AMERICAN MATHEMATICAL SOCIETY
www.ams.org

Ryunosuke Ozawa, Norihiko Suzuki

Stability of Talagrand's inequality under concentration topology

Proceedings of the American Mathematical Society
DOI: 10.1090/proc/13580

Accepted Manuscript

STABILITY OF TALAGRAND'S INEQUALITY UNDER CONCENTRATION TOPOLOGY

RYUNOSUKE OZAWA AND NORIHIKO SUZUKI

Abstract. In this paper, we study the compatibility between Talagrand's inequality and the concentration topology, i.e., if a sequence of mm-spaces satisfying Talagrand's inequality converges with respect to the observable distance, then the limit space satisfies Talagrand's inequality.

1. Introduction

Gromov [4, Chapter $3.\frac{1}{2}_+$] introduced the observable distance function d_{conc} on the set $\mathcal{X}$ of isomorphism classes of mm-spaces (metric measure spaces). This comes from the idea of measure concentration phenomenon which is stated as that any 1-Lipschitz function on an mm-space is close to a constant function on a Borel set with almost full measure. The observable distance function is defined by the difference between the sets of 1-Lipschitz functions on two mm-spaces. The topology generated by the observable distance function admits a convergence sequence of Riemannian manifolds of unbounded dimension. For example, the sequence $\{S^n\}_{n=1}^{\infty}$ of n-dimensional unit spheres d_{conc}-converges to one-point mm-space.

Talagrand's inequality is one of the functional approaches to the concentration phenomenon. An mm-space (X, d_X, μ_X) satisfies Talagrand's inequality $(T_p(K))$ if we have

$$W_p(\nu, \mu_X)^2 \leq \frac{2}{K} \operatorname{Ent}(\nu|\mu_X)$$

for any $\nu \in \mathcal{P}_p(X)$. Here, W_p is the L^p-Wasserstein distance function, $\operatorname{Ent}(\nu|\mu_X)$ is the relative entropy of ν with respect to μ_X, and $\mathcal{P}_p(X)$ is the set of Borel probability measures with finite p^{th} moment. The case $p = 2$ was first proved by Talagrand [8]. He proved that n-dimensional Gaussian space satisfies Talagrand's inequality $(T_2(1))$ for any $n \in \mathbb{N}$. After that, Sturm [7] and Lott-Villani [5] introduced the curvature-dimension condition $\mathrm{CD}(K, \infty)$ for mm-spaces. This is a generalized notion of Ricci curvature bound from below by $K \in \mathbb{R}$. Lott-Villani

Date: November 1, 2016.
2010 *Mathematics Subject Classification.* Primary: 53C23; Secondary: 60E15.
Key words and phrases. metric measure space, observable distance, Talagrand's inequality.

[5] proved that the curvature-dimension condition $\mathrm{CD}(K,\infty)$ implies Talagrand's inequality $(T_2(K))$.

In this paper, we study the compatibility between d_{conc}-convergence and Talagrand's inequality. Our main theorem stated as follows.

Theorem 1.1. *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of mm-spaces satisfying Talagrand's inequality $(T_p(K))$ for $K > 0$ and p with $1 \le p < \infty$. If X_n concentrates to an mm-space Y as $n \to \infty$, then Y also satisfies Talagrand's inequality $(T_p(K))$.*

Acknowledgment. During the work on this paper, the first author has been supported by JSPS KAKENHI Grant Number 24224002 and post-doctoral program at Max Planck Institute for Mathematics.

2. Preliminaries

In this section, we give the definitions and properties stated in [4, Chapter $3\frac{1}{2}{}_+$], [6], and [9, 10].

2.1. **Observable distance function.**

Definition 2.1 (mm-Space)**.** A triple $X = (X, d_X, \mu_X)$ is called an *mm-space* (*metric measure space*) if (X, d_X) is a complete separable metric space and if μ_X is a Borel probability measure on X.

