

AN ELEMENTARY PROOF OF THE TRIANGLE INEQUALITY FOR THE WASSERSTEIN METRIC

PHILIPPE CLEMENT AND WOLFGANG DESCH

ABSTRACT. We give an elementary proof for the triangle inequality of the p -Wasserstein metric for probability measures on separable metric spaces. Unlike known approaches, our proof does not rely on the disintegration theorem in its full generality, therefore the additional assumption that the underlying space is Radon can be omitted. We also supply a proof not depending on disintegration, that the Wasserstein metric is complete on Polish spaces.

1. INTRODUCTION

In [1] the Wasserstein metric W_p (with $p \geq 1$) is defined for probability measures on a Radon space. The proof of the triangle inequality relies on the disintegration theorem [1, Theorem 5.3.1]. The aim of this paper is to give an elementary proof of the triangle inequality for a general separable metric space.

To be more precise, we introduce the following notations and definitions (according to [1]): Let (X, d) be a separable metric space with its Borel σ -algebra $\mathcal{B}(X)$. $\mathcal{P}(X)$ denotes the set of Borel probability measures. If (Y, d) is another separable metric space, if $\mu \in \mathcal{P}(X)$ and $f : X \rightarrow Y$ is a measurable map, then $f_{\#}\mu$ is the image measure on $\mathcal{B}(Y)$, i.e. $(f_{\#}\mu)(A) = \mu(f^{-1}(A))$ for $A \in \mathcal{B}(Y)$. In particular, considering the product $X_1 \times X_2$ of two separable metric spaces, we define the canonical projections $\pi^i : X_1 \times X_2 \rightarrow X_i$. If $\gamma \in \mathcal{P}(X_1 \times X_2)$, then $\pi^i_{\#}\gamma$ (for $i = 1, 2$) are the marginal distributions of γ . Similarly, for products of three spaces, $\pi^{i,j} : X_1 \times X_2 \times X_3 \rightarrow X_i \times X_j$ denotes the canonical projection. If $\mu^i \in \mathcal{P}(X_i)$, we define

$$\Gamma(\mu^1, \mu^2) = \{\gamma \in \mathcal{P}(X_1 \times X_2) : \pi^i_{\#}\gamma = \mu^i \text{ for } i = 1, 2\}.$$

Let $1 \leq p < \infty$. By $\mathcal{P}_p(X)$ we denote the set of all $\mu \in \mathcal{P}(X)$ such that $\int_X d(x, y)^p d\mu(x) < \infty$ for some (equivalently, all) $y \in X$. For $\mu^1, \mu^2 \in \mathcal{P}_p(X)$ we define the Wasserstein metric as in [2, Section 11.8, Problem 7]:

$$(1.1) \quad W_p(\mu^1, \mu^2) = \inf_{\gamma \in \Gamma(\mu^1, \mu^2)} \left[\int_{X \times X} d(x_1, x_2)^p d\gamma(x_1, x_2) \right]^{1/p}.$$

In [2, Section 11.8.3] it is shown for $p = 1$ that W_1 is a metric. Moreover, [2, Section 11.8.3, Problem 9] consists in proving that W_p is a metric for $p \geq 1$, provided (X, d) is a Polish space. In [1, Section 7.1] it is proved that W_p is a metric

1991 *Mathematics Subject Classification.* 60B05.

Key words and phrases. Wasserstein metric, triangle inequality, probability measures on metric spaces.

for $p \geq 1$ under the more general assumption that (X, d) is a separable Radon space. In this note (Section 2) we prove that the triangle inequality can be proved by more elementary means, thus extending its validity to general separable metric spaces. The strategy of our proof follows quite closely that of [2]. However, in the case of a countable metric space, the disintegration can be done in a very straightforward way without requiring higher tools of measure theory. The generalisation from countable to separable metric spaces is done by an approximation procedure.

In [1], the disintegration theorem is also used to prove the completeness of $(\mathcal{P}_p(X), W_p)$ when (X, d) is complete. In Section 3 of this note we give a proof which does not rely on the disintegration theorem but on the completeness of $(\mathcal{P}(X), \beta)$, where β is the dual bounded Lipschitz metric (see [2, p. 394]).

