Wasserstein 距離にまつわる確率解析[1]

桑田 和正 (Kuwada, Kazumasa)[2]
(お茶の水女子大学 (Ochanomizu University))

1 What is Wasserstein distance?

Let (M, d) be a Polish space (complete separable metric space) and $\mathcal{P}(M)$ the set of probability measures on $(M, \mathcal{B}(M))$. For $\mu_0, \mu_1 \in \mathcal{P}(M)$, let $\Pi(\mu_0, \mu_1)$ be the set of all couplings between μ_0 and μ_1. That is,

$$\Pi(\mu_0, \mu_1) := \left\{ \pi \in \mathcal{P}(M \times M) \;\middle|\; \begin{array}{l} \pi(A \times M) = \mu_0(A), \\ \pi(M \times A) = \mu_1(A) \end{array} \text{ for each } A \in \mathcal{B}(M) \right\}.$$

Under these notations, let us introduce L^p-Wasserstein distance $W_p(\mu_0, \mu_1)$ for $p \in [1, \infty]$ and $\mu_0, \mu_1 \in \mathcal{P}(M)$ as follows:

$$W_p(\mu_0, \mu_1) := \inf_{\pi \in \Pi(\mu_0, \mu_1)} \|d\|_{L^p(\pi)}.$$

In general, $W_p(\mu_0, \mu_1) = \infty$ can occur. To prevent it, we sometimes restrict W_p on $\mathcal{P}_p(M) \subset \mathcal{P}(M)$, the set of probability measures on M with a finite p-th moment (in terms of the distance function), defined as follows:

$$\mathcal{P}_p(M) := \left\{ \mu \in \mathcal{P}(M) \;\middle|\; \int_M d(x, y)^p \mu(dy) < \infty \text{ for some/any } x \in M \right\}.$$

The Wasserstein distance can be expressed by means of couplings between random variables. Let X_0 and X_1 be M-valued random variables and we denote distributions of X_i by μ_i for $i = 0, 1$. Then we say that an $M \times M$-valued random variable (Y_0, Y_1) is a coupling between X_0 and X_1 when the distribution of Y_i is the same as that of X_i for $i = 0, 1$. By using this terminology, we can express the Wasserstein distance as follows:

$$W_p(\mu_0, \mu_1) = \inf \left\{ \mathbb{E}[d(Y_0, Y_1)^p]^{1/p} \;\middle|\; (Y_0, Y_1) \text{ is a coupling between } X_0 \text{ and } X_1 \right\}.$$

As stated in Villani's book [37], some notions corresponding to Wasserstein distance have been appeared in several fields in mathematics and hence there are many different names, e.g. Monge-Kantorovich distance, Tanaka distance, Kantorovich-Rubinstein distance, minimal metrics. One reason why it happens could consist in the fact that it is very useful especially in measuring the rate of convergence of probability measures. Indeed, the Wasserstein distance enjoys the following nice properties (see [3, 37, 38] for instance; see [20, 33, 34] also)

- When $p < \infty$, "$\lim_{n \to \infty} W_p(\mu_n, \mu) = 0$" is equivalent to the following two conditions:
  (1) $\mu_n \to \mu$ (weak convergence of probability measures).
  (2) $\sup_{n \in \mathbb{N}} \int_M d(x, y)^p \mu_n(dy) < \infty$ for some/any $x \in M$.

[1]研究集会「確率論シンポジウム」(2011 年 12 月 19 日-12 月 22 日) 講演予稿

[2]URL: `http://www.math.ocha.ac.jp/kuwada` e-mail: `kuwada.kazumasa@ocha.ac.jp`

In particular, if d is bounded, W_p is a distance function being compatible with the topology of the weak convergence on $\mathcal{P}(X)$.

