^V\big)^n \big(e^{-V/2} f\big).$$

Similarly $(Pe^V)^n f := e^{-V/2}(P^V)^n(e^{V/2} f)$. Then

$$\Lambda_p(V|P) = \lim_{n\to\infty} \frac{1}{n} \log \big\| \big(e^V P\big)^n \big\|_p = \lim_{n\to\infty} \frac{1}{n} \log \big\| \big(Pe^V\big)^n \big\|_p. \tag{4.3}$$

LEMMA 4.1. *For any initial distribution $\beta \ll \mu$, let Λ be one of $\Lambda_\beta(\cdot|P)$ or $\Lambda_p(\cdot|P)$ where $1 \le p \le \infty$. Then for any $\nu \in M_b(\mathcal{X})$, the space of all bounded and signed measures on $(\mathcal{X}, \mathcal{B})$, the Legendre transform of Λ is given by*

$$\Lambda^*(\nu) := \sup\big\{\nu(V) - \Lambda(V);\, V \in b\mathcal{B}\big\} = \begin{cases} I(\nu|P, \mu) & \text{if } \nu \in \mathcal{M}_1(\mathcal{X}), \\ +\infty & \text{otherwise.} \end{cases} \tag{4.4}$$

In particular for all $\nu \in M_1(\mathcal{X})$,

$$I(\nu|P, \mu) = I\big(\nu|P^*, \mu\big), \tag{4.5}$$

where P^* is the adjoint operator of P on $L^2(\mu)$.

PROOF. (4.4) is essentially contained in Wu [61]. In fact in [61], Proposition B.9 and remarks, it is proved that for any $\beta \ll \mu$, (4.4) holds for $\Lambda = \Lambda_\beta(\cdot|P)$ or $\Lambda_\infty(\cdot|P)$. Now for any $p \in [1, +\infty)$ fixed, as $\Lambda_\mu(V|P) \le \Lambda_p(V|P)$, we have

$$\big(\Lambda_p(\cdot|P)\big)^*(\nu) \le (\Lambda_\mu(\cdot|P)^*(\nu) = \bar{I}(\nu|P, \mu), \quad \nu \in \mathcal{M}_b(\mathcal{X}),$$

where $\bar{I}(\nu|P, \mu)$ defined over $\mathcal{M}_b(\mathcal{X})$ is defined by the r.h.s. of (4.4). For the converse inequality, notice that for any $1 \le u \in b\mathcal{B}$, letting $V_u := \log \frac{u}{Pu}$, we have

$$P^{V_u}(\sqrt{uPu}) = \sqrt{\frac{u}{Pu}} P\left(\sqrt{\frac{u}{Pu}} \cdot \sqrt{uPu}\right) = \sqrt{uPu}.$$

That means $\sqrt{uPu}$ (≥ 1) is a positive eigenfunction of P^{V_u} associated with the eigenvalue 1, on $L^p(\mu)$. Therefore by Perron–Frobenius theorem for positive operators, the spectral radius of P^{V_u} on $L^p(\mu)$ is 1, that is (Gelfand's formula),

$$\lim_{n\to\infty} \frac{1}{n} \log \big\| \big(P^{V_u}\big)^n \big\|_p = \log 1 = 0.$$

Thus $\Lambda_p(V_u|P) = 0$ and

$$\big(\Lambda_p(\cdot|P)\big)^*(\nu) \ge \sup_{1 \le u \in b\mathcal{B}} \int V_u\, d\nu - \Lambda_p(V_u|P) = \sup_{1 \le u \in b\mathcal{B}} \int \log \frac{u}{Pu}\, d\nu = I(\nu|P, \mu).$$

We have so proved (4.4) for $\Lambda = \Lambda_p(\cdot|P)$.

Since $(P^V)^* = (P^*)^V$,

$$\Lambda_\mu(V|P) = \limsup_{n\to\infty} \frac{1}{n} \log\langle 1, (P^V)^n 1\rangle_\mu$$
$$= \limsup_{n\to\infty} \frac{1}{n} \log\langle [(P^*)^V]^n 1, 1\rangle_\mu = \Lambda_\mu(V|P^*),$$

hence $I(\nu|P,\mu) = [\Lambda_\mu(\cdot|P)]^*(\nu) = [\Lambda_\mu(\cdot|P^*)]^*(\nu) = I(\nu|P^*,\mu)$, that is, (4.5) holds. $\square$

LEMMA 4.2. *Let $P_1, \dots, P_n$ be Markov operators on $L^2(\mu)$. Then*

$$I(\nu|P_1\cdots P_n;\mu) \le \sum_{k=1}^n I(\nu|P_k;\mu). \tag{4.6}$$

In particular

$$I^\sigma(\nu) := \frac{1}{2} I(\nu|PP^*;\mu) \le I(\nu|P;\mu). \tag{4.7}$$

PROOF. Let $Q_0 = I$ and for each $1 \le k \le n$, let $Q_k = P_{n-k+1}\cdots P_n$ which is again Markovian. For any $\nu \in M_1(\mathcal{X})$, $\nu \ll \mu$ and $1 \le u \in b\mathcal{B}$, we have by the definition of $I(\nu|P_k,\mu)$,

$$\int \log \frac{u}{Q_n u}\, d\nu = \sum_{k=1}^n \int \log \frac{Q_{k-1}u}{P_{n-k+1}(Q_{k-1}u)}\, d\nu \le \sum_{k=1}^n I(\nu|P_{n-k+1};\mu),$$

where (4.6) follows by taking the supremum over all $1 \le u \in b\mathcal{B}$. Finally (4.7) is a direct consequence of (4.6) by (4.5). $\square$

The following lemma, due to Lei [29], Lemma 3.1(b), under the uniform integrability condition of P on $L^2(\mu)$ (a notion introduced in [61]), plays a basic role in Theorem 2.4.

LEMMA 4.3. *Assume that P is μ-symmetric. Then*

$$\log\|P^V\|_2 = \Lambda_2(V|P) = \sup_{\nu\in M_1(\mathcal{X})} \big(\nu(V) - I(\nu|P,\mu)\big) \quad \forall V \in b\mathcal{B} \tag{4.8}$$

and in particular, for any initial measure $\beta \ll \mu$ with $d\beta/d\mu \in L^2(\mu)$, $r > 0$ and $n \ge 1$,

$$\mathbb{P}_\beta\big(\tilde{L}_n(V) > \mu(V) + r\big) \le \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\big(-nI_V\big(\mu(V)+r\big)\big), \quad r > 0, \tag{4.9}$$

where

$$I_V(x) := \inf\{I(\nu)|\nu(V) = x\}, \quad x \in \mathbb{R}. \tag{4.10}$$

The continuous time counterpart of this lemma is due to [60].

PROOF. We begin with the proof of (4.8). At first since $P^V = e^{V/2}Pe^{V/2}$ is symmetric, its norm coincides with its spectral radius, where the first equality in (4.8) follows. Next by [62], Lemma 5.3,

$$\Lambda_2(V) := \Lambda_2(V|P) = \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}_\mu f(X_0)g(X_n)\exp\big(n\tilde{L}_n(V)\big) \tag{4.11}$$

for any couple (f, g) of nonnegative functions so that $\mu(f), \mu(g) > 0$ (under the ergodicity assumption). By Lemma 4.1, the Legendre transform Λ_2^* of Λ_2 is given by

$$\Lambda_2^*(\nu) := \sup\{\nu(V) - \Lambda_2(V); V \in b\mathcal{B}\} = \begin{cases} I(\nu) & \text{if } \nu \in \mathcal{M}_1(\mathcal{X}); \\ +\infty & \text{otherwise} \end{cases}$$

for any $\nu \in \mathcal{M}_b(\mathcal{X})$. If the large deviation principle of L_n in the τ-topology holds (i.e., the case under the uniform integrability condition by [61]), the r.h.s. of (4.11) with $f = g = 1$ coincides with the r.h.s. of (4.8) by Varadhan's Laplace principle. Our problem here is without the uniform integrability condition.

If Λ_2 were lower semicontinuous (l.s.c. in short) on $b\mathcal{B}$ w.r.t. the weak topology $\sigma(b\mathcal{B}, \mathcal{M}_b(\mathcal{X}))$, that is, the weakest topology so that $V \to \nu(V)$ is continuous for every $\nu \in \mathcal{M}_b(\mathcal{X})$, then by the convexity of Λ_2 on $b\mathcal{B}$ and the Fenchel–Legendre theorem,

$$\Lambda_2(V) = \Lambda_2^{**}(V) = \sup\{\nu(V) - I(\nu); \nu \in \mathcal{M}_1(\mathcal{X})\}$$

the desired second equality in (4.8).

Let us prove the lower semicontinuity of Λ_2 on $b\mathcal{B}$ w.r.t. the weak topology $\sigma(b\mathcal{B}, \mathcal{M}_b(\mathcal{X}))$. We recall [35], Lemma 3.4.

