RIMS Kôkyûroku Bessatsu
B43 (2013), 61–80

Gradient estimate for Markov kernels, Wasserstein control and Hopf-Lax formula

By

Kazumasa KUWADA*

Abstract

We extend the duality between gradient estimates of the Markov kernel and Wasserstein controls of that studied by the author (2010). Especially, the gauge norm-Orlicz norm type duality holds on Polish geodesic space without any assumption on the Markov kernel. For the proof of the duality, we proceed analysis of Hopf-Lax semigroups. Some sorts of stability of these estimates are also studied. As an application of a stability result, we show a gradient estimate for a semigroup of Markov kernels yields the corresponding estimate for subordinated semigroups.

§ 1. Introduction

As an effective way of measuring the rate of convergence to equilibrium of (possibly nonlinear) diffusions, Wasserstein distances have been used in the literature (see e.g. [10, 27, 28] and references therein). Among them, an exponential control in time of Wasserstein distances between heat distributions ((6.4) below) has been investigated extensively since it is deeply interacted with other research fields such as differential geometry, partial differential equations, functional inequalities and probability theory. As a part of such connections, a control of the L^2-Wasserstein distance links the presence of a lower Ricci curvature bound in the sense of Sturm and Lott-Villani [21, 25] with Bakry-Émery's gradient estimate [5, 11, 16]. Moreover, those two conditions are *equivalent* to the Wasserstein control [2, 5, 29]. Such an equivalence as well as Bakry-Émery's gradient estimate was known mostly on essentially smooth spaces and hence the Wasserstein distance played a prominent role to extend the theory to more singular

Received February 3, 2013.
2000 Mathematics Subject Classification(s): 39B62, 49N15, 49L99, 47D06, 70H20
Key Words: *Gradient estimate, Wasserstein distance, Hopf-Lax formula, Markov kernel, stability*
Partially supported by the Grant in Aid for Young Scientists (B) 22740083
*Graduate School of Humanities and Sciences, Ochanomizu University, Tokyo 112-8610, Japan.
e-mail: `kuwada.kazumasa@ocha.ac.jp`

spaces than Riemannian manifolds. As those results explain, a control of the Wasserstein distance now possesses other significant meanings even apart from the convergence rate. As one of such researches, the article [17] formulates and establishes a duality between a control of the Wasserstein distance for Markov kernels and a gradient estimate of the Bakry-Émery type in a fairly general framework. As pointed out there, those duality holds not only for heat semigroups and it does not rely on any curvature bounds. Thus it might provide us new tools to proceed geometric analysis even in the absence of uniform lower curvature bounds.

The main purpose of this article is to extend the duality result in [17]. The primal emphasis is put on removing technical assumptions. As a consequence, the same result always holds on geodesic metric spaces without any further assumptions on the Markov kernel or on the underlying space. Among others, we do not require the local Poincaré inequality, the volume doubling condition and moreover any reference measure on the underlying space. These conditions are rather weak in the sense that those spaces which satisfy them are sufficiently ample even in the class of singular spaces. However, even on a smooth space as Riemannian manifolds without boundary, those conditions are not always satisfied. We also extend the result in other three respects. First, we weaken the assumption on the distance function. Our new condition fits well with analysis on (a class of) infinite dimensional spaces. Second, we replace L^p-L^q type duality with more general gauge norm-Orlicz norm type duality (See [23] for these norms. See [26] for the Orlicz-Wasserstein distance). It enables us to deal with more subtle situations where L^p-spaces are not sufficient. Third, we separate the parameter space of the Markov kernel from the underlying space. Though it is a rather minor extension from a technical point of view, it broadens the range of the theory, as we will see in examples. While the proof goes along the same line as in [17], we must modify some arguments because of the generality of our framework. For instance, according to these extensions, we also extend the theory of Hopf-Lax or Hamilton-Jacobi semigroups from that in [3, 4, 12]. to the one which fits with our framework. In [3, 4], they only consider the L^p-case, and in [12], their (topological) condition on the underlying space is more restrictive. Such an extension would be of independent interest since the Hopf-Lax semigroup has several applications in analysis on metric spaces (see e.g. [3, 4, 6, 8, 9, 12]). Note that the local Poincaré inequality and the volume doubling condition are used in [17] to employ the existing theory of Hopf-Lax semigroups in [9, 20]. However, we does not require them already in [3, 4, 12].

As another achievement of this article, we provide some remarks on stability results. Compared with the gradient estimate, the Wasserstein control is more stable under several operations such as convergence of Markov kernels, tensorization and averaging. Thus, based on our duality, we can obtain the same stability for the corresponding gradient estimates. Though each stability result seems rather elementary from the viewpoint of analysis of Wasserstein distances, it seems non-trivial for the gradient estimate if we cannot employ our duality. For instance, we can use a gradient estimate for a semigroup of Markov kernels to obtain a gradient estimate for subordinated semigroups.

We now mention some of related results which are not discussed yet. Ollivier [22] initiated geometric analysis based on a control of the (L^1-)Wasserstein distance for Markov kernels by regarding it as a *definition* of generalized lower Ricci curvature bound. We also refer to [14, 15] and references therein for further developments and related results. Our gradient estimate is originally studied by Bakry and Émery for diffusion semigroups, and it has been a source of several important functional inequalities such as Poincaré, log-Sobolev and isoperimetric inequalities (see e.g. [7,8]). Note that a similar but different approach as ours to our duality result is provided in [8].

Now we demonstrate the organization of the paper. In the next section, we will state our framework, notations and the main result (Theorem 2.2). In section 3, we study the Hopf-Lax semigroup along with the same line as in [3,12]. Here we prove that the Hopf-Lax semigroup solves the Hamilton-Jacobi equation in an appropriate sense even in our general framework (Theorem 3.7 and Theorem 3.8). Note that, in the proof of Theorem 2.2, we only use a partial result (a part of the assertion of Theorem 3.6). Theorem 2.2 will be proved in section 4. In section 5, we will exhibit stability results. In section 6, we provide three examples. Two of them explain that the separation of the parameter space of the Markov kernel from the underlying space is meaningful. The last example is an application of our stability result to a gradient estimate for subordinated semigroups.

§ 2. Framework and the main result

Let X be a Polish topological space. Let $d : X \times X \to [0, \infty]$ be an extended distance in the sense of [3]. That is, d satisfies all properties of distance function except for finiteness, it is lower semi-continuous and the convergence with respect to d implies the convergence in X. Let $\Phi : [0, \infty) \to [0, \infty)$ be a C^1-convex increasing function satisfying $\Phi(0) = 0$, $\Phi(x) > 0$ for $x > 0$ and $\lim_{u\to\infty} \Phi(u)/u = \infty$. We denote the Legendre conjugate of Φ by Φ^*. That is, $\Phi^*(v) := \sup_{u\geq 0}[uv - \Phi(u)]$ for $v \geq 0$. Note that $\Phi^*(v) < \infty$ for any $v \in [0, \infty)$. We set $\Phi_p(u) := p^{-1}u^p$ for $p \in [1, \infty)$ and $\Phi_\infty(u) := \lim_{p\to\infty} \Phi_p(u)$. Note that $\Phi_p^* = \Phi_{p_*}$ holds for $p \in [1, \infty]$, where p_* is the Hölder conjugate of p, i.e. $p^{-1} + p_*^{-1} = 1$. We can easily verify that Φ_p satisfies the assumption on Φ if and only if $p \in (1, \infty)$. In what follows, p always stands for a real number in $[1, \infty]$ and p_* is the Hölder conjugate of p otherwise stated explicitly.

We denote the space of probability measures on X by $\mathscr{P}(X)$. For $\mu, \nu \in \mathscr{P}(X)$, we denote the set of couplings of μ and ν by $\Pi(\mu, \nu)$. For $\mu, \nu \in \mathscr{P}(X)$, let us define

L^Φ-Wasserstein distance $W_\Phi(\mu,\nu)$ as follows:

$$W_\Phi(\mu,\nu) := \inf_{\pi\in\Pi(\mu,\nu)} \|d\|_{L^\Phi(\pi)}. \tag{2.1}$$

Here $\|\cdot\|_{L^\Phi(\pi)}$ means the gauge norm. That is,

$$\|d\|_{L^\Phi(\pi)} = \inf\left\{\lambda>0 \;\middle|\; \int_{X\times X} \Phi\left(\frac{d}{\lambda}\right) d\pi \le 1\right\}.$$

For simplicity of notations, we denote $p^{1/p} W_{\Phi_p}$ and W_{Φ_∞} by W_p and W_∞ respectively. Note that W_p coincides with the usual L^p-Wasserstein (extended) distance. More precisely, $W_p(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \|d\|_{L^p(\pi)}$ holds. For a measurable function $f : X \to \mathbb{R}$, let us denote the local Lipschitz constant of f with respect to d by $|\nabla f| \in \mathscr{B}(X)$. That is,

$$|\nabla f|(x) := \lim_{r\to 0} \sup_{\substack{y\in X \\ d(x,y)\in(0,r)}} \left[\frac{|f(y)-f(x)|}{d(y,x)}\right]. \tag{2.2}$$

Note that $|\nabla f|$ is universally measurable (see [3, Lemma 2.4]). Let $\tilde{X}$ be another Polish space and $\tilde{d} : \tilde{X}\times\tilde{X} \to [0,\infty]$ an extended distance on $\tilde{X}$. We also use the notations $\tilde{W}_\Phi(\tilde{\mu},\tilde{\nu})$ for $\tilde{\mu},\tilde{\nu} \in \mathscr{P}(X)$ or $|\tilde{\nabla}\tilde{f}|$ for $\tilde{f} : \tilde{X} \to \mathbb{R}$ defined similarly as in (2.1) and (2.2).