Acknowledgement: The authors wish to thank L. Ambrosio, G. Savaré and C. Villani for their valuable comments. In particular a suggestion by G. Savaré helped to make the technicalities in Section 2 shorter and the proof of Proposition 2.1 more transparent.

2. THE TRIANGLE INEQUALITY

We begin with a proof of the triangle inequality in the case where (X, d) is a countable metric space.

Proposition 2.1. *Let (X, d) be a countable metric space and $1 \leq p < \infty$. Let $\mu^1, \mu^2, \mu^3 \in \mathcal{P}_p(X)$, $\gamma^{1,2} \in \Gamma(\mu^1, \mu^2)$ and $\gamma^{2,3} \in \Gamma(\mu^2, \mu^3)$. Then there exists some $\gamma^{1,3} \in \Gamma(\mu^1, \mu^3)$ such that*

$$\begin{aligned} & \left[\int_{X \times X} d(x_1, x_3)^p d\gamma^{1,3}(x_1, x_3) \right]^{1/p} \\ & \leq \left[\int_{X \times X} d(x_1, x_2)^p d\gamma^{1,2}(x_1, x_2) \right]^{1/p} + \left[\int_{X \times X} d(x_2, x_3)^p d\gamma^{2,3}(x_2, x_3) \right]^{1/p}. \end{aligned}$$

In particular, $W_p(\mu^1, \mu^3) \leq W_p(\mu^1, \mu^2) + W_p(\mu^2, \mu^3)$.

Proof. Let $X = \{v_1, v_2, \dots\}$. For shorthand we denote $\mu_k^i = \mu^i(\{v_k\})$ and $\gamma_{k,l}^{i,j} = \gamma^{i,j}(\{(v_k, v_l)\})$ whenever $i, j \in \{1, 2, 3\}$ and $k, l \in \mathbb{N}$. We define (in accordance with the notation just introduced) the measure

$$\begin{aligned} \gamma &= \sum_{k,m,n} \gamma_{k,m,n} (\delta_{v_k} \times \delta_{v_m} \times \delta_{v_n}) \\ \text{with } \gamma_{k,m,n} &= \begin{cases} \frac{\gamma_{k,m}^{1,2} \gamma_{m,n}^{2,3}}{\mu_m^2} & \text{if } \mu_m^2 \neq 0, \\ 0 & \text{if } \mu_m^2 = 0. \end{cases} \end{aligned}$$

Since the first marginal of $\gamma^{2,3}$ equals μ^2 , we obtain

$$\pi^{1,2} \# \gamma(\{(v_k, v_m)\}) = \begin{cases} \sum_n \frac{\gamma_{k,m}^{1,2} \gamma_{m,n}^{2,3}}{\mu_m^2} = \gamma_{k,m}^{1,2} & \text{if } \mu_m^2 \neq 0 \\ 0 = \gamma_{k,m}^{1,2} & \text{if } \mu_m^2 = 0. \end{cases}$$

Similarly, $\pi^{2,3}\#\gamma = \gamma^{2,3}$. Since the marginals are probability measures, we infer that γ is a probability measure. Moreover it follows for $j = 1, 2, 3$ that $\pi^j\#\gamma = \mu^j$. We define

$$\gamma^{1,3} = \pi^{1,3}\#\gamma$$

which is again a probability measure and has marginals μ^1 and μ^3 .

Now use the definition of $\gamma^{1,3}$ and Minkowski's inequality to estimate

$$\begin{aligned} & \left[\int_{X \times X} d(x_1, x_3)^p d\gamma^{1,3}(x_1, x_3) \right]^{1/p} \\ &= \left[\int_{X \times X \times X} d(x_1, x_3)^p d\gamma(x_1, x_2, x_3) \right]^{1/p} \\ &\leq \left[\int_{X \times X \times X} [d(x_1, x_2) + d(x_2, x_3)]^p d\gamma(x_1, x_2, x_3) \right]^{1/p} \\ &\leq \left[\int_{X \times X \times X} d(x_1, x_2)^p d\gamma(x_1, x_2, x_3) \right]^{1/p} + \\ &\quad + \left[\int_{X \times X \times X} d(x_2, x_3)^p d\gamma(x_1, x_2, x_3) \right]^{1/p} \\ &= \left[\int_{X \times X} d(x_1, x_2)^p d\gamma^{1,2}(x_1, x_2) \right]^{1/p} + \left[\int_{X \times X} d(x_2, x_3)^p d\gamma^{2,3}(x_2, x_3) \right]^{1/p}. \end{aligned}$$

Thus $\gamma^{1,3}$ satisfies the desired estimate. \square

The extension from the case of a countable metric space to a separable space is done by an approximation procedure. We will need several lemmas.