Let $\Lambda : b\mathcal{B} \to (-\infty, +\infty]$ be a convex functional such that

(i) *if $V_1 = V_2$, μ-a.s. on $\mathcal{X}$, $\Lambda(V_1) = \Lambda(V_2)$;*
(ii) *if $V_1 \le V_2$, then $\Lambda(V_1) \le \Lambda(V_2)$.*

Then Λ is lower semicontinuous on $b\mathcal{B}$ w.r.t. the weak topology $\sigma(b\mathcal{B}, \mathcal{M}_b(\mathcal{X}))$, iff Λ is increasingly continuous, that is, for any nondecreasing sequence (V_n) in $b\mathcal{B}$ such that $\sup_n V_n = V \in b\mathcal{B}$, $\Lambda(V_n) \to \Lambda(V)$.

In the actual symmetric case, we have by the Perron–Frobenius theorem, the largest spectral point $e^{\Lambda_2(V)}$ of P^V satisfies

$$e^{\Lambda_2(V)} = \sup\{\langle f, P^V f\rangle_\mu; f \ge 0, \|f\|_2 = 1\},$$

where the increasing continuity of Λ_2 follows immediately.

Next let us derive (4.9) from (4.8), by following Lei [29]. At first by the definition (4.10) of I_V, we have for any $\lambda \in \mathbb{R}$,

$$\begin{aligned} \Lambda_2(\lambda V) = \log\|P^{\lambda V}\|_2 &= \sup\{\lambda\nu(V) - I(\nu); \nu \in M_1(\mathcal{X})\} \\ &= \sup_{r \in \mathbb{R}}(\lambda r - I_V(r)) = (I_V)^*(\lambda). \end{aligned}$$

Consequently for any $\lambda > 0$ and $r > 0$,

$$\begin{aligned} \mathbb{P}_\beta(\tilde{L}_n(V) > \mu(V) + r) &\le e^{-\lambda n(\mu(V)+r)}\mathbb{E}_\beta e^{\lambda n \tilde{L}_n(V)} \\ &= e^{-n\lambda(\mu(V)+r)}\left\langle \frac{d\beta}{d\mu}, (P^{\lambda V})^n 1\right\rangle_\mu \\ &\le e^{-n\lambda(\mu(V)+r)}\left\|\frac{d\beta}{d\mu}\right\|_2 \|P^{\lambda V}\|_2^n \\ &= \left\|\frac{d\beta}{d\mu}\right\|_2 \exp(-n[\lambda(\mu(V)+r) - I_V^*(\lambda)]). \end{aligned}$$

Taking the infimum over all $\lambda > 0$ we get

$$\mathbb{P}_\beta(\tilde{L}_n(V) > \mu(V) + r) \le \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\Big(-n \sup_{\lambda > 0}[\lambda(\mu(V)+r) - (I_V)^*(\lambda)]\Big).$$

As $I_V \geq 0$ is convex and $I_V(\mu(V)) = 0$, I_V is nondecreasing on $[\mu(V), +\infty)$. And we have for any $r > 0$,

$$\sup_{\lambda>0}[\lambda(\mu(V)+r) - (I_V)^*(\lambda)] = \sup_{\lambda\in\mathbb{R}}[\lambda(\mu(V)+r) - (I_V)^*(\lambda)]$$
$$= (I_V)^{**}(\mu(V)+r) = I_V((\mu(V)+r)-),$$

where the first equality follows by the fact that $(I_V^*)(\lambda) = \Lambda_2(\lambda V) \geq \lambda\mu(V)$ (Jensen's inequality), the second is derived by the Fenchel–Legendre theorem and the fact that $I_V(x-) := \lim_{\varepsilon\to 0+} I_V(x-\varepsilon)$ is exactly the lower semicontinuous modification of I_V for $x \in (\mu(V), +\infty)$. So by the nondecreasingness of I_V on $[\mu(V), +\infty)$, we have for all $r, \delta > 0$, $I_V((\mu(V)+r+\delta)-) \geq I_V(\mu(V)+r)$, and then

$$\mathbb{P}_\beta(\tilde{L}_n(V) > \mu(V)+r) = \lim_{\delta\to 0+} \mathbb{P}_\beta(\tilde{L}_n(V) > \mu(V)+r+\delta)$$
$$\leq \lim_{\delta\to 0+} \left\| \frac{d\beta}{d\mu} \right\|_2 \exp(-nI_V((\mu(V)+r+\delta)-))$$
$$\leq \left\| \frac{d\beta}{d\mu} \right\|_2 \exp(-nI_V(\mu(V)+r)). \qquad \square$$

4.2. *Comparison with a continuous time Markov process.* Consider the continuous time Markov process (Y_t) with generator $\mathcal{L} = P^\sigma - I$, that is, its transition probability semigroup is given by $e^{t(P^\sigma - I)}$. This process can be constructed very easily: $Y_t = Z_{N(t)}$ where (Z_n) is the Markov chain with transition $P^\sigma = (P+P^*)/2$, $N(t)$ is a Poisson process with parameter 1, independent of $(Z_n)_{n\geq 0}$.

The associated Donsker–Varadhan information $I_c : M_1(\mathcal{X}) \to [0, +\infty]$ is given by:

$$I_c(\nu) := \sup_{1\leq u\in b\mathcal{B}} \int \frac{(I-P^\sigma)u}{u}\, d\nu \tag{4.12}$$

if $\nu \ll \mu$ and $I_c(\nu) := +\infty$ otherwise. In the symmetric case, we have (see [61], Corollary B.11): for $\nu = f\mu$,

$$I_c(\nu) = \langle\sqrt{f}, (I - P^\sigma)\sqrt{f}\rangle_\mu. \tag{4.13}$$

That is the definition given in (2.33). We have the following simple observation.

LEMMA 4.4. *For any* $\nu \in M_1(\mathcal{X})$,

$$I_c(\nu) \leq I(\nu|P^\sigma, \mu) \leq I(\nu|P, \mu). \tag{4.14}$$

From the point of view of large deviations, in the symmetric case, (4.14) means that the mixing rate of the discrete time Markov chain (X_n) is more rapid than that of the corresponding continuous time Markov process (Y_t).

PROOF. The second inequality in (4.14) follows by (4.5) and the the convexity of $I(\nu|P, \mu)$ in P. For the first inequality, we observe that for any $1 \leq u \in b\mathcal{B}$, by the elementary inequality that $\log(1+x) \leq x$ for $x > -1$,

$$\log\frac{u}{P^\sigma u} = -\log\left(1 + \frac{(P^\sigma - I)u}{u}\right) \geq \frac{(I-P^\sigma)u}{u},$$

hence (4.14) follows. $\square$

4.3. *An upper bound of the second eigenvalue.* Throughout this paragraph P is μ-symmetric. The largest (first) eigenvalue $\lambda_0(P)$ of P in $L^2(\mu)$ is 1. The second one denoted by $\lambda_1(P)$, being the supremum of $\sigma(P|_{L^2(\mu)})\backslash\{1\}$ where $\sigma(P|_{L^2(\mu)})$ is the spectrum of P on $L^2(\mu)$, is given by Rayleigh's formula,

$$\lambda_1(P) = \sup_{0\neq u\in L_0^2(\mu)} \frac{\langle u, Pu\rangle_\mu}{\langle u, u\rangle_\mu}, \tag{4.15}$$

where $L_0^2(\mu) := \{u \in L^2(\mu); \mu(u) = 0\}$.

LEMMA 4.5. *Assume that P is μ-symmetric. Then*

$$\lambda_1(P) \le \sup_{A\in\mathcal{B}:\mu(A)\in(0,1/2]} \lambda_0(A), \tag{4.16}$$

where $\lambda_0(A)$ is the supremum of the spectrum $\sigma(1_A P 1_A|_{L^2(A,\mu)})$ of $1_A P 1_A$ on $L^2(A,\mu)$. Furthermore

$$\lambda_0(A) = \mu\text{-}\operatorname*{esssup}_{x\in A} \limsup_{n\to\infty} [\mathbb{P}_x(\sigma_A = n)]^{1/n}, \tag{4.17}$$

where $\sigma_A := \inf\{n \ge 0; X_n \notin A\}$ is the first exit time of A.

Its counterpart in the continuous time case, due to Lawler, is well known. See also [5] for extensions.