We assume the following in some occasions. We state it explicitly when we do so.

Assumption 2.1.

(i) The extended distance d is a geodesic extended distance. It means that, for every $x, y \in X$ with $d(x,y) < \infty$, there is a curve $\gamma : [0,1] \to X$ with $\gamma(0) = x$ and $\gamma(1) = y$ such that $d(\gamma(s),\gamma(t)) = |t-s| d(x,y)$

(ii) The extended distance $\tilde{d}$ is a geodesic extended distance.

Remark 1. In Assumption 2.1, we can weaken "geodesic extended distance" to length extended distance in all our results except Remark 4 (See [3] for length extended distance). We assumed the stronger "geodesic" assumption just for simplicity of presentation.

We call the curve γ appeared in the definition of Assumption 2.1 (i) a d-minimal geodesic. We also use the term "$\tilde{d}$-minimal geodesic" under Assumption 2.1 (ii).

For each $\tilde{x} \in \tilde{X}$, let $P_{\tilde{x}} \in \mathscr{P}(X)$. We suppose that P is a Markov kernel, that is, for each $A \in \mathscr{B}(X)$, $\tilde{x} \mapsto P_{\tilde{x}}(A)$ is measurable. For a measurable function $f : X \to \mathbb{R}$ and $\tilde{\mu} \in \mathscr{P}(\tilde{X})$, we denote the action of P to f and the dual action to $\tilde{mu}$ by Pf and $P^*\tilde{\mu}$ respectively. We denote the space of bounded measurable functions on X which are Lipschitz with respect to d by $\mathrm{Lip}_b(X)$. Note that $\mathrm{Lip}_b(X) \subset C(X)$ may not hold if d is not continuous.

We are interested in the relation between the following two properties:

(i) For every $\tilde{x}, \tilde{y} \in \tilde{X}$,

$$(W(\Phi)) \qquad W_\Phi(P_{\tilde{x}}, P_{\tilde{y}}) \le \tilde{d}(\tilde{x}, \tilde{y}).$$

(ii) For every $\tilde{x} \in \tilde{X}$ and $f \in \mathrm{Lip}_b(X)$,

$$(G(\Phi^*)) \qquad |\tilde{\nabla} Pf|(\tilde{x}) \le \| |\nabla f| \|_{\tilde{L}^{\Phi^*}(P_{\tilde{x}})},$$

where $\| \cdot \|_{\tilde{L}^{\Phi^*}(\mu)}$ is the Orlicz norm associated with Φ^* for $\mu \in \mathscr{P}(X)$. That is,

$$\|f\|_{\tilde{L}^{\Phi^*}(\mu)} = \sup \left\{ \int_X fg \, d\mu \;\middle|\; g : X \to \mathbb{R} \text{ measurable}, \int_X \Phi(g) d\mu \le 1 \right\}.$$

We sometimes consider a similar condition where W_Φ is replaced with W_p. We denote it by (W_p) instead of $(W(\Phi))$. Similarly, when $p > 1$, we denote the condition where $\| |\nabla f| \|_{\tilde{L}^{\Phi^*}(P_{\tilde{x}})}$ in $(G(\Phi^*))$ is replaced with $P(|\nabla f|^{p_*})^{1/p_*}$ by (G_p). When $p = 1$, the condition (G_∞) is given in the following:

$$(G_\infty) \qquad \sup_{\tilde{x} \ne \tilde{y}} \left[\frac{Pf(\tilde{x}) - Pf(\tilde{y})}{\tilde{d}(\tilde{x}, \tilde{y})} \right] \le \sup_{x \ne y} \left[\frac{f(x) - f(y)}{d(x, y)} \right].$$

Now we are ready to state our main theorem.

Theorem 2.2.

(i) $(W(\Phi))$ *implies* $(G(\Phi^*))$.

(ii) *Suppose Assumption 2.1. Then* $(G(\Phi^*))$ *implies* $(W(\Phi))$.

Note that we can slightly extend the result in Theorem 2.2 in different ways. For simplicity of presentation, we will state them separately in remarks below. See Remark 2, Remark 3 and Remark 4. A typical and well-studied situation in Theorem 2.2 is the case $\tilde{X} = X$ and $\tilde{d} = Cd$ with a constant $C > 0$ (cf.(6.4)). We can easily see that Theorem 2.2 also asserts the duality between (W_p) and (G_{p_*}) for $p \in (1, \infty)$.

Remark 2. As for the duality between (W_p) and (G_{p_*}), the cases $p = 1, \infty$ does not seem to be dealt in Theorem 2.2. However, we can easily deduce them. When $p = 1$, the same proof as in [17] works. Note that, unlike Theorem 2.2 (ii), we do not require Assumption 2.1 (i) in this case. When $p = \infty$, it is reduced to the case $p \in (1, \infty)$ as we did in [17]. The key ingredient there is [17, Lemma 3.3] and the corresponding result (Lemma 4.1 (iii) below) also holds in our framework.

§ 3. Hopf-Lax semigroups

In this section, we fix $f : X \to \mathbb{R} \cup \{\infty\}$. Let us consider the Hopf-Lax semigroup $Q_t f$ associated with Φ. For $x, y \in X$ and $t > 0$, we define $F(t, x, y)$ and $Q_t f(x) \in \mathbb{R} \cup \{\pm\infty\}$ by

$$F(t,x,y) := f(y) + t\Phi\left(\frac{d(x,y)}{t}\right), \qquad Q_t f(x) := \inf_{y \in X} F(t,x,y). \tag{3.1}$$

Conventionally, we use the notation $Q_0 f := f$. Note that $Q_t f(x)$ is non-increasing in t since Φ is convex and $\Phi(0) = 0$. Set $\mathcal{D}(f) := \{x \in X \mid F(1,x,y) < \infty \text{ for some } y \in X\}$ and $t_*(x) := \sup\{t > 0 \mid Q_t f(x) > -\infty\}$. Then $x \in \mathcal{D}(f)$ and $0 < t < t_*(x)$ imply $Q_t f(x) \in \mathbb{R}$. For $x \in \mathcal{D}(f)$ and $0 < t < t_*(x)$, let us define $D^+(x,t)$ and $D^-(x,t)$ by

$$D^+(x,t) := \sup_{(y_n)_n} \limsup_{n\to\infty} d(x, y_n), \qquad D^-(x,t) := \inf_{(y_n)_n} \liminf_{n\to\infty} d(x,y_n).$$

where $(y_n)_n$ in the above supremum or infimum runs over all minimizing sequences of $F(t, x, \cdot)$. Note that these supremum or infimum is attained. Indeed, it follows from a diagonal argument. We begin with basic properties of t_* and $D^\pm$.

Lemma 3.1.

(i) *$t_*(x) = t_*(y)$ holds for $x, y \in X$ with $d(x,y) < \infty$.*

(ii) *D^+ is locally bounded in the sense that for $x \in X$, $R > 0$ and $t_0 \in (0, t_*(x))$, there is $M > 0$ such that $D^+(y,s) \le M$ for $y \in X$ with $d(x,y) \le R$ and $s \in (0, t_0]$.*

Proof. By the convexity of Φ, for $z \in X$ and $s < t$,

$$t\Phi\left(\frac{d(x,z)}{t}\right) \le s\Phi\left(\frac{d(y,z)}{s}\right) + (t-s)\Phi\left(\frac{d(x,y)}{t-s}\right). \tag{3.2}$$

It easily implies $Q_t f(x) \le Q_s f(y) + (t-s)\Phi(d(x,y)/(t-s))$. Thus $t_*(y) \le t_*(x)$ follows by taking $s < t_*(x)$ arbitrarily and $t \in (s, t_*(x))$. The opposite inequality also follows in a symmetric way and hence the first assertion holds.