Definition 2.2 (mm-Isomorphism)**.** Two mm-spaces X and Y are said to be *mm-isomorphic* to each other if there exists an isometry $f : \operatorname{supp}\mu_X \to \operatorname{supp}\mu_Y$ such that $f_*\mu_X = \mu_Y$, where $f_*\mu_X$ is the push-foward measure of μ_X by f. Such an f is called an *mm-isomorphism.*

Note that X is mm-isomorphic to $(\operatorname{supp}(\mu_X), d_X, \mu_X)$. Denote by $\mathcal{X}$ the set of mm-isomorphism classes of mm-spaces.

Let $I := [\,0,1\,]$ and X be an mm-space. A Borel measurable map $\varphi : I \to X$ is called a *parameter of* X if φ satisfies $\varphi_*\mathcal{L} = \mu_X$, where $\mathcal{L}$ is the Lebesgue measure. Any mm-space has a parameter (see [6, Proposition 4.1]). For two μ_X-measurable functions $f, g : X \to \mathbb{R}$, we define the *Ky Fan distance between f and g* by

$$d_{\mathrm{KF}}(f,g) := \inf\{\,\varepsilon > 0 \mid \mu_X(\{\,x \in X \mid |\,f(x) - g(x)\,| > \varepsilon\,\}) \le \varepsilon\,\}.$$

The distance function d_{KF} is called the *Ky Fan metric* on the set of μ_X-measurable functions on X. Note that the Ky Fan metric is a metrization of convergence in measure of μ_X-measurable functions.

Definition 2.3 (Observable distance)**.** Denote by $\mathcal{L}ip_1(X)$ the set of 1-Lipschitz continuous functions on an mm-space X. For any parameter φ of X, we set $\varphi^*\mathcal{L}ip_1(X) := \{\,f \circ \varphi \mid f \in \mathcal{L}ip_1(X)\,\}$. We define the *observable distance* $d_{\mathrm{conc}}(X,Y)$ *between two mm-spaces X and Y* by

$$d_{\mathrm{conc}}(X,Y) := \inf_{\varphi,\psi} d_{\mathrm{H}}(\varphi^*\mathcal{L}ip_1(X), \psi^*\mathcal{L}ip_1(Y)),$$

where $\varphi : I \to X$ and $\psi : I \to Y$ run over all parameters of X and Y, respectively, and where d_{H} is the Hausdorff distance function with respect to the Ky Fan metric d_{KF}. We say that a sequence of mm-spaces X_n, $n = 1, 2, \dots$, *concentrates* to an mm-space Y if X_n d_{conc}-converges to Y as $n \to \infty$.

The observable distance d_{conc} is a metric on $\mathcal{X}$ (see [4, Section $3\frac{1}{2}$.45] and [6, Theorem 5.16]). We call the topology on $\mathcal{X}$ induced by d_{conc} the *concentration topology.*

Proposition 2.4 ([3, Proposition 3.5, Proposition 3.11, Lemma 5.4], [6, Lemma 5.27, Corollary 5.35, Proposition 9.31])**.** *Let X_n and Y be mm-spaces, $n = 1, 2, \dots$. If X_n concentrates to Y as $n \to \infty$, then there exist Borel measurable maps $p_n : X_n \to Y$, positive real numbers ε_n with $\varepsilon_n \to 0$ as $n \to \infty$ and Borel subsets $\tilde{X}_n \subset X_n$ with $\mu_{X_n}(\tilde{X}_n) \geq 1 - \varepsilon_n$ such that*

1. $d_{\mathrm{H}}(\mathcal{L}ip_1(X_n), p_n^*\mathcal{L}ip_1(Y)) \leq \varepsilon_n$,
2. $(p_n)_*\mu_{X_n}$ *converges weakly to* μ_Y *as* $n \to \infty$,
3. $d_Y(p_n(x_n), p_n(x_n')) \leq d_{X_n}(x_n, x_n') + \varepsilon_n$ *for any* $x_n, x_n' \in \tilde{X}_n$,
4. $\limsup_{n\to\infty} \sup_{x_n \in X_n \setminus \tilde{X}_n} d_Y(p_n(x_n), y_0) < +\infty$ *for any* $y_0 \in Y$.

We call $\tilde{X}_n$ the non-exceptional domain of p_n for an additive error ε_n.

