SEPARABILITY AND COMPLETENESS FOR THE WASSERSTEIN DISTANCE

FRANÇOIS BOLLEY

ABSTRACT. In this note we prove in an elementary way that the Wasserstein distances, which play a basic role in optimal transportation issues, turn some spaces of probability measures into separable complete metric spaces.

INTRODUCTION

Let (X, d) be a separable complete metric space and $\mathcal{P}$ be the set of Borel probability measures on X. Given $p > 0$, the $\mathbb{R} \cup \{+\infty\}$-valued map W_p defined on $\mathcal{P} \times \mathcal{P}$ by

$$\begin{aligned} W_p(\mu, \nu) &= \inf_\pi \Big(\iint_{X \times X} d(x,y)^p \, d\pi(x,y) \Big)^{\frac{1}{p}} && \text{if} \quad 1 \leq p \\ W_p(\mu, \nu) &= \inf_\pi \iint_{X \times X} d(x,y)^p \, d\pi(x,y) && \text{if} \quad 0 < p < 1, \end{aligned}$$

where π runs over the set of probability measures on $X \times X$ with marginals μ and ν, defines a metric on the subset $\mathcal{P}_p$ of measures μ in $\mathcal{P}$ such that $\int_X d(x_0, x)^p \, d\mu(x)$ be finite for some (and hence any) x_0 in X: it is called the Wasserstein distance of order p (see [1], [4], [5] or [6] for instance).

These distances are strongly linked to the theory of optimal transportation and have been widely used in various applications to partial differential equations, functional inequalities and probability theory. Some of them involve probability measures on infinite dimensional spaces such as the Wiener space of $\mathbb{R}^d$-valued continuous functions on the interval $[0, T]$ (as in [3] for instance) or some sets of probability measures on a phase space; this is a motivation to the general framework considered in this note, in which we shall prove:

Theorem. *If (X, d) is a separable complete metric space and p a positive number, then the metric space $(\mathcal{P}_p, W_p)$ is separable and complete.*

The completeness property has been proven in [4] by comparing the W_p distances with the weaker Prohorov distance for which the property is known, and in [1] by means of a deep result by Kolmogorov. Here we shall give a more direct and elementary argument.

Let us actually note that these kind of properties can be studied as in [2] within the following broader scope of weighted spaces of probability measures.

Let (X, τ) be a topological space and ω be a real-valued continuous function on X, bounded by below by a positive constant, and let $\mathcal{P}_\omega$ denote the set of Borel probability measures μ on X such that $\int_X \omega(x) \, d\mu(x)$ be finite. We equip $\mathcal{P}_\omega$ with the natural weak topology defined by the set $\mathcal{C}_{b\omega}$ of real-valued continuous functions f on X such that $\omega^{-1} f$ be bounded on X: this topology, which will be denoted w-$\mathcal{C}_{b\omega}$, is defined by the seminorms

$$\mu \mapsto \sup_{i=1,\dots,n} \Big| \int_X f_i(x) \, d\mu(x) \Big|$$

for any finite family $f_1, \ldots, f_n$ of functions in $\mathcal{C}_{b\omega}$.

Then one can prove that if the topological space (X, τ) is separable (resp. separable and metrizable, resp. separable, metrizable and topologically complete), then so is $(\mathcal{P}_\omega, w\text{-}\mathcal{C}_{b\omega})$. Conversely if $(\mathcal{P}_\omega, w\text{-}\mathcal{C}_{b\omega})$ is separable (resp. separable, metrizable and topologically complete), then so is (X, τ) if (X, τ) is a priori metrizable.

As in the case when $\omega = 1$, that is, without weight, where they are known, these properties can be proven either by building some explicit distances on the considered spaces of probability measures, or by abstract functional methods as in [2].

In the case when (X, d) is a separable complete metric space and $\omega = 1 + d(x_0, \cdot)^p$ for $p > 0$, then the $w\text{-}\mathcal{C}_{b\omega}$ topology on the set $\mathcal{P}_\omega = \mathcal{P}_p$ is metrized by the distance W_p: in particular $(\mathcal{P}_p, W_p)$ is separable and topologically complete.

The following two sections are devoted to a direct proof of the above theorem, which in particular ensures that $(\mathcal{P}_p, W_p)$ is complete.

1. Separability

In this section we prove that the metric space $(\mathcal{P}_p, W_p)$ is separable if (X, d) is a separable complete metric space and p is a positive number.