For the second assertion, take $y \in X$ with $d(x,y) \le R$, $s \in (0, t_0]$ and $t_0 < t_1 < t_2 < t_*(x)$. Then, by using (3.2) with $s = t_1$ and $t = t_2$, for $z \in X$ with $d(y,z) < \infty$,

$$F(s,y,z) \ge s\Phi\left(\frac{d(y,z)}{s}\right) - t_1\Phi\left(\frac{d(y,z)}{t_1}\right) + Q_{t_2} f(x) - (t_2 - t_1)\Phi\left(\frac{R}{t_2 - t_1}\right).$$

Take $z_0 \in X$ so that it satisfies $d(x, z_0) < \infty$ and $f(z_0) < \infty$ (such z_0 exists since $x \in \mathcal{D}(f)$). Note that $t'\Phi(u/t') - t_1\Phi(u/t_1)$ is non-decreasing when $t' < t_1$ since Φ is convex (cf. (3.8) below). Thus the last inequality yields

$$
\begin{aligned}
(3.3)\quad s\Phi\left(\frac{D^+(y,s)}{s}\right) - t_1\Phi\left(\frac{D^+(y,s)}{t_1}\right) &\le Q_s f(y) - Q_{t_2} f(x) + (t_2 - t_1)\Phi\left(\frac{R}{t_2 - t_1}\right) \\
&\le f(z_0) + s\Phi\left(\frac{R + d(x,z_0)}{s}\right) - Q_{t_2} f(x) + (t_2 - t_1)\Phi\left(\frac{R}{t_2 - t_1}\right).
\end{aligned}
$$

We claim that there is $\delta > 0$ being independent of y and s such that $D^+(y,s) \le R + d(x,z_0) + 1$ holds when $0 < s < \delta$. Indeed, if it is not the case, there is a sequence $(s_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ with $s_1 < t_0$, $s_n \downarrow 0$ and $D^+(y,s_n) > R + d(x,z_0) + 1$ for each $n \in \mathbb{N}$. Since $s_n\Phi(u/s_n) - t_1\Phi(u/t_1)$ and $u^{-1}\Phi(u)$ are non-decreasing in u, (3.3) yields

$$
(R + d(x,z_0) + 1)\,\frac{s_n}{R + d(x,z_0)}\Phi\left(\frac{R + d(x,z_0)}{s_n}\right) \le s_n\Phi\left(\frac{R + d(x,z_0)}{s_n}\right) + C_1,
$$

where $C_1 > 0$ is a constant independent of $n \in \mathbb{N}$. Since Φ is superlinear at infinity, the last inequality implies the contradiction by tending $n \to \infty$. Hence the claim holds. Thus it suffices to show the assertion only when $s \ge \delta$. In this case, we can replace s in the right hand side of (3.3) with δ and s in the left hand side of (3.3) with t_0. Therefore, the proof will be completed once we show $t_0\Phi(u/t_0) - t_1\Phi(u/t_1) \to \infty$ as $u \to \infty$. Since Φ is convex and superlinear at infinity, $\Phi'(u) \to \infty$ as $u \to \infty$. Thus we can apply Lemma 3.2 below with $\alpha = t_1^{-1}$, $\beta = t_0^{-1}$ and $g = \Phi'$ to conclude the assertion. □

Lemma 3.2. *Let $g : [0,\infty) \to [0,\infty)$ be a non-decreasing with $\lim_{u\to\infty} g(u) = \infty$. Then, for $\beta > \alpha > 0$,*

$$
\int_1^\infty (g(\beta u) - g(\alpha u))\,du = \infty.
$$

Proof. Note that the integrand is non-negative by assumption. Set $\eta := \beta/\alpha$ and $\eta_k := \beta^k/\alpha^{k-1}$. Then, by the Abel method on summation by parts,

$$
\begin{aligned}
\int_1^{\eta^n} (g(\beta u) - g(\alpha u))du &= \sum_{k=1}^n \int_{\eta^{k-1}}^{\eta^k} (g(\beta u) - g(\alpha u))du \\
&= \sum_{k=1}^n \eta^{k-1} \int_1^\eta (g(\eta_k u) - g(\eta_{k-1} u))\,du \\
&= \eta^{n-1}\int_1^\eta g(\eta_n u)\,du - \int_1^\eta g(\alpha u)\,du - (\eta - 1)\sum_{k=2}^n \eta^{k-2}\int_1^\eta g(\eta_{k-1} u)\,du \\
&= \int_1^\eta g(\eta_n u)\,du - \int_1^\eta g(\alpha u)\,du + (\eta - 1)\sum_{k=2}^n \eta^{k-2}\int_1^\eta (g(\eta_n u) - g(\eta_{k-1} u))\,du \\
&\ge \int_1^\eta g(\eta_{n-1} u)\,du - \int_1^\eta g(\alpha u)\,du.
\end{aligned}
$$

Therefore the conclusion follows by tending $n \to \infty$. □

Lemma 3.3. *Let $x \in \mathcal{D}(f)$ and $t \in (0, t_*(x))$. For $x_n \in X$ and $t_n \in (0, t_*(x))$ with $d(x, x_n) \to 0$ and $t_n \to t$ as $n \to \infty$, we have*

$$D^-(x,t) \le \liminf_{n\to\infty} D^-(x_n, t_n), \quad D^+(x,t) \ge \limsup_{n\to\infty} D^+(x_n, t_n).$$

Proof. For each $n \in \mathbb{N}$, let $(y_{n,k})_{k\in\mathbb{N}}$ be a minimizing sequence of $F(t_n, x_n, \cdot)$ such that $\lim_{k\to\infty} d(x_n, y_{n,k}) = D^+(x_n, t_n)$. Then, for each $n \in \mathbb{N}$, we can take $k_n \in \mathbb{N}$ such that $k_{n+1} > k_n$ and

$$F(t_n, x_n, y_{n,k_n}) \le Q_{t_n} f(x_n) + \frac{1}{n}, \qquad |d(x_n, y_{n,k_n}) - D^+(x_n, t_n)| \le \frac{1}{n}. \tag{3.4}$$

By virtue of Lemma 3.1 (ii), (3.4) yields

$$\lim_{n\to\infty} \left| t_n \Phi\left(\frac{d(x_n, y_{n,k_n})}{t_n}\right) - t\Phi\left(\frac{d(x, y_{n,k_n})}{t}\right) \right| = 0.$$

This fact together with the upper semi-continuity $\limsup_{n\to\infty} Q_{t_n} f(x_n) \le Q_t f(x)$ and (3.4) yields that $(y_{n,k_n})_{n\in\mathbb{N}}$ is a minimizing sequence of $F(t, x, \cdot)$. Then we have

$$D^+(x,t) \ge \limsup_{n\to\infty} d(x, y_{n,k_n}) = \limsup_{n\to\infty} d(x_n, y_{n,k_n}) \ge \limsup_{n\to\infty} D^+(x_n, t_n),$$

where the last inequality comes from (3.4). Hence the assertion for D^+ is proved. We can show the assertion for D^- in a similar way. □

From now on, we will turn to discuss the Hamilton-Jacobi equation associated with $Q_t f$. Our first goal is to show the sub-solution property of $Q_t f$ (Theorem 3.6).

Proposition 3.4. *For $x \in \mathcal{D}(f)$ and $t \in (0, t_*(x))$, we have*

$$\frac{d^+}{dt} Q_t f(x) = -\Phi^*\left(\Phi'\left(\frac{D^+(x,t)}{t}\right)\right), \qquad \frac{d^-}{dt} Q_t f(x) = -\Phi^*\left(\Phi'\left(\frac{D^-(x,t)}{t}\right)\right).$$

In particular, $Q_t f(x)$ is differentiable at t if and only if $D^+(x,t) = D^-(x,t)$.

Proof. Take $s \in (0, t_*(x))$. Let $(y_n)_n$ be a minimizing sequence of $F(t, x, \cdot)$. Then

$$\begin{aligned} Q_s f(x) - Q_t f(x) &\le \liminf_{n\to\infty} [F(s, x, y_n) - F(t, x, y_n)] \\ &= \liminf_{n\to\infty} \left[(s-t)\Phi\left(\frac{d(x,y_n)}{t}\right) + s\left(\Phi\left(\frac{d(x,y_n)}{s}\right) - \Phi\left(\frac{d(x,y_n)}{t}\right)\right)\right]. \end{aligned}$$

It yields

$$Q_s f(x) - Q_t f(x) \le (s-t)\Phi\left(\frac{D^+(x,t)}{t}\right) + s\left(\Phi\left(\frac{D^+(x,t)}{s}\right) - \Phi\left(\frac{D^+(x,t)}{t}\right)\right).$$

Note that the Legendre duality implies $\Phi^*(\Phi'(w)) = \Phi'(w)w - \Phi(w)$ for $w \ge 0$. Hence we obtain the upper bound of the right derivative in t of $Q_t f(x)$ from the last inequality. For the corresponding lower bound, for $s > t$, let us take a minimizing sequence $(y'_n)_{n\in\mathbb{N}}$ of $F(s,x,\cdot)$ satisfying $d(x,y'_n) \to D^+(x,s)$ as $n \to \infty$. By a similar argument as above but using $(y'_n)_n$ instead,

$$Q_s f(x) - Q_t f(x) \ge s\Phi\left(\frac{D^+(x,s)}{s}\right) - t\Phi\left(\frac{D^+(x,s)}{t}\right). \tag{3.5}$$

Take $\varepsilon > 0$. Since $D^+(x,\cdot)$ is upper semi-continuous by Lemma 3.3, $D^+(x,s) \le D^+(x,t) + \varepsilon$ holds if $s - t > 0$ is sufficiently small. In addition, the convexity of Φ yields that $s\Phi(u/s) - t\Phi(u/t)$ is non-increasing in u if $s > t$ (cf. (3.8) below). Thus, we can replace $D^+(x,s)$ in (3.5) with $D^+(x,t) + \varepsilon$ if $s - t > 0$ is sufficiently small. It yields the lower bound of the right derivative by obtaining a bound from the last inequality and tending $\varepsilon \downarrow 0$ after that. The result for the left derivative can be shown similarly. □

Proposition 3.5. *For $x \in \mathcal{D}(f)$ and $t \in (0, t_*(x))$,*

$$|\nabla Q_t f|(x) \le \Phi'\left(\frac{D^+(x,t)}{t}\right), \qquad |\nabla^+ Q_t f|(x) \le \Phi'\left(\frac{D^-(x,t)}{t}\right),$$

where $|\nabla^+ f|$ is defined by replacing $|f(y) - f(x)|$ in the definition of $|\nabla f|$ in (2.2) with $[f(y) - f(x)]_+$.