- W_p is a distance function on $\mathcal{P}_p(X)$.
- There is another variational expression of the Wasserstein distance by means of an integration against test functions (Kantorovich duality; see Theorem 1 below).
- We can obtain an (upper) bound of the Wasserstein distance by constructing a coupling of distributions explicitly.
- The Wasserstein distance is "stable" under "perturbation" of the underlying space or the distance function. Although this statement is not mathematically rigorous, this is (heuristically) important viewpoint.
- The property of the Wasserstein distance strongly reflects that of the metric structure of the underlying space. For instance, $(\mathcal{P}_p(X), W_p)$ is a Polish space for $1 \le p < \infty$. As another example, if d is a geodesic distance, the same is true for W_p, $p \in (1, \infty)$ (see Theorem 2 below).

It should be remarked that the Wasserstein distance is closely related with the Monge-Kantorovich mass transportation problem. Here we introduce two properties of the Wasserstein distance as a special case of the general theory of optimal transportation.

Theorem 1 (Kantorovich duality; e.g. [38, Theorem 5.10])

$$W_p(\mu_0, \mu_1)^p = \sup \left\{ \int_M f^* d\mu_0 - \int_M f d\mu_1 \;\middle|\; f \in C_b^{\mathrm{Lip}}(M) \right\},$$

where $f^*(x) := \inf\{f(y) + d(x,y)^p\}$. In particular,

$$W_1(\mu_0, \mu_1) = \sup \left\{ \int_M f d(\mu_0 - \mu_1) \;\middle|\; f : M \to \mathbb{R} \text{ 1-Lipschitz} \right\}.$$

The latter one is called the Kantorovich-Rubinstein formula.

To state the second property, we review the notion of geodesic distance. We call the distance d a geodesic distance when there exists a curve $\gamma : [0,1] \to M$ (of constant speed) from x to y whose length realizes the distance between x and y for any $x, y \in M$. More precisely, the curve $\gamma : [0,1] \to M$ appeared in the last sentence satisfies $\gamma(0) = x$, $\gamma(1) = y$ and $d(\gamma(s), \gamma(t)) = d(x,y)|s - t|$ for each $s, t \in [0,1]$. We call such a curve geodesic (of constant speed). Let $\Gamma([0,1]; M)$ be the set of constant speed geodesics $\gamma : [0,1] \to M$ with the topology of uniform convergence. We denote the evaluation map $\Gamma([0,1]; M) \to M$ at $t \in [0,1]$ by e_t. That is, $e_t(\gamma) := \gamma(t)$ for $\gamma \in \Gamma([0,1]; M)$. We denote the push-forward of a measure μ by a map f by $f^{\#}\mu$.

Theorem 2 (Displacement interpolation; e.g. [38, Corollary 7.22])

Suppose d to be a geodesic distance and $p \in (1, \infty)$. For $\mu_0, \mu_1 \in \mathcal{P}_p(M)$, There exists $\Xi \in \mathcal{P}(\Gamma([0,1]; M))$ such that $e_i^{\#}\Xi = \mu_i$ for $i = 0, 1$ and

$$W_p(e_t^{\#}\Xi, e_s^{\#}\Xi) = \left\{ \int_{\Gamma([0,1];M)} d(e_t(\gamma), e_s(\gamma))^p \Xi(d\gamma) \right\}^{1/p} = |s - t| W_p(\mu_0, \mu_1).$$

In particular, $(e_t^{\#}\Xi)_{t\in[0,1]}$ is a constant-speed geodesic on $(\mathcal{P}_p(X), W_p)$ joining μ_0 and μ_1, and hence W_p is a geodesic distance on $\mathcal{P}_p(M)$. Ξ is called a dynamical coupling of μ_0 and μ_1.

As mentioned above, the Wasserstein distance has appeared in several different contexts. In what follows, among them, we will concentrate on its connection with couplings of diffusion processes. For other topics, I just demonstrate some (incomplete) references.

- [32] for a general reference on optimal transportation other than [37, 38].
- [19, 35] for a general reference of coupling methods.
- [10, 11] for a recent development in coupling methods for SPDE.
- [9] for transport inequalities, which is a functional inequality involving the Wasserstein distance.