PROOF. At first (4.17) is contained in [62], Theorem 5.5(b), (the condition (A1) therein is not used for this conclusion as seen from its proof). For (4.16) we may assume that $\lambda_1 = \lambda_1(P) > 0$ (trivial otherwise). Assume at first that $\lambda_1 := \lambda_1(P)$ is an eigenvalue, that is, there were some $0 \neq u \in L_0^2(\mu)$ such that

$$Pu = \lambda_1 u.$$

We may assume $\mu(u > 0) \le \frac{1}{2}$ (consider $-u$ otherwise). Setting $A := [u > 0]$, we have $\mu(A) > 0$ and

$$\lambda_1 1_A u = \lambda_1 u^+ = (Pu)^+ \le Pu^+ = P(1_A u)$$

which yields to $\lambda_1^n(1_A u) \le (1_A P 1_A)^n(1_A u)$ for all $n \ge 1$. Thus

$$\lambda_1 \le \lim_{n\to\infty} \left(\|(1_A P 1_A)^n\|_2\right)^{1/n} = \lambda_0(A).$$

In the general case, consider an increasing sequence of sub-algebras $(\mathcal{B}_n)$ of $\mathcal{B}$, each of which is finitely generated, and $P_n g := E_n P E_n g$, where $E_n g = \mathbb{E}^\mu(g|\mathcal{B}_n)$. By the martingale convergence theorem, $P_n g \to Pg$ in $L^2(\mu)$ for all $g \in L^2(\mu)$. Thus by Rayleigh's formula (4.15) and the fact that P_n can be represented as a finite symmetric matrix,

$$\begin{aligned}\lambda_1(P) &\le \liminf_{n\to\infty} \lambda_1(P_n) \le \liminf_{n\to\infty} \sup_{A\in\mathcal{B}_n:\mu(A)\in(0,1/2]} \lambda_0(1_A P_n 1_A)\\ &\le \sup_{A\in\mathcal{B}:\mu(A)\in(0,1/2]} \lambda_0(1_A P 1_A)\end{aligned}$$

because $\lambda_0(1_A P_n 1_A) = \lambda_0(E_n 1_A P 1_A E_n) \le \lambda_0(1_A P 1_A)$ for all $A \in \mathcal{B}_n$. $\square$

5. Proofs of Theorem 2.4, Proposition 2.10 and Theorem 2.12.

5.1. *Proofs of Theorem 2.4 and Corollary 2.6.*

PROOF OF THEOREM 2.4. The proof of this result is similar to Gozlan–Léonard's [22], Theorems 2 and 15, in the i.i.d. case, and essentially the same as that of [26], Theorem 2.4, for the continuous time symmetric Markov processes, except using Lei's lemma 4.3 in place of [26], Lemma 6.1. For the convenience of the reader we give the detailed proof.

Step 1. In this part we show (without the convexity of α),

$$\text{(a)} \implies \text{(b)} \implies (\text{b}'); \qquad \text{(a)} \implies \text{(c)} \implies (\text{c}').$$

• (a) $\Rightarrow$ (b): For $u \in b\mathcal{B}$, let

$$I_u(r) := \inf\{I(\nu)|\nu(u) = r\} \tag{5.1}$$

we notice that $(\alpha - T_{\mathcal{V}}I)$ implies that for any $(u, v) \in \mathcal{V}$,

$$I_u\big(\mu(v) + r\big) \ge \alpha(r) \quad \forall r \in \mathbb{R}. \tag{5.2}$$

Indeed it is trivial for $r \le 0$ (as $\alpha(r) = 0$ for $r \le 0$). Now for any $r > 0$ and $\nu \in M_1(\mathcal{X})$ such that $\nu(u) = \mu(v) + r$, $T_{\mathcal{V}}I$ implies that

$$I(\nu) \ge \alpha\big(T_{\mathcal{V}}(\nu, \mu)\big) \ge \alpha\big(\nu(u) - \mu(v)\big) = \alpha(r),$$

where (5.2) follows. Now by Lemma 4.3, for $\lambda > 0$,

$$\begin{aligned}\log\|P^{\lambda u}\|_2 &= \sup\{\lambda\nu(u) - I(\nu)|\nu \in M_1(\mathcal{X})\} \\ &= \sup_{a\in\mathbb{R}}[\lambda a - I_u(a)] \le \sup_{r\in\mathbb{R}}[\lambda\big(\mu(v) + r\big) - \alpha(r)]\} = \lambda\mu(v) + \alpha^*(\lambda)\end{aligned}$$

(that still holds for $\lambda < 0$ for $\alpha^*(\lambda) = +\infty$). Then $\|(P^{\lambda u})^n\|_2 \le \|P^{\lambda u}\|_2^n \le e^{n[\lambda\mu(v)+\alpha^*(\lambda)]}$, that is, the statement (b).

• (a) $\Rightarrow$ (c): This follows from (4.9) in Lemma 4.3 and (5.2), for $\mu(u) \le \mu(v)$.

• (b) $\Rightarrow$ (b$'$) and (c) $\Rightarrow$ (c$'$): They are trivial.

Step 2. (c$'$) $\Rightarrow$ (a) *and* (b$'$) $\Rightarrow$ (c$'$).

• (c$'$) $\Rightarrow$ (a) (without the μ-symmetry of P). By the large deviation lower bound in [61], Theorem B.1, we have for any initial probability measure $\beta \ll \mu$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\beta\big(\tilde{L}_n(u) > \mu(v) + r\big) \ge -\inf\{I(\nu); \nu(u) > \mu(v) + r\}.$$

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

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

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

$$\alpha(r_0 - 1/n) \le I(\nu),$$

hence $T_{\mathcal{V}}I$ follows by letting $n \to \infty$ and by the left-continuity of α.

• (b$'$) $\Rightarrow$ (c$'$): Without loss of generality, we assume that $\beta = \mu$. As α is assumed to be convex and is lower semicontinuous, $\sup_{\lambda>0}(\lambda r - \alpha^*(\lambda)) = \alpha^{**}(r) = \alpha(r)$ for $r \ge 0$. By the Chebyshev's inequality and (b$'$), (c$'$) is proved. This completes the proof of the theorem. $\square$

PROOF OF COROLLARY 2.6. As $(W_1I(C))$ is the same as $(\alpha - T_{\mathcal{V}}I)$ with $\alpha(r) = r^2/(2C)$ and $\mathcal{V} := \{(u, u); u \text{ is bounded and } \|u\|_{\text{Lip}} \le 1\}$, part (1) is a direct consequence of Theorem 2.4.

For part (2), notice that $(W_2I(C))$ is the same as $(\alpha - T_{\mathcal{V}}I)$ with $\alpha(r) = r/(2C)$ and $\mathcal{V} := \{(u, v); u, v \text{ are bounded and } u(x) - v(y) \le d^2(x, y), \forall x, y \in \mathcal{X}\}$. Then

$$\alpha^*(\lambda) = \begin{cases} 0 & \text{if } 0 \le \lambda \le \dfrac{1}{2C}; \\ +\infty & \text{if } \lambda > \dfrac{1}{2C}. \end{cases}$$

Since inf-convolution Qv is upper semicontinuous and sup-convolution Su is lower semicontinuous (as $d(x, y)$ is continuous on $\mathcal{X}^2$ by our assumption here), they are measurable. Now we can obtain part (2) directly by noting that for any $(u, v) \in \mathcal{V}$, $u \le Qv$ and $v \ge Su$. □

5.2. *Proof of Proposition* 2.10.

PROOF OF PROPOSITION 2.10. (a) is contained in Lemma 4.2.

(b) For all $V, g \in b\mathcal{B}$, as $0 \le \langle g - \mu(g), P[g - \mu(g)]\rangle_\mu = \langle g, Pg\rangle_\mu - [\mu(g)]^2$ by the assumed nonnegative definiteness of P, we have

$$[\mu(e^V)]^2 \le \langle e^V, Pe^V\rangle_\mu = \mu[P^{2V}1] \le \|P^{2V}\|_2.$$

But $\log \|P^{2V}\|_2 = \Lambda_\mu(2V|P, \mu)$ in the actual symmetric case, by Lemma 4.3 (the first equality in (4.8)). We get by the well-known variational formula of entropy (of Donsker–Varadhan)

$$\begin{aligned} H(\nu|\mu) &= \sup_{V \in b\mathcal{B}} \big(\nu(V) - \log \mu(e^V)\big) \\ &\ge \sup_{V \in b\mathcal{B}} \left(\nu(V) - \frac{1}{2}\Lambda_\mu(2V|P, \mu)\right) = \frac{1}{2} I(\nu|P, \mu). \end{aligned}$$

(c) Applying part (b) to PP^* which is symmetric and nonnegative definite on $L^2(\mu)$, we have

$$2I^\sigma(\nu|\mu) = I(\nu|PP^*, \mu) \le 2H(\nu|\mu)$$

the desired result. □

5.3. *Proof of Theorem* 2.12. *Part I.* We prove at first (2.19) under $(\alpha - T_{\mathcal{V}}I)$, that is, the condition (2.20) with $N = 1$.

Step 1. We notice that

$$\|e^u P\|_2 = \|P^*(e^u \cdot)\|_2 = \sqrt{\|e^u PP^*(e^u \cdot)\|_2} = \sqrt{\|(PP^*)^{2u}\|_2} \tag{5.3}$$

and by Lemma 4.3 (and the symmetry of PP^*)

$$\begin{aligned} 2\log\|e^u P\|_2 &= \log\|(PP^*)^{2u}\|_2 \\ &= \sup\{2\nu(u) - I(\nu|PP^*, \mu); \nu \in M_1(\mathcal{X})\} \\ &= 2\sup\{\nu(u) - I^\sigma(\nu); \nu \in M_1(\mathcal{X})\} =: 2(I^\sigma)^*(u). \end{aligned} \tag{5.4}$$

Thus if α is moreover convex, by the proof of Theorem 2.4, $(\alpha - T_{\mathcal{V}}I^\sigma)$ is equivalent to

$$\|e^{\lambda u} P\|_2 \le \exp(\lambda\mu(v) + \alpha^*(\lambda)), \quad \lambda \ge 0, (u, v) \in \mathcal{V}. \tag{5.5}$$

