Electronic Journal of Probability

Electron. J. Probab. **25** (2020), no. 1, 1–16.
ISSN: 1083-6489 https://doi.org/10.1214/19-EJP403

The Poincaré inequality and quadratic transportation-variance inequalities

Yuan Liu[*†]

Abstract

It is known that the Poincaré inequality is equivalent to the quadratic transportation-variance inequality (namely $W_2^2(f\mu,\mu) \leqslant C_V \mathrm{Var}_\mu(f)$), see Jourdain [10] and most recently Ledoux [12]. We give two alternative proofs to this fact. In particular, we achieve a smaller C_V than before, which equals the double of Poincaré constant. Applying the same argument leads to more characterizations of the Poincaré inequality. Our method also yields a by-product as the equivalence between the logarithmic Sobolev inequality and strict contraction of heat flow in Wasserstein space provided that the Bakry-Émery curvature has a lower bound (here the control constants may depend on the curvature bound).

Next, we present a comparison inequality between $W_2^2(f\mu,\mu)$ and its centralization $W_2^2(f_c\mu,\mu)$ for $f_c = \frac{|\sqrt{f}-\mu(\sqrt{f})|^2}{\mathrm{Var}_\mu(\sqrt{f})}$, which may be viewed as some special counterpart of the Rothaus' lemma for relative entropy. Then it yields some new bound of $W_2^2(f\mu,\mu)$ associated to the variance of $\sqrt{f}$ rather than f. As a by-product, we have another proof to derive the quadratic transportation-information inequality from Lyapunov condition, avoiding the Bobkov-Götze's characterization of the Talagrand's inequality.

Keywords: Poincaré inequality; transportation-variance inequality; quadratic Wasserstein distance; quadratic transportation-information inequality.
AMS MSC 2010: 26D10; 60E15; 60J60.
Submitted to EJP on June 18, 2019, final version accepted on December 8, 2019.

1 Introduction

The aim of this paper is to investigate some links between the Poincaré inequality (PI for short) and various comparison inequalities of quadratic Wasserstein distance with variance. Some conclusions might be extended to abstract settings of metric measure spaces, nevertheless for simplicity, our basic framework is specified as follows. Let E be a connected complete Riemannian manifold of finite dimension, d the geodesic

[*]Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China. E-mail: liuyuan@amss.ac.cn
[†]University of Chinese Academy of Sciences, Beijing 100049, China.

distance, $\mathrm{d}x$ the volume measure, $\mathcal{P}(E)$ the collection of all probability measures on E, $\mu(\mathrm{d}x) = e^{-V(x)}\mathrm{d}x \in \mathcal{P}(E)$ with $V \in C^1(E)$, $\mathrm{L} = \Delta - \nabla V \cdot \nabla$ the μ-symmetric diffusion operator with domain $\mathbb{D}(\mathrm{L})$, and $\Gamma(f,g) = \nabla f \cdot \nabla g$ the carré du champ operator with domain $\mathbb{D}(\Gamma)$, satisfying the integration by parts formula

$$\int \Gamma(f,g)\,\mathrm{d}\mu = -\int f\mathrm{L}g\,\mathrm{d}\mu,\ \forall f \in \mathbb{D}(\Gamma), g \in \mathbb{D}(\mathrm{L}).$$

Define the L^p Wasserstein (transportation) distance (also called Kantorovich metric) between $\nu, \mu \in \mathcal{P}(E)$ for any $p \geqslant 1$ by

$$W_p(\nu,\mu) = \left(\inf_{\pi \in \mathcal{C}(\nu,\mu)} \int_{E\times E} d^p(x,y)\pi(\mathrm{d}x,\mathrm{d}y)\right)^{1/p},$$

where $\mathcal{C}(\nu,\mu)$ denotes the set of any coupling π on $E \times E$ with marginals ν and μ respectively. Throughout this paper we focus on quadratic Wasserstein distance, so it is convenient to assume μ has a finite moment of order 2. The reader is referred to several constant references as Bakry-Gentil-Ledoux [2] and Villani [16, 17] for detailed presentations.

Our motivation partially arises from the problem of how to characterize the exponential decay of quadratic Wasserstein distance along heat flow. It is known that the exponential decay of heat semigroup $P_t = \exp(t\mathrm{L})$ in L^2-norm is equivalent to PI, which reads for any $f \in \mathbb{D}(\Gamma) \cap L^2(\mu)$

$$\mathrm{Var}_\mu(P_t f) \leqslant e^{-2t/C_P}\mathrm{Var}_\mu(f) \iff \mathrm{Var}_\mu(f) \leqslant C_P \int \Gamma(f,f)\mathrm{d}\mu$$

(simply denote by $\mu(h) = \int h\mathrm{d}\mu$ the expectation and by $\mathrm{Var}_\mu(f) = \mu(f^2) - (\mu(f))^2$ the variance). Similarly, the exponential decay of P_t in relative entropy is equivalent to the logarithmic Sobolev inequality (LSI for short), which reads for any $f > 0$ with $\sqrt{f} \in \mathbb{D}(\Gamma)$

$$\mathrm{Ent}_\mu(P_t f) \leqslant e^{-2t/C_{LS}}\mathrm{Ent}_\mu(f) \iff \mathrm{Ent}_\mu(f) \leqslant \frac{1}{2}C_{LS}\mathrm{I}_\mu(f)$$

(denote by $\mathrm{Ent}_\mu(f) = \int f\log f\mathrm{d}\mu$ the relative entropy and by $\mathrm{I}_\mu(f) = \int \frac{\Gamma(f,f)}{f}\mathrm{d}\mu$ the Fisher information). Somehow, we think it is tough to give a proper answer to the same question in Wasserstein space, namely to find some equivalent inequality characterizing $W_2^2(P_t\nu,\mu) \leqslant e^{-2\kappa t}W_2^2(\nu,\mu)$ (or up to a multiple) with $\kappa > 0$ for any $\nu = f\mu \in \mathcal{P}(E)$. When we turn to some weak replacements, one natural candidate is to compare W_2 with variance, which can be quickly derived from the control inequality of weighted total variation (see [16, Proposition 7.10]) and Hölder inequality that

$$W_2^2(\nu,\mu) \leqslant 2||d^2(x_0,\cdot)(\nu-\mu)||_{\mathrm{TV}} \leqslant 2\int d^2(x_0,\cdot)\,|f-1|\,\mathrm{d}\mu \leqslant C\sqrt{\mathrm{Var}_\mu(f)}$$

if $d^4(x_0,\cdot)$ is μ-integrable. At least, it follows the integrability of $W_2^2(P_t\nu,\mu)$ for $t \in [0,\infty)$ provided that PI holds true, which is helpful to the semigroup analysis more or less.

If μ fulfills the Talagrand's inequality ($\mathrm{W_2H}$ for short), namely the control of relative entropy on $W_2(\nu,\mu)$ as

$$W_2^2(\nu,\mu) \leqslant 2C_T\mathrm{Ent}_\mu(f),$$

it follows from the preliminary inequality $\mathrm{Ent}_\mu(f) \leqslant p\,(\mathrm{Var}_\mu(f))^{\frac{1}{p}}$ for $p \geqslant 1$ that

$$W_2^2(\nu,\mu) \leqslant 2C_T p\,(\mathrm{Var}_\mu(f))^{\frac{1}{p}}.$$

In particular, for $p=2$ it covers $W_2^2(\nu,\mu)\leqslant C\sqrt{\mathrm{Var}_\mu(f)}$, and for $p=1$ it gives

$$W_2^2(\nu,\mu)\leqslant 2C_T\mathrm{Var}_\mu(f), \tag{1.1}$$

which suggests an improved decay rate of W_2 along heat flow. Since $\mathrm{W_2H}$ implies PI with $C_P\leqslant C_T$ (see [2] for example), it is natural to ask what about the relation between PI and a transportation-variance inequality like (1.1). Indeed, Jourdain [10] proved their equivalence in dimension one. Ding [6] claimed a general inequality between W_2 and the so called Rényi-Tsallis divergence of order α, which equals the variance for $\alpha=2$ (somehow, it is obscure for us to check Remark 3.3 therein for small variance, maybe we misunderstand something). Then Ledoux [12] provided a very streamlined proof to show a general result that PI is equivalent to the quadratic transportation-variance inequality ($\mathrm{W_2V}$ for short)

$$W_2^2(\nu,\mu)\leqslant C_V\mathrm{Var}_\mu(f)$$

for $C_V\leqslant 4C_P$. We give two alternative proofs to this fact and achieve a smaller constant as $C_V\leqslant 2C_P$. Conversely, various perturbation techniques ensure PI with a constant no more than C_V if assume $\mathrm{W_2V}$ (see [12]). Precisely, our first main result is the following.