2 Wasserstein contraction and equivalent conditions

As a simple example, let us consider the following SDE on $\mathbb{R}^m$:

$$\begin{aligned} dX^x(t) &= \sqrt{2}dB_t - \nabla V(X_t^x)dt, \\ X^x(0) &= x, \end{aligned}$$

where $V \in C^\infty(\mathbb{R}^m)$ with $\nabla^2 V \geq K$ for some $K \in \mathbb{R}$. Then a coupling argument easily yields the following estimate of the Wasserstein distance: for $p \in [1, \infty]$,

$$W_p((X^x(t))^{\#}\mathbb{P}, (X^y(t))^{\#}\mathbb{P}) \leq e^{-Kt} d(x, y).$$

A condition of this kind is known to be equivalent to several geometric or analytic conditions for Brownian motions on a Riemannian manifold:

Theorem 3 (see [39] and references therein)

Let M be a complete Riemannian manifold and P_t the heat semigroup on M. Then, for $K \in \mathbb{R}$, the following conditions are equivalent:

(a) $W_2(P_t^*\mu_0, P_t^*\mu_1) \leq e^{-Kt} W_2(\mu_0, \mu_1)$ for $t > 0$ and $\mu_0, \mu_1 \in \mathcal{P}_2(X)$.

(b) $\mathrm{Ric} \geq K$, where Ric stands for the Ricci curvature.

(c) $|\nabla P_t f|(x)^2 \leq e^{-2Kt} P_t(|\nabla f|^2)(x)$ for $t > 0$, $x \in M$ and $f \in C^{\mathrm{Lip}}(M)$.

(d) Ent is K-convex with respect to W_2, i.e., for any W_2-geodesic μ_t in $\mathcal{P}_2(M)$,

$$\mathrm{Ent}(\mu_t) \leq (1-t)\,\mathrm{Ent}(\mu_0) + t\,\mathrm{Ent}(\mu_1) - \frac{K}{2} t(1-t) W_2(\mu_0, \mu_1)^2.$$

Here $\mathrm{Ent} : \mathcal{P}(X) \to \bar{\mathbb{R}}$ is the relative entropy defined by

$$\mathrm{Ent}(\mu) := \int_M \rho(x) \log \rho(x)\, dx$$

when $\mu(dx) = \rho(x)\, dx$ with $\displaystyle\int_M \rho(x)[\log \rho(x)]_+ dx < \infty$ and ∞ otherwise.

In differential geometry, the condition (b) has been studied well, especially in the case $K > 0$ or $K = 0$ (see e.g. [30]). The condition (c) is so-called Bakry-Émery's (L^2-)gradient estimate. It has several applications in analysis, especially in functional inequalities (see e.g. [5, 18]). The condition (d) is closely related to the theory of gradient flow on $(\mathcal{P}_2(M), W_2)$, which will be studied below.

3 Coupling by parallel transport

In Theorem 3, one direct way to prove (a) from (b) is to construct a coupling of two Brownian motions by parallel transport of infinitesimal motions. The resulted coupling $(B_t^{(1)}, B_t^{(2)})$ satisfies

$$d(B^{(1)}(t), B^{(2)}(t)) \le e^{-Kt} d(B^{(1)}(0), B^{(2)}(0)) \tag{1}$$

for all $t \ge 0$ $\mathbb{P}$-almost surely when $\mathrm{Ric} \ge K$. By taking an expectation, this inequality yields not only (a), or the W_2-contraction, but a W_p-contraction with $p \in [1, \infty]$. For the construction, we can make it by solving a (degenerated) SDE on $M \times M$ *if the distance function d has no other singularity than that on the diagonal set $\{(x, x) \mid x \in M\}$*. Therefore, almost all technical difficulties arises from the singular points of the distance function, or the cut locus. See [14, 41] and references therein for more details.