That is Remark 2.13. By (5.4),

$$\begin{aligned} \log\|(PP^*)^{2\lambda u}\|_2 &= \sup\{2\lambda\nu(u) - I(\nu|PP^*, \mu); \nu \in M_1(\mathcal{X})\} \\ &= 2\sup_{r \in \mathbb{R}}(\lambda r - I_u^\sigma(r)) = 2(I_u^\sigma)^*(\lambda), \end{aligned}$$

where

$$I_u^\sigma(r) := \inf\left\{ I^\sigma(\nu) = \frac{1}{2} I(\nu | PP^*, \mu) \,\middle|\, \nu \in M_1(\mathcal{X}), \nu(u) = r \right\}. \tag{5.6}$$

Consequently for any $\lambda > 0$ and $r > \mu(u)$,

$$\begin{aligned}
\mathbb{P}_\beta\big(L_n(u) > r\big) &\le e^{-\lambda n r} \mathbb{E}_\beta e^{\lambda n L_n(u)} = e^{-n\lambda r} \left\langle \frac{d\beta}{d\mu}, \big(e^{\lambda u} P\big)^n 1 \right\rangle_\mu \\
&\le e^{-n\lambda r} \left\| \frac{d\beta}{d\mu} \right\|_2 \| e^{\lambda u} P \|_2^n = e^{-n\lambda r} \left\| \frac{d\beta}{d\mu} \right\|_2 \cdot \| (PP^*)^{2\lambda u} \|_2^{n/2} \\
&= \left\| \frac{d\beta}{d\mu} \right\|_2 \exp\big(-n\big[\lambda r - (I_u^\sigma)^*(\lambda)\big]\big).
\end{aligned}$$

With this crucial inequality in hand, we get by repeating the proof of (4.8) $\Longrightarrow$ (4.9) in Lemma 4.3,

$$\mathbb{P}_\beta\big(L_n(u) > r\big) \le \left\| \frac{d\beta}{d\mu} \right\|_2 \exp\big(-n I_u^\sigma(r)\big). \tag{5.7}$$

Step 2. Having (5.7) in hand, we can now prove easily the desired result (2.19). In fact, for any $(u, v) \in \mathcal{V}$ and $r > 0$, noting that $I_u^\sigma(\mu(v) + r) \ge \alpha(r)$ by our assumed $T_\mathcal{V} I^\sigma$ (as in the proof of (5.2)) we have by (5.7),

$$\mathbb{P}_\beta\big(L_n(u) > \mu(v) + r\big) \le \left\| \frac{d\beta}{d\mu} \right\|_2 \exp\big(-n I_u^\sigma(\mu(v) + r)\big) \le \left\| \frac{d\beta}{d\mu} \right\|_2 e^{-n\alpha(r)}$$

the desired (2.19).

Part II. The general $N \ge 1$ case. We work now under the condition (2.20) for general $N \ge 1$.

For each $n \ge N$ let $I_j = \{k = lN + j \le n - 1; l = 0, \dots, [(n-1)/N]\}$ ($[x]$ being the greatest integer bounded by x), $j = 0, 1, \dots, N - 1$. By Hölder's inequality,

$$\begin{aligned}
&\log \mathbb{E}_\beta \exp\left(\sum_{k=0}^{n-1} u(X_k) \right) \\
&\quad = \log \mathbb{E}_\beta \exp\left(\frac{1}{N} \sum_{j=0}^{N-1} \sum_{k \in I_j} N u(X_k) \right) \\
&\quad \le \frac{1}{N} \sum_{j=0}^{N-1} \log \mathbb{E}_\beta \exp\left(\sum_{k \in I_j} N u(X_k) \right) \\
&\quad = \frac{1}{N} \sum_{j=0}^{N-1} \log \beta P^j \big[\big(e^{Nu} P^N\big)^{|I_j|} 1 \big] \\
&\quad \le \frac{1}{N} \sum_{j=0}^{N-1} \left(\log \left\| \frac{d(\beta P^j)}{d\mu} \right\|_2 + |I_j| \sup_{\nu \in M_1(\mathcal{X})} \left\{ \nu(Nu) - \frac{1}{2} I\big(\nu | P^N (P^*)^N, \mu\big) \right\} \right) \\
&\quad \le \log \left\| \frac{d\beta}{d\mu} \right\|_2 + n \sup_{\nu \in M_1(\mathcal{X})} \left\{ \nu(u) - \frac{1}{2N} I\big(\nu | P^N (P^*)^N, \mu\big) \right\},
\end{aligned}$$

where the fourth line inequality follows by applying (5.4) to $e^{Nu} P^N$. The remained proof is exactly the same as $N = 1$ case above, so omitted.

6. Relations with spectral gaps and hypercontractivity.

6.1. *Poincaré inequality is equivalent to $W_1 I$: Proof of Theorem 2.14.*

PROOF OF THEOREM 2.14.

• (a) Let $\mathcal{V} = \{(u, u) \in (b\mathcal{B})^2; \|u\|_{\mathrm{osc}} \le 1\}$, since $\|\nu - \mu\|_{\mathrm{TV}} = 2W_1(\nu, \mu) = 2T_{\mathcal{V}}(\nu, \mu)$ and $\|u\|_{\mathrm{Lip}} = \|u\|_{\mathrm{osc}}$ w.r.t. the discrete metric $d(x, y) = 1_{x \neq y}$, (2.23) follows by (2.22) by Corollary 2.6. The transport inequality (2.22) holds because the León and Perron inequality (2.28) holds for the general state space Markov chain by Miasojedow [39] and Fan et al. [47] (see the detailed proof therein), according to Theorem 2.4[$(c') \Longrightarrow (a)$], as explained in Remark 2.15(i).

• (b) The condition (2.24) says exactly that

$$\lambda_1\big(P^N (P^*)^N\big) = \lambda_1\big((P^*)^N P^N\big) \le \delta^2 < 1.$$

Applying part (a) to $P^N (P^*)^N$, we get

$$\|\nu - \mu\|_{\mathrm{TV}}^2 \le 2\frac{1 + \delta^2}{1 - \delta^2} I\big(\nu | P^N (P^*)^N, \mu\big)$$

that is the inequality (2.25). Now it remains to apply Theorem 2.12 to conclude the concentration inequality (2.26).

• (c) In the actual symmetric case and with the notation in Lemma 4.5, we have

$$c_P = \frac{1}{1 - \lambda_1(P)} \le \sup_{A:\mu(A) \in (0,1/2]} \frac{1}{1 - \lambda_0(A)}.$$

Fix $A \in \mathcal{B}$ with $0 < \mu(A) \le 1/2$. Since for any $a > \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\sigma_A > n)$,

$$h_a(x) := \sum_{n=1}^{\infty} \mathbb{P}_x(\sigma_A > n) e^{-an} \in L^1(\mu)$$

$h_a(x) < +\infty$, μ-a.s. Thus

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x(\sigma_A > n) \le a, \qquad \mu\text{-a.s.}$$

and then $\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x(\sigma_A > n) \le \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\sigma_A > n)$, μ-a.s. Now by (4.17) in Lemma 4.5, we have

$$\begin{aligned}
\log \lambda_0(A) &= \mu\text{-}\operatorname*{esssup}_{x \in A} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x(\sigma_A = n + 1) \\
&\le \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu(\sigma_A \ge n + 1) \\
&= \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_\mu\big(\tilde{L}_n(A^c) = 0\big) \\
&\le - \lim_{\varepsilon\to 0+} \inf\big\{I(\nu) | \nu \in M_1(\mathcal{X}), \nu\big(A^c\big) \le \varepsilon\big\},
\end{aligned}$$

where the last inequality follows by Lemma 4.3 (the inequality (4.9)) by setting $V = -1_{A^c}$ and $r = \mu(A^c) - \varepsilon$.

Now by our assumed $(\alpha - T_{\mathcal{V}}I)$, for any $\nu \in M_1(\mathcal{X})$ with $\nu(A^c) \le \varepsilon$, $\|\nu - \mu\|_{\mathrm{TV}} \ge 2[\mu(A^c) - \nu(A^c)] > 1 - 2\varepsilon$, so $I(\nu) \ge \alpha(\|\nu - \mu\|_{\mathrm{TV}}) \ge \alpha(1-2\varepsilon)$. Hence $\log \lambda_0(A) \le -\alpha(1)$ by the left-continuity of α. Plugging it into the previous estimate of c_P, we obtain finally

$$c_P \le \frac{1}{1 - e^{-\alpha(1)}}$$

the desired result. □

REMARK 6.1. Recall the following inequality in [26]: for $\forall\mu$-probability density f, that is, $f \ge 0$ and $\mu(f) = 1$:

$$\|f\mu - \mu\|_{\mathrm{TV}}^2 \le 4\,\mathrm{Var}_\mu(\sqrt{f}) \le 4c_P I_c(\nu) \le 4c_P I(\nu) \tag{6.1}$$

which is slightly less sharp than that of Léon–Perron in part (a) of Theorem 2.14.