Proof. Let $x' \in X$ and $(y_n)_{n\in\mathbb{N}}$ a minimizing sequence of $F(t,x,\cdot)$. Then the convexity of Φ yields

$$\begin{aligned}
Q_t f(x') - Q_t f(x) &\le \liminf_{n\to\infty} \left[F(t,x',y_n) - F(t,x,y_n)\right] \\
&= \liminf_{n\to\infty} \left[t\Phi\left(\frac{d(x',y_n)}{t}\right) - t\Phi\left(\frac{d(x,y_n)}{t}\right)\right] \\
&\le \liminf_{n\to\infty} \left(d(x',y_n) - d(x,y_n)\right)\Phi'\left(\frac{d(x',y_n)}{t}\right) \\
&\le d(x',x) \liminf_{n\to\infty} \Phi'\left(\frac{d(x,y'_n)}{t}\right).
\end{aligned}$$

Thus we obtain

$$Q_t f(x') - Q_t f(x) \le d(x',x)\Phi'\left(\frac{D^-(x,t)}{t}\right).$$

This estimate easily implies the latter assertion. For the former one, by the same argument with the exchange of the role of x' and x,

$$|Q_t f(x) - Q_t f(x')| \le d(x,x')\Phi'\left(\frac{\max\{D^+(x,t), D^+(x',t)\}}{t}\right).$$

Since $D^+(\cdot,t)$ is upper semi-continuous by Lemma 3.3, the conclusion follows. □

Now the following immediately follows from Proposition 3.4 and Proposition 3.5.

Theorem 3.6. *For $x \in \mathcal{D}(f)$ and $t \in (0, t_*(x))$,*

$$\frac{d^+}{dt} Q_t f(x) + \Phi^*(|\nabla Q_t f|(x)) \le 0, \qquad \frac{d^-}{dt} Q_t f(x) + \Phi^*\left(|\nabla^+ Q_t f|(x)\right) \le 0. \tag{3.6}$$

The rest of this section is devoted to the differentiability in t of $Q_t f$ and the equality in (3.6).

Theorem 3.7. *Suppose that Φ is strictly convex. Then $D^+(x,t) \le D^-(x,s)$ holds for $x \in \mathcal{D}(f)$ and $0 < t < s < t_*(x)$. In particular, $D^+(x,t) = D^-(x,t)$ holds and hence $Q_t f(x)$ is differentiable in t with at most countably many exceptions for each fixed $x \in X$.*

Proof. Let us take minimizing sequences $(y_n)_n$ and $(y_n')_n$ of $F(s,x,\cdot)$ and $F(t,x,\cdot)$ respectively satisfying $d(x,y_n) \to D^-(x,s)$ and $d(x,y_n') \to D^+(x,t)$ as $n \to \infty$. We may assume $f(y_n) < \infty$ and $f(y_n') < \infty$ for all $n \in \mathbb{N}$. Take $\varepsilon > 0$ arbitrary. Then, for sufficiently large $n \in \mathbb{N}$,

$$\begin{aligned} F(s,x,y_n) &\le Q_s f(x) + \varepsilon \le F(s,x,y_n') + \varepsilon, \\ F(t,x,y_n') &\le Q_t f(x) + \varepsilon \le F(t,x,y_n) + \varepsilon. \end{aligned}$$

By summing them up, letting $n \to \infty$ and $\varepsilon \downarrow 0$, we obtain

$$t\left(\Phi\left(\frac{D^+(x,t)}{t}\right) - \Phi\left(\frac{D^-(x,s)}{t}\right)\right) \le s\left(\Phi\left(\frac{D^+(x,t)}{s}\right) - \Phi\left(\frac{D^-(x,s)}{s}\right)\right). \tag{3.7}$$

Here we implicitly used Lemma 3.1 (ii) to ensure the finiteness of $D^\pm$. Now we prove the assertion by contradiction. Suppose that $D^-(x,s) < D^+(x,t)$ holds. Since Φ is convex, $D^-(x,s)/t > D^-(x,s)/s$ and $D^+(x,t)/t > D^+(x,t)/s$ yield

$$\begin{aligned} &\left(\frac{D^+(x,t) - D^-(x,s)}{s}\right)^{-1} \left(\Phi\left(\frac{D^+(x,t)}{s}\right) - \Phi\left(\frac{D^-(x,s)}{s}\right)\right) \\ &\qquad \le \left(\frac{D^+(x,t) - D^-(x,s)}{t}\right)^{-1} \left(\Phi\left(\frac{D^+(x,t)}{t}\right) - \Phi\left(\frac{D^-(x,s)}{t}\right)\right). \end{aligned} \tag{3.8}$$

Thus the equality must hold in (3.7), but it is absurd since Φ is strictly convex. Hence $D^+(x,s) \le D^-(x,t)$. The assertion for the coincidence of D^+ and D^- is easy because $D^-(x,t) \le D^+(x,t) < \infty$. Then the assertion for the differentiability of $Q_t f(x)$ in t is immediate from Proposition 3.4. □

Theorem 3.8. *Suppose Assumption 2.1* (i). *Then*

$$|\nabla^- Q_t f|(x) = |\nabla Q_t f|(x) = \Phi' \left(\frac{D^+(x,t)}{t} \right),$$

where $|\nabla^- f|$ is defined by replacing $|f(y) - f(x)|$ in the definition of $|\nabla f|$ in (2.2) *with $[f(y) - f(x)]_-$. As a result, the equality holds in the first inequality of* (3.6) *for every $t \in (0, t_*(x))$.*

Proof. It suffices to consider the case $D^+(x,t) > 0$ since the conclusion immediately follows from Proposition 3.5 if $D^+(x,t) = 0$. Take a minimizing sequence $(y_n)_{n\in\mathbb{N}}$ of $F(t,x,\cdot)$ satisfying $d(x,y_n) \to D^+(x,t)$ as $n \to \infty$. We may assume $d(x,y_n) > 0$ for all $n \in \mathbb{N}$. Take a d-minimal geodesic $\gamma_n : [0,1] \to X$ with $\gamma_n(0) = x$, $\gamma_n(1) = y_n$ for each $n \in \mathbb{N}$. Then $d(x, \gamma_n(1/n)) \to 0$ as $n \to \infty$. Thus the mean value theorem yields

$$\begin{aligned}
|\nabla^- Q_t f|(x) &\geq \limsup_{n\to\infty} \frac{Q_t f(x) - Q_t f(\gamma_n(1/n))}{d(x, \gamma_n(1/n))} \\
&\geq \limsup_{n\to\infty} \frac{F(t,x,y_n) - F(t, \gamma_n(1/n), y_n)}{d(x, \gamma_n(1/n))} \\
&= \limsup_{n\to\infty} \frac{nt}{d(x,y_n)} \left(\Phi\left(\frac{d(x,y_n)}{t} \right) - \Phi\left(\frac{(1-n^{-1}) d(x,y_n)}{t} \right) \right) \\
&= \Phi'\left(\frac{D^+(x,t)}{t} \right).
\end{aligned}$$

Therefore the conclusion follows from this estimate and Proposition 3.5. □

§ 4. Proof of Theorem 2.2

To begin with, we gather extensions of well known properties for L^p-Wasserstein distance to W_Φ associated with the extended distance d. We refer to [28, Chapter 4] for basic properties of optimal transportation costs which is used in the proof of the following auxiliary lemma.

Lemma 4.1.

(i) *W_Φ is sequentially lower semi-continuous with respect to the weak convergence of probability measures. That is, for sequences $(\mu_n)_{n\in\mathbb{N}}$ and $(\nu_n)_{n\in\mathbb{N}}$ in $\mathscr{P}(X)$ which weakly converge to $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(X)$ respectively,*

$$W_\Phi(\mu,\nu) \leq \liminf_{n\to\infty} W_\Phi(\mu_n, \nu_n).$$

(ii) *For each $\mu, \nu \in \mathscr{P}(X)$, a minimizer of $W_\Phi(\mu,\nu)$ exists. That is, there is $\pi \in \Pi(\mu,\nu)$ such that $W_\Phi(\mu,\nu) = \|d\|_{L^\Phi(\pi)}$ holds.*

(iii) *For each* $\mu, \nu \in \mathscr{P}(X)$, $\lim_{p\to\infty} W_p(\mu,\nu) = W_\infty(\mu,\nu)$.