Theorem 1.1. *Let $\nu=f\mu\in\mathcal{P}(E)$. The Poincaré inequality implies next every inequality:*

1. $W_2^2(\nu,\mu)\leqslant 2C_P\sqrt{\mathrm{Var}_\mu(f)}\cdot\sqrt{\mathrm{Ent}_\mu(f)}$.
2. $W_2^2(\nu,\mu)\leqslant 2C_P\mathrm{Var}_\mu(f)$.
3. $W_2^2(\nu,\mu)\leqslant 2C_P\inf\limits_{p\geqslant 1}\left\{p^2\left(\mathrm{Var}_\mu(f)\right)^{\frac{1}{p}}\right\}$.
4. $W_2^2(\nu,\mu)\leqslant 2C_P\inf\limits_{p\geqslant 1}\left\{p^2\big(C_P\mu(\Gamma(f,f))\big)^{\frac{1}{p}}\right\}$.
5. $W_2^2(\nu,\mu)\leqslant 2C_P^2\mu(\Gamma(f,f))$.

Conversely, the above every one implies the Poincaré inequality with constant $\sqrt{2}C_P$.

Remark 1.2. If assume (1) or (5) prior to PI, the perturbation technique ensures PI with constant $\sqrt{2}C_P$. Note that the same technique doesn't work for (2) directly.

There are two approaches to this end, and both are contributed to get the inequality (see also (2.1) below)

$$W_2^2(\nu,\mu)\leqslant 2\sqrt{\mathrm{Ent}_\mu(f)}\int_0^\infty\sqrt{\mathrm{Ent}_\mu(P_tf)}\mathrm{d}t.$$

The first approach is a shortcut based on the interpolation technique developed by Kuwada [11] and further by [12]. The other one appeals to the derivative formula of $W_2^2(P_tf\mu,\mu)$ in t (almost everywhere), which is slightly different from what Otto-Villani employed in [15, Lemma 2]. Our method doesn't involve the theory of solving Fokker-Planck equation on Riemannian manifolds, so we have a by-product as reproving their lemma for nice initial data but avoiding the curvature condition.

Another by-product is to show the equivalence between the LSI and strict contraction of heat flow in Wasserstein space (here we actually mean a strictly exponential decay of $W_2(P_tf\mu,\mu)$ with some multiple in front) provided that the Bakry-Émery curvature has a lower bound. One can compare the following with the well known characterization of curvature-dimension condition through the heat flow contraction (see [2, Theorem 9.7.2] for this fact and [2, Subsection 3.4.5] for precise definition of curvature-dimension condition $CD(\rho,\infty)$).

Proposition 1.3. *Assume V is a smooth potential such that the curvature-dimension condition $CD(\rho,\infty)$ holds for $\rho\in\mathbb{R}$. Then the next two statements are equivalent:*

1. *there exist two constants $C > 0$ and $\kappa > 0$ such that for all $t > 0$ and any $\nu = f\mu \in \mathcal{P}(E)$*
$$W_2(P_t\nu, \mu) \leqslant Ce^{-\kappa t} W_2(\nu, \mu);$$
2. *there exists a constant $C_{LS} > 0$ such that the LSI holds.*

Remark 1.4. The constants involved here may depend on ρ. If the LSI holds, we have $\kappa = 1/C_{LS}$. Very recently, Wang [19] discussed exponential contraction in any W_p $(p \geqslant 1)$ for a class of diffusion semigroups and gave the implication from (2) to (1) as well.

Next, we are interested in the comparison of $W_2^2(\nu, \mu)$ to $\mathrm{Var}_\mu(\sqrt{f})$ rather than $\mathrm{Var}_\mu(f)$. In general, one can't expect a strong inequality as $W_2^2(\nu, \mu) \leqslant C\mathrm{Var}_\mu(\sqrt{f})$, since from PI it follows $W_2^2(\nu, \mu) \leqslant \frac{1}{4}CC_P\mathrm{I}_\mu(f)$, which is called the quadratic transportation-information inequality ($\mathrm{W_2I}$ for short, see [9]), and it is known that $\mathrm{W_2I}$ is strictly stronger than PI and even than $\mathrm{W_2H}$. Actually what we present first is a new inequality between the Wasserstein distance and its "centralization", which may be viewed as a special counterpart of the Rothaus' lemma for relative entropy (see [2, Lemma 5.1.4]), namely for any $a \in \mathbb{R}$
$$\mathrm{Ent}_\mu\left((h + a)^2\right) \leqslant \mathrm{Ent}_\mu(h^2) + 2\mu(h^2).$$
Precisely we have

Theorem 1.5. *Let $\nu = f\mu$, $c = \mu(\sqrt{f})$ and $\sigma^2 = \mathrm{Var}_\mu(\sqrt{f})$. Let $f_c = \frac{|\sqrt{f}-c|^2}{\sigma^2}$. If the Poincaré inequality holds, then there exists two constants C_1 and C_2 such that*
$$W_2^2(\nu, \mu) \leqslant C_1\sigma^2 W_2^2(f_c\mu, \mu) + C_2\sigma^2.$$

Remark 1.6. For instance, we can take $C_1 = 2$ and $C_2 = 96C_P$. Actually our method implies that C_1 can approach 1 but should be strictly greater than 1. Moreover, f_c can be extended to $f_\theta = \frac{|\sqrt{f}-\theta|^2}{\mu((\sqrt{f}-\theta)^2)}$ for any $\theta \in (0, 2c)$ associated with two constants $C_1(\theta)$ and $C_2(\theta)$ depending on θ.

As consequence, when E has a finite diameter, it follows by the definition of W_2
$$W_2^2(\nu, \mu) \leqslant \sigma^2\left(C_1(\mathrm{diam}E)^2 + C_2\right), \tag{1.2}$$
which can't be directly concluded by Theorem 1.1 we think. Then it quickly derives $\mathrm{W_2I}$ from PI again. Moreover, a LSI holds by using the HWI inequality in [15, 16, 2] under the curvature-dimension condition $CD(\rho, \infty)$, with the control constant $C_{LS} = \lambda\big((1-\frac{\rho}{4}\lambda)\vee 1\big)$ for $\lambda = \sqrt{C_P(C_1(\mathrm{diam}E)^2 + C_2)}$. There is a lot of literature concerning LSI, for example one can compare the above (1.2) with [18, Theorem 1.4] about the constant estimate on compact manifolds by means of semigroup analysis.

When E is unbounded, we have at least by using [16, Proposition 7.10] that
$$W_2^2(\nu, \mu) \leqslant C\left(\sigma^2 + \int d^2(x_0, \cdot)(\sqrt{f} - c)^2\mathrm{d}\mu\right). \tag{1.3}$$
It gives a direct way to derive $\mathrm{W_2I}$ from the so-called Lyapunov condition. Recall [13], the Lyapunov condition here means there exists such a function $W > 0$ satisfying that W^{-1} is locally bounded and for some $c > 0, b \geqslant 0$ and $x_0 \in E$ holds in the sense of distribution
$$\mathrm{L}W \leqslant \left(-cd^2(x, x_0) + b\right)W. \tag{1.4}$$
Partial proof in [13] applied the Bobkov-Götze's characterization of $\mathrm{W_2H}$, namely there is a constant $C > 0$ such that $\mu\left(\exp(Q_Ch)\right) \leqslant \exp\left(\mu(h)\right)$ for all $h \in L^\infty(\mu)$, where Q_C denotes the infimum-convolution operator and Q_Ch solves the Hamilton-Jacobi equation $\frac{\mathrm{d}}{\mathrm{d}t}Q_th + \frac{1}{2}|\nabla Q_th|^2 = 0$ for initial data h, see [2, 3] for example. Nevertheless, facing the

stability problem for W_2H under bounded perturbation, one needs various additional curvature conditions so far, for example see [8, 14]. When we turn to the same problem for W_2I, it would be more robust if we can find a direct method to derive W_2I from (1.4) with no appearance of W_2H. Actually, Theorem 1.5 takes on such a role.

The paper is organized as follows. In next Section 2, we give a quick proof to Theorem 1.1. In Section 3 and 4, we compute the derivative of quadratic Wasserstein distance along heat flow, and then complete the other proof of Theorem 1.1. The equivalence of the LSI and strict contraction of heat flow in Wasserstein space is shown in Section 5. Section 6 is devoted to the comparison inequality about centralization of quadratic Wasserstein distance, and Section 7 provides a direct proof of W_2I under the Lyapunov condition.