This idea of constructing a coupling by parallel transport works in a more general framework. An important extension is done on manifolds whose metric $g(t)$ depends on time parameter t and evolves as a backward Ricci flow, i.e.,

$$\partial_t g(t) = 2 \,\mathrm{Ric}_{g(t)}, \quad t \in [0, T]. \tag{2}$$

Note that the condition (2) would imply a sort of time-dependent analogue of non-negative Ricci curvature. As a direct extension of (1), we can obtain

$$d_t(B^{(1)}(t), B^{(2)}(t)) \le d_0(B^{(1)}(0), B^{(2)}(0))$$

for $t \ge 0$ $\mathbb{P}$-almost surely, where d_s is the distance function associated with the metric at time s (see [4, 13]; see [23] also). In addition, there is analogous but different contraction property, which is associated with the Perelman's $\mathcal{L}$-functional instead of the squared distance. Let $R_{g(\tau)}$ be the scalar curvature with respect to the metric $g(\tau)$. For a C^1-curve $\gamma : [\tau_1, \tau_2] \to M$,

$$\mathcal{L}(\gamma) := \int_{\tau_1}^{\tau_2} \sqrt{\tau} \left(|\dot{\gamma}(\tau)|_{g(\tau)}^2 + R_{g(\tau)}(\gamma(\tau)) \right) d\tau.$$

By using this functional $\mathcal{L}$, we define $L(\tau_1, x; \tau_2, y)$ by

$$L(\tau_1, x; \tau_2, y) := \inf \left\{ \mathcal{L}(\gamma) \mid \gamma : [\tau_1, \tau_2] \to M, \gamma(\tau_1) = x, \gamma(\tau_2) = y \right\}.$$

Let us define a normalized L-function as follows:

$$\Theta_t(x, y) := 2(\sqrt{\bar{\tau}_2 t} - \sqrt{\bar{\tau}_1 t}) L(\bar{\tau}_1 t, x; \bar{\tau}_2 t, y) - 2m(\sqrt{\bar{\tau}_2 t} - \sqrt{\bar{\tau}_1 t})^2,$$

where $\bar{\tau}_1 < \bar{\tau}_2$ are normalizing constants and $m := \dim M$.

Theorem 4 ([16]; see [36] also)

Suppose

$$\inf_{\substack{X\in TM\\ t\in[0,T]}} \frac{\mathrm{Ric}_{g(t)}(X,X)}{g(t)(X,X)} > -\infty.$$

Then there exists a coupling of $g(\tau)$-Brownian motions $(X_0(\tau), X_1(\tau))_\tau$ s.t.

$$(\Theta_t(X_0(\bar\tau_1 t), X_1(\bar\tau_2 t)))_{t\in[1,T/\bar\tau_2]}$$

is a supermartingale.

4 Dual approach

There is a deep connection between the condition (a) and the condition (c) beyond the framework in Theorem 3. Again, let M be a Polish space. Instead of the heat semigroup, we consider a Markov kernel $p(x,\cdot) \in \mathcal{P}(M)$ $(x \in M)$. We denote the action of this Markov kernel to functions and to probability measures by P and P^* respectively. Let us define a "modulus of gradient" $|\nabla f|$ of a function on M as the local Lipschitz constant. That is,

$$|\nabla f|(x) := \limsup_{y\to x} \frac{|f(x)-f(y)|}{d(x,y)}.$$

Theorem 5 ([K.] cf. [15])

For $p, q \in [1,\infty]$ with $\frac{1}{p}+\frac{1}{q} = 1$, and a constant $C > 0$, the following are equivalent:

(i) $W_p(P^*\mu_0, P^*\mu_1) \le CW_p(\mu_0,\mu_1)$ for any $\mu_0,\mu_1 \in \mathcal{P}(M)$.

(ii) $|\nabla Pf|(x) \le CP(|\nabla f|^q)(x)^{1/q}$ for any $x \in M$ and $f \in C^{\mathrm{Lip}}(M)$, where the right hand side must be replaced with $\|\nabla f\|_\infty$ when $q = \infty$.

By using this duality, we can obtain an estimate of Wasserstein distance like (i) for a semigroup associated with a hypoelliptic diffusion (see [15] for details). It should be remarked that no coupling methods are known to obtain such an estimate for those hypoelliptic diffusion processes.