Proof. (i) For $\lambda' > 0$ and $\mu', \nu' \in \mathscr{P}(X)$, let us define $A(\mu',\nu',\lambda')$ by

$$A(\mu',\nu',\lambda') := \inf_{\pi\in\Pi(\mu',\nu')} \int_{X\times X} \Phi\left(\frac{d}{\lambda'}\right) d\pi.$$

Since $\Phi(d/\lambda')$ is non-negative and lower semi-continuous, a minimizer of $A(\mu',\nu',\lambda')$ always exists. Note that $A(\mu',\nu',\lambda') \le 1$ is equivalent to $W_\Phi(\mu',\nu') \le \lambda'$, which follows from the corresponding fact for the gauge norm (see [23]).

Let us take $\lambda > 0$ so that $A(\mu,\nu,\lambda) > 1$. By the lower semi-continuity of optimal transportation cost for the cost function $\Phi(d/\lambda)$,

$$A(\mu,\nu,\lambda) \le \liminf_{n\to\infty} A(\mu_n,\nu_n,\lambda). \tag{4.1}$$

Thus we have $A(\mu_n,\nu_n,\lambda) > 1$ for sufficiently large $n \in \mathbb{N}$. As a result, we obtain $\lambda \le \liminf_{n\to\infty} W_\Phi(\mu_n,\nu_n)$ and hence the conclusion follows by letting $\lambda \uparrow W_\Phi(\mu,\nu)$.

(ii) It directly follows from (i) and the fact that $\Pi(\mu,\nu)$ is compact with respect to the topology of weak convergence inherited from $\mathscr{P}(X\times X)$.

(iii) We can show it in the same way as [17, Lemma 3.2] by using the lower semi-continuity of the optimal transportation cost (cf. (4.1)). $\square$

In the sequel, we will enter the proof of Theorem 2.2. We refer to [23] for basic facts on the gauge norm and the Orlicz norm which are used in the proof.

Proof. of Theorem 2.2 (i). Let $\tilde{y} \in \tilde{X}$ with $0 < \tilde{d}(\tilde{x},\tilde{y}) < \infty$ and $\pi \in \Pi(P_{\tilde{x}}, P_{\tilde{y}})$ a minimizer of $W_\Phi(P_{\tilde{x}}, P_{\tilde{y}})$. For $r > 0$, we define $G_r f : X \to \mathbb{R}$ by

$$G_r f(z) := \sup\left\{ \frac{|f(z)-f(w)|}{d(z,w)} \;\middle|\; w \in X, d(z,w) \in (0,r) \right\}.$$

Then we have

$$\begin{aligned}
\frac{|Pf(\tilde{x}) - Pf(\tilde{y})|}{\tilde{d}(\tilde{x},\tilde{y})} &= \frac{1}{\tilde{d}(\tilde{x},\tilde{y})} \left| \int_{X\times X} (f(z)-f(w))\pi(dzdw) \right| \\
&\le \int_{X\times X} \frac{G_r f(z) d(z,w)}{\tilde{d}(\tilde{x},\tilde{y})} \pi(dzdw) + \frac{2\|f\|_\infty \pi(d\ge r)}{\tilde{d}(\tilde{x},\tilde{y})}.
\end{aligned} \tag{4.2}$$

For the second term of the right hand side of (4.2), The Chebyshev inequality together with the choice of π and $(W(\Phi))$ implies

$$\begin{aligned}
\pi(d\ge r) &\le \Phi\left(\frac{r}{\|d\|_{L^\Phi(\pi)}}\right)^{-1} \int_{X\times X} \Phi\left(\frac{d}{\|d\|_{L^\Phi(\pi)}}\right) d\pi \\
&\le \Phi\left(\frac{r}{\|d\|_{L^\Phi(\pi)}}\right)^{-1} = \Phi\left(\frac{r}{W_\Phi(P_{\tilde{x}},P_{\tilde{y}})}\right)^{-1} \le \Phi\left(\frac{r}{\tilde{d}(\tilde{x},\tilde{y})}\right)^{-1}.
\end{aligned} \tag{4.3}$$

For the first term in the right hand side of (4.2), the Hölder inequality for the gauge norm and the Orlicz norm together with $(W(\Phi))$ yields

$$\int_{X\times X} \frac{G_r f(z) d(z,w)}{\tilde{d}(\tilde{x},\tilde{y})} \pi(dzdw) \le \left(\| |\nabla f| \|_{\tilde{L}^{\Phi^*}(P_{\tilde{x}})} + \| G_r f - |\nabla f| \|_{\tilde{L}^{\Phi^*}(P_{\tilde{x}})} \right). \tag{4.4}$$

Note that we have

$$\limsup_{r\downarrow 0} \| G_r f - |\nabla f| \|_{\tilde{L}^{\Phi^*}(P_{\tilde{x}})} \le 2 \limsup_{r\downarrow 0} \| G_r f - |\nabla f| \|_{L^{\Phi^*}(P_{\tilde{x}})} = 0.$$

Here the first inequality comes from the general relation between the Orlicz norm and the gauge norm, and the second follows from the usual monotone convergence theorem for a decreasing sequence of functions. Take $\varepsilon > 0$ and set $r = \tilde{d}(\tilde{x},\tilde{y})\Phi^{-1}(\varepsilon^{-1}\tilde{d}(\tilde{x},\tilde{y})^{-1})$. Note that r tends to 0 as $\tilde{d}(\tilde{x},\tilde{y}) \to 0$ since Φ is superlinear at infinity. Thus, with this choice of r, by plugging (4.3) and (4.4) into (4.2) and by letting $\tilde{y} \to \tilde{x}$ with respect to $\tilde{d}$, we obtain the conclusion since $\varepsilon > 0$ is arbitrary. □

For proving the opposite implication, we prepare some additional properties of the Hopf-Lax semigroup $Q_t f$.

Lemma 4.2. *Under Assumption 2.1* (i), *for $f \in \mathrm{Lip}_b(X)$, $x, y \in X$ and $t, s > 0$,*

$$|Q_t f(x) - Q_t f(y)| \le \mathrm{Lip}(f) d(x,y), \qquad |Q_t f(x) - Q_s f(x)| \le \Phi^*(\mathrm{Lip}(f))|t-s|,$$

where $\mathrm{Lip}(f)$ *is the (global) Lipschitz constant of f with respect to d.*

We can prove this assertion in the same way as in the proof of [9, Theorem 2.1 (iv)]. Thus we omit the proof.

Lemma 4.3. *Suppose Assumption 2.1* (i) *and $(G(\Phi^*))$. Let $f \in \mathrm{Lip}_b(X)$ and $\tilde{\gamma} : [0,1] \to \tilde{X}$ a $\tilde{d}$-minimal geodesic. Then $PQ_t f(\tilde{\gamma}(t))$ is Lipschitz in $t \in [0,1]$.*

Proof. Note that $|\nabla f|$ is a d-upper gradient if $f : X \to \mathbb{R}$ is Lipschitz with respect to d (see [3, Section 2.3], for instance). Thus, under Assumption 2.1 (i), $(G(\Phi^*))$ and Lemma 4.2 yield that $PQ_t f$ is Lipschitz with respect to $\tilde{d}$ if $f \in \mathrm{Lip}_b(X)$. Moreover, $|\nabla PQ_t f|$ is bounded uniformly in t. Thus we can easily show $PQ_t f(\tilde{\gamma}(t))$ is Lipschitz in $t \in (0,1]$ (cf. the proof of [17, Proposition 3.7]). Thus only the continuity at $t = 0$ is left. Since the pointwise convergence of $Q_t f$ to f follows in the same way as [12, Proposition A.3 (3)], we can show $PQ_t f(\tilde{\gamma}(0)) \to Pf(\tilde{\gamma}(0))$ as $t \to 0$ and it implies the conclusion. □

Now we are ready to finish the proof of Theorem 2.2 (ii).