2 The first proof of Theorem 1.1

Recall that, for any bounded Lipschitz function h, define its infimum-convolution for any $t > 0$ by

$$Q_t h(x) := \inf_y \left\{ h(y) + \frac{1}{2t} d^2(x,y) \right\},$$

which solves the Hamilton-Jacobi equation (see for example [2, Section 9.4], [7, Section 3.3], [16, Section 5.4])

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} u + \frac{1}{2}|\nabla u|^2 = 0, \\ u(x,0) = h(x). \end{cases}$$

According to [11, 12], for any decreasing function $\lambda \in C^1[0,+\infty)$ with $\lambda(0) = 1$ and $\lim_{t\to\infty} \lambda(t) = 0$, one has a semigroup interpolation by virtue of Hamilton-Jacobi equation, integration by parts and the Hölder inequality that

$$\begin{aligned} \int_E Q_1 hf \mathrm{d}\mu - \int_E h \mathrm{d}\mu &= \int_E \int_0^\infty -\frac{\mathrm{d}}{\mathrm{d}t} Q_\lambda h P_t f \mathrm{d}t \mathrm{d}\mu \\ &= \int_E \int_0^\infty \frac{1}{2}\lambda' |\nabla Q_\lambda h|^2 P_t f - Q_\lambda h \cdot \mathrm{L} P_t f \mathrm{d}t \mathrm{d}\mu \\ &= \int_0^\infty \int_E \frac{1}{2}\lambda' |\nabla Q_\lambda h|^2 P_t f + \nabla Q_\lambda h \cdot \nabla P_t f \mathrm{d}\mu \mathrm{d}t \\ &\leqslant \int_0^\infty -\frac{\mathrm{I}_\mu(P_t f)}{2\lambda'} \mathrm{d}t. \end{aligned}$$

Using the Kantorovich dual (see [2, Section 9.2], [16, Chapter 1]) yields for $\nu = f\mu$

$$W_2^2(\nu,\mu) = 2\sup_h \left\{ \int_E Q_1 hf \mathrm{d}\mu - \int_E h \mathrm{d}\mu \right\} \leqslant \int_0^\infty -\frac{\mathrm{I}_\mu(P_t f)}{\lambda'} \mathrm{d}t.$$

It is flexible to choose a nice λ to prove Theorem 1.1. For instance, if $\sqrt{\mathrm{Ent}_\mu(P_t f)}$ is integrable on $[0,\infty)$, let $\lambda(t) = \frac{\int_t^\infty \sqrt{\mathrm{Ent}_\mu(P_t f)}\mathrm{d}t}{\int_0^\infty \sqrt{\mathrm{Ent}_\mu(P_t f)}\mathrm{d}t}$, then it follows

$$\begin{aligned} W_2^2(\nu,\mu) &\leqslant \int_0^\infty \frac{\mathrm{I}_\mu(P_t f)}{\sqrt{\mathrm{Ent}_\mu(P_t f)}} \mathrm{d}t \cdot \int_0^\infty \sqrt{\mathrm{Ent}_\mu(P_t f)} \mathrm{d}t \\ &= 2\sqrt{\mathrm{Ent}_\mu(f)} \int_0^\infty \sqrt{\mathrm{Ent}_\mu(P_t f)} \mathrm{d}t. \end{aligned} \tag{2.1}$$

We will revisit (2.1) in Section 4 by means of derivative estimate of Wasserstein distance.

Proof. It consists of two parts.

Part 1. First of all, using the inequality $\log x \leqslant x-1$ yields that

$$\mathrm{Ent}_\mu(f) = \int f \log \frac{f}{\mu(f)} \mathrm{d}\mu \leqslant \int f \cdot \frac{f-\mu(f)}{\mu(f)} \mathrm{d}\mu = \frac{1}{\mu(f)} \mathrm{Var}_\mu(f).$$

For $\mu(f)=1$, we have $\mathrm{Ent}_\mu(f) \leqslant \mathrm{Var}_\mu(f)$. If PI holds with a constant C_P, we have further

$$\mathrm{Ent}_\mu(P_t f) \leqslant \mathrm{Var}_\mu(P_t f) \leqslant e^{-\frac{2}{C_P}t} \mathrm{Var}_\mu(f),$$

and then $\sqrt{\mathrm{Ent}_\mu(P_t f)}$ is integrable on $[0,\infty)$. It follows from (2.1) that

$$\begin{aligned} W_2^2(\nu,\mu) &\leqslant 2\sqrt{\mathrm{Ent}_\mu(f)} \int_0^\infty \sqrt{\mathrm{Ent}_\mu(P_t f)} \mathrm{d}t \\ &\leqslant 2\sqrt{\mathrm{Ent}_\mu(f)} \int_0^\infty e^{-\frac{1}{C_P}t} \sqrt{\mathrm{Var}_\mu(f)} \mathrm{d}t \ = \ 2C_P \sqrt{\mathrm{Ent}_\mu(f)} \sqrt{\mathrm{Var}_\mu(f)}. \end{aligned}$$

Inversely, assume there exists some $C>0$ such that

$$W_2^2(\nu,\mu) \leqslant 2C \sqrt{\mathrm{Ent}_\mu(f)} \sqrt{\mathrm{Var}_\mu(f)}. \tag{2.2}$$

Various perturbation techniques give PI with a constant $\sqrt{2}C$, see [12, 17] and the references therein. For completeness, we write down a sketch.

Let h be Lipschitz and bounded with $\mu(h)=0$. Let $f_t = 1+\lambda t h$ for $t \approx 0$ and some parameter $\lambda > 0$. It follows from (2.2) that

$$2\int Q_1(th) f_t \mathrm{d}\mu \leqslant W_2^2(f_t\mu,\mu) \leqslant 2C\sqrt{\mathrm{Ent}_\mu(f_t)} \cdot \sqrt{\mathrm{Var}_\mu(f_t)}.$$

Substituting the Taylor's expansion $Q_1(th) = tQ_t h = ht - \frac{1}{2}|\nabla h|^2 t^2 + o(t^2)$ at $t=0$ into the above inequality yields

$$-\mu(\Gamma(h,h)) + 2\lambda\mu(h^2) \leqslant \sqrt{2}C\lambda^2\mu(h^2), \tag{2.3}$$

which implies PI by taking $\lambda = \frac{\sqrt{2}}{2C}$. We obtain the equivalence between PI and (2.2) now.

Part 2. When we bound relative entropy by other functionals, it should lead to new types of transportation-variance inequalities. Indeed, for any $p \geqslant 1$ holds by Jensen's inequality (recall $\mu(f)=1$ here) that

$$\begin{aligned} \mathrm{Ent}_\mu(f) &= \int f \log f \mathrm{d}\mu \\ &\leqslant \log \mu(f^2) = \log(\mathrm{Var}_\mu(f)+1) \\ &\leqslant p\log((\mathrm{Var}_\mu(f))^{\frac{1}{p}}+1) \ \leqslant \ p(\mathrm{Var}_\mu(f))^{\frac{1}{p}}. \end{aligned}$$

If PI holds, it follows similarly from (2.1)

$$\begin{aligned} W_2^2(\nu,\mu) &\leqslant 2\sqrt{\mathrm{Ent}_\mu(f)} \int_0^\infty \sqrt{\mathrm{Ent}_\mu(P_t f)} \mathrm{d}t \\ &\leqslant 2p\mathrm{Var}_\mu^{\frac{1}{2p}}(f) \int_0^\infty \mathrm{Var}_\mu^{\frac{1}{2p}}(P_t f)\mathrm{d}t \ \leqslant \ 2C_P p^2 \mathrm{Var}_\mu^{\frac{1}{p}}(f), \end{aligned}$$

which covers the second inequality in Theorem 1.1 for $p=1$ and also gives the third one

$$W_2^2(\nu,\mu) \leqslant 2C_P \inf_{p\geqslant 1} \left\{ p^2 (\mathrm{Var}_\mu(f))^{\frac{1}{p}} \right\}.$$

Using PI again yields

$$W_2^2(\nu,\mu) \leqslant 2C_P \inf_{p\geqslant 1}\left\{p^2(\mathrm{Var}_\mu(f))^{\frac{1}{p}}\right\} \leqslant 2C_P \inf_{p\geqslant 1}\left\{p^2\big(C_P\mu(\Gamma(f,f))\big)^{\frac{1}{p}}\right\},$$

which gives the fourth inequality in Theorem 1.1. It follows the fifth inequality by taking $p=1$ that

$$W_2^2(\nu,\mu) \leqslant 2C_P^2\mu(\Gamma(f,f)). \tag{2.4}$$

Inversely, still following the routine of perturbation technique, (2.4) implies PI too. More precisely, recall the first part, we have a similar result as (2.3) that

$$-\mu(\Gamma(h,h)) + 2\lambda\mu(h^2) \leqslant 2C_P^2\lambda^2\mu(\Gamma(h,h)),$$

which implies PI with a constant $\sqrt{2}C_P$ by taking $\lambda = (\sqrt{2}C_P)^{-1}$. □

3 Derivative of quadratic Wasserstein distance along heat flow

In this section, we compute the derivative formula of $W_2(\nu_t,\mu)$ for $\frac{\mathrm{d}\nu_t}{\mathrm{d}\mu} = P_tf$. Recall that, in our notation, Otto-Villani [15, Lemma 2] (see [16, Subsection 9.3.4] also) was actually concerned to the upper right-hand derivative of $W_2(\nu,\nu_t)$ and found a bound as

$$\frac{\mathrm{d}}{\mathrm{d}t}^{+} W_2(\nu,\nu_t) \leqslant \limsup_{s\to 0+} W_2(\nu_t,\nu_{t+s})/s \leqslant \sqrt{\mathrm{I}_\mu(P_tf)}, \tag{3.1}$$

provided that $V\in C^2(\mathbb{R}^n)$ and $\mathrm{D}^2V \geqslant \rho\mathrm{I}$ for some $\rho\in\mathbb{R}$ (namely the curvature-dimension condition $CD(\rho,\infty)$). The difference between $W_2(\nu_t,\mu)$ and $W_2(\nu,\nu_t)$ is that the former might be integrable for $t\in[0,+\infty)$.