Remark 2.5. (1) By the inner regularity of μ_{X_n}, we may assume $\tilde{X}_n$ is a compact set.
(2) The conditions (1) and (2) of Proposition 2.4 imply the d_{conc}-convergence (see [3, Proposition 3.5], [6, Corollary 5.36]).

2.2. **Talagrand's inequality.** Let X be a complete separable metric space. A Borel probability measure π on X^2 is a *coupling* of two Borel probability measures ν_0 and ν_1 on X if π satisfies $(\mathrm{proj}_0)_*\pi = \nu_0$ and $(\mathrm{proj}_1)_*\pi = \nu_1$, where $\mathrm{proj}_i : X \times X \to X$, $i = 0, 1$, are the projections defined by $\mathrm{proj}_0(x_0, x_1) = x_0$, $\mathrm{proj}_1(x_0, x_1) = x_1$.

Definition 2.6 (Wasserstein distance)**.** Let (X, d_X) be a complete separable metric space and $p \in [\,1, \infty\,)$. For two Borel probability measures μ and ν on X, we define the *L^p-Wasserstein distance between μ and ν* by

$$W_p(\mu, \nu) := \inf_\pi \left(\int_{X \times X} d_X(x, x')^p \, d\pi(x, x') \right)^{1/p}, \tag{2.1}$$

where π runs over all couplings of μ and ν.

Denote by $\mathcal{P}_p(X)$ the set of Borel probability measures μ satisfying

$$W_p(\mu, \delta_{x_0})^p = \int_X d_X(x, x_0)^p \, d\mu(x) < \infty$$

for the Dirac measure δ_{x_0} of some point $x_0 \in X$. The L_p-Wasserstein distance W_p is a metric on $\mathcal{P}_p(X)$ (see [9, Theorem 7.3] and [10, Chapter 6]).

Remark 2.7. (1) There exists a minimizer for the infimum in (2.1). We will call it optimal coupling of ν_0 and ν_1 (see [10, Theorem 4.1]).
(2) The topology generated by the Wasserstein distance is stronger than the weak topology. If a complete separable metric space X is bounded, then the topology generated by the Wasserstein distance and the weak topology coincide to each other (see [9, Theorem 7.12] and [10, Theorem 6.9]).

Definition 2.8 (Relative entropy)**.** Let X be an mm-space and ν a Borel probability measure on X. The *relative entropy* $\mathrm{Ent}(\nu|\mu_X)$ of ν with respect to μ_X is defined as follows. If ν is absolutely continuous with respect to μ_X, then

$$\mathrm{Ent}(\nu|\mu_X) := \int_X \frac{d\nu}{d\mu_X} \log\left(\frac{d\nu}{d\mu_X}\right) d\mu_X,$$

otherwise $\mathrm{Ent}(\nu|\mu_X) := \infty$.

For an mm-space X, we denote $\mathcal{P}^{cb}(X)$ by the set of Borel probability measures ν on X with compact support which is absolutely continuous with respect to μ_X and the Radon-Nikodym derivative is essentially bounded on X. Note that $\mathcal{P}^{cb}(X)$ is a dense subset in $(\mathcal{P}_p(X), W_p)$.

Lemma 2.9 ([6, Lemma 9.20])**.** *Let X be an mm-space and $\nu \in \mathcal{P}_p(X)$ with $\mathrm{Ent}(\nu|\mu_X) < \infty$. Then, for any $\varepsilon > 0$, there exists $\tilde{\nu} \in \mathcal{P}^{cb}(X)$ such that*

$$W_p(\tilde{\nu}, \nu) < \varepsilon, \quad \textit{and} \quad |\,\mathrm{Ent}(\tilde{\nu}|\nu_X) - \mathrm{Ent}(\nu|\mu_X)\,| < \varepsilon.$$

Definition 2.10 (Talagrand's inequality)**.** Let X be an mm-space. X satisfies *Talagrand's inequality* $(T_p(K))$ for positive real numbers K and p with $1 \leq p < \infty$ if we have

$$W_p(\nu, \mu_X)^2 \leq \frac{2}{K} \mathrm{Ent}(\nu|\mu_X) \qquad (T_p(K))$$

for any $\nu \in \mathcal{P}_p(X)$.