6.2. *Lipschtzian spectral gap criteria.*

PROOF OF THEOREM 2.16. The first inequality in (2.35) is established in [24]. For the completeness we give the details. Given a probability measure $\nu = f\mu$ and $h = \sqrt{f}$, for any bounded and d-Lipschitzian function g with $\|g\|_{\mathrm{Lip}} \le 1$ and $\mu(g) = 0$, letting G be the solution of Poisson equation $(I - P)G = g$ with $\mu(G) = 0$, we have

$$\begin{aligned}
\nu(g) &= \langle g, f\rangle_\mu = \big\langle (I-P)G, f\big\rangle_\mu \\
&= \frac{1}{2}\iint_{\mathcal{X}^2} \big(G(y) - G(x)\big)\big(h(y) - h(x)\big)\big(h(x) + h(y)\big)\mu(dx)P(x,dy) \\
&= \iint_{\mathcal{X}^2} \big(G(y) - G(x)\big)\big(h(y) - h(x)\big)h(x)\mu(dx)P(x,dy) \\
&\le \sqrt{\iint_{\mathcal{X}^2} \big(h(y) - h(x)\big)^2 \mu(dx)P(x,dy)} \\
&\quad \bullet \sqrt{\iint_{\mathcal{X}^2} \big(G(y) - G(x)\big)^2 h^2(x)\mu(dx)P(x,dy)}.
\end{aligned}$$

The first square root is exactly $\sqrt{2I_c(\nu)}$, and the second is $\sqrt{2\int \Gamma(G)h^2\,d\mu}$. In other words we have proved

$$\nu(g) \le 2\sqrt{I_c(\nu)} \cdot \sqrt{\int \Gamma(G)\,d\nu}. \tag{6.2}$$

Since

$$\int_{\mathcal{X}} \Gamma(G)(x)\nu(dx) \le \big\|\Gamma(G)\big\|_\infty \le \big[c_{\mathrm{Lip}}(d,\rho)M\big]^2,$$

thus taking the supremum over all such g, we get

$$W_1(\nu,\mu)^2 \le 4\big(Mc_{\mathrm{Lip}}(d,\rho)\big)^2 I_c(\nu).$$

As $I_c(\nu) \le I(\nu)$ by Lemma 4.4, the second inequality in (2.35) holds. Finally (2.36) follows from (2.35) by Corollary 2.6. □

PROOF OF COROLLARY 2.18. At first let g, G be as in the proof of Theorem 2.16. Since the oscillation $\|\Gamma(G)\|_{\mathrm{osc}}$ of $\Gamma(G)$ is less than or equal to $M^2(c_{\mathrm{Lip}}(d,\rho))^2$, by (6.1) we have

$$\begin{aligned}\int_{\mathcal{X}} \Gamma[G](x)\nu(dx) &\le \mu(\Gamma(G)) + \left(Mc_{\mathrm{Lip}}(d,\rho)\right)^2\sqrt{c_P I_c(\nu)}\\ &= \frac{1}{2}V_c(g) + \left(Mc_{\mathrm{Lip}}(d,\rho)\right)^2\sqrt{c_P I_c(\nu)}\end{aligned}$$

substituting this estimate into (6.2), we get

$$\nu(g) \le \sqrt{I_c(\nu)}\cdot\sqrt{2V_c(g) + 4\left(Mc_{\mathrm{Lip}}(d,\rho)\right)^2\sqrt{c_P I_c(\nu)}}$$

which is the first inequality in (2.38). The second inequality in (2.38) holds for $I_c(\nu) \le I(\nu)$. Finally the last Bernstein-type concentration inequality in this corollary follows from (2.38) by Theorem 2.4. $\square$

PROOF OF PROPOSITION 2.19. We shall apply the technique of forward-backward martingale decomposition in the second named author's previous work [59]. Given a Lipschitzian function g with $\mu(g) = 0$, $\|g\|_{\mathrm{Lip}} = 1$, let G be the solution of the Poisson equation $(I - P)G = g$. Consider

$$M_n := \sum_{k=1}^{n}[G(X_k) - PG(X_{k-1})], \qquad M_n^* = \sum_{k=0}^{n-1}[G(X_k) - PG(X_{k+1})]$$

which are respectively $\mathbb{P}_\mu$ forward and backward martingales. We see

$$\begin{aligned}M_n + M_n^* &= \left((I-P)G(X_0) + (I-P)G(X_n)\right) + 2\sum_{k=1}^{n-1}(I-P)G(X_k)\\ &= 2n\tilde{L}_n(g).\end{aligned}$$

Since M_n, M_n^* have the same law, we have by Jensen's inequality that for any convex function Φ on $\mathbb{R}$,

$$\mathbb{E}_\mu\Phi\left(n\tilde{L}_n(g)\right) = \mathbb{E}_\mu\Phi\left(\frac{M_n + M_n^*}{2}\right) \le \mathbb{E}_\mu\Phi(M_n). \tag{6.3}$$

Applying it to $\Phi(x) = e^{\lambda x}$, $\lambda > 0$, we get

$$\mathbb{E}_\mu e^{\lambda n\tilde{L}_n(g)} \le \mathbb{E}_\mu e^{\lambda M_n}.$$

But by the assumed W_1H-inequality for $P(x,\cdot)$

$$\begin{aligned}\mathbb{E}_\mu\left(e^{\lambda M_n}|\mathcal{F}_{n-1}\right) &\le e^{\lambda M_{n-1}}\exp\left(\lambda^2 c_H(P,\rho)\|G\|^2_{\mathrm{Lip}(\rho)}/2\right)\\ &\le e^{\lambda M_{n-1}}\exp\left(\lambda^2 c_H(P,\rho)\left(c_{\mathrm{Lip}}(d,\rho)\right)^2/2\right)\end{aligned}$$

and successively we get

$$\mathbb{E}_\mu e^{\lambda M_n} \le \exp\left(n\lambda^2 c_H(P,\rho)\left(c_{\mathrm{Lip}}(d,\rho)\right)^2/2\right).$$

Now it remains to apply Theorem 2.4((b′) $\Longrightarrow$ (a)). $\square$

6.3. *Hyperbounedness implies $T_{\mathcal{V}}I$.* Recall that the norm of P from $L^p(\mu)$ to $L^q(\mu)$ is $\|P\|_{p,q} := \sup_{f:\|f\|_p=1} \|Pf\|_q$, where $1 \le p, q \le \infty$.

LEMMA 6.2. *Assume that for some $p > 2$, $\|P\|_{2,p} < +\infty$. Then for any $N \ge 1$,*

$$\|P^N\|_{2,p_N} \le \big(\|P\|_{2,p}\big)^{\frac{2}{p_0}+\frac{2}{p_1}+\cdots+\frac{2}{p_{N-1}}} \le \big(\|P\|_{2,p}\big)^{\frac{p}{p-2}},$$

$$\left\|\frac{d(\beta P^N)}{d\mu}\right\|_2 \le \big(\|P\|_{2,p}\big)^{\frac{p}{p-2}} \left\|\frac{d\beta}{d\mu}\right\|_{q_N},$$

where $p_N := p^N/2^{N-1}$, $q_N = \frac{1}{1-2^{N-1}/p^N}$ $(N \ge 0)$ as in Theorem 2.26.

PROOF. Since $\|P\|_{\infty,\infty} = 1$, by Riesz–Thorin's theorem we have

$$\|P\|_{p_k,p_{k+1}} \le \|P\|_{2,p}^{2/p_k}, \quad k \ge 1$$

and it holds trivially for $k = 0$ (since $p_0 = 2$). Thus

$$\|P^N\|_{2,p_N} = \|P^N\|_{p_0,p_N} \le \prod_{k=0}^{N-1} \|P\|_{p_k,p_{k+1}} \le \big(\|P\|_{2,p}\big)^{\frac{2}{p_0}+\frac{2}{p_1}+\cdots+\frac{2}{p_{N-1}}}.$$

Since $\|(P^*)^N\|_{q_N,2} = \|P^N\|_{2,p_N}$ and $\frac{d(\beta P^N)}{d\mu} = (P^*)^N \frac{d\beta}{d\mu}$, we conclude this lemma. $\square$

PROOF OF THEOREM 2.26. By (5.3), $\|e^V P\|_2^2 = \|(PP^*)^{2V}\|_2$. For any $g \in L^2(\mu)$ with $\|g\|_2 \le 1$, by Hölder's inequality we have for $c(p) := p/(p-2)$ (the conjugate number of $p/2$),

$$\int (e^V Pg)^2\, d\mu \le \|Pg\|_p^2 \left(\int e^{2c(p)V}\, d\mu\right)^{1/c(p)} \le \|P\|_{2,p}^2 \left(\int e^{2c(p)V}\, d\mu\right)^{1/c(p)},$$

where it follows that

$$\Lambda_\mu(2V|PP^*) = \log\big\|(PP^*)^{2V}\big\|_2 \le \frac{1}{c(p)} \log \int e^{2c(p)V}\, d\mu + 2\log\|P\|_{2,p}.$$

Thus by Lemma 4.1,

$$\begin{aligned} I^\sigma(\nu) &= \frac{1}{2} \sup_{V \in b\mathcal{B}} \{\nu(2V) - \log\big\|(PP^*)^{2V}\big\|_2\} \\ &\ge \sup_{V \in b\mathcal{B}} \left\{\nu(V) - \frac{1}{2c(p)} \log \int e^{2c(p)V}\, d\mu\right\} - \log\|P\|_{2,p} \\ &= \frac{1}{2c(p)} H(\nu|\mu) - \log\|P\|_{2,p}, \end{aligned}$$

where the desired inequality (2.44) follows.