If $(x_n)_n$ is a sequence dense in (X, d) we actually prove that the countable set of measures of the form $\displaystyle\sum_{n=1}^N b_n \delta_{x_n}$, where N is an integer number, the b_n's are nonnegative rational numbers with unit sum and δ_x stands for the point mass at x, is dense in $(\mathcal{P}_p, W_p)$.

Let indeed μ be a given measure in $\mathcal{P}_p$ and ε be a given positive number.

1. We first approach μ by a measure $\mu_1 = \displaystyle\sum_{n=1}^{+\infty} a_n \delta_{x_n}$ where the a_n's are nonnegative real numbers with $\displaystyle\sum_{n=1}^{+\infty} a_n = 1$.

For this we note that X is covered by the balls $B(x_n, \varepsilon^{\max(1,1/p)})$ with centers x_n and radius $\varepsilon^{\max(1,1/p)}$, and is partitioned by the sets $\tilde{B}_n = B(x_n, \varepsilon^{\max(1,1/p)}) \setminus \displaystyle\bigcup_{k \leq n-1} B(x_k, \varepsilon^{\max(1,1/p)})$, so that the $a_n = \mu[\tilde{B}_n]$ have unit sum. Moreover sending each point in $\tilde{B}_n$ onto x_n for each n defines a transport map between μ and $\mu_1 = \displaystyle\sum_{n=1}^{+\infty} a_n \delta_{x_n}$ with cost

$$\sum_{n=1}^{+\infty} \int_{\tilde{B}_n} |x - x_n|^p \, d\mu(x) \leq \sum_{n=1}^{+\infty} a_n \, \varepsilon^{p \max(1,1/p)} = \varepsilon^{\max(p,1)};$$

consequently

$$W_p(\mu, \mu_1) \leq \varepsilon.$$

2. Then we approach μ_1 by a measure $\mu_2 = \displaystyle\sum_{n=1}^{N} b_n \delta_{x_n}$ where the b_n's are nonnegative *rational* numbers with $\displaystyle\sum_{n=1}^{N} b_n = 1$.

First of all μ_1 belongs to $\mathcal{P}_p$ since it is at finite W_p distance from the measure μ in $\mathcal{P}_p$. Hence

$$\sum_{n=1}^{+\infty} a_n |x_n - x_1|^p = W_p(\mu_1, \delta_{x_1})^{\max(p,1)}$$

is finite since δ_{x_1} also belongs to $\mathcal{P}_p$. In particular there exists an integer N such that

$$\sum_{n=N+1}^{+\infty} a_n |x_n - x_1|^p \leq \varepsilon^{\max(p,1)}.$$

For each $2 \leq n \leq N$ we now let b_n be a nonnegative rational number such that

$$0 \;\leq\; a_n - b_n \;\leq\; \varepsilon^{\max(p,1)} \Big(\sum_{j=1}^{N} a_j |x_j - x_1|^p \Big)^{-1} a_n,$$

and such that

$$b_1 = a_1 + \sum_{n=2}^{N} (a_n - b_n) + \sum_{n=N+1}^{+\infty} a_n$$

be rational: in particular the b_n's have unit sum. Moreover one can transport μ_1 onto $\mu_2 = \displaystyle\sum_{n=1}^{N} b_n \delta_{x_n}$ by keeping a b_n mass at x_n for each $n \leq N$ and sending the remaining $a_n - b_n$ mass from x_n onto x_1, and sending the whole a_n mass from x_n onto x_1 for each $n \geq N+1$; the associated cost is

$$\sum_{n=1}^{N} (a_n - b_n) |x_n - x_1|^p + \sum_{n=N+1}^{+\infty} a_n |x_n - x_1|^p \;\leq\; 2\,\varepsilon^{\max(p,1)},$$

so that

$$W_p(\mu_1, \mu_2) \;\leq\; 2\,\varepsilon.$$

3. To sum up we have approached μ by a measure μ_2 in $\mathcal{P}_p$, of the expected form and which is at most $3\,\varepsilon$ distant in W_p metric.

2. Completeness

In this section we prove that the metric space $(\mathcal{P}_p, W_p)$ is complete if (X, d) is a separable complete metric space and p is a positive number.

Let indeed $(\mu_n)_n$ be a Cauchy sequence in $(\mathcal{P}_p, W_p)$.

1. For $p \geq 1$ we first prove that $(\mu_n)_n$ is uniformly tight by adapting a classical proof of Ulam lemma.

Let ε be a given positive number.