For this duality, it is essential that the action of the Markov kernel to functions or probability measures are linear. Indeed, there naturally appears a nonlinear heat equation on a Finsler manifold [27]. In these cases, we can formulate conditions corresponding to (a)-(d) in Theorem 3. Then (b) and (d) are equivalent and it implies (c) [25, 27]. However, (a) typically does not hold [26].

5 Heat distribution as a gradient flow on $\mathcal{P}_2(M)$

An important connection between the theory of Optimal transportation and stochastic analysis is in the fact that the heat distribution $(\mu_t)_{t\ge0}$, which is obviously a curve in $\mathcal{P}(M)$, can be regarded as a gradient curve of the relative entropy Ent on $(\mathcal{P}_2(M), W_2)$. This viewpoint was proposed by Otto [12, 28, 29]. See [3, 38] for a rigorous treatment of the gradient flow in a fairly general framework. Based on this viewpoint, heuristic argument yields an easy derivation of (a) from (d) in Theorem 3 (see below). Actually, for the heat distributions on compact Alexandrov spaces, the property (a) is shown by making this heuristic argument rigorous. That is,

Theorem 6 ([31] for (i), [24] for (ii), [8] for (iii))

Let M be a compact m-dimensional Alexandrov space with curvature bounded from below by k. Then the following holds:

(i) The condition (d) in Theorem 3 holds for $K = (m-1)k$.

(ii) Any pair of gradient curves $(\mu_t^{(1)}, \mu_t^{(2)})_t$ of Ent satisfies

$$W_2(\mu_t^{(1)}, \mu_t^{(2)}) \le e^{-Kt} W_2(\mu_0^{(1)}, \mu_0^{(2)}).$$

(iii) Any gradient curve of μ_t equals $P_t^* \mu_0$, where P_t is the heat semigroup associated with the canonical Dirichlet form on $L^2(M)$ constructed in [17].

By passing through the duality argument in Theorem 5, we obtain the Lipschitz continuity of the heat kernel or eigenfunction of the Laplacian as an application of Theorem 6 (iii). Since the set of singular points where usual differential calculus cannot work can be dense on Alexandrov spaces, such a regularity is totally non-trivial.

Along the argument in [8] with several extensions of techniques, this approach is generalized in the framework of metric measure spaces in [1, 2] with a lower Ricci curvature bound in the sense of Sturm-Lott-Villani [20, 33, 34]. Under an additional assumption, they constructed a Dirichlet form based via a Cheeger type energy. Moreover, they obtained the property (a) for the associated heat semigroup by identifying it with the gradient of relative entropy. Note that such an approach are also done for Markov chains by introducing a suitable distance function [7, 21].

To explain a heuristic reason why Ent produces the heat flow, we restrict ourselves into the case $M = \mathbb{R}^m$. To formulate a gradient flow, we requires a notion of tangent space over $\mathcal{P}_2(\mathbb{R}^m)$ and Riemannian metric on it. Let $\mathcal{P}_2^{\mathrm{ac}}(M)$ be a subset of $\mathcal{P}_2(M)$ consisting of probability measures being absolutely continuous with respect to the Lebesgue measure. For $\mu_0, \mu_1 \in \mathcal{P}_2^{\mathrm{ac}}(M)$, there exists a convex function φ on $\mathbb{R}^m$ such that $(\nabla\varphi)^\# \mu_0 = \mu_1$ and

$$\int_{\mathbb{R}^m} |x - \nabla\varphi(x)|^2 \mu_0(dx) = W_2(\mu_0, \mu_1)^2.$$

Moreover, a map $\xi : \mathbb{R}^m \to \Gamma([0,1]; \mathbb{R}^m)$ defined by $\xi(x) := ((1-t)x + t\nabla\varphi(x))_{t \ge 0}$ induces a dynamical coupling of μ_0 and μ_1 in the sense of Theorem 2 by pushing forward μ_0 by ξ (see [6, 22]; see [37, 38] also). From this property it seems to be natural to define the tangent space $T_\mu \mathcal{P}(M)$ and a Riemannian metric σ on it as follows:

$$T_\mu \mathcal{P}(M) := \overline{\{\nabla\varphi \mid \varphi \in C^\infty(\mathbb{R}^m)\}}^{L^2(\mu)},$$
$$\sigma(\nabla\varphi, \nabla\psi)(\mu) := \int_{\mathbb{R}^m} \langle \nabla\varphi, \nabla\psi \rangle d\mu.$$