According to [16, Exercise 2.36], there exists $h_t\in L^1(\mu)$ such that $\mu(h_t)=0$ and $Q_1h_t\in L^1(\nu_t)$, and the conjugate pair (Q_1h_t,h_t) attains the supremum as

$$W_2^2(\nu_t,\mu) = 2\sup_{\mu(\phi)=0}\int Q_1\phi\,\mathrm{d}\nu_t = 2\int Q_1h_t\mathrm{d}\nu_t = 2\int Q_1h_tP_tf\mathrm{d}\mu. \tag{3.2}$$

Given nice initial data, we obtain the derivative formula for $W_2^2(\nu_t,\mu)$ in almost all t with no condition on curvature.

Lemma 3.1. *Assume $f\in\mathbb{D}(\mathrm{L})$ has a positive lower bound. Assume $\mathrm{L}f$ is bounded. Then for almost all $t>0$, there exists some $h_t\in L^1(\mu)$ satisfying (3.2) and*

$$\frac{\mathrm{d}}{\mathrm{d}t}W_2^2(\nu_t,\mu) = 2\int Q_1h_t\ \mathrm{L}P_tf\mathrm{d}\mu.$$

Moreover $|\frac{\mathrm{d}}{\mathrm{d}t}W_2^2(\nu_t,\mu)| \leqslant 2W(\nu_t,\mu)\sqrt{\mathrm{I}_\mu(P_tf)}$.

Proof. It consists of four steps. Note that $L^1(\nu_t)\subset L^1(\mu)$ in our case since f has a positive lower bound and then $\nu_t(|h|)\geqslant \inf f\cdot\mu(|h|)$. The assumption of $\mathrm{L}f\in L^\infty(E)$ is reasonable due to that the resolvent operator R_λ sends $C_\mathrm{b}(E)$ into $C_\mathrm{b}(E)\cap\mathbb{D}(\mathrm{L})$ and $\mathrm{L} = -R_\lambda^{-1}+\lambda\mathrm{I}$ (see for example Evans [7, Subsection 7.4.1]).

Step 1. To show the continuity of $W_2(\nu_t,\mu)$ in t.

Using the control inequality of weighted total variation (see [16, Proposition 7.10]) yields that for any $t,t'>0$

$$\begin{aligned} W_2^2(\nu_{t'},\nu_t) &\leqslant 2\int d^2(x_0,\cdot)\,|P_{t'}f-P_tf|\,\mathrm{d}\mu \\ &= 2\int d^2(x_0,\cdot)\left|\int_t^{t'}\mathrm{L}P_sf\mathrm{d}s\right|\mathrm{d}\mu \leqslant 2|t'-t|\cdot||\mathrm{L}f||_\infty\cdot\mu(d^2(x_0,\cdot)). \end{aligned}$$

It follows from the triangle inequality $|W_2(\nu_{t'},\mu) - W_2(\nu_t,\mu)| \leqslant W_2(\nu_{t'},\nu_t)$ that $W_2(\nu_t,\mu)$ is continuous in t.

Step 2. To choose a conjugate pair (Q_1h_t, h_t) satisfying (3.2) and some auxiliary "maximality" (which will be introduced in (3.3) and applied for next step).

First of all, let $(Q_1\tilde{h}_t, \tilde{h}_t) \in L^1(\nu_t) \times L^1(\mu)$ satisfy $\mu(\tilde{h}_t) = 0$ and

$$W_2^2(\nu_t,\mu) = 2\int Q_1\tilde{h}_t \mathrm{d}\nu_t.$$

$Q_1\tilde{h}_t$ may not have a gradient, so we take a sequence of bounded Lipschitz functions $\{\tilde{h}_{k,t}\}_{k\in\mathbb{N}}$ such that $\mu(\tilde{h}_{k,t}) = 0$ and $(Q_1\tilde{h}_{k,t}, \tilde{h}_{k,t})$ tends to $(Q_1\tilde{h}_t, \tilde{h}_t)$ in $L^1(\nu_t) \times L^1(\mu)$ as $k \to \infty$. Then $Q_1\tilde{h}_{k,t}$ is bounded Lipschitz too (see [7, Subsection 3.3.2]), and there exists $u_k \in [0,1]$ such that

$$\int Q_1\big((1-u_k)\tilde{h}_{k,t}\big)\mathrm{d}\nu_t = \sup_{0\leqslant u\leqslant 1}\int Q_1\big((1-u)\tilde{h}_{k,t}\big)\mathrm{d}\nu_t. \tag{3.3}$$

Denote $h_{k,t} = (1-u_k)\tilde{h}_{k,t}$.

Without loss of generality, assume $u_\infty = \lim_{k\to\infty} u_k \in [0,1]$, denote

$$h_t := (1-u_\infty)\tilde{h}_t = \lim_{k\to\infty}(1-u_k)\tilde{h}_{k,t} = \lim_{k\to\infty} h_{k,t} \in L^1(\mu). \tag{3.4}$$

We want to show that (Q_1h_t, h_t) is also a conjugate pair satisfying $W_2^2(\nu_t,\mu) = 2\int Q_1h_t\mathrm{d}\nu_t$. The difference between (Q_1h_t, h_t) and $(Q_1\tilde{h}_t, \tilde{h}_t)$ is that the former can be approximated by a special sequence of bounded Lipschitz pairs with the property (3.3).

To this end, by the definition of infimum convolution, we have first

$$h_{k,t} \geqslant Q_1h_{k,t} = (1-u_k)Q_{1-u_k}\tilde{h}_{k,t} \geqslant (1-u_k)Q_1\tilde{h}_{k,t},$$

which means that $Q_1h_{k,t}$ falls between two L^1-convergent sequences. By virtue of the Prokhorov theorem (namely the tightness argument) together with the fact of $L^1(\nu_t) \subset L^1(\mu)$, one can extract a subsequence of $Q_1h_{k,t}$ (denoted by itself for the ease of notation) converging in $L^1(\nu_t)$. Denote $\phi_t = \lim_{k\to\infty} Q_1h_{k,t}$, which satisfies

$$\phi_t(x) - h_t(y) \leqslant \frac{1}{2}d^2(x,y)$$

almost everywhere and then $\phi_t(x) \leqslant Q_1h_t(x)$ and (since $\mu(h_t) = 0$)

$$2\nu_t(\phi_t) \leqslant 2\int Q_1h_t\mathrm{d}\nu_t \leqslant W_2^2(\nu_t,\mu).$$

On the other hand, due to the definition of $h_{k,t}$ in (3.3), it follows

$$2\nu_t(\phi_t) = \lim_{k\to\infty} 2\nu_t(Q_1h_{k,t}) \geqslant \lim_{k\to\infty} 2\nu_t(Q_1\tilde{h}_{k,t}) = W_2^2(\nu_t,\mu).$$

Hence, (ϕ_t, h_t) attains the supremum of the dual Kantorovich problem too. Moreover, it follows $\phi_t = Q_1h_t$ almost everywhere with respect to ν_t and μ as well since f has a positive lower bound.

Step 3. To estimate upper and lower derivatives of $W_2^2(\nu_t,\mu)$.