Part (a) The first transport-information inequality is obvious by (2.44). For (2.45), recall the classical Bennett's inequality in the i.i.d. case: if $\mu(u) = 0$, $|u| \le 1$,

$$\mathbb{P}_\mu(L_n(u) > r) \le \exp\left(-n\frac{r}{2}\log\left(1 + \frac{r}{\mu(u^2)}\right)\right) \quad \forall n, r > 0$$

which, by Gozlan–Léonard's theorem (i.e., Theorem 2.4 in the i.i.d. case), is equivalent to

$$\alpha\big(\nu(u) - \mu(u)\big) \le H(\nu|\mu), \qquad \alpha(r) := \frac{r}{2}\log\left(1 + \frac{r}{\mu(u^2)}\right).$$

Hence we have $\alpha(\nu(u)-\mu(u)) \le 2c(p)I^\sigma(\nu)$, $\forall \nu \in M_1(\mathcal{X})$, which yields (2.45) by Theorem 2.12 (applied to $\mathcal{V}=\{(u,u)\}$) and Lemma 6.2.

Part (b). For (2.47), we may assume that g is bounded by approximation. Given a bounded measurable function g, let $u = g-\mu(g)$. For any $\nu = f\mu$, we have

$$
\begin{aligned}
\nu(u) = \langle u, f\rangle_\mu &= \frac{1}{2}\iint \big(u(y)-u(x)\big)\big(f(y)-f(x)\big)\, d\mu(x)\, d\mu(y) \\
&= \iint \big(u(y)-u(x)\big)\big(\sqrt{f(y)}-\sqrt{f(x)}\big)\sqrt{f(x)}\, d\mu(x)\, d\mu(y) \\
&\le \sqrt{\iint \big(u(y)-u(x)\big)^2 f(x)\, d\mu(x)\, d\mu(y)\cdot \iint \big(\sqrt{f(y)}-\sqrt{f(x)}\big)^2\, d\mu(x)\, d\mu(y)} \\
&= \sqrt{2\mathrm{Var}_\mu(\sqrt{f})\Big[\mu(u^2)+\int u(x)^2 f(x)\, d\mu(x)\Big]}.
\end{aligned}
\tag{6.4}
$$

By Donsker–Varadhan's variation formula of entropy and (2.44), we have for any $\delta>0$,

$$
\begin{aligned}
\int u(x)^2 f(x)\, d\mu(x) &\le \frac{1}{\delta}\Big[H(\nu|\mu)+\log\int e^{\delta u^2}\, d\mu\Big] \\
&\le \frac{2p}{\delta(p-2)}\big[I^\sigma(\nu)+\log\|P\|_{2,p}\big]+b(\delta),
\end{aligned}
$$

where $b(\delta) := \frac{1}{\delta}\log\int e^{\delta u^2}\, d\mu$. Substituting it into (6.4), we get

$$
\nu(u) \le \sqrt{2\mathrm{Var}_\mu(\sqrt{f})\Big\{\mu(u^2)+\frac{2p}{\delta(p-2)}\big[I^\sigma(\nu)+\log\|P\|_{2,p}\big]+b(\delta)\Big\}}. \tag{6.5}
$$

Let us bound $\mathrm{Var}_\mu(\sqrt{f})$ in terms of $I^\sigma(\nu)$. At first $\mathrm{Var}_\mu(\sqrt{f}) = 1-\mu(\sqrt{f})^2 \le 1$. Since $\|P\|_{L_0^2(\mu)} = \|P^*\|_{L_0^2(\mu)} = \sqrt{\|PP^*\|_{L_0^2(\mu)}}$, by the contraction condition (2.46) we have $\|PP^*\|_{L_0^2(\mu)} \le \gamma^2 < 1$. Thus

$$
\begin{aligned}
\mathrm{Var}_\mu(\sqrt{f}) &\le \frac{1}{1-\gamma^2}\big\langle \sqrt{f}-\mu(\sqrt{f}), (I-PP^*)\big(\sqrt{f}-\mu(\sqrt{f})\big)\big\rangle_\mu \\
&= \frac{1}{1-\gamma^2}\big\langle \sqrt{f}, (I-PP^*)\sqrt{f}\big\rangle_\mu \\
&\le \frac{1}{1-\gamma^2} I\big(\nu|PP^*\big) = \frac{2}{1-\gamma^2} I^\sigma(\nu),
\end{aligned}
$$

where the inequality in the third line follows by applying Lemma 4.4 to the symmetric kernel PP^*. Hence we obtain $\mathrm{Var}_\mu(\sqrt{f}) \le \min\{1, \frac{2}{1-\gamma^2} I^\sigma(\nu)\}$. Plugging it into the inequality above we get

$$
\begin{aligned}
&\nu(u) \\
&\quad \le \sqrt{2\Big[\frac{2p}{\delta(p-2)}+\frac{2}{1-\gamma^2}\Big\{\mu(u^2)+b(\delta)+\frac{2p}{\delta(p-2)}\log\|P\|_{2,p}\Big\}\Big]\cdot I^\sigma(\nu)}
\end{aligned}
\tag{6.6}
$$

which is exactly (2.47). By Theorem 2.12 applied to $\mathcal{V}=\{(g,g)\}$, the inequality above implies that for all $n, r>0$,

$$
\begin{aligned}
\mathbb{P}_\beta\big(L_n(g)\circ\theta^N > \mu(g)+r\big) &= \mathbb{P}_{\beta P^N}\big(L_n(g)>\mu(g)+r\big) \\
&\le \Big\|\frac{d\beta P^N}{d\mu}\Big\|_2 \exp\Big(-n\frac{r^2}{2C_G(g)}\Big).
\end{aligned}
$$

As $\|\frac{d\beta P^N}{d\mu}\|_2 \le \|\frac{d\beta}{d\mu}\|_{q_N} \|P\|_{2,p}^{\frac{p}{p-2}}$ by Lemma 6.2, we get the concentration inequality in part (b).

Taking the supremum in (6.6) over all bounded Lipschitzian functions g such that $\|g\|_{\mathrm{Lip}} = 1$, we get the desired $W_1 I^\sigma$-inequality (2.49) with the sub-Gaussian constant C_G^σ satisfying (2.50). □

PROOF OF THEOREM 2.27. $\|P\|_{2,\infty} = \sup_{p>2} \|P\|_{2,p} = \lim_{p\to\infty} \|P\|_{2,p}$, so the $W_1 I^\sigma$-inequality (2.49) holds with the constant C_G^σ bounded by the limit of the r.h.s. of (2.50) as $p \to \infty$. Now the concentration inequality (2.51) follows by Theorem 2.12 and the fact that

$$\left\|\frac{d(\beta P)}{d\mu}\right\|_2 = \|P^*\|_{1,2} \cdot \left\|\frac{d\beta}{d\mu}\right\|_1 = \|P\|_{2,\infty}. \qquad \square$$

7. Lyapunov function criterion: Proof of Theorem 2.33. The starting point is the following large deviation result.

LEMMA 7.1. *For every measurable function $U \ge 1$ such that $\log \frac{U}{PU} \ge -b$, μ-a.e. for some constant $b > 0$, then*

$$\int \log \frac{U}{PU}\, d\nu \le I(\nu) \quad \forall \nu \in M_1(\mathcal{X}). \tag{7.1}$$

PROOF. When U is bounded, this is contained in the definition (2.1) of $I(\nu) = I(\nu|P,\mu)$. Now for U unbounded, considering $U \wedge N$ for each $N \ge 1$, we have

$$I(\nu) \ge \int \log \frac{U \wedge N}{P(U \wedge N)}\, d\nu.$$

By Fatou's lemma, it remains to show that $\{\log \frac{U\wedge N}{P(U\wedge N)}, N \ge 1\}$ is bounded from below by some constant. Since $P(U \wedge N) \le (PU) \wedge N$ and $U \ge (PU)e^{-b}$, we see that if $PU \le N$,

$$\log \frac{U \wedge N}{P(U \wedge N)} \ge \log \frac{U \wedge N}{(PU) \wedge N} \ge \log \frac{e^{-b} PU}{PU} = -b$$

and if $PU > N$,

$$\log \frac{U \wedge N}{P(U \wedge N)} \ge \log \frac{U \wedge N}{N \wedge PU} \ge \log \frac{N e^{-b}}{N} = -b$$

that is the desired lower boundedness. □

PROOF OF THEOREM 2.33. As noticed in Remark 2.36, it is enough to show (2.54). To that end we may assume that $I(\nu) = I(\nu|P,\mu) < +\infty$, then ν is absolutely continuous w.r.t. μ with density f. Now for any $u : \mathcal{X} \to \mathbb{R}$ such that $u^2 \le \phi$, we have by Cauchy–Schwarz's inequality

$$\begin{aligned}
\int u\, d(\nu - \mu) &= \int u(f-1)\, d\mu \\
&\le \left(\int u^2(\sqrt{f}+1)^2\, d\mu \int (\sqrt{f}-1)^2\, d\mu\right)^{1/2} \\
&\le \left(2\int u^2(f+1)\, d\mu \cdot 2[1 - \mu(\sqrt{f})]\right)^{1/2}.
\end{aligned}$$

But by Lemma 7.1,

$$\int u^2(f+1)\,d\mu \le \mu(u^2) + \int \phi f\,d\mu \le \mu(\phi) + \int\left(b + \log\frac{U}{PU}\right) f\,d\mu$$
$$\le \mu(\phi) + b + I(\nu)$$

and $1-\mu(\sqrt{f}) \le 1-(\mu(\sqrt{f}))^2 = \mathrm{Var}_\mu(\sqrt{f}) \le \min\{1, c_P I(\nu)\}$ by the Poincaré inequality and Lemma 4.4, we get thus

$$\left(\int u\,d(\nu-\mu)\right)^2 \le 4\big(\mu(\phi)+b+I(\nu)\big)\min\{1, c_P I(\nu)\},$$

hence (2.54) follows by taking the supremum over all u so that $|u| \le \sqrt{\phi}$. $\square$

Added Remark. One of the referees informed us of the paper [57] by S. Watanabe and M. Hayashi, that we were unaware of. Watanabe and Hayashi established not only the concentration inequality (upper bound) in [57], Theorem 8.1, but also the lower bound [57], Theorem 8.2, for the Markov chains in the finite states space. Their upper bound is sharp both in the range of large deviations and in that of moderate deviations. Let us do some comparisons of [57] with our work.