Proof. of Theorem 2.2 (ii). Set $\lambda := \tilde{d}(\tilde{x}, \tilde{y})$. By the Kantorovich duality,

$$\begin{aligned}\inf_{\pi\in\Pi(P_{\tilde{x}},P_{\tilde{y}})}\int_{X\times X}\Phi\left(\frac{d}{\lambda}\right)d\pi &= \sup_{f\in C_b(X)}\left[PQ_1f(\tilde{y})-Pf(\tilde{x})\right]\\ &= \sup_{f\in \mathrm{Lip}_b(X)}\left[PQ_1f(\tilde{y})-Pf(\tilde{x})\right],\end{aligned}\tag{4.5}$$

where Q_tf is the Hopf-Lax semigroup associated with $\Phi(\cdot/\lambda)$ instead of Φ in (3.1). For the second equality, see [13] or the proof of [28, Theorem 5.11], for instance. Let $f \in \mathrm{Lip}_b(X)$. By Assumption 2.1 (ii), there exists a $\tilde{d}$-geodesic $\tilde{\gamma} : [0,1] \to \tilde{X}$ with $\tilde{\gamma}(0) = \tilde{x}$ and $\tilde{\gamma}(1) = \tilde{y}$. By Lemma 4.3, $PQ_t\tilde{f}(\tilde{\gamma}(t))$ is differentiable in t a.e. with respect to the Lebesgue measure and the derivative is bounded. Thus we have

$$PQ_1f(\tilde{y})-Pf(\tilde{x}) = PQ_1f(\tilde{\gamma}(1)) - PQ_0f(\tilde{\gamma}(0)) = \int_0^1 \frac{\partial}{\partial t}PQ_tf(\tilde{\gamma}(t))dt. \tag{4.6}$$

By virtue of Lemma 4.2, we can apply [6, Lemma 4.3.4] to obtain

$$\begin{aligned}\frac{\partial}{\partial t}PQ_tf(\tilde{\gamma}(t)) \le \limsup_{h\downarrow 0}&\frac{PQ_{t+h}f(\tilde{\gamma}(t))-PQ_tf(\tilde{\gamma}(t))}{h}\\ &+\limsup_{h\downarrow 0}\frac{PQ_tf(\tilde{\gamma}(t))-PQ_tf(\tilde{\gamma}(t-h))}{h}\end{aligned}\tag{4.7}$$

for a.e. $t \in (0,1)$. Our assumption $(G(\Phi^*))$ yields

$$\begin{aligned}\limsup_{h\downarrow 0}\frac{PQ_tf(\tilde{\gamma}(t))-PQ_tf(\tilde{\gamma}(t-h))}{h} &\le \lambda|\tilde{\nabla}PQ_tf|(\tilde{\gamma}(t))\\ &\le \|\lambda|\nabla Q_tf|\|_{\tilde{L}^{\Phi^*}(P_{\tilde{\gamma}(t)})}.\end{aligned}\tag{4.8}$$

For the first term in the right hand side of (4.7), Theorem 3.6 yields

$$\limsup_{h\downarrow 0}\frac{PQ_{t+h}f(\tilde{\gamma}(t))-PQ_tf(\tilde{\gamma}(t))}{h} \le -P(\Phi^*(\lambda|\nabla Q_tf|))(\tilde{\gamma}(t)) \tag{4.9}$$

by the Dominated convergence theorem and Lemma 4.2. By plugging (4.8) and (4.9) into (4.7),

$$\frac{\partial}{\partial t}PQ_tf(\tilde{\gamma}(t)) \le \|\lambda|\nabla Q_tf|\|_{\tilde{L}^{\Phi^*}(P_{\tilde{\gamma}(t)})} - P(\Phi^*(\lambda|\nabla Q_tf|))(\tilde{\gamma}(t)). \tag{4.10}$$

By virtue of the definition of the Orlicz norm, the Hausdorff-Young inequality yields

$$\|\lambda|\nabla Q_tf|\|_{\tilde{L}^{\Phi^*}(P_{\tilde{\gamma}(t)})} \le P(\Phi^*(\lambda|\nabla Q_tf|))(\tilde{\gamma}(t)) + 1$$

By combining this estimate with (4.10), (4.6) and (4.5), we obtain

$$\inf_{\pi\in\Pi(P_{\tilde{x}},P_{\tilde{y}})}\int_{X\times X}\Phi\left(\frac{d}{\lambda}\right)d\pi \le 1.$$

It means $W_\Phi(P_{\tilde{x}}, P_{\tilde{y}}) \le \lambda$ and hence the conclusion holds. □

Remark 3. We can easily show that (W_p) implies

$$W_p(P^*\tilde{\mu}, P^*\tilde{\nu}) \le \tilde{W}_p(\tilde{\mu}, \tilde{\nu}) \tag{4.11}$$

for any $\tilde{\mu}, \tilde{\nu} \in \mathscr{P}(\tilde{X})$ by applying [28, Lemma 4.8] (see [17, Lemma 3.3] also). In particular, if $\tilde{X} = X$ and $\tilde{d} = Cd$ for some constant $C > 0$, we can obtain the corresponding estimate for the iteration P^n of the Markov kernel P. It is not clear whether the same argument works for W_Φ or not.

Remark 4. When we can obtain (W_{p_*}) in a functional analytic way, it sometimes occurs that it holds only $\tilde{\mathfrak{m}}$-a.e. for some base measure $\tilde{\mathfrak{m}}$ with $\mathrm{supp}(\tilde{\mathfrak{m}}) = \tilde{X}$. Even in such a case, we can obtain (4.11) if the following additional assumption holds: There exists a probability measure $\tilde{\pi}^*$ on the space of $\tilde{d}$-minimal geodesics in $\tilde{X}$ such that $(e_0 \times e_1)_\sharp \tilde{\pi}^* \in \Pi(\tilde{\mu}, \tilde{\nu})$ is optimal and $(e_t)_\sharp \tilde{\pi}^* \ll \tilde{\mathfrak{m}}$ for each $t > 0$, where $e_t(\tilde{\gamma}) := \tilde{\gamma}(t)$ is the evaluation map. Indeed, the Fubini theorem implies that, for $\tilde{\pi}^*$-a.e. $\tilde{\gamma}$, (W_{p_*}) holds at $\tilde{\gamma}(t)$ for a.e. t. Thus we can apply the same argument as in the proof of Theorem 2.2 (ii) for $\tilde{\pi}^*$-a.e. $\tilde{\gamma}$ instead of just one $\tilde{d}$-minimal geodesic. For example, if $\tilde{d}$ is a genuine distance being compatible with the topology of $\tilde{X}$ and the Ricci curvature is bounded from below on $\tilde{X}$ in a generalized sense, then this additional assumption holds whenever $\tilde{\mu}, \tilde{\nu} \ll \tilde{\mathfrak{m}}$ (see [21, 25]).

Remark 5. As we can observe in the proof of Theorem 2.2, the duality between $(W(\Phi))$ and $(G(\Phi^*))$ is *local*. More precisely, in order to obtain $(G(\Phi^*))$ at $\tilde{x}$, we requires $(W(\Phi))$ for $\tilde{y} \in \tilde{X}$ where $\tilde{d}(\tilde{x}, \tilde{y})$ is small. Similarly, the proof of $(W(\Phi))$ requires $(G(\Phi^*))$ only on a $\tilde{d}$-minimal geodesic joining $\tilde{x}$ and $\tilde{y}$.

§ 5. Stabilities

We begin with the stability of $(W(\Phi))$ or $(G(\Phi^*))$ for weak convergence of Markov kernels. It immediately follows from Theorem 2.2 and Lemma 4.1 (i).

Corollary 5.1. *Let $P^{(n)}_{\tilde{x}}$ be a sequence of Markov kernels on X parametrized by $\tilde{x} \in \tilde{X}$. Suppose that, for each $\tilde{x} \in \tilde{X}$, $P^{(n)}_{\tilde{x}}$ converges to a Markov kernel $P_{\tilde{x}}$ as $n \to \infty$ with respect to the topology of the weak convergence of probability measures. If $(W(\Phi))$ holds for $P^{(n)}$ for each $n \in \mathbb{N}$, then the same holds for $P_{\tilde{x}}$. In particular, under Assumption 2.1, if $(G(\Phi^*))$ holds for $P^{(n)}$ for each $n \in \mathbb{N}$, then the same holds for $P_{\tilde{x}}$.*

Remark 6. It could be possible to extend Corollary 5.1 to the case the underlying space is varying. Let (X_n, d_n) be a sequence of compact *metric* spaces which converges to a metric space (X, d) in the Gromov-Hausdorff sense. If a sequence of Markov kernels $P^{(n)}_{\tilde{x}} \in \mathscr{P}(X_n)$ converges to a Markov kernel $P_{\tilde{x}} \in \mathscr{P}(X)$ associated with the convergence of spaces, then the same stability should hold (cf. [2] or [19, Section 7]).

For Theorem 5.3 below which deals with the tensorization property, we state the following lemma. It asserts the stability for a push-forward by a 1-Lipschitz map.

Lemma 5.2. *Let $\hat{X}$ be a Polish space equipped with an extended distance $\hat{d}$ and $\varphi : X \to \hat{X}$ a 1-Lipschitz map with respect to $\hat{d}$ and d. We define a new Markov kernel $\hat{P}_{\tilde{x}}$ on $\hat{X}$ by the push-forward: $\hat{P}_{\tilde{x}} := \varphi_\sharp P_{\tilde{x}}$. Suppose that $(W(\Phi))$ or $(G(\Phi^*))$ holds. Then $\hat{P}$ also enjoys the corresponding property.*

We omit the proof of Lemma 5.2 since we can show it by a simple straightforward argument. Note that we do not require Theorem 2.2 for the proof.