For (Q_1h_t, h_t), we have

$$\begin{aligned} \underline{D}_t^+ \quad &:= \quad \liminf_{s\to 0+} \frac{W_2^2(\nu_{t+s},\mu) - W_2^2(\nu_t,\mu)}{s} \\ &\geqslant \quad \lim_{s\to 0+} \frac{2}{s}\left(\int Q_1h_t\mathrm{d}\nu_{t+s} - \int Q_1h_t\mathrm{d}\nu_t\right) \;=\; 2\int Q_1h_t\; \mathrm{L}P_tf\mathrm{d}\mu. \end{aligned} \tag{3.5}$$

Similarly, we have

$$\begin{aligned}\overline{D_t}^- &:= \limsup_{s\to 0+}\frac{W_2^2(\nu_t,\mu)-W_2^2(\nu_{t-s},\mu)}{s}\\ &\leqslant \lim_{s\to 0+}\frac{2}{s}\left(\int Q_1h_t\mathrm{d}\nu_t-\int Q_1h_t\mathrm{d}\nu_{t-s}\right) \;=\; 2\int Q_1h_t\,\mathrm{L}P_tf\mathrm{d}\mu. \end{aligned}\tag{3.6}$$

Recall the approximating sequence $(Q_1h_{k,t},h_{k,t})$ for (Q_1h_t,h_t) in Step 2, using the formula of integration by parts and the Hölder inequality yields that

$$\left|\int Q_1h_{k,t}\,\mathrm{L}P_tf\mathrm{d}\mu\right| = \left|\int \nabla Q_1h_{k,t}\,\nabla P_tf\mathrm{d}\mu\right| \leqslant \sqrt{\int |\nabla Q_1h_{k,t}|^2\mathrm{d}\nu_t}\cdot\sqrt{\mathrm{I}_\mu(P_tf)}.$$

Since $Q_sh_{k,t}$ solves the Hamilton-Jacobi equation $\frac{\mathrm{d}}{\mathrm{d}s}Q_sh_{k,t}+\frac{1}{2}|\nabla Q_sh_{k,t}|^2=0$ (see [7, Subsection 3.3.2]), we have by (3.3) (namely the integral "maximality" for $Q_1h_{k,t}$) that

$$\begin{aligned}\int|\nabla Q_1h_{k,t}|^2\mathrm{d}\nu_t &= \lim_{u\to 0+}2\int\frac{Q_{1-u}h_{k,t}-Q_1h_{k,t}}{u}\mathrm{d}\nu_t\\ &= \lim_{u\to 0+}2\int\frac{\frac{1}{1-u}Q_1\big((1-u)h_{k,t}\big)-Q_1h_{k,t}}{u}\mathrm{d}\nu_t\\ &\leqslant \lim_{u\to 0+}2\cdot\frac{\frac{1}{1-u}-1}{u}\cdot\int Q_1h_{k,t}\mathrm{d}\nu_t \;\leqslant\; W_2^2(\nu_t,\mu),\end{aligned}$$

which implies by taking $k\to\infty$

$$2\left|\int Q_1h_t\,\mathrm{L}P_tf\mathrm{d}\mu\right| = \lim_{k\to+\infty}2\left|\int Q_1h_{k,t}\,\mathrm{L}P_tf\mathrm{d}\mu\right| \leqslant 2W_2(\nu_t,\mu)\sqrt{\mathrm{I}_\mu(P_tf)} =: A_t. \tag{3.7}$$

Note that A_t is continuous in t.

Step 4. To show the Lipschitz property of $W_2^2(\nu_t,\mu)$.

For convenience, denote $F(t)=W_2^2(\nu_t,\mu)$. Heuristically, using (3.5) and (3.7) yields a local estimate that for any $t>0$ there exists $s>0$ such that $F(t+s)-F(t)\geqslant -O(s)$. It follows $F(b)-F(a)\geqslant -O(b-a)$ for any interval $[a,b]\subset\mathbb{R}^+$ if one could "find" a finite partition of $[a,b]$ and sum up all the local estimates. Similarly, using (3.6) and (3.7) yields $F(b)-F(a)\leqslant O(b-a)$, and then gives the local Lipschitz property.

The rest of the proof is basically a careful application of Borel-Lebesgue covering theorem. Fix arbitrary $\varepsilon>0$. Let $K=\sup_{t\in[a,b]}A_t+\varepsilon$. For any $t\in[a,b]$, there exists some $\eta_t\in(0,b-a]$ by using (3.5) and (3.7) such that for all $s\in(0,\eta_t]$

$$F(t+s)-F(t) > s\left(2\int Q_1h_t\,\mathrm{L}P_tf\mathrm{d}\mu-\varepsilon\right)\geqslant -s(A_t+\varepsilon)\geqslant -sK\geqslant -\eta_tK.$$

On the other hand, the continuity of $F(t)$ implies there exists $\tilde{\eta}_t\in(0,\eta_t]$ such that for all $-s\in[-\tilde{\eta}_t,0]$

$$|F(t)-F(t-s)|<\eta_tK.$$

Then the open interval $I_t=(t-\tilde{\eta}_t,t+\eta_t)$ is of length no less than η_t and no more than $2\eta_t$, and holds for any $t_2\geqslant t\geqslant t_1$ or $t\geqslant t_2\geqslant t_1$ in I_t

$$F(t_2)-F(t_1) > -2\eta_tK \geqslant -2|I_t|K. \tag{3.8}$$

(Notice that we don't know whether (3.8) is true for $t_2\geqslant t_1>t$.)

The collection of all I_t becomes an open covering of $[a,b]$, which implies a finite sub-covering $\mathcal{I}$. To reduce overlaps, we have to do some selection. Starting from $t_0=a$, one can successively take the i-th open interval I_{t_i} from $\mathcal{I}$ for $i=1,2\ldots$ satisfying next two properties:

(1). $I_{t_i} \cap I_{t_{i-1}} \neq \emptyset$, and I_{t_i} contains the right-hand endpoint of $I_{t_{i-1}}$.
(2). If there is another $I_{t_*} \in \mathcal{I}$ intersecting with $I_{t_{i-1}}$, then $I_{t_*} \subset \bigcup_{j \leqslant i} I_{t_j}$, namely the right-hand endpoint of I_{t_*} doesn't exceed I_{t_i}. It means I_{t_i} is the most effective cover than any other I_{t_*}.

This procedure will stop at time N once I_{t_N} contains b.

Now, we have a chain $I_{t_0}, I_{t_1}, \ldots, I_{t_N}$ satisfying that each element only intersects with its neighbors, which means their overlap is at most 2-fold for every point in $[a, b]$. Let $t_{i-1,i} \in I_{t_{i-1}} \cap I_{t_i}$ satisfy $t_{i-1,i} \leqslant t_i$ for $i = 1, \ldots, N$ and $a \leqslant t_{0,1} \leqslant t_{1,2} \cdots \leqslant t_{N-1,N} \leqslant b$. It must occur either $t_{i-1,i} \leqslant t_i \leqslant t_{i,i+1}$ or $t_{i-1,i} \leqslant t_{i,i+1} \leqslant t_i$ for each i. In any case, we obtain an interpolation by (3.8)

$$\begin{aligned} F(b) - F(a) &= F(b) - F(t_{N-1,N}) + \sum_{i=1}^{N-1} F(t_{i,i+1}) - F(t_{i-1,i}) + F(t_{0,1}) - F(a) \\ &\geqslant -2|I_{t_N}|K - \sum_{i=1}^{N-1} 2|I_{t_i}|K - 2|I_{t_0}|K \ \geqslant\ -8(b-a)K. \end{aligned}$$

Similarly, it follows from (3.6) and (3.7) that

$$F(b) - F(a) \leqslant 8(b-a)K.$$

Combining the above estimates yields that $F(t) = W_2^2(\nu_t, \mu)$ is locally Lipschitz and then has a derivative for almost all $t > 0$ as

$$\frac{\mathrm{d}}{\mathrm{d}t} W_2^2(\nu_t, \mu) = 2 \int Q_1 h_t \, \mathrm{L} P_t f \mathrm{d}\mu.$$

It follows that for almost all $t > 0$

$$\left| \frac{\mathrm{d}}{\mathrm{d}t} W_2^2(\nu_t, \mu) \right| \leqslant 2 W_2(\nu_t, \mu) \sqrt{\mathrm{I}_\mu(P_t f)},$$

which can be rewritten to

$$\left| \frac{\mathrm{d}}{\mathrm{d}t} W_2(\nu_t, \mu) \right| \leqslant \sqrt{\mathrm{I}_\mu(P_t f)}.$$

The proof is completed. □

Remark 3.2. It is interesting to ask further that whether $h_t = \tilde{h}_t$ almost everywhere (namely $u_\infty = 0$ in (3.4)). For any positive α and β with $\alpha + \beta = 1$, we have $\alpha Q_1 \tilde{h}_t + \beta Q_1 h_t \leqslant Q_1\big(\alpha \tilde{h}_t + \beta h_t\big)$ and

$$W_2^2(\nu_t, \mu) = 2 \int \alpha Q_1 \tilde{h}_t + \beta Q_1 h_t \mathrm{d}\nu_t \leqslant 2 \int Q_1\big(\alpha \tilde{h}_t + \beta h_t\big) \mathrm{d}\nu_t \leqslant W_2^2(\nu_t, \mu),$$

which implies $\alpha Q_1 \tilde{h}_t + \beta Q_1 h_t = Q_1\big(\alpha \tilde{h}_t + \beta h_t\big)$ almost everywhere. It follows that for almost every $x \in E$ and $h = \tilde{h}_t$ or h_t or $\alpha \tilde{h}_t + \beta h_t$, $Q_1 h(x)$ can take its value at the same critical point y_x such that $Q_1 h(x) = h(y_x) + \frac{1}{2} d^2(x, y_x)$ (or the same point sequence $\{y_x^{(n)}\}$ such that $Q_1 h(x) = \lim_{n \to +\infty} h(y_x^{(n)}) + \frac{1}{2} d^2(x, y_x^{(n)})$). If $u_\infty \neq 0$ and $\tilde{h}_t$ is bounded and differentiable, we have $\nabla h_t(y_x) = \nabla \tilde{h}_t(y_x) = x - y_x$ and then $\nabla h_t(y_x) = \nabla \tilde{h}_t(y_x) \equiv 0$ since $h_t = (1 - u_\infty)\tilde{h}_t$, which means $\tilde{h}_t$ has to be a constant function and furthermore $\tilde{h}_t \equiv 0$ for $\mu(\tilde{h}_t) = 0$. This suggests that $h_t = \tilde{h}_t$ is true, however, it seems complicated to deal with L^1 functions.