On this formal Riemannian structure, a natural family of curves emanating from $\mu \in \mathcal{P}(M)$ are given by the push-forward of μ by a one-parameter semigroup of smooth maps φ_t on $\mathbb{R}^m$. Then the resulted curve $\mu_t = \varphi_t^{\#}\mu$ satisfies the following continuity equation in the weak sense:

$$\partial_t \mu_t - \mathrm{div}_{\mu_t}(\nabla \varphi_t)\mu_t = 0. \tag{3}$$

From these observation, we can conclude that

$$\nabla \operatorname{Ent}(\mu) = \frac{\nabla \rho}{\rho} \quad \text{if } \mu(dx) = \rho(x)dx.$$

By combining it with (3), we can easily verify that a gradient curve of Ent solves the heat equation in the weak sense.

6 Curvature-dimension conditions

As an extension of the condition (c), F.-Y. Wang [40] introduced the following functional inequality

$$|\nabla P_t f|(x)^2 \le e^{-2Kt} P_t(|\nabla f|^2)(x) - \frac{1 - e^{-2Kt}}{KN}(\Delta P_t f(x))^2 \tag{4}$$

on a complete Riemannian manifold M. He proved that it is equivalent to the Bakry-Émery's curvature-dimension inequality

$$\frac{1}{2}\Delta |\nabla f|^2 - \langle \nabla \Delta f, \nabla f \rangle \ge K |\nabla f|^2 + \frac{1}{N}(\Delta f)^2,$$

which is known to be equivalent to $\mathrm{Ric} \ge K$ and $\dim M \le N$.

By following a duality argument in section 4, we obtain an equivalent estimate of the L^2-Wasserstein distance between heat distributions as follows:

Theorem 7 [K.]

For the heat semigroup P_t on a complete Riemannian manifold M, the inequality (4) is equivalent to the following inequality: For $s > t$ and $\mu_0, \mu_1 \in \mathcal{P}(M)$,

$$W_2(P_t^* \mu_0, P_s^* \mu_1)^2 \le \frac{e^{-2Kt} - e^{-2Ks}}{2K(s-t)} W_2(\mu_0, \mu_1)^2 + (s-t)\int_t^s \frac{NK}{e^{2Ku} - 1} du.$$

Note that we can recover the condition (a) in Theorem 3 from this inequality by taking a limit $s \downarrow t$.

References

[1] 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. arXiv:1106.2090.

[2] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. preprint. arXiv:1109.0222.

[3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Birkhäuser Verlag, Basel, second edition, 2008.

[4] M. Arnaudon, K.A. Coulibaly, and A. Thalmaier. Horizontal diffusion in C^1-path space. to appear in Séminaire de Probabilités, Lecture Notes in Mathematics (2009); arXiv:0904.2762.

[5] D. Bakry. On Sobolev and logarithmic Sobolev inequalities for Markov semigroups. In *New trends in stochasitic analysis*, pages 43–75, Charingworth, 1994, 1997. World Sci. Publ. River Edge, NJ.

[6] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991.

[7] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. Preprint. Available at: arXiv:1111.2687.

[8] N. Gigli, K. Kuwada, and S. Ohta. Heat flow on Alexandrov spaces. Preprint (2010). Available at: `http://www.math.kyoto-u.ac.jp/~sohta/`.

[9] N. Gozlan and C. Léonard. Transport inequalities: A survey. *Markov Processes Relat. Fields*, 16:635–736, 2010.

[10] M. Hairer. Coupling stochastic PDEs. In *XIVth International Congress on Mathematical Physics*, pages 281–289, Hackensack, NJ, 2005. World Sci. Publ.

[11] M. Hairer and J. C. Mattingly. Spectral gaps in Wasserstein distances and 2D stochastic Navier-Stokes equations. *Ann. Probab.*, 36(6):2050–2091, 2008.