Sturm [7] and Lott-Villani [5] introduced the curvature-dimension condition $\mathrm{CD}(K, \infty)$. This is a generalized notion of Ricci curvature bound from below by $K \in \mathbb{R}$ (see [7, Theorem 4.9] and [5, Theorem 7.3]). Lott-Villani proved the following.

Example 2.11 ([5, Theorem 6.1])**.** Let $K > 0$ and X be an mm-space satisfying $\mathrm{CD}(K, \infty)$. Then X satisfies Talagrand's inequality $(T_2(K))$. In particular, if M is a complete Riemannian manifold with $\mathrm{Ric}_M \geq K$, then M satisfies Talagrand's inequality $(T_2(K))$.

Remark 2.12. Consider the n-dimensional standard Gaussian measure γ^n on $(\mathbb{R}^n, \|\cdot\|_2)$. Since $(\mathbb{R}^n, \|\cdot\|_2, \gamma^n)$ satisfies $\mathrm{CD}(1,\infty)$ (see [7, Example 4.10]), this space satisfies Talagrand's inequality $(T_2(1))$. This coincides with Talagrand's result (see [8, Theorem 1.1]).

Combining Csiszár-Kullback-Pinsker's inequality (see e.g. [1, Theorem 8.2.7]) and [10, Theorem 6.15], we obtain the following example.

Example 2.13. Let K_n be the complete graph with n vertices, unit distance and uniform probability distribution. Then K_n satisfies Talagrand's inequality $(T_1(1/4))$.

3. Proof of Theorem 1.1

For a Borel subset B of an mm-space X with positive measure, we define a Borel probability measure μ_B by

$$\mu_B := \frac{\mu_X|_B}{\mu_X(B)}.$$

For two Borel measures μ and ν on a metric space X, we write $\mu \leq \nu$ if $\mu(B) \leq \nu(B)$ for any Borel set B of X.

Lemma 3.1. *Let X be an mm-space. If we assume that every $\nu \in \mathcal{P}^{cb}(X)$ satisfies the condition of the definition of Talagrand's inequality $(T_p(K))$, then we have the following.*

(1) $\mu_X \in \mathcal{P}_p(X)$.
(2) *X satisfies Talagrand's inequality $(T_p(K))$.*

Proof. We prove (1). Let $C \subset X$ be a compact set with $\mu_X(C) > 0$ and $x_0 \in X$. Then, we obtain

$$\begin{aligned} W_p(\mu_X, \delta_{x_0}) &\leq W_p(\mu_X, \mu_C) + W_p(\mu_C, \delta_{x_0}) \\ &\leq \sqrt{\frac{2}{K}\,\mathrm{Ent}(\mu_C|\mu_X)} + \sup_{x\in C} d_X(x, x_0) \\ &= \sqrt{\frac{2}{K}\log\frac{1}{\mu_X(C)}} + \sup_{x\in C} d_X(x, x_0) \\ &< \infty. \end{aligned}$$

(2) follows from Lemma 2.9. □

Lemma 3.2 ([3, Lemma 3.13], [6, Lemma 9.33])**.** *Let X_n and Y be mm-spaces, $n = 1, 2, \dots$. Assume that a sequence of Borel measurable maps $p_n : X_n \to Y$ and a sequence $\{\varepsilon'_n\}_{n=1}^\infty$ of positive real numbers with $\varepsilon'_n \to 0$ satisfy (1)–(3) of Proposition 2.4. For a real number $\delta > 0$, we give two Borel subsets $B_0, B_1 \subset Y$ such that*

$$\operatorname{diam} B_i \leq \delta, \quad \mu_Y(B_i) > 0, \quad \textit{and} \quad \mu_Y(\partial B_i) = 0$$

for $i = 0, 1$, and set

$$\tilde{B}_i := p_n^{-1}(B_i) \cap \tilde{X}_n \subset X_n,$$

where $\tilde{X}_n$ is a non-exceptional domain of p_n. Then, there exist a sequence $\{\varepsilon_n\}_{n=1}^\infty$ of positive real numbers with $\varepsilon_n \to 0$, Borel probability measures $\tilde{\xi}_0^n$, $\tilde{\xi}_1^n$ on X_n and couplings $\tilde{\pi}_n$ between $\tilde{\xi}_0^n$ and $\tilde{\xi}_1^n$, $n = 1, 2, \ldots$, such that, for every sufficiently large natural number n,