The same argument is also effective in reproving Lemma 2 in [15] as

$$\left| \frac{\mathrm{d}}{\mathrm{d}t} W_2(\nu, \nu_t) \right| \leqslant \sqrt{\mathrm{I}_\mu(P_t f)},$$

which avoids using the second inequality in (3.1).

4 The second proof of Theorem 1.1

Proof. Assume PI holds with a constant C_P. Recall that

$$\mathrm{Ent}_\mu(P_t f) \leqslant \mathrm{Var}_\mu(P_t f) \leqslant \exp\{-2t/C_P\}\mathrm{Var}_\mu(f),$$

which implies $\mathrm{Ent}_\mu(P_t f) \to 0$ for $t \to \infty$. Using the same method in the second part of [15, Lemma 3] yields $W_2(\nu_t, \mu) \to 0$ too. More precisely, $W_2(\nu_t, \mu)$ decays exponentially fast due to that for any continuous ξ with $|\xi(x)| \leqslant C(d^2(x_0, x) + 1)$,

$$\begin{aligned}\left|\int \xi \mathrm{d}\nu_t - \int \xi \mathrm{d}\mu\right| &\leqslant C\int |P_t f - 1|(d^2(x_0, \cdot) + 1)d\mu \\ &\leqslant C\sqrt{\mathrm{Var}_\mu(P_t f)}\sqrt{\mu((d^2(x_0,\cdot)+1)^2)},\end{aligned}$$

where the integrability of $d^4(x_0, \cdot)$ comes from PI as well.

For simplicity, assume f fulfills all the conditions in Lemma 3.1, then we have by using the Hölder inequality to get (2.1) again

$$\begin{aligned}W_2^2(\nu,\mu) &= \left(\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}s} W_2(\nu_s,\mu)\mathrm{d}s\right)^2 \leqslant \left(\int_0^\infty \sqrt{\mathrm{I}_\mu(P_s f)}\mathrm{d}s\right)^2 \\ &= \left(\int_0^\infty \frac{\sqrt{\mathrm{I}_\mu(P_s f)}}{\sqrt[4]{\mathrm{Ent}_\mu(P_s f)}} \cdot \sqrt[4]{\mathrm{Ent}_\mu(P_s f)}\,\mathrm{d}s\right)^2 \leqslant 2\sqrt{\mathrm{Ent}_\mu(f)}\int_0^\infty \sqrt{\mathrm{Ent}_\mu(P_t f)}\mathrm{d}t.\end{aligned}$$

The following steps are the same as those in Section 2. ☐

Alternatively, using Lemma 3.1 and Hölder inequality yields also for any $t > 0$

$$\begin{aligned}W_2^2(\nu_t,\mu) = \int_t^\infty \frac{\mathrm{d}}{\mathrm{d}s} W_2^2(\nu_s,\mu)\mathrm{d}s &\leqslant 2\int_t^\infty W_2(\nu_s,\mu)\sqrt{\mathrm{I}_\mu(P_s f)}\mathrm{d}s \\ &\leqslant 2\sqrt{\int_t^\infty W_2^2(\nu_s,\mu)\mathrm{d}s} \cdot \sqrt{\int_t^\infty \mathrm{I}_\mu(P_s f)\mathrm{d}s} \\ &= 2\sqrt{\int_t^\infty W_2^2(\nu_s,\mu)\mathrm{d}s} \cdot \sqrt{\mathrm{Ent}_\mu(P_t f)} \qquad (4.1)\end{aligned}$$

(4.1) looks like (2.1), which is still useful to prove Theorem 1.1 as follows.

Denote $\mathcal{W}_t = \sqrt{\int_t^\infty W_2^2(\nu_s,\mu)\mathrm{d}s}$ (it is finite since $W_2(\nu_t, \mu)$ decays exponentially fast), (4.1) can be rewritten to

$$-\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{W}_t \leqslant \sqrt{\mathrm{Ent}_\mu(P_t f)} \leqslant \sqrt{\mathrm{Var}_\mu(P_t f)} \leqslant \exp\{-t/C_P\}\sqrt{\mathrm{Var}_\mu(f)},$$

and then

$$\begin{aligned}\mathcal{W}_t = \int_t^\infty -\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{W}_s \mathrm{d}s &\leqslant \int_t^\infty \exp\{-s/C_P\}\mathrm{d}s\sqrt{\mathrm{Var}_\mu(f)} \\ &= C_P\exp\{-t/C_P\}\sqrt{\mathrm{Var}_\mu(f)}.\end{aligned}$$

Substituting this estimate back to (4.1) for $t = 0$ gives us

$$\begin{aligned}W_2^2(\nu,\mu) &\leqslant 2\sqrt{\int_0^\infty 2C_P \mathrm{Var}_\mu(P_s f)\mathrm{d}s} \cdot \sqrt{\mathrm{Ent}_\mu(f)} \\ &\leqslant 2\sqrt{\int_0^\infty 2C_P\exp\left(-\frac{2}{C_P}s\right)\mathrm{Var}_\mu(f)\mathrm{d}s} \cdot \sqrt{\mathrm{Ent}_\mu(f)} \\ &= 2C_P\sqrt{\mathrm{Var}_\mu(f)} \cdot \sqrt{\mathrm{Ent}_\mu(f)}.\end{aligned}$$

The following steps are the same as before.

By the way, if one is concerned to the quantity $W_2^2(\tilde{\nu}_t,\mu)$ for $\frac{\mathrm{d}\tilde{\nu}_t}{\mathrm{d}\mu}=\frac{|P_t\sqrt{f}|^2}{\mu(|P_t\sqrt{f}|^2)}$, it also decays exponentially fast provided that PI holds. Firstly we have for any $g^2\mu\in\mathcal{P}(E)$ (denote $m=\mu(g)$ and $\sigma_t^2=\mu\left((P_tg-m)^2\right)$)

$$\mathrm{Var}_\mu(g^2)\leqslant\int|g^2-m^2|^2\mathrm{d}\mu\leqslant 2\int|g-m|^4\mathrm{d}\mu+8m^2\int|g-m|^2\mathrm{d}\mu.$$

Then it follows from PI that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mu\left((P_tg-m)^4\right) &= -12\mu\left((P_tg-m)^2\left|\nabla P_tg\right|^2\right)\\ &\leqslant -3C_P^{-1}\mu\left(\left((P_tg-m)^2-\sigma_t^2\right)^2\right)\\ &= -3C_P^{-1}\left[\mu\left((P_tg-m)^4\right)-\sigma_t^4\right],\end{aligned}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma_t^4=-4\sigma_t^2\mu\left(|\nabla P_tg|^2\right)\leqslant -4C_P^{-1}\sigma_t^4.$$

Set $\Lambda_t=\mu\left((P_tg-m)^4\right)+\lambda\sigma_t^4$ with the parameter λ, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\Lambda_t\leqslant C_P^{-1}\left(-3\Lambda_t+(3-\lambda)\sigma_t^4\right),$$

which implies by taking $\lambda=3$ that

$$\frac{\mathrm{d}}{\mathrm{d}t}\Lambda_t\leqslant -3C_P^{-1}\Lambda_t$$

and then $\Lambda_t\leqslant\exp\left(-3t/C_P\right)\Lambda_0$.

Hence using Theorem 1.1 yields for $g=\sqrt{f}$ that

$$\begin{aligned}W_2^2\left(\tilde{\nu}_t,\mu\right) &\leqslant 2C_P\mathrm{Var}_\mu(\frac{\mathrm{d}\tilde{\nu}_t}{\mathrm{d}\mu})\leqslant\frac{2C_P}{\left(\mu(|P_tg|^2)\right)^2}\mathrm{Var}_\mu\left((P_tg)^2\right)\\ &\leqslant\frac{4C_P}{m^4}(\Lambda_t+4m^2\sigma_t^2)\leqslant\frac{4C_P}{m^4}(e^{-3t/C_P}\Lambda_0+e^{-2t/C_P}4m^2\sigma_0^2),\end{aligned}$$

where the total rate is no more than e^{-2t/C_P}.

5 The logarithmic Sobolev inequality and strict contraction of heat flow in Wasserstein space

In this section, we prove Proposition 1.3. The curvature-dimension condition plays a fundamental role such that we can compare several functionals for heat flow at different times. The derivative estimate in previous section is also useful.

Proof. Assume V is a smooth potential satisfying the curvature-dimension condition $CD(\rho,\infty)$.