[12] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[13] K. Kuwada. Convergence of time-inhomogeneous geodesic random walks and its application to coupling methods. to appear in Ann. Probab.

[14] K. Kuwada. Couplings of the Brownian motion via discrete apporoximation under lower Ricci curvature bounds. In *Probabilistic Approach to Geometry*, Adv. Stud. Pure Math. 57, pages 273–292, Tokyo, 2010. Math. Soc. Japan.

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

[16] K. Kuwada and R. Philipowski. Coupling of Brownian motion and Perelman's $\mathcal{L}$-functional. *J. Funct. Anal.*, 260(9):2742–2766, 2011.

[17] K. Kuwae, Y. Machigashira, and T. Shioya. Sobolev spaces, Laplacian and heat kernel on Alexandrov spaces. *Math. Z.*, 238(2):269–316, 2001.

[18] M. Ledoux. The geometry of Markov diffusion generators. *Ann. Fac. Sci. Toulouse Math. (6)*, 9(2):305–366, 2000.

[19] T. Lindvall. *Lectures on the coupling method*. John Wiley & Sons, Chichester and New York, 1992.

[20] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. Math.*, 169(3):903–991, 2009.

[21] J. Maas. Gradient flow of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.

[22] R.J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80(2):309–323, 1995.

[23] R.J. McCann and P. Topping. Ricci flow, entropy and optimal transportation. *Amer. J. Math.*, 132:711–730, 2010.

[24] S. Ohta. Gradient flows on Wasserstein spaces over compact Alexandrov spaces. *Amer. J. Math*, 131(2):475–516, 2009.

[25] S. Ohta and K.-Th. Sturm. Bochner-Weitzenböck formula and Li-Yau estimates on Finsler manifolds. Preprint (2011). Available at arXiv:1105.0983.

[26] S. Ohta and K.-Th. Sturm. Non-contraction of heat flow on Minkowski spaces. Preprint (2010). Available at arXiv:1009.2312.

[27] S. Ohta and K.-Th. Sturm. Heat flow on Finsler manifolds. *Comm. Pure Appl. Math.*, 62:1386–1433, 2009.

[28] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1–2):101–174, 2001.

[29] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

[30] P. Petersen. *Riemannian geometry*. Graduate Texts in Mathematics, 171. Springer-Verlag, New York, second edition, 2006.

[31] A. Petrunin. Alexandrov meets Lott-Villani-Sturm. Preprint (2009). Available at arXiv:1003.5948.

[32] S. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I: Theory, Vol. II: Applications*. Probability and its applications. Springer-Verlag, New York, 1998.

[33] K.-Th. Sturm. On the geometry of metric measure spaces. I. *Acta. Math.*, 196(1):65–131, 2006.

[34] K.-Th. Sturm. On the geometry of metric measure spaces. II. *Acta. Math.*, 196(1):133–177, 2006.

[35] H. Thorisson. *Coupling, stationarity, and regeneration*. Probability and its applications. Springer-Verlag, New York, 2000.

[36] P. Topping. $\mathcal{L}$-optimal transportation for Ricci flow. *J. Reine Angew. Math.*, 636:93–122, 2009.

[37] C. Villani. *Topics in optimal transportations*. Graduate studies in mathematics, 58. American mathematical society, Providence, RI, 2003.

[38] C. Villani. *Optimal transport, old and new*. Grundlehren der Mathematischen Wissenschaften, 338. Springer-Verlag, 2008.

[39] M.-K. von Renesse and K.-Th. Sturm. Transport inequalities, gradient estimates, entropy and Ricci curvature. *Comm. Pure. Appl. Math.*, 58(7):923–940, 2005.

[40] F.-Y. Wang. Equivalent semigroup properties for curvature-dimension condition. preprint. arXiv:1012.5686.

[41] F.-Y. Wang. *Functional inequalities, Markov semigroups, and spectral theory.* Mathematics Monograph Series 4. Science Press, Beijing, China, 2005.