We note that $(\mu_n)_n$ is Cauchy in $(\mathcal{P}_1, W_1)$ since $W_1 \leq W_p$. Hence there exists N such that $W_1(\mu_n, \mu_N) \leq \varepsilon^2$ for any $n \geq N$ so that, for any n, there exists $j \leq N$ such that

$$W_1(\mu_n, \mu_j) \leq \varepsilon^2. \tag{1}$$

The finite family $(\mu_j)_{j\leq N}$ is uniformly tight by Ulam lemma, so there exist a compact set K such that

$$\mu_j(K) \geq 1-\varepsilon$$

for any $j \leq N$, whence q points $x_1, \ldots, x_q$ in X such that

$$\mu_j\,(U) \geq 1-\varepsilon \tag{2}$$

for any $j \leq N$, where $U = \bigcup_{k=1}^{q} B(x_k, \varepsilon)$.

Then let ϕ be the $\frac{1}{\varepsilon}$-Lipschitz function defined on X by $\phi(x) = \Big(1 - \dfrac{d(x,U)}{\varepsilon}\Big)^+$. Given j and n, if π is any joint measure on $X \times X$ with marginals μ_j and μ_n, then

$$\int_X \phi(x)\,d\mu_j(x) - \int_X \phi(y)\,d\mu_n(y) = \iint_{X\times X} (\phi(x) - \phi(y))\,d\pi(x,y) \leq \frac{1}{\varepsilon}\iint_{X\times X} d(x,y)\,d\pi(x,y);$$

hence

$$\int_X \phi(x)\,d\mu_j(x) \;-\; \int_X \phi(y)\,d\mu_n(y) \;\leq\; \frac{1}{\varepsilon}\,W_1(\mu_j,\mu_n).$$

On the other hand $\mathbf{1}_U \;\leq\; \phi \;\leq\; \mathbf{1}_{U^\varepsilon}$ where $U^\varepsilon = \{x; d(x,U) < \varepsilon\}$, so

$$\int_X \phi(x)\,d\mu_j(x) \geq \mu_j(U) \qquad \text{and} \qquad \int_X \phi(y)\,d\mu_n(y) \leq \mu_n(U^\varepsilon).$$

Consequently

$$\mu_n(U^\varepsilon) \;\geq\; \mu_j(U) - \frac{1}{\varepsilon}\,W_1(\mu_j,\mu_n). \tag{3}$$

Thus, by (1), (2) and (3), for any $\varepsilon > 0$ we have found q points $x_1, \ldots, x_q$ such that

$$\mu_n\big(X \setminus \bigcup_{k=1}^{q} B(x_k, 2\,\varepsilon)\big) \;\leq\; 2\,\varepsilon$$

for any n since $U^\varepsilon \subset \bigcup_{k=1}^{q} B(x_k, 2\,\varepsilon)$.

Therefore, replacing ε by $\varepsilon\,2^{-m-1}$ where m is any integer, there exist $q(m)$ points $x_1^m, \ldots, x_{q(m)}^m$ in X such that

$$\mu_n\big(X \setminus \bigcup_{k=1}^{q(m)} B\left(x_k^m, \varepsilon\,2^{-m}\right)\big) \;\leq\; \varepsilon\,2^{-m}$$

for any n. In particular the set

$$S \;=\; \bigcap_{m=1}^{+\infty}\bigcup_{k=1}^{q(m)} B\left(x_k^m, \varepsilon\,2^{-m}\right)$$

is such that

$$\mu_n(X\setminus S) \leq \sum_{m=1}^{+\infty} \mu_n\big(X \setminus \bigcup_{k=1}^{q(m)} B\left(x_k^m, \varepsilon\,2^{-m}\right)\big) \leq \sum_{m=1}^{+\infty} \varepsilon\,2^{-m} \;=\; \varepsilon$$

for any n.

On the other hand, for any ρ, and choosing m such that $\varepsilon\, 2^{-m} \leq \rho$, the set S can be covered by the $q(m)$ balls $B(x_k^m, \varepsilon\, 2^{-m})$ with radius $\varepsilon\, 2^{-m} \leq \rho$: in other words it is totally bounded, so that its closure $\overline{S}$ is compact since X is complete.

To sum up, the set $\overline{S}$ is compact and satisfies $\mu_n(X \setminus \overline{S}) \leq \varepsilon$ for any n: this means that the sequence $(\mu_n)_n$ is indeed uniformly tight.