If the LSI holds, it is known that the entropy along heat flow decays exponentially fast. Moreover, the Talagrand inequality comes true (see [15] or [2, Theorem 9.6.1]), namely for any positive bounded f and any $t>T>0$

$$W_2^2(P_tf\mu,\mu)\leqslant 2C_{LS}\mathrm{Ent}_\mu(P_tf)\leqslant 2C_{LS}e^{-2(t-T)/C_{LS}}\mathrm{Ent}_\mu(P_Tf).$$

On the other hand, based on the logarithmic Harnack inequlity (see [2, Remark 5.6.2])

$$P_T(\log f)(x) \leqslant \log P_T f(y) + \frac{\rho d(x,y)^2}{2(e^{2\rho T}-1)},$$

it follows from the the same argument as [2, Page 446] that

$$\mathrm{Ent}_\mu(P_T f) \leqslant \frac{1}{2\beta(T)} W_2^2(f\mu,\mu),$$

where $\frac{1}{\beta(T)} = \frac{\rho}{1-e^{-2\rho T}} - \rho$ ($= \frac{1}{2T}$ for $\rho = 0$). Combining the above estimates yields

$$W_2^2(P_t f\mu,\mu) \leqslant \gamma(T) e^{-2t/C_{LS}} W_2^2(f\mu,\mu)$$

by letting $\gamma(T) = \frac{C_{LS} e^{2T/C_{LS}}}{\beta(T)}$, which attains its minimum at $T_0 = \frac{1}{2|\rho|}\log(1 + C_{LS}|\rho|)$. So now we obtain the exponential decay for $t > T_0$.

For $0 < t \leqslant T_0$, there is a general bound according to the heat flow contraction in Wasserstein space (see [2, Theorem 9.7.2]) as

$$W_2^2(P_t f\mu,\mu) \leqslant e^{-2\rho t} W_2^2(f\mu,\mu) = e^{(2C_{LS}^{-1}-2\rho)t} e^{-2t/C_{LS}} W_2^2(f\mu,\mu).$$

Combining two regions gives us a control constant $C := \sqrt{\max\{\gamma(T_0), e^{(2C_{LS}^{-1}-2\rho)T_0}, 1\}}$ such that for all $t > 0$ and $\kappa := C_{LS}^{-1}$

$$W_2(P_t f\mu,\mu) \leqslant C e^{-\kappa t} W_2(f\mu,\mu).$$

Conversely, if $W_2(P_t f\mu,\mu) \leqslant C e^{-\kappa t} W_2(f\mu,\mu)$, there exists t (independent of f) such that $\eta := C e^{-\kappa t} < 1$. Using the derivative estimate for nice f (see Lemma 3.1) yields

$$\begin{aligned} W_2^2(f\mu,\mu) &= W_2^2(f\mu,\mu) - W_2^2(P_t f\mu,\mu) + W_2^2(P_t f\mu,\mu) \\ &\leqslant \int_0^t 2W_2(P_s f\mu,\mu)\sqrt{\mathrm{I}_\mu(P_s f)}\mathrm{d}s + \eta^2 W_2^2(f\mu,\mu) \\ &\leqslant 2\left(\int_0^t W_2^2(P_s f\mu,\mu)\mathrm{d}s\right)^{\frac12}\left(\int_0^t \mathrm{I}_\mu(P_s f)\mathrm{d}s\right)^{\frac12} + \eta^2 W_2^2(f\mu,\mu). \end{aligned}$$

Based on the heat flow contraction and information contraction (see [2, Eq. 5.7.4])

$$\mathrm{I}_\mu(P_s f) \leqslant e^{-2\rho s}\mathrm{I}_\mu(f),$$

we have further

$$\begin{aligned} W_2^2(f\mu,\mu) &\leqslant 2\left(\int_0^t C^2 e^{-2\kappa t} W_2^2(f\mu,\mu)\mathrm{d}s\right)^{\frac12}\left(\int_0^t e^{-2\rho s}\mathrm{I}_\mu(f)\mathrm{d}s\right)^{\frac12} + \eta^2 W_2^2(f\mu,\mu) \\ &\leqslant (\varepsilon+\eta^2) W_2^2(f\mu,\mu) + \frac{C^2(1-e^{-2\kappa t})(1-e^{-2\rho t})}{4\kappa\rho\varepsilon}\mathrm{I}_\mu(f), \end{aligned}$$

where the last step comes from the Cauchy-Schwarz inequality for any $\varepsilon > 0$. It follows $\mathrm{W_2I}$ by taking $\varepsilon = \eta = \frac12$ explicitly that

$$W_2^2(f\mu,\mu) \leqslant \frac{2C^2(1-e^{-2\rho t})}{\kappa\rho}\mathrm{I}_\mu(f).$$

Since $\mathrm{W_2I}$ is equivalent to LSI under $CD(\rho,\infty)$ by virtue of the HWI inequality (see [15] or [2, Subsection 9.3])

$$\mathrm{Ent}_\mu(f) \leqslant W_2(f\mu,\mu)\sqrt{\mathrm{I}_\mu(f)} - \frac{\rho}{2}W_2^2(f\mu,\mu),$$

we complete the proof. □

6 Centralization of quadratic Wasserstein distance

Recall the notation $c = \mu(\sqrt{f})$ and $\sigma^2 = \mathrm{Var}_\mu(\sqrt{f})$, now we prove Theorem 1.5.

Proof. For any bounded Lipschitz h with $\mu(h) = 0$, let $m_t = \mu(Q_t h)$, we have

$$\begin{aligned}\mu(Q_t hf) &= \int Q_t h(\sqrt{f}-c)^2 \mathrm{d}\mu + 2c\int Q_t h(\sqrt{f}-c)\mathrm{d}\mu + c^2 \int Q_t h \mathrm{d}\mu \\ &= \int Q_t h(\sqrt{f}-c)^2 \mathrm{d}\mu + 2c\int (Q_t h - m_t)(\sqrt{f}-c)\mathrm{d}\mu + c^2 m_t.\end{aligned}$$

Taking any interval $[a,b] \subset \mathbb{R}^+$ and any nonnegative $\phi \in C^1([a,b])$, we integrate both sides to get

$$\begin{aligned}\mathbb{I}_0 &:= \int_a^b \mu(Q_t hf)\phi \mathrm{d}t \\ = \int_a^b \mu(Q_t h(\sqrt{f}-c)^2)\phi \mathrm{d}t &+ 2c\int_a^b \mu\big((Q_t h - m_t)(\sqrt{f}-c)\big)\phi \mathrm{d}t + \ c^2 \int_a^b m_t \phi \mathrm{d}t.\end{aligned}$$

For convenience, denote the right-hand three terms by $\mathbb{I}_1, \mathbb{I}_2, \mathbb{I}_3$ respectively. Using the Cauchy-Schwarz, Hölder and Poincaré inequalities yields for any $\lambda > 0$

$$\begin{aligned}\mathbb{I}_2 &= 2c\int \left(\int_a^b (Q_t h - m_t)\phi(t)\mathrm{d}t\right)(\sqrt{f}-c)\mathrm{d}\mu \\ &\leqslant \lambda c^2 \int \left(\int_a^b (Q_t h - m_t)\phi(t)\mathrm{d}t\right)^2 \mathrm{d}\mu + \frac{1}{\lambda}\mu((\sqrt{f}-c)^2) \\ &\leqslant \lambda c^2 (b-a)\int\int_a^b (Q_t h - m_t)^2 \phi^2(t)\mathrm{d}t\mathrm{d}\mu + \frac{1}{\lambda}\sigma^2 \\ &= \lambda c^2(b-a)\int_a^b \mu\left((Q_t h - m_t)^2\right)\phi^2(t)\mathrm{d}t + \frac{1}{\lambda}\sigma^2 \\ &\leqslant \lambda c^2 (b-a) C_P \int_a^b \mu\left(|\nabla Q_t h|^2\right)\phi^2(t)\mathrm{d}t + \frac{1}{\lambda}\sigma^2 \\ &= 2\lambda c^2(b-a)C_P \int\int_a^b -\frac{\mathrm{d}}{\mathrm{d}t}Q_t h\ \phi^2(t)\mathrm{d}t\mathrm{d}\mu + \frac{1}{\lambda}\sigma^2,\end{aligned}$$

where the last step comes from the Hamilton-Jacobi equation. Using the integration by parts gives

$$\int_a^b -\frac{\mathrm{d}}{\mathrm{d}t}Q_t h\ \phi^2(t)\mathrm{d}t = Q_a h\phi^2(a) - Q_b h \phi^2(b) + \int_a^b Q_t h \cdot 2\phi\phi' \mathrm{d}t.$$

If $\phi(a) = \phi(b) = 0$, we have further

$$\mathbb{I}_2 \leqslant 4\lambda c^2 (b-a) C_P \int_a^b m_t \phi\phi' \mathrm{d}t + \frac{1}{\lambda}\sigma^2,$$

and then

$$\mathbb{I}_2 + \mathbb{I}_3 \leqslant c^2 \int_a^b m_t \phi \left[4\lambda(b-a)C_P\phi' + 1\right]\mathrm{d}t + \frac{1}{\lambda}\sigma^2.$$

Now we want to drop the first integral on the right side of above inequality. For instance, take $a = \frac{1}{2}$, $b = 1$, $\phi(t) = (t-a)(b-t)$ (satisfying $\phi(a) = \phi(b) = 0$, $\phi \geqslant 0$ and $|\phi'| \leqslant \frac{1}{2}$), and $\lambda = C_P^{-1}$, then for $t \in [a,b]$, the quantity

$$\psi := (4\lambda(b-a)C_P\phi' + 1) \geqslant 0,$$

which implies $\int_a^b m_t \phi\psi \mathrm{d}t \leqslant 0$ since the monotonicity of Q_t in t gives $m_t = \mu(Q_t h) \leqslant \mu(h) = 0$. Hence $\mathbb{I}_2 + \mathbb{I}_3 \leqslant C_P \sigma^2$.