(1) $\tilde{\xi}_i^n \leq (1+O(\delta^{1/2}))\mu_{\tilde{B}_i}$ $(i = 0, 1)$, *where $O(\cdot)$ is a Landau symbol,*
(2) $d_{X_n}(x_0, x_1) \geq d_Y(B_0, B_1) - \varepsilon_n$ *for any* $x_i \in \tilde{B}_i$, $i = 0, 1$,
(3) $\operatorname{supp} \tilde{\pi}^n \subset \{ (x_n, x_n') \in X_n^2 \mid d_{X_n}(x_n, x_n') \leq d_Y(B_0, B_1) + \delta^{1/2} \}$,
(4) $-\varepsilon_n \leq W_p(\tilde{\xi}_0^n, \tilde{\xi}_1^n) - d_Y(B_0, B_1) \leq \delta^{1/2}$ *for any* $p \geq 1$.

Let $\theta(\cdot) : \mathbb{R} \to \mathbb{R}$ be a function such that $\theta(\varepsilon) \to 0$ as $\varepsilon \to 0$.

Proof of Theorem 1.1. By Lemma 3.1, it suffices to prove

$$W_p(\nu, \mu_Y)^2 \leq \frac{2}{K} \operatorname{Ent}(\nu|\mu_Y) \tag{3.1}$$

for any $\nu \in \mathcal{P}^{cb}(Y)$. Let $p_n : X_n \to Y$, $n = 1, 2, \ldots$, be Borel measurable maps as in Proposition 2.4. To prove the theorem, we first prove the inequality

$$W_p(\mu, \nu)^2 \leq \frac{2}{K}(\sqrt{\operatorname{Ent}(\mu|\mu_Y)} + \sqrt{\operatorname{Ent}(\nu|\mu_Y)})^2. \tag{3.2}$$

for any $\mu, \nu \in \mathcal{P}^{cb}(Y)$.

We take any $\mu, \nu \in \mathcal{P}^{cb}(Y)$ and fix them. For any natural number m, there are finite disjoint Borel subsets $B_j \subset Y$, $j = 1, 2, \ldots, J$, such that $\bigcup_{j=1}^J \overline{B_j} = \operatorname{supp} \mu \cup \operatorname{supp} \nu$, $\operatorname{diam} B_j \leq m^{-1}$, $\mu_Y(B_j) > 0$, and $\mu_Y(\partial B_j) = 0$ for any j. For each $(j, k) \in \{1, \ldots, J\}^2$, we apply Lemma 3.2 for B_j and B_k to obtain Borel probability measures $\tilde{\xi}_{jk}^{mn} \in \mathcal{P}^{cb}(X_n)$, $n = 1, 2, \ldots$, such that

$$\tilde{\xi}_{jk}^{mn} \leq (1 + \theta(m^{-1}))\mu_{\tilde{B}_j}, \tag{3.3}$$

for any sufficiently large natural number n. By the diagonal argument, we may assume that $(p_n)_*\tilde{\xi}_{jk}^{mn}$ converges weakly to a Borel probability measure $\tilde{\xi}_{jk}^m \in \mathcal{P}^{cb}(Y)$ as $n \to \infty$ for each $(j, k, m) \in \{1, \ldots, J\}^2 \times \mathbb{N}$. Let π be an optimal coupling for $W_p(\mu, \nu)$. We define

$$w_{jk} := \pi(B_j \times B_k)$$

$$\tilde{\mu}^{mn} := \sum_{j,k=1}^J w_{jk}\tilde{\xi}_{jk}^{mn}, \qquad \tilde{\nu}^{mn} := \sum_{j,k=1}^J w_{jk}\tilde{\xi}_{kj}^{mn} \quad \in \mathcal{P}^{cb}(X_n),$$

$$\tilde{\mu}^{m} := \sum_{j,k=1}^J w_{jk}\tilde{\xi}_{jk}^{m}, \qquad \tilde{\nu}^{m} := \sum_{j,k=1}^J w_{jk}\tilde{\xi}_{kj}^{m} \quad \in \mathcal{P}^{cb}(Y).$$