Lemma 2.2. *Let (X, d) be a separable metric space and \tilde{X} be a countable dense subset of X . Then, for each $\epsilon > 0$, there exists a Borel measurable map $f : X \rightarrow \tilde{X}$ such that $d(x, f(x)) < \epsilon$ for each $x \in X$. The sets $S_i = f^{-1}(v_i)$ form a partition of X .*

Proof. Let $\tilde{X} = \{v_1, v_2, \dots\}$. We define inductively

$$\begin{aligned} S_1 &= B(v_1, \epsilon), \\ S_i &= B(v_i, \epsilon) \setminus \bigcup_{j < i} S_j. \end{aligned}$$

This is a partition of X by Borel sets. Finally, we define

$$f(x) = v_i \quad \text{whenever } x \in S_i.$$

\square

Remark 2.3. *By skipping the indices where $S_i = \emptyset$ and renumbering, we can obtain a partition of X into nonempty sets S_i .*

Lemma 2.4. *Let X be a separable metric space and \tilde{X} be a countable dense subset of X . Let $\epsilon > 0$ and f be given according to Lemma 2.2. Moreover, let $\gamma \in \mathcal{P}(X \times X)$ and $\tilde{\gamma} \in \mathcal{P}(\tilde{X} \times \tilde{X})$ be such that $\tilde{\gamma} = (f \times f)\#\gamma$. Then the following assertions hold:*

- (1) The marginals satisfy (for $i = 1, 2$) $\pi^i \# \tilde{\gamma} = f \# (\pi^i \gamma)$.
(2) $\left| \left[\int_{X \times X} d(x, y)^p d\gamma(x, y) \right]^{1/p} - \left[\int_{\tilde{X} \times \tilde{X}} d(x, y)^p d\tilde{\gamma}(x, y) \right]^{1/p} \right| \leq 2\epsilon$.

Proof. Let $\tilde{U} \subset \tilde{X}$. Then

$$(\pi^1 \# \tilde{\gamma})(\tilde{U}) = \tilde{\gamma}(U \times \tilde{X}) = \gamma(f^{-1}(\tilde{U}) \times X) = (\pi^1 \# \gamma)(f^{-1}(\tilde{U})) = [f \# (\pi^1 \# \gamma)](\tilde{U}).$$

Of course, the same proof holds for π^2 . To obtain the estimate, we utilize the fact that $\tilde{\gamma} = (f \times f) \# \gamma$ and Minkowski's inequality:

$$\begin{aligned} & \left| \left[\int_{X \times X} d(x, y)^p d\gamma(x, y) \right]^{1/p} - \left[\int_{\tilde{X} \times \tilde{X}} d(x, y)^p d\tilde{\gamma}(x, y) \right]^{1/p} \right| \\ &= \left| \left[\int_{X \times X} d(x, y)^p d\gamma(x, y) \right]^{1/p} - \left[\int_{X \times X} d(f(x), f(y))^p d\gamma(x, y) \right]^{1/p} \right| \\ &\leq \left[\int_{X \times X} |d(x, y) - d(f(x), f(y))|^p d\gamma(x, y) \right]^{1/p} \\ &\leq \left[\int_{X \times X} (2\epsilon)^p d\gamma(x, y) \right]^{1/p} = 2\epsilon. \end{aligned}$$

□

Theorem 2.5. Let (X, d) be a separable metric space and $1 \leq p < \infty$. Let $\mu^1, \mu^2, \mu^3 \in \mathcal{P}_p(X)$. Then $W_p(\mu^1, \mu^3) \leq W_p(\mu^1, \mu^2) + W_p(\mu^2, \mu^3)$.