Finally, combining all above estimates yields

$$\mathbb{I}_0 \leqslant \mathbb{I}_1 + C_P \sigma^2.$$

Denote $M = \int_a^b \phi \mathrm{d}t = \frac{1}{48}$, it follows

$$M \cdot \mu(Q_b h f) \leqslant \mathbb{I}_0 \leqslant \mathbb{I}_1 + C_P\sigma^2 \leqslant M \cdot \mu(Q_a h(\sqrt{f} - c)^2) + C_P \sigma^2,$$

which implies by the Kantorovich dual of W_2-distance that

$$\frac{M}{2b} W_2^2(f\mu, \mu) \leqslant \frac{M}{2a} \sigma^2 W_2^2 \left(\frac{(\sqrt{f} - c)^2}{\sigma^2} \mu, \mu \right) + C_P \sigma^2.$$

The proof is completed. □

When we check the proof, for any θ still holds

$$\begin{aligned}
\mu(Q_t h f) &= \mu(Q_t h(\sqrt{f} - \theta)^2) + 2\theta \mu\big(Q_t h(\sqrt{f} - \theta)\big) + \theta^2 \mu(Q_t h) \\
&= \mu(Q_t h(\sqrt{f} - \theta)^2) + 2\theta \mu\big((Q_t h - m_t)(\sqrt{f} - \theta)\big) + (2\theta c - \theta^2)\mu(Q_t h) \\
&= \mu(Q_t h(\sqrt{f} - \theta)^2) + 2\theta \mu\big((Q_t h - m_t)(\sqrt{f} - c)\big) + (2\theta c - \theta^2)\mu(Q_t h).
\end{aligned}$$

Denote $\sigma_\theta^2 = \mu((\sqrt{f} - \theta)^2)$. Once $\theta \in (0, 2c)$, we have by the same argument

$$W_2^2(f\mu, \mu) \leqslant C_1(\theta) \sigma_\theta^2 W_2^2 \left(\frac{(\sqrt{f} - \theta)^2}{\sigma_\theta^2} \mu, \mu \right) + C_2(\theta) C_P \sigma^2,$$

where $C_1(\theta)$ and $C_2(\theta)$ are two constants depending on θ.

7 Application to quadratic transportation-information inequality

According to [4, 5, 13], the Lyapunov condition (1.4) implies that there are two constants $C_3, C_4 > 0$ such that

$$\int d^2(x_0, \cdot) h^2 \mathrm{d}\mu \leqslant C_3 \int |\nabla h|^2 \mathrm{d}\mu + C_4 \int h^2 \mathrm{d}\mu, \tag{7.1}$$

and then implies $\mathrm{W_2I}$ by [13], which partially depends on two facts that (7.1) implies $\mathrm{W_2H}$ and $\mathrm{W_2H}$ has a Bobkov-Götze's characterization.

Now there appears another way. For unbounded manifolds, (7.1) implies there exists some $r > 0$ such that

$$\int d^2(x_0, \cdot) h^2 \mathrm{d}\mu \leqslant C_5 \int |\nabla h|^2 \mathrm{d}\mu + C_6 \int_{d(x_0, \cdot) \leqslant r} h^2 \mathrm{d}\mu,$$

which leads to PI by [1]. Then using Theorem 1.5 and (7.1) and PI yields

$$W_2^2(\nu, \mu) \leqslant 2C_1 \int \left(d^2(x_0, \cdot) + \mu\left(d^2(x_0, \cdot)\right)\right) (\sqrt{f} - c)^2 \mathrm{d}\mu + C_2 \sigma^2 \leqslant C_7 \mathrm{I}_\mu(\nu|\mu),$$

where we use the fact that for any x and any bounded h with $\mu(h) = 0$ holds

$$Q_1 h(x) \leqslant \int h(y) + \frac{1}{2} d^2(x, y) \mathrm{d}\mu(y) \leqslant d^2(x_0, \cdot) + \mu\left(d^2(x_0, \cdot)\right).$$

Hence we reach $\mathrm{W_2I}$.

References

[1] Bakry, D., Barthe, F., Cattiaux, P., Guillin, A.: A simple proof of the Poincaré inequality for a large class of measures including the logconcave case, *Electron. Commun. Probab.* **13** (2008), 60–66. MR-2386063

[2] Bakry D., Gentil I., and Ledoux M.: Analysis and geometry of Markov diffusion operators. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], **348**. Springer, Cham, 2014. MR-3155209

[3] Bobkov S. G., Götze F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities, *J. Funct. Anal.* **163** (1999), 1–28. MR-1682772

[4] Cattiaux P., Guillin A.: Functional Inequalities via Lyapunov conditions. In *Optimal transportation, Theory and applications*, London Mathematical Society Lecture Notes Series, **413**, 274–287. Cambridge Univ. Press, 2014. MR-3328999

[5] Cattiaux P., Guillin A., and Wu L.-M.: A note on Talagrands transportation inequality and logarithmic Sobolev inequality, *Proba. Theory Relat. Fields* **148** (2010), no. 1-2, 285–304 MR-2653230

[6] Ding Y.: A note on quadratic transportation and divergence inequality, *Statist. Probab. Lett.* **100** (2015), 115–123. MR-3324082

[7] Evans L. C.: Partial differential equations. Second edition. Graduate Studies in Mathematics, **19**. American Mathematical Society, Providence, RI, 2010. MR-2597943

[8] Gozlan N., Roberto C., and Samson P. M.: A new characterization of Talagrand's transport-entropy inequalities and applications, *Ann. Probab.* **39** (2011), no. 3, 857–880. MR-2789577

[9] Guillin A., Léonard C., Wu L.-M., and Yao N.: Transportation information inequalities for Markov processes, *Probab. Theory Relat. Fields* **144** (2009), no. 3-4, 669–696. MR-2496446

[10] Jourdain B.: Equivalence of the Poincaré inequality with a transport-chi-square inequality in dimension one, *Electron. Commun. Probab.* **17** (2012), no. 43, 1–12. MR-2981899

[11] Kuwada K.: Duality on gradient estimates and Wasserstein controls, *J. Funct. Anal.* **258** (2010), 3758–3774. MR-2606871

[12] Ledoux M.: Remarks on some transportation cost inequalities, preprint (2018), see the website https://perso.math.univ-toulouse.fr/ledoux/publications-3/.

[13] Liu Y.: A new characterization of quadratic transportation-information inequalities, *Probab. Theory Related Fields* **168** (2017), 675–689. MR-3663628

[14] Milman E.: Properties of isoperimetric, functional and transport-entropy inequalities via concentration, *Probab. Theory Related Fields* **152** (2012), no. 3-4, 475–507. MR-2892954

[15] Otto F., Villani C.: Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, *J. Funct. Anal.* **173** (2000), no. 2, 361–400. MR-1760620

[16] Villani C.: Topics in Optimal Transportation. Graduate Studies in Mathematics **58**, American Mathematical Society, Providence RI, 2003. MR-1964483

[17] Villani C.: Optimal Transport: old and new. Grundlehren der Mathematischen Wissenschaf-ten **338**, Springer-Verlag, Berlin, 2009. MR-2459454

[18] Wang F.-Y.: Logarithmic Sobolev inequalities on noncompact Riemannian manifolds, *Probab. Theory and Relat. Fields* **109** (1997), no. 3, 417–424. MR-1481127

[19] Wang F.-Y.: Exponential contraction in Wasserstein distances for diffusion semigroups with negative curvature, *Potential Anal.*, to appear.

Acknowledgments. I am so grateful to the anonymous referee for his/her conscientious reading with many suggestions and comments on the first version of this paper. I also thank Prof. Li-Ming Wu and Prof. Feng-Yu Wang very much for their kindly comments during the conferences held in WHU and BNU respectively. The author is supported by NSFC (no. 11431014, no. 11688101), AMSS research grant (no. Y129161ZZ1), and Key Laboratory of Random Complex Structures and Data, Academy of Mathematics and Systems Science, Chinese Academy of Sciences (No. 2008DP173182).