Theorem 5.3. *Let Λ be an at most countable set. For each $i \in \Lambda$, let X_i and $\tilde{X}_i$ be Polish spaces equipped with extended distances d_i and $\tilde{d}_i$. Let $P^{(i)}_{\tilde{x}_i}$ be a Markov kernel on X_i parametrized by $\tilde{x}_i \in \tilde{X}_i$. Set $\boldsymbol{X} := \prod_{i\in\Lambda} X_i$ and $\tilde{\boldsymbol{X}} := \prod_{i\in\Lambda} \tilde{X}_i$. Let $\boldsymbol{d}_{(p)}$ and $\tilde{\boldsymbol{d}}_{(p)}$ be l^p-product extended distances on $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ respectively. That is, for $\boldsymbol{x} = (x_i)_{i\in\Lambda}$ and $\boldsymbol{y} = (y_i)_{i\in\mathbb{N}}$,*

$$\boldsymbol{d}_{(p)}(\boldsymbol{x}, \boldsymbol{y}) := \|(d_i(x_i, y_i))_{i\in\Lambda}\|_{l^p}.$$

Let $\boldsymbol{P}_{\tilde{\boldsymbol{x}}} := \otimes_{i\in\Lambda} P^{(i)}_{\tilde{x}_i}$ be the product Markov kernel on $\boldsymbol{X}$ parametrized by $\tilde{\boldsymbol{x}} = (\tilde{x}_i)_{i\in\Lambda} \in \tilde{\boldsymbol{X}}$. Then the following are equivalent:

(i) *$P^{(i)}$ enjoys (W_p) for each $i \in \Lambda$.*

(ii) *$\boldsymbol{P}$ enjoys (W_p) with respect to $\boldsymbol{d}_{(p)}$ and $\tilde{\boldsymbol{d}}_{(p)}$. That is,*

$$W_p\left(\boldsymbol{P}_{\tilde{\boldsymbol{x}}}, \boldsymbol{P}_{\tilde{\boldsymbol{y}}}\right) \le \tilde{\boldsymbol{d}}_{(p)}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$$

for any $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}} \in \tilde{\boldsymbol{X}}$, where W_p is defined on $\boldsymbol{X}$ with respect to $\boldsymbol{d}_{(p)}$.

In particular, the corresponding equivalence holds for the gradient estimate (W_{p_*}) under Assumption 2.1 for X_i and $\tilde{X}_i$ $(i \in \Lambda)$.

We remark that $\boldsymbol{d}_{(p)}$ can become an extended distance even if all of $(d_i)_{i\in\Lambda}$ are genuine distances when Λ is not finite.

Proof. By virtue of Remark 2, it suffices to consider the case $p \in [1, \infty)$. The implication from (ii) to (i) immediately follows from Lemma 5.2 since the canonical projection $\eta_i : \boldsymbol{X} \to X_i$ $(i \in \Lambda)$ is 1-Lipschitz with respect to $\boldsymbol{d}_{(p)}$ and d_i.

Let us consider the implication from (i) to (ii). Let $\tilde{\boldsymbol{x}} = (\tilde{x}_i)_{i\in\Lambda}, \tilde{\boldsymbol{y}} = (\tilde{y}_i)_{i\in\Lambda} \in \tilde{\boldsymbol{X}}$ and take an optimal $\pi_i \in \Pi(P_{\tilde{x}_i}, P_{\tilde{y}_i})$ for each $i \in \Lambda$ and set $\boldsymbol{\pi} := \otimes_{i\in\Lambda}\pi_i$. Note that $\boldsymbol{\pi}$ can be interpreted as an element of $\Pi(\boldsymbol{P}_{\tilde{\boldsymbol{x}}}, \boldsymbol{P}_{\tilde{\boldsymbol{y}}})$. Under this interpretation,

$$W_p(\boldsymbol{P}_{\tilde{\boldsymbol{x}}}, \boldsymbol{P}_{\tilde{\boldsymbol{y}}})^p \le \int_{\boldsymbol{X}\times\boldsymbol{X}} \boldsymbol{d}_{(p)}^p \, d\boldsymbol{\pi} = \sum_{i\in\Lambda} \int_{X_i\times X_i} d_i^p \, d\pi_i = \sum_{i\in\Lambda} W_p(P_{\tilde{x}_i}, P_{\tilde{y}_i})^p \le \boldsymbol{d}_{(p)}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})^p$$

by using the condition (i). Hence the assertion holds. □

The last result in this section is the stability for averaging. It follows similarly as in Remark 3 when $p < \infty$. The case $p = \infty$ will be dealt according to Remark 2.

Corollary 5.4. *Let $(S, \mathscr{S}, \xi)$ be a probability space and $(P^{(\alpha)})_{\alpha \in S}$ a family of Markov kernels parametrized on $\tilde{X}$. Let $(\tilde{d}_\alpha)_{\alpha \in S}$ be a family of extended distances on $\tilde{X}$. Assume that, for each $A \in \mathscr{B}(X)$, $(\tilde{x}, \alpha) \mapsto P^{(\alpha)}_{\tilde{x}}(A)$ is a measurable map from $(\tilde{X} \times S, \mathscr{B}(\tilde{X}) \otimes \mathscr{S})$ to $([0,1], \mathscr{B}([0,1]))$. We define the Markov kernel P and the extended distance $\tilde{d}$ on $\tilde{X}$ by*

$$P_{\tilde{x}}(A) := \int_S P^{(\alpha)}_{\tilde{x}}(A) \xi(d\alpha), \qquad \tilde{d}(\tilde{x}, \tilde{y}) := \|\tilde{d}_\cdot(\tilde{x}, \tilde{y})\|_{L^p(\xi)}.$$

(i) *Suppose (W_p) for P^α and $\tilde{d}_\alpha$ for a.e. $\alpha \in S$. Then (W_p) holds for P and $\tilde{d}$.*

(ii) *Suppose Assumption 2.1* (i) *and Assumption 2.1* (ii) *for $\tilde{d}_\alpha$ for a.e. $\alpha \in S$. Suppose (W_{p_*}) for $P^{(\alpha)}$ and $\tilde{d}_\alpha$ for a.e. $\alpha \in S$. Then (W_{p_*}) holds for P and $\tilde{d}$.*

We remark that Assumption 2.1 (ii) for $\tilde{d}$ is not required in Corollary 5.4 (ii).

§ 6. Examples

We first demonstrate that the Hölder continuity estimate for solutions to the Dirichlet problem falls into our framework.

Example 6.1. Let D be a regular bounded domain in $\mathbb{R}^m$, $m \geq 2$. We denote the Euclidean distance by ρ. Let us denote the harmonic measure over D by $(H_{\tilde{x}})_{\tilde{x} \in D}$. That is, $H_{\tilde{x}}$ is a Markov kernel on ∂D parametrized by $\tilde{x} \in D$ such that, given $f : \partial D \to \mathbb{R}$ bounded and measurable, Hf gives a solution to the Dirichlet problem $\Delta u = 0$ on D and $u|_{\partial D} = f$. Let $\alpha \in (0,1)$. In [1], the following property is studied in detail: there exists a constant $C > 0$ such that, for any bounded measurable function $f : \partial D \to \mathbb{R}$,

$$\|Hf\|_\infty + \|Hf\|_{C^{0,\alpha}} \leq C \left(\|f\|_\infty + \|f\|_{C^{0,\alpha}} \right), \tag{6.1}$$

where $\|\cdot\|_{C^{0,\alpha}}$ is the Hölder constant of the exponent α. We show that (6.1) can be interpreted as a variant of (G_∞) under an appropriate choice of X, $\tilde{X}$, d, $\tilde{d}$ and P. Let $\star$ and $\tilde{\star}$ be points separated from $\mathbb{R}^m$ and set $X = \partial D \cup \{\star\}$ and $\tilde{X} = D \cup \{\tilde{\star}\}$. We define a distance function d on X respectively by $d|_{\partial D \times \partial D} := \rho^\alpha$ and $d(x, \star) := \operatorname{diam}(\partial D)^\alpha$ for $x \in \partial D$. We also define a distance function $\tilde{d}'$ on $\tilde{X}$ in the same manner. Let $C' > 0$ be a constant and set $\tilde{d} := C' \tilde{d}'$. A Markov kernel $(P_{\tilde{x}})_{\tilde{x} \in \tilde{X}}$ on X is defined by $P_{\tilde{x}} = H_{\tilde{x}}$ when $\tilde{x} \in D$ and $P_{\tilde{\star}} = \delta_\star$. Now we claim that (6.1) is equivalent to (G_∞) up to a choice of constants. For $f : \partial D \to \mathbb{R}$, we extend it to $\hat{f} : X \to \mathbb{R}$ by $\hat{f}|_{\partial D} = f|_{\partial D}$ and $\hat{f}(\star) = 0$. Then (G_∞) for $\hat{f}$ means

$$\max\left\{ \frac{1}{\operatorname{diam}(D)} \|Hf\|_\infty, \|Hf\|_{C^{0,\alpha}} \right\} \le C' \max\left\{ \frac{1}{\operatorname{diam}(\partial D)} \|f\|_\infty, \|f\|_{C^{0,\alpha}} \right\}. \tag{6.2}$$

Obviously, (6.2) is equivalent to (6.1) up to a choice of C and C'. Note that (G_∞) is invariant under adding a constant to f. Thus if (G_∞) holds for those $f : X \to \mathbb{R}$ with $f(\star) = 0$, then (G_∞) holds for all $f : X \to \mathbb{R}$. These observations easily imply the claim. Though both d and $\tilde{d}$ do *not* satisfy Assumption 2.1 in this case, we can employ the duality since $p = 1$; See Remark 2.