Proof. Let $\tilde{X} = \{v_1, v_2, \dots\}$ be a dense countable subset of X , let $\epsilon > 0$ and f be given according to Lemma 2.2. Let $S_i = f^{-1}(\{v_i\})$. For $i = 1, 2, 3$ we define $\tilde{\mu}^i = f \# \mu^i$.

Now, for the pairs $(i, j) \in \{(1, 2), (2, 3)\}$, let $\gamma^{i,j} \in \Gamma(\mu^i, \mu^j)$ be such that

$$\left[\int_{X \times X} d(x_i, x_j)^p d\gamma^{i,j}(x_i, x_j) \right]^{1/p} < W_p(\mu^i, \mu^j) + \epsilon.$$

On $\tilde{X} \times \tilde{X}$ we define the measures $\tilde{\gamma}^{i,j} = (f \times f) \# \gamma^{i,j}$. By Lemma 2.4(1) we infer that $\tilde{\gamma}^{i,j} \in \Gamma(\tilde{\mu}^i, \tilde{\mu}^j)$. By Lemma 2.4(2) we have the estimates

$$\left[\int_{\tilde{X} \times \tilde{X}} d(x_i, x_j)^p d\tilde{\gamma}^{i,j} \right]^{1/p} \leq \left[\int_{X \times X} d(x_i, x_j)^p d\gamma^{i,j} \right]^{1/p} + 2\epsilon \leq W_p(\mu^i, \mu^j) + 2\epsilon.$$

Proposition 2.1 implies that there exists $\tilde{\gamma}^{1,3} \in \Gamma(\tilde{\mu}^1, \tilde{\mu}^2)$ such that

$$\begin{aligned} & \left[\int_{\tilde{X} \times \tilde{X}} d(x_1, x_3)^p d\tilde{\gamma}^{1,3} \right]^{1/p} \\ &\leq \left[\int_{\tilde{X} \times \tilde{X}} d(x_1, x_2)^p d\tilde{\gamma}^{1,2} \right]^{1/p} + \left[\int_{\tilde{X} \times \tilde{X}} d(x_2, x_3)^p d\tilde{\gamma}^{2,3} \right]^{1/p} \\ &\leq W_p(\mu^1, \mu^2) + W_p(\mu^2, \mu^3) + 4\epsilon. \end{aligned}$$

For shorthand we write $\tilde{\gamma}_{k,n}^{1,3} = \tilde{\gamma}^{1,3}(\{(v_k, v_n)\})$ and $\tilde{\mu}_m^i = \tilde{\mu}^i(\{v_m\}) = \mu^i(S_m)$. On $X \times X$ we define the measure

$$\gamma^{1,3}(U) = \sum_{k,n \in \mathbb{N}, \tilde{\mu}_k^1 \neq 0, \tilde{\mu}_n^3 \neq 0} \frac{\tilde{\gamma}_{k,n}^{1,3}}{\tilde{\mu}_k^1 \tilde{\mu}_n^3} (\mu^1 \times \mu^3)[U \cap (S_k \times S_n)].$$

We will show below that $\gamma^{1,3} \in \Gamma(\mu^1, \mu^3)$ and $(f \times f)_\# \gamma^{1,3} = \tilde{\gamma}^{1,3}$. Then Lemma 2.4(2) implies that

$$\begin{aligned} & \left[\int_{X \times X} d(x_1, x_3)^p d\gamma^{1,3} \right]^{1/p} \leq \left[\int_{\tilde{X} \times \tilde{X}} d(x_1, x_3)^p d\tilde{\gamma}^{1,3} \right]^{1/p} + 2\epsilon \\ & \leq W_p(\mu^1, \mu^2) + W_p(\mu^2, \mu^3) + 6\epsilon. \end{aligned}$$

From this we conclude that $W_p(\mu^1, \mu^3) \leq W_p(\mu^1, \mu^2) + W_p(\mu^2, \mu^3) + 6\epsilon$. With $\epsilon \rightarrow 0$ we infer the desired triangle inequality.