Then, $(p_n)_*\tilde{\mu}^{mn}$ and $(p_n)_*\tilde{\nu}^{mn}$ converge weakly to $\tilde{\mu}^m$ and $\tilde{\nu}^m$, respectively, as $n \to \infty$. $\tilde{\mu}^m$ and $\tilde{\nu}^m$ converge weakly to μ and ν, respectively, as $m \to \infty$. Moreover, $W_p((p_n)_*\tilde{\mu}^{mn}, \mu)$, $W_p((p_n)_*\tilde{\nu}^{mn}, \nu) \to 0$ as $n \to \infty$ and then $m \to \infty$.

Let $\tilde{\pi}$ be an optimal coupling for $W_p(\tilde{\mu}^{mn}, \tilde{\nu}^{mn})$. By $\operatorname{supp}\tilde{\mu}^{mn}$, $\operatorname{supp}\tilde{\nu}^{mn} \subset \tilde{X}_n$, and Proposition 2.4 (3), we have

$$\begin{aligned} W_p((p_n)_*\tilde{\mu}^{mn}, (p_n)_*\tilde{\nu}^{mn})^p &\leq \int_{X_n \times X_n} d_Y(p_n(x_n), p_n(x_n'))^p \, d\tilde{\pi}(x_n, x_n') \\ &\leq \int_{X_n \times X_n} (d_{X_n}(x_n, x_n') + \varepsilon_n)^p \, d\tilde{\pi}(x_n, x_n') \\ &\leq (W_p(\tilde{\mu}^{mn}, \tilde{\nu}^{mn}) + \varepsilon_n)^p. \end{aligned}$$

Then, we have

$$\begin{aligned} W_p(\mu, \nu) &= \lim_{m\to\infty} \lim_{n\to\infty} W_p((p_n)_*\tilde{\mu}^{mn}, (p_n)_*\tilde{\nu}^{mn}) \\ &\leq \liminf_{m\to\infty} \liminf_{n\to\infty} W_p(\tilde{\mu}^{mn}, \tilde{\nu}^{mn}). \end{aligned} \tag{3.4}$$

By (3.3), we have

$$\begin{aligned} \frac{d\tilde{\mu}^{mn}}{d\mu_{X_n}} &= \sum_{j,k=1}^{J} w_{jk} \frac{d\tilde{\xi}_{jk}^{mn}}{d\mu_{X_n}} \\ &\leq (1 + \theta(m^{-1})) \sum_{j,k=1}^{J} \frac{w_{jk}}{\mu_{X_n}(\tilde{B}_j)} \chi_{\tilde{B}_j} \\ &= (1 + \theta(m^{-1})) \sum_{j=1}^{J} \frac{\mu(B_j)}{\mu_{X_n}(\tilde{B}_j)} \chi_{\tilde{B}_j}. \end{aligned}$$

In particular, we have $\tilde{\mu}^{mn}(\tilde{B}_j) \leq (1+\theta(m^{-1}))\mu(B_j)$. The monotonicity of $f(x) = \log x$ and the previous inequality imply

$$\begin{aligned} &\operatorname{Ent}(\tilde{\mu}^{mn}|\mu_{X_n}) \\ &= \int_{X_n} \log\left(\frac{d\tilde{\mu}^{mn}}{d\mu_{X_n}}(x_n)\right) d\tilde{\mu}^{mn}(x_n) \\ &\leq \int_{X_n} \log\left((1 + \theta(m^{-1})) \sum_{j=1}^{J} \frac{\mu(B_j)}{\mu_{X_n}(\tilde{B}_j)} \chi_{\tilde{B}_j}(x_n)\right) d\tilde{\mu}^{mn}(x_n) \\ &= \sum_{j=1}^{J} \tilde{\mu}^{mn}(\tilde{B}_j) \log\left((1 + \theta(m^{-1})) \frac{\mu(B_j)}{\mu_{X_n}(\tilde{B}_j)}\right) \\ &\leq (1 + \theta(m^{-1})) \sum_{j=1}^{J} \mu(B_j) \log \frac{\mu(B_j)}{\mu_{X_n}(\tilde{B}_j)} + \theta(m^{-1}). \end{aligned}$$