(a) For $\mathcal{V} = \{(u,u)\}$ where $u \in b\mathcal{B}$ is fixed, by Lemma 4.3 (due to Lei [29] under the uniform integrability of P on $L^2(\mu)$), we have

$$\mathbb{P}_\beta\big(\tilde{L}_n(u) > \mu(u)+r\big) \le \left\|\frac{d\beta}{d\mu}\right\|_2 \cdot e^{-nI_u(\mu(u)+r)}, \quad r>0, n\ge 1, \tag{7.2}$$

where $I_u(x) := \inf\{I(\nu); \nu(u) = x\}$, $x \in \mathbb{R}$. Equivalently $(\alpha - T_{\mathcal{V}} I)$ holds for $\alpha(r) = I_u((\mu(u)+r)-) := \lim_{\varepsilon\to 0+} I_u((\mu(u)+r)-\varepsilon)$ for $r>0$ (and $\alpha(r)=0$ for $r \le 0$ by our convention).

As I_u is the rate function governing the large deviation principle of $\tilde{L}_n(u)$, this concentration inequality is sharp in the domain of large deviations. When the spectral gap of P on $L^2(\mu)$ exists, by [65], the moderate large deviation principle for $L_n(u)$ or $\tilde{L}_n(u)$ holds, and

$$I_u'(\mu(u)) = 0, \qquad I_u''(\mu(u)) = \frac{1}{\sigma^2(u)},$$

where $\sigma^2(u) = \lim_{n\to\infty} n\mathrm{Var}_{\mathbb{P}_\mu}(L_n(u))$ is the asymptotic variance in the central limit theorem. Hence the concentration inequality above is also sharp in the domain of moderate deviations.

In comparison with the concentration inequality [57], Theorem 8.1, of S. Watanabe and M. Hayashi in the finite states space case, which, applied to $L_n(u)\circ\theta$, is read as: for all $r>0$, $\lambda>0$, $n\in\mathbb{N}^*$,

$$\mathbb{P}_\beta\big(L_n(u)\circ\theta > \mu(u)+r\big) \le \beta(v_\lambda)\cdot\exp\big(-n\big[\lambda\big(\mu(u)+r\big) - \Lambda_u(\lambda)\big]\big), \tag{7.3}$$

where v_λ is the eigenfunction with minimum equal to 1 of $e^{\lambda u}P$ associated with its largest eigenvalue $e^{\Lambda_u(\lambda)}$. Their exponentially (in n) small term, if optimized over $\lambda>0$, equals to $e^{-n\alpha(r)}$. But their constant factor term $\beta(v_\lambda)$, instead of our $\|d\beta/d\mu\|_2$, depends upon the eigenfunction v_λ. The latter is very difficult to be controlled in the general states space (already difficult in the case of finite states space). We emphasize that their result works in the nonreversible case and is also valid for $\frac{1}{n}\sum_{k=1}^n g(X_k, X_{k+1})$ (note that $(Y_k := (X_k, X_{k+1}))_{k\in\mathbb{N}}$ forms a nonreversible Markov chain even if (X_k) is reversible). The concentration inequality (7.2) succeeds in avoiding their difficult-to-bound constant factor (because of the use of $\tilde{L}_n(u)$ instead of $L_n(u)$), but only in the symmetric case.

(b) Our investigation is in the spirit of transport inequalities and our purpose is the classical concentration inequalities such as those of Hoeffding, Bernstein, Bennett etc. That is why our deviation function $\alpha(r)$ is simple and explicit (even quadratic in the most part of our investigation as in Hoeffding's inequality), valid for a class of functions. Their deviation function is in a variational form (Legendre transform), only for $\mathcal{V} = \{(u, u)\}$, a single function.

(c) Our method is completely different from theirs: our starting point is Theorem 2.4 and then all our studies are concentrated on proving the transport-information inequalities; their method consists in the control of the log-Laplace transform, as in [11, 21, 30, 31] when the spectral gap exists.

(d) Their lower bound, new and original, is not at all studied in our paper because of the limitation of our method.

Acknowledgements. We are grateful to the two referees for their conscientious comments, suggestions and encouragements, which improve the presentation and the readability of our paper.

N.-Y. Wang is supported by National Natural Science Foundation of China under Grant No. 11601170. L. Wu dedicates this paper to the memory of his wife Xia Jing.

REFERENCES

[1] BOBKOV, S. G., GENTIL, I. and LEDOUX, M. (2001). Hypercontractivity of Hamilton–Jacobi equations. *J. Math. Pures Appl.* (9) **80** 669–696. MR1846020 https://doi.org/10.1016/S0021-7824(01)01208-9

[2] BOBKOV, S. G. and GÖTZE, F. (1999). Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** 1–28. MR1682772 https://doi.org/10.1006/jfan.1998.3326

[3] BOBKOV, S. G., HOUDRÉ, C. and TETALI, P. (2006). The subgaussian constant and concentration inequalities. *Israel J. Math.* **156** 255–283. MR2282379 https://doi.org/10.1007/BF02773835

[4] BOLLEY, F. and VILLANI, C. (2005). Weighted Csiszár–Kullback–Pinsker inequalities and applications to transportation inequalities. *Ann. Fac. Sci. Toulouse Math.* (6) **14** 331–352. MR2172583

[5] CHEN, M.-F. (2005). *Eigenvalues, Inequalities, and Ergodic Theory. Probability and Its Applications (New York)*. Springer London, Ltd., London. MR2105651

[6] CHEN, M.-F. (2007). Exponential convergence rate in entropy. *Front. Math. China* **2** 329–358. MR2327010 https://doi.org/10.1007/s11464-007-0022-5

[7] CHEN, M. F. (1992). *From Markov Chains to Nonequilibrium Particle Systems*. World Scientific Co., Inc., River Edge, NJ. MR1168209 https://doi.org/10.1142/1389

[8] DEL MORAL, P., LEDOUX, M. and MICLO, L. (2003). On contraction properties of Markov kernels. *Probab. Theory Related Fields* **126** 395–420. MR1992499 https://doi.org/10.1007/s00440-003-0270-6

[9] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics (New York)* **38**. Springer, New York. MR1619036 https://doi.org/10.1007/978-1-4612-5320-4

[10] DEUSCHEL, J.-D. and STROOCK, D. W. (1989). *Large Deviations*. *Pure and Applied Mathematics* **137**. Academic Press, Boston, MA. MR0997938

[11] DINWOODIE, I. H. (1995). A probability inequality for the occupation measure of a reversible Markov chain. *Ann. Appl. Probab.* **5** 37–43. MR1325039

[12] DJELLOUT, H., GUILLIN, A. and WU, L. (2004). Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.* **32** 2702–2732. MR2078555 https://doi.org/10.1214/009117904000000531

[13] DJELLOUT, H. and WU, L. (2011). Lipschitzian norm estimate of one-dimensional Poisson equations and applications. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 450–465. MR2814418 https://doi.org/10.1214/10-AIHP360

[14] DOBRUŠIN, R. L. (1968). Description of a random field by means of conditional probabilities and conditions for its regularity. *Teor. Veroyatn. Primen.* **13** 201–229. MR0231434

[15] DOBRUŠIN, R. L. (1970). Definition of a system of random variables by means of conditional distributions. *Teor. Veroyatn. Primen.* **15** 469–497. MR0298716

[16] DONSKER, M. D. and VARADHAN, S. R. S. (1975). Asymptotic evaluation of certain Markov process expectations for large time. I. II. *Comm. Pure Appl. Math.* **28** 1–47; ibid. 28 (1975), 279–301. MR0386024 https://doi.org/10.1002/cpa.3160280102

[17] DONSKER, M. D. and VARADHAN, S. R. S. (1976). Asymptotic evaluation of certain Markov process expectations for large time. III. *Comm. Pure Appl. Math.* **29** 389–461. MR0428471 https://doi.org/10.1002/cpa.3160290405

[18] DONSKER, M. D. and VARADHAN, S. R. S. (1983). Asymptotic evaluation of certain Markov process expectations for large time. IV. *Comm. Pure Appl. Math.* **36** 183–212. MR0690656 https://doi.org/10.1002/cpa.3160360204