Now we show that $\gamma^{1,3} \in \Gamma(\mu^1, \mu^3)$. We will use the definition of $\gamma^{1,3}$ and the marginals of $\tilde{\gamma}^{1,3}$. In the following computation the sum has to be understood over all indices where the denominators are nonzero. Notice that $\mu_k^1 = 0$ or $\mu_n^3 = 0$ imply $\gamma_{k,n}^{1,3} = 0$.

$$\begin{aligned} \pi^1_\# \gamma^{1,3}(V) &= \gamma^{1,3}(V \times X) = \sum_{k,n=1}^{\infty} \frac{\tilde{\gamma}_{k,n}^{1,3}}{\tilde{\mu}_k^1 \tilde{\mu}_n^3} (\mu^1 \times \mu^3)[(V \cap S_k) \times S_n] \\ &= \sum_{k,n=1}^{\infty} \frac{\tilde{\gamma}_{k,n}^{1,3}}{\tilde{\mu}_k^1} \mu^1(V \cap S_k) = \sum_{k=1}^{\infty} \frac{1}{\tilde{\mu}_k^1} \mu^1(V \cap S_k) \sum_{n=1}^{\infty} \tilde{\gamma}_{k,n}^{1,3} \\ &= \sum_{k=1}^{\infty} \mu^1(V \cap S_k) = \mu^1(V). \end{aligned}$$

To show that $\tilde{\gamma}^{1,3} = (f \times f)_\# \gamma^{1,3}$, consider $\tilde{U} \subset \tilde{X} \times \tilde{X}$. Then $(f \times f)^{-1}(\tilde{U}) = \bigcup_{k,n:(v_k, v_n) \in \tilde{U}} S_k \times S_n$. Consequently,

$$\begin{aligned} (f \times f)_\# \gamma^{1,3}(\tilde{U}) &= \gamma^{1,3} \left(\bigcup_{k,n:(v_k, v_n) \in \tilde{U}} S_k \times S_n \right) \\ &= \sum_{k,n:(v_k, v_n) \in \tilde{U}} \frac{\tilde{\gamma}_{k,n}^{1,3}}{\tilde{\mu}_k^1 \tilde{\mu}_n^3} (\mu^1 \times \mu^3)(S_k \times S_n) = \sum_{k,n:(v_k, v_n) \in \tilde{U}} \tilde{\gamma}_{k,n}^{1,3} = \tilde{\gamma}^{1,3}(\tilde{U}). \end{aligned}$$

Thus the proof is finished. \square

3. COMPLETENESS

Here we give an alternative proof of the completeness of $(\mathcal{P}_p(X), W_p)$ whenever X is complete (see [1, Propostion 7.1.5]).

Throughout this section, let $1 \leq p < \infty$. We recall the definition of the dual bounded Lipschitz metric for $\mu, \nu \in \mathcal{P}(X)$:

$$(3.1) \quad \beta(\mu, \nu) = \sup \left| \int_X f(x) d\mu(x) - \int_X f(x) d\nu(x) \right|$$

where the supremum is taken over all bounded and Lipschitz continuous $f : X \rightarrow \mathbb{R}$ such that $\|f\|_\infty + [f]_{\text{Lip}} \leq 1$. Convergence with respect to β is equivalent to narrow convergence [2, 11.3.3].

Moreover, for $\mu, \nu \in \mathcal{P}_p(X)$

$$(3.2) \quad \beta(\mu, \nu) \leq W_p(\mu, \nu).$$

Indeed, let f be bounded and Lipschitz with $[f]_{\text{Lip}} \leq 1$. For $\gamma \in \Gamma(\mu, \nu)$ we have

$$\begin{aligned} & \left| \int_X f(x) d\mu(x) - \int_X f(x) d\nu(x) \right| = \left| \int_{X \times X} (f(x) - f(y)) d\gamma(x, y) \right| \\ & \leq \int_{X \times X} |f(x) - f(y)| d\gamma(x, y) \leq \int_{X \times X} d(x, y) d\gamma(x, y) \\ & \leq \left[\int_{X \times X} d(x, y)^p d\gamma(x, y) \right]^{1/p}. \end{aligned}$$

Let μ_n be a Cauchy sequence in $(\mathcal{P}_p(X), W_p)$. From (3.2) it is also a Cauchy sequence in $(\mathcal{P}(X), \beta)$. Since (X, d) is complete, $(\mathcal{P}(X), \beta)$ is also complete by [2, Corollary 11.5.5]. Let μ denote the limit of μ_n with respect to the metric β .