Since B_j satisfies $\mu_Y(\partial B_j) = 0$, Proposition 2.4 (2) and the portmanteau theorem (see [2, Corollary 8.2.10]) imply

$$\lim_{n\to\infty} \mu_{X_n}(\tilde{B}_j) = \lim_{n\to\infty} \mu_{X_n}(p_n^{-1}(B_j) \cap \tilde{X}_n) = \mu_Y(B_j)$$

and then we obtain

$$\limsup_{n\to\infty} \operatorname{Ent}(\tilde{\mu}^{mn}|\mu_{X_n}) \le (1+\theta(m^{-1})) \sum_{j=1}^{J} \mu(B_j) \log \frac{\mu(B_j)}{\mu_Y(B_j)} + \theta(m^{-1}). \tag{3.5}$$

Define a probability measure $\overline{\mu}^m$ by

$$\overline{\mu}^m := \sum_{j=1}^{J} \frac{\mu(B_j)}{\mu_Y(B_j)} \mu_Y|_{B_j}.$$

Jensen's inequality implies

$$\begin{aligned}
&\operatorname{Ent}(\mu|\mu_Y) \\
&= \int_Y \frac{d\mu}{d\mu_Y}(y) \log \frac{d\mu}{d\mu_Y}(y)\, d\mu_Y(y) \\
&= \sum_{j=1}^{J} \int_{B_j} \frac{d\mu}{d\mu_Y}(y) \log \frac{d\mu}{d\mu_Y}(y)\, d\mu_Y(y) \\
&\ge \sum_{j=1}^{J} \left(\int_{B_j} \frac{d\mu}{d\mu_Y}(y)\, d\mu_Y(y) \right) \log \left(\frac{1}{\mu_Y(B_j)} \int_{B_j} \frac{d\mu}{d\mu_Y}(y)\, d\mu_Y(y) \right) \\
&= \sum_{j=1}^{J} \mu(B_j) \log \frac{\mu(B_j)}{\mu_Y(B_j)} \\
&= \operatorname{Ent}(\overline{\mu}^m|\mu_Y).
\end{aligned}$$

Combining this inequality and (3.5) and taking the limit as $n \to \infty$, we obtain

$$\limsup_{m\to\infty} \limsup_{n\to\infty} \operatorname{Ent}(\tilde{\mu}^{mn}|\mu_{X_n}) \le \operatorname{Ent}(\mu|\mu_Y). \tag{3.6}$$

In the same way, we also obtain

$$\limsup_{m\to\infty} \limsup_{n\to\infty} \operatorname{Ent}(\tilde{\nu}^{mn}|\mu_{X_n}) \le \operatorname{Ent}(\nu|\mu_Y). \tag{3.7}$$

The triangle inequality and Talagrand's inequality on X_n imply

$$\begin{aligned}
W_p(\tilde{\mu}^{mn}, \tilde{\nu}^{mn}) &\le W_p(\tilde{\mu}^{mn}, \mu_{X_n}) + W_p(\mu_{X_n}, \tilde{\nu}^{mn}) \\
&\le \sqrt{\frac{2}{K}} \left(\sqrt{\operatorname{Ent}(\tilde{\mu}^{mn}|\mu_{X_n})} + \sqrt{\operatorname{Ent}(\tilde{\nu}^{mn}|\mu_{X_n})}\right),
\end{aligned}$$

which together with (3.4), (3.6), and (3.7) imply (3.2).

Let us next prove that μ_Y belongs to $\mathcal{P}_p(Y)$. We take an optimal coupling $\overline{\pi}$ for $W_p(\mu_{X_n}, \tilde{\mu}^{mn})$. By Proposition 2.4 (4) and $\tilde{\mu}^{mn}(X_n \setminus \tilde{X}_n) = 0$, there exists a constant $D > 0$ such that

$$d_Y(p_n(x_n), p_n(x'_n)) \leq D$$

for $\overline{\pi}|_{(X_n \setminus \tilde{X}_n) \times X_n}$-a.e. $(x_n, x'_n) \in X_n^2$. This together with Proposition 2.4 (3) and Talagrand's inequality on X_n imply