Example 6.2. Let $\tilde{X}$ be a complete Riemannian manifold of $\dim \tilde{X} \ge 2$ without boundary. We denote the Riemannian distance on $\tilde{X}$ by $\tilde{d}$. Suppose that the Ricci curvature on $\tilde{X}$ is bounded from below by a constant $K \in \mathbb{R}$. It is well-known that for each $\tilde{x}_1, \tilde{x}_2 \in \tilde{X}$ there is a coupling of Brownian motions $(B^{(1)}(t), B^{(2)}(t))_{t \ge 0}$ starting from $(\tilde{x}_1, \tilde{x}_2)$ such that

$$\tilde{d}(B^{(1)}(t), B^{(2)}(t)) \le \mathrm{e}^{-Kt} \tilde{d}(\tilde{x}, \tilde{y}) \tag{6.3}$$

almost surely for each $t \ge 0$ (see e.g. [18, 30]). Let $X := C([0, \infty) \to \tilde{X})$ with the topology of compact uniform convergence. We define an (extended) distance d on X by $d(w, w') := \sup_{t \ge 0} \mathrm{e}^{Kt} \tilde{d}(w(t), w'(t))$. Let us define a Markov kernel $(P_{\tilde{x}})_{\tilde{x} \in \tilde{X}}$ as a Wiener measure. That is, $P_{\tilde{x}} \in \mathscr{P}(X)$ is the law of the Brownian motion on $\tilde{X}$ starting from $\tilde{x}$. Then we can easily verify that (6.3) yields (W_p) with $p = \infty$.

Example 6.3. The following estimate for a diffusion semigroup $P^{(t)}$ of Markov kernels on X is studied well in the literature (e.g. [5,18,29]): There is a constant $K \in \mathbb{R}$ and $p \in (1, \infty)$ such that, for $x, y \in X$ and $t > 0$,

$$W_p(P_x^{(t)}, P_y^{(t)}) \le \mathrm{e}^{-Kt} d(x, y). \tag{6.4}$$

It is regarded as a characterization of the presence of a lower Ricci curvature bound by K (Actually, (6.3) easily implies (6.4)). A subordination of $(P^{(t)})_{t \ge 0}$ by a subordinator $\zeta_t \in \mathscr{P}([0, \infty))$, $t \ge 0$ (see e.g. [24]) is an example of the averaging of (6.4) in the sense of Corollary 5.4. Actually, this estimate falls into the framework of Corollary 5.4 by choosing $\tilde{X} = X$ and $\tilde{d}_t := \mathrm{e}^{-Kt} d$. Thus we immediately obtain the following result.

Corollary 6.4. *Let X, d and $P_x^{(t)}$ be as in Example 6.3 and suppose* (6.4). *Let $\zeta_t \in \mathscr{P}([0, \infty))$, $t \ge 0$ be a subordinator with the Laplace exponent ψ. That is, for $z \ge 0$,*

$$\int_0^\infty \mathrm{e}^{-zs} \zeta_t(ds) = \exp\left(-t\psi(z)\right). \tag{6.5}$$

Let $P_x^{\zeta,(t)}$ *be the subordination of* $(P_x^{(t)})_{t\ge 0}$ *by* $(\zeta_t)_{t\ge 0}$*. That is, for* $t \ge 0$*,*

$$P_x^{\zeta,(t)} := \int_0^\infty P_x^{(s)}\, \zeta_t(ds).$$

When $K < 0$*, we assume that the left hand side of* (6.5) *is finite even when* $z = pK$ *and we denote the right hand side of it by using the same symbol* $\psi(pK)$*. Then we have*

$$W_p(P_x^{\zeta,(t)}, P_y^{\zeta,(t)}) \le \mathrm{e}^{-t\psi(pK)/p} d(x,y).$$

In particular, for β*-resolvent kernel* $R_x^{(\beta)} := \int_0^\infty \mathrm{e}^{-\beta s} P_x^{(s)}\, ds$ *with* $\beta > -pK$*,*

$$W_p(\beta R_x^{(\beta)}, \beta R_y^{(\beta)}) \le \left(\frac{\beta}{\beta + pK}\right)^{1/p} d(x,y).$$

Note that the corresponding stability of the gradient estimate (W_{p_*}) under subordination does not seem obvious when $K \ne 0$ in (6.4).

References

[1] H. Aikawa, *Hölder continuity of the Dirichlet solution for a general domain*, Bull. London Math. Soc. **34** (2002), 691–702.

[2] L. Ambrosio, N. Gigli, and G. Savaré, *Bakry-Émery curvature-dimension condition and Riemannian Ricci curvature bounds*, Preprint. Available at: arXiv: 1209.5786.

[3] L. Ambrosio, N. Gigli, and G. Savaré, *Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below*, Preprint. Available at: arXiv:1106.2090.

[4] ———, *Density of Lipschitz functions and equivalence of weak gradients in metric measure spaces*, Preprint. Available at: arXiv:1111.3730.

[5] ———, *Metric measure spaces with Riemannian Ricci curvature bounded from below*, Preprint. Available at: arXiv:1109.0222.

[6] ———, *Gradient flows in metric spaces and in the space of probability measures*, second ed., Birkhäuser Verlag, Basel, 2008.

[7] D. Bakry, *Functional inequalities for Markov semigroups*, Probability measures on groups: recent directions and trends (Mumbai), Tata Inst. Fund. Res., 2006, pp. 91–147.

[8] D. Bakry, I. Gentil, and M. Ledoux, *On Harnack inequalities and optimal transport*, Preprint. Available at: arXiv:1210.4650.

[9] Z.M. Balogh, A. Engoulatov, L. Hunziker, and O.E. Maasalo, *Functional inequalities and Hamilton-Jacobi equations in geodesic spaces*, Potential Anal. **36** (2012), no. 2, 317–337.

[10] François Bolley, Ivan Gentil, and Arnaud Guillin, *Convergence to equilibrium in Wasserstein distance for Fokker–Planck equations*, J. Funct. Anal. **263** (2012), no. 8, 2430–2457.

[11] N. Gigli, K. Kuwada, and S. Ohta, *Heat flow on Alexandrov spaces*, Comm. Pure Appl. Math. **66** (2013), no. 3, 307–331.

[12] N. Gozlan, C. Roberto, and P.-M. Samson, *Hamilton-jacobi equations on metric spaces and transport entropy inequalities*, To appear in Rev. Mat. Iberoam. Available at: arXiv:1203.2783.

[13] F. Hirsch, *Measurable metrics, intrinsic metrics, and Lipschitz functions*, Current trends in potential theory (Theta, Bucharest), Theta Ser. Adv. Math. 4, 2005, pp. 47–61.
[14] Y. Kitabeppu, *Lower bound of coarse Ricci curvature on metric measure spaces and eigenvalues of Laplacian*, Preprint. Available at: arXiv:1112.5820.
[15] E. Kokubo and K. Kuwae, *On nonlinear spectral gap for symmetric Markov chains with coarse Ricci curvatures*, Preprint.
[16] P. Koskela and Y. Zhou, *Geometry and analysis of Dirichlet forms*, Adv. Math. **231** (2012), no. 5, 2755–2801.
[17] K. Kuwada, *Duality on gradient estimates and Wasserstein controls*, J. Funct. Anal. **258** (2010), no. 11, 3758–3774.
[18] ———, *Convergence of time-inhomogeneous geodesic random walks and its application to coupling methods*, Ann. Probab. **40** (2012), no. 5, 1945–1979.
[19] K. Kuwada and K.-T. Sturm, *Monotonicity of time-dependent transportation costs and coupling by reflection*, To appear in Potential Anal.
[20] J. Lott and C. Villani, *Hamilton-Jacobi semigroup on length spaces and applications*, J. Math. Pures Appl. (9) **88** (2007), no. 3, 219–229.
[21] ———, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. Math. **169** (2009), no. 3, 903–991.
[22] Y. Ollivier, *Ricci curvature of Markov chains on metric spaces*, J. Funct. Anal. **256** (2009), no. 3, 810–864.
[23] M. M. Rao and Z. D. Ren, *Theory of orlicz spaces*, CRC press, 1991.
[24] K. Sato, *Lévy processes and infinitely divisible distributions*, Cambridge studies in advanced mathematics, 68, Cambridge university press, Cambridge, 1999.
[25] K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta. Math. **196** (2006), no. 1, 65–131.
[26] ———, *Generalized Orlicz spaces and Wasserstein distances for convex–concave scale functions*, Bull. Sci. Math. **135** (2011), no. 6-7, 795–802.
[27] C. Villani, *Topics in optimal transportations*, Graduate studies in mathematics, 58, Amer. Math. Soc., Providence, RI, 2003.
[28] ———, *Optimal transport, old and new*, Grundlehren der Mathematischen Wissenschaften, 338, Springer-Verlag, 2008.
[29] M.-K. von Renesse and K.-T. Sturm, *Transport inequalities, gradient estimates, entropy and Ricci curvature*, Comm. Pure. Appl. Math. **58** (2005), no. 7, 923–940.
[30] F.-Y. Wang, *On estimation of the logarithmic Sobolev constant and gradient estimates of heat semigroups*, Probab. Theory Related Fields **108** (1997), no. 1, 87–101.