Given $\epsilon > 0$, let $N \geq 1$ be such that $W_p(\mu_m, \mu_n) \leq \epsilon$ for all $m, n \geq N$. We claim that given $\bar{x} \in X$, $n \geq N$

$$(3.3) \quad \int_X d(x, \bar{x})^p d\mu(x) \leq 2^p \epsilon^p + 2^p \int_X d(y, \bar{x})^p d\mu_n(y),$$

$$(3.4) \quad W_p(\mu, \mu_n) \leq \epsilon.$$

This implies that $\mu \in \mathcal{P}_p(X)$ and $W_p(\mu_n, \mu) \rightarrow 0$.

By [1, Section 7.1], since (X, d) is complete, the infimum in (1.1) is in fact a minimum. For $m, n \geq 1$ let $\gamma_{m,n} \in \Gamma(\mu_m, \mu_n)$ be such that

$$(3.5) \quad W_p^p(\mu_m, \mu_n) = \int_{X \times X} d(x, y)^p d\gamma_{m,n}(x, y).$$

Since $\mu_m \rightarrow \mu$ narrowly, [1, Lemma 5.2.2] implies that for each n there exist $\gamma_n \in \mathcal{P}(X \times X)$ and a subsequence $\gamma_{m_k, n}$ such that

$$(3.6) \quad \gamma_{m_k, n} \rightarrow \gamma_n \text{ narrowly in } \mathcal{P}(X \times X).$$

Since the maps $\pi_i^\#$ ($i = 1, 2$) are continuous with respect to narrow convergence, we infer that

$$(3.7) \quad \gamma_n \in \Gamma(\mu, \mu_n).$$

Let $\bar{x} \in X$. Then, for $m_k, n \geq N$,

$$\begin{aligned} & \int_X d(x, \bar{x})^p d\mu_{m_k}(x) = \int_{X \times X} d(x, \bar{x})^p d\gamma_{m_k, n} \\ & \leq 2^p \int_{X \times X} d(x, y)^p d\gamma_{m_k, n} + 2^p \int_{X \times X} d(y, \bar{x})^p d\gamma_{m_k, n} \\ & \leq 2^p \epsilon^p + 2^p \int_X d(y, \bar{x})^p d\mu_n. \end{aligned}$$

In the last step we have utilized (3.5). Since μ_{m_k} converges narrowly to μ , claim (3.3) follows from the portmanteau theorem (see [1, (5.1.15)]). Similarly,

$$\begin{aligned} & \int_{X \times X} d(x, y)^p d\gamma_n(x, y) \leq \liminf_{k \rightarrow \infty} \int_{X \times X} d(x, y)^p d\gamma_{m_k, n}(x, y) \\ & = \liminf_{k \rightarrow \infty} W_p^p(\mu_{m_k}, \mu_n) \leq \epsilon^p. \end{aligned}$$

By (3.7) we have

$$W_p^p(\mu, \mu_n) \leq \int_{X \times X} d(x, y)^p d\gamma_n(x, y) \leq \epsilon^p.$$

This proves claim (3.4). Thus our proof is complete.

REFERENCES

- [1] L. Ambrosio, N. Gigli, G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Birkhäuser, 2005.
- [2] R. M. Dudley, *Real Analysis and Probability*, Cambridge Studies in Advanced Mathematics 74, Cambridge University Press 2002.

P. C.: MATHEMATICAL INSTITUTE, LEIDEN UNIVERSITY, P.O.BOX 9512, NL-2300 RA LEIDEN,
THE NETHERLANDS

E-mail address: philippecllem@gmail.com

W. D.: INSTITUT FÜR MATHEMATIK UND WISSENSCHAFTLICHES RECHNEN, KARL-FRANZENS-
UNIVERSITÄT GRAZ, HEINRICHSTRASSE 36, 8010 GRAZ, AUSTRIA

E-mail address: georg.desch@uni-graz.at