$$\begin{aligned}
&W_p((p_n)_*\mu_{X_n}, (p_n)_*\tilde{\mu}^{mn})^p \\
&\leq \int_{\tilde{X}_n \times \tilde{X}_n} (d_{X_n}(x_n, x'_n) + \varepsilon_n)^p \, d\overline{\pi}(x_n, x'_n) \\
&\quad + \int_{(X_n \setminus \tilde{X}_n) \times \tilde{X}_n} d_Y(p_n(x_n), p_n(x'_n))^p \, d\overline{\pi}(x_n, x'_n) \\
&\leq (W_p(\mu_{X_n}, \tilde{\mu}^{mn}) + \varepsilon_n)^p + D^p \varepsilon_n \\
&\leq \left(\sqrt{\frac{2}{K} \operatorname{Ent}(\tilde{\mu}^{mn} | \mu_{X_n})} + \varepsilon_n \right)^p + D^p \varepsilon_n.
\end{aligned}$$

By the inequality just before and (3.6), we have

$$\limsup_{m\to\infty} \limsup_{n\to\infty} W_p((p_n)_*\mu_{X_n}, (p_n)_*\tilde{\mu}^{mn}) \leq \sqrt{\frac{2}{K} \operatorname{Ent}(\mu | \mu_Y)}. \tag{3.8}$$

We take any point $y_0 \in Y$ and fix this. Fatou's lemma, Proposition 2.4 (2), and $W_p((p_n)_*\tilde{\mu}^{mn}, \mu) \to 0$ as $n, m \to \infty$ together imply

$$\begin{aligned}
\int_Y d_Y(y, y_0)^p \, d\mu_Y(y) &\leq \liminf_{R\to\infty} \int_Y (d_Y(y, y_0) \wedge R)^p \, d\mu_Y(y) \\
&= \liminf_{R\to\infty} \lim_{n\to\infty} \int_Y (d_Y(y, y_0) \wedge R)^p \, d(p_n)_*\mu_{X_n}(y) \\
&\leq \liminf_{n\to\infty} \int_Y d_Y(y, y_0)^p \, d(p_n)_*\mu_{X_n}(y) \\
&= \liminf_{n\to\infty} W_p((p_n)_*\mu_{X_n}, \delta_{y_0})^p \\
&\leq \left(\sqrt{\frac{2}{K} \operatorname{Ent}(\mu | \mu_Y)} + W_p(\mu, \delta_{y_0}) \right)^p \\
&< \infty.
\end{aligned}$$

This means μ_Y belongs to $\mathcal{P}_p(Y)$. We apply Lemma 2.9 for μ_Y and then obtain the inequality (3.1). This completes the proof. $\square$

References

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer, *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses [Panoramas and Syntheses], vol. 10, Société Mathématique de France, Paris, 2000 (French, with French summary). With a preface by Dominique Bakry and Michel Ledoux.

[2] V. I. Bogachev, *Measure theory. Vol. I, II*, Springer-Verlag, Berlin, 2007.
[3] K. Funano and T. Shioya, *Concentration, Ricci curvature, and eigenvalues of Laplacian*, Geom. Funct. Anal. **23** (2013), no. 3, 888–936.
[4] M. Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston Inc., Boston, MA, 1999.
[5] J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2) **169** (2009), no. 3, 903–991.
[6] T. Shioya, *Metric measure geometry*, IRMA Lectures in Mathematics and Theoretical Physics, vol. 25, EMS Publishing House, Zürich, 2016. Gromov's theory of convergence and concentration of metrics and measures.
[7] K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta Math. **196** (2006), no. 1, 65–131.
[8] M. Talagrand, *Transportation cost for Gaussian and other product measures*, Geom. Funct. Anal. **6** (1996), no. 3, 587–600.
[9] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.
[10] ———, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009. Old and new.

Max Planck Institute for Mathematics, Vivatsgasse 7, 53111 Bonn, Germany
E-mail address: ozawa@mpim-bonn.mpg.de

Mathematical Institute, Tohoku University, Sendai 980-8578, Japan
E-mail address: sb2m20@math.tohoku.ac.jp