2. We deduce from step 1 that $(\mu_n)_n$ converges in $(\mathcal{P}_p, W_p)$ in the case when $p \geq 1$.

Indeed $(\mu_n)_n$ is uniformly tight by step 1, so by Prohorov theorem there exists a subsequence $(\mu_{n'})_{n'}$ of $(\mu_n)_n$ converging to a probability measure μ on X for the narrow weak topology.

The distance $W_p(\mu, \mu_{n'})$ actually tends to 0 as n' goes to infinity. Let indeed $\pi_{n'm'}$ be a probability measure on $X \times X$ with marginals $\mu_{n'}$ and $\mu_{m'}$, optimal in the sense that

$$\iint_{X\times X} d(x,y)^p \, d\pi_{n'm'}(x,y) = W_p(\mu_{n'}, \mu_{m'})^p.$$

The sequence $(\mu_{n'})_{n'}$ is uniformly tight, hence so is $(\pi_{n'm'})_{n'}$ for given m'. Thus by Prohorov theorem again there exists a subsequence $(\pi_{n''m'})_{n''}$ of $(\pi_{n'm'})_{n'}$ converging to a probability measure $\pi_{m'}$ on $X \times X$ for the narrow weak topology. Then by semicontinuity

$$\iint_{X\times X} d(x,y)^p \, d\pi_{m'}(x,y) \leq \liminf_{n''\to+\infty} \iint_{X\times X} d(x,y)^p \, d\pi_{n''m'}(x,y) = \liminf_{n''\to+\infty} W_p(\mu_{n''}, \mu_{m'})^p. \tag{4}$$

But on one hand $\pi_{n''m'}$ has marginals $\mu_{n''}$ and $\mu_{m'}$, so at the limit (in n'') $\pi_{m'}$ has marginals μ and $\mu_{m'}$; hence

$$W_p(\mu, \mu_{m'})^p \leq \iint_{X\times X} d(x,y)^p \, d\pi_{m'}(x,y) \tag{5}$$

for any m'.

On the other hand the sequence $(\mu_{n'})_{n'}$ is Cauchy for the distance W_p, so for any $\varepsilon > 0$ and n'', m' large enough

$$W_p(\mu_{n''}, \mu_{m'}) \leq \varepsilon. \tag{6}$$

It finally follows from (4), (5) and (6) that

$$W_p(\mu, \mu_{m'}) \leq \varepsilon$$

for m' large enough, which means that μ belongs to $\mathcal{P}_p$ and that $W_p(\mu, \mu_{n'})$ indeed tends to 0 as n' goes to infinity.

Finally $W_p(\mu_n, \mu)$ tends to 0 as n goes to infinity since the whole sequence $(\mu_n)_n$ is Cauchy in $(\mathcal{P}_p, W_p)$.

3. We deduce from step 2 that $(\mu_n)_n$ converges in $(\mathcal{P}_p, W_p)$ in the case when $0 < p < 1$.

Indeed d^p is a distance on X which defines the same topology as d and (X, d^p) is complete if so is (X, d). Moreover $\mathcal{P}_p(X, d) = \mathcal{P}_1(X, d^p)$ and $W_p(X, d) = W_1(X, d^p)$ in obvious notation.

Thus, given an exponent $p \in]0, 1[$ and a metric d on X, the results associated with the exponent p and the metric d stem from the results proven in step 2 for the exponent 1 and the metric d^p.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the spaces of probability measures.* Birkhäuser, Basel, 2005.

[2] F. Bolley. Applications du transport optimal à des problèmes de limites de champ moyen. Thèse de doctorat, Ecole Normale Supérieure de Lyon. Available at http://www.lsp.ups-tlse.fr/Fp/Bolley, 2005.

[3] F. Bolley. Quantitative concentration inequalities on sample path space for mean field interaction. Preprint available at http://www.lsp.ups-tlse.fr/Fp/Bolley, 2005.

[4] S. T. Rachev. *Probability metrics and the stability of stochastic models.* John Wiley and Sons, Chichester, 1991.

[5] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol I and II.* Springer, New York, 1998.

[6] C. Villani. *Topics in optimal transportation*, volume 58 of *Grad. Stud. Math.* AMS, Providence, 2003.

Ecole Normale Supérieure de Lyon, Umpa (UMR 5669), 46 allée d'Italie, F-69364 Lyon cedex 7

Current address: Institut de Mathématiques, LSP (UMR C5583), Université Paul Sabatier, Route de Narbonne, F-31062 Toulouse cedex 9

E-mail address: bolley@cict.fr