[19] GAO, F., GUILLIN, A. and WU, L. (2014). Bernstein-type concentration inequalities for symmetric Markov processes. *Theory Probab. Appl.* **58** 358–382. MR3403002 https://doi.org/10.1137/S0040585X97986667

[20] GEORGII, H.-O. (1988). *Gibbs Measures and Phase Transitions*. *De Gruyter Studies in Mathematics* **9**. de Gruyter, Berlin. MR0956646 https://doi.org/10.1515/9783110850147

[21] GILLMAN, D. (1993). A Chernoff bound for random walks on expander graphs. In 34*th Annual Symposium on Foundations of Computer Science* (*Palo Alto*, *CA*, 1993) 680–691. IEEE Comput. Soc. Press, Los Alamitos, CA. MR1328459 https://doi.org/10.1109/SFCS.1993.366819

[22] GOZLAN, N. and LÉONARD, C. (2007). A large deviation approach to some transportation cost inequalities. *Probab. Theory Related Fields* **139** 235–283. MR2322697 https://doi.org/10.1007/s00440-006-0045-y

[23] GOZLAN, N. and LÉONARD, C. (2010). Transport inequalities. A survey. *Markov Process. Related Fields* **16** 635–736. MR2895086

[24] GUILLIN, A., JOULIN, A., LÉONARD, C. and WU, L. (2007). Transport-information inequalities for Markov processes (III): Jumps processes case. Preprint.

[25] GUILLIN, A., LÉONARD, C., WANG, F. Y. and WU, L. (2007). Transport-information inequalities for Markov processes (II): Relations with other functional inequalities. Preprint.

[26] GUILLIN, A., LÉONARD, C., WU, L. and YAO, N. (2009). Transportation-information inequalities for Markov processes. *Probab. Theory Related Fields* **144** 669–695. MR2496446 https://doi.org/10.1007/s00440-008-0159-5

[27] JOULIN, A. and OLLIVIER, Y. (2010). Curvature, concentration and error estimates for Markov chain Monte Carlo. *Ann. Probab.* **38** 2418–2442. MR2683634 https://doi.org/10.1214/10-AOP541

[28] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347

[29] LEI, L. (2006). Large deviations of the kernel density estimator in $L^1(\mathbb{R}^d)$ for reversible Markov processes. *Bernoulli* **12** 65–83. MR2202321

[30] LEÓN, C. A. and PERRON, F. (2004). Optimal Hoeffding bounds for discrete reversible Markov chains. *Ann. Appl. Probab.* **14** 958–970. MR2052909 https://doi.org/10.1214/105051604000000170

[31] LEZAUD, P. (1998). Chernoff-type bound for finite Markov chains. *Ann. Appl. Probab.* **8** 849–867. MR1627795 https://doi.org/10.1214/aoap/1028903453

[32] LEZAUD, P. (2001). Chernoff and Berry–Esséen inequalities for Markov processes. *ESAIM Probab. Stat.* **5** 183–201. MR1875670 https://doi.org/10.1051/ps:2001108

[33] LIN, Y., LU, L. and YAU, S.-T. (2011). Ricci curvature of graphs. *Tohoku Math. J.* (2) **63** 605–627. MR2872958 https://doi.org/10.2748/tmj/1325886283

[34] LIU, W. and MA, Y. (2009). Spectral gap and convex concentration inequalities for birth-death processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 58–69. MR2500228 https://doi.org/10.1214/07-AIHP149

[35] LIU, W. and WU, L. (2009). Identification of the rate function for large deviations of an irreducible Markov chain. *Electron. Commun. Probab.* **14** 540–551. MR2564488 https://doi.org/10.1214/ECP.v14-1512

[36] MARTON, K. (1996). A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.* **6** 556–571. MR1392329 https://doi.org/10.1007/BF02249263

[37] MARTON, K. (1996). Bounding $\overline{d}$-distance by informational divergence: A method to prove measure concentration. *Ann. Probab.* **24** 857–866. MR1404531 https://doi.org/10.1214/aop/1039639365

[38] MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. *Communications and Control Engineering Series*. Springer London, Ltd., London. MR1287609 https://doi.org/10.1007/978-1-4471-3267-7

[39] MIASOJEDOW, B. (2014). Hoeffding's inequalities for geometrically ergodic Markov chains on general state space. *Statist. Probab. Lett.* **87** 115–120. MR3168944 https://doi.org/10.1016/j.spl.2014.01.013

[40] MICLO, L. (2015). On hyperboundedness and spectrum of Markov operators. *Invent. Math.* **200** 311–343. MR3323580 https://doi.org/10.1007/s00222-014-0538-8

[41] MILMAN, E. (2012). Properties of isoperimetric, functional and transport-entropy inequalities via concentration. *Probab. Theory Related Fields* **152** 475–507. MR2892954 https://doi.org/10.1007/s00440-010-0328-1

[42] OLLIVIER, Y. (2009). Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* **256** 810–864. MR2484937 https://doi.org/10.1016/j.jfa.2008.11.001

[43] OTTO, F. and VILLANI, C. (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** 361–400. MR1760620 https://doi.org/10.1006/jfan.1999.3557

[44] PAULIN, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20** no. 79, 32. MR3383563 https://doi.org/10.1214/EJP.v20-4039

[45] PAULIN, D. (2018). Concentration inequalities for Markov chains by Marton couplings and spectral methods. Preprint. Available at arXiv:1212.2015.

[46] PAULIN, D. (2016). Mixing and concentration by Ricci curvature. *J. Funct. Anal.* **270** 1623–1662. MR3452712 https://doi.org/10.1016/j.jfa.2015.12.010

[47] FAN, J., JIANG, B. and SUN, Q. (2018). Hoeffding's lemma for Markov chains and its applications to statistical learning. Preprint. Available at arXiv:1802.00211.

[48] RIO, E. (2000). Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes. *C. R. Acad. Sci. Paris Sér. I Math.* **330** 905–908. MR1771956 https://doi.org/10.1016/S0764-4442(00)00290-1

[49] SALOFF-COSTE, L. (1997). Lectures on finite Markov chains. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1996). *Lecture Notes in Math.* **1665** 301–413. Springer, Berlin. MR1490046 https://doi.org/10.1007/BFb0092621

[50] SAMSON, P.-M. (2000). Concentration of measure inequalities for Markov chains and Φ-mixing processes. *Ann. Probab.* **28** 416–461. MR1756011 https://doi.org/10.1214/aop/1019160125

[51] TALAGRAND, M. (1996). Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.* **6** 587–600. MR1392331 https://doi.org/10.1007/BF02249265

[52] VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483 https://doi.org/10.1090/gsm/058

[53] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454 https://doi.org/10.1007/978-3-540-71050-9

[54] WANG, F.-Y. (2004). Probability distance inequalities on Riemannian manifolds and path spaces. *J. Funct. Anal.* **206** 167–190. MR2024350 https://doi.org/10.1016/S0022-1236(02)00100-3

[55] WANG, F.-Y. (2014). Criteria of spectral gap for Markov operators. *J. Funct. Anal.* **266** 2137–2152. MR3150155 https://doi.org/10.1016/j.jfa.2013.11.016

[56] WANG, F. Y. (2005). *Functional Inequalities*, *Markov Semigroups and Spectral Theory*. Sci. Press, Beijing.

[57] WATANABE, S. and HAYASHI, M. (2017). Finite-length analysis on tail probability for Markov chain and application to simple hypothesis testing. *Ann. Appl. Probab.* **27** 811–845. MR3655854 https://doi.org/10.1214/16-AAP1216

[58] WINTENBERGER, O. (2015). Weak transport inequalities and applications to exponential and oracle inequalities. *Electron. J. Probab.* **20** no. 114, 27. MR3418546 https://doi.org/10.1214/EJP.v20-3558

[59] WU, L. (1999). Forward-backward martingale decomposition and compactness results for additive functionals of stationary ergodic Markov processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **35** 121–141. MR1678517 https://doi.org/10.1016/S0246-0203(99)80008-9

[60] WU, L. (2000). A deviation inequality for non-reversible Markov processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **36** 435–445. MR1785390 https://doi.org/10.1016/S0246-0203(00)00135-7

[61] WU, L. (2000). Uniformly integrable operators and large deviations for Markov processes. *J. Funct. Anal.* **172** 301–376. MR1753178 https://doi.org/10.1006/jfan.1999.3544

[62] WU, L. (2004). Essential spectral radius for Markov semigroups. I. Discrete time case. *Probab. Theory Related Fields* **128** 255–321. MR2031227 https://doi.org/10.1007/s00440-003-0304-0

[63] WU, L. (2006). Poincaré and transportation inequalities for Gibbs measures under the Dobrushin uniqueness condition. *Ann. Probab.* **34** 1960–1989. MR2271488 https://doi.org/10.1214/009117906000000368

[64] WU, L. (2009). Gradient estimates of Poisson equations on Riemannian manifolds and applications. *J. Funct. Anal.* **257** 4015–4033. MR2557733 https://doi.org/10.1016/j.jfa.2009.07.013

[65] WU, L. M. (1995). Moderate deviations of dependent random variables related to CLT. *Ann. Probab.* **23** 420–445. MR1330777