A Hitchhiker’s guide to Wasserstein distances

Giuliano Basso
June 5, 2015

The main references of this section are [Edw11] and [Kel85]. For measure theoretic notions we refer to [Bog07]. In the following we introduce some notation. Let \((X, d)\) denote a metric space and let \(\mathcal{B}(X)\) denote the Borel \(\sigma\)-algebra of \((X, d)\). Suppose that \(\mu : \mathcal{B}(X) \to \mathbb{R}\) is a signed finite measure on the measurable space \((X, \mathcal{B}(X))\), that is, \(\mu(\emptyset) = 0\) and \(\mu\) is countably additive.

Recall that the map \(|\mu| : \mathcal{B}(X) \to \mathbb{R}\) given by the assignment

\[
B \mapsto \sup \left\{ \sum_{k=1}^{n} |\mu(B_k)| : n \geq 1, B_k \in \mathcal{B}(X) \text{ and } B = \bigcup_{k \in \{1, \ldots, n\}} B_k \right\}
\]

is a non-negative measure on \((X, \mathcal{B}(X))\). The measure \(|\mu|\) is called the total variation of \(\mu\). We say that \(\mu\) is a Radon measure if \(\mu\) is inner regular, that is, we have

\[
|\mu|(B) = \sup \{ |\mu|(K) : K \subset B, \text{ K compact} \}
\]

for all Borel measurable sets \(B\) contained in \(\mathcal{B}(X)\). Let \(M(X)\) denote the set that consists of all signed finite measures on the measurable space \((X, \mathcal{B}(X))\) that are Radon measures. It can be shown that \(M(X)\) is a real vector space with the sum and the scalar multiplication defined in the obvious way. Let \(M^+(X)\) be the subset of \(M(X)\) consisting of all non-negative Radon measures on \((X, \mathcal{B}(X))\). Moreover, we define the set \(P(X) := \{ \mu \in M^+(X) : \mu(X) = 1 \}\).

Suppose now that \(X\) is non-empty and fix a point \(x_0 \in X\). For every real number \(p \in [1, \infty)\) we define the set \(M^+_p(X)\) as the collection of all measures \(\mu\) in \(M^+(X)\) such that

\[
\int_X d(x, x_0)^p \mu(dx) < +\infty.
\]

Note that the definition of \(M^+_p(X)\) is independent of the point \(x_0\). If \(p \in [1, \infty)\) is a real number and \(\mu\) is an element of \(M_p(X)\), then we say that \(\mu\) has a finite \(p\)-th moment. Finally, we define for every real number \(p \in [1, \infty)\) the set

\[
P_p(X) := \{ \mu \in M^+_p(X) : \mu(X) = 1 \}\,
that is, \( P_p(X) \) is the set that consists of all Radon measures on the measurable space \( (X; \mathcal{B}(X)) \) with finite \( p \)-th moment that are probability measures. Given two Radon measures \( \mu, \nu \in M^+(X) \) with \( \mu(X) = \nu(X) \) we call
\[
\Lambda(\mu, \nu) := \{ \pi \in M^+(X \times X) : \text{for all } A \in \mathcal{B} : \pi(A \times X) = \mu(A), \pi(X \times A) = \nu(A) \} 
\]
the set of all Radon couplings of \( \mu \) and \( \nu \).

**Definition 1.1** Let \( (X, d) \) denote a metric space and let \( p \in [1, \infty) \) be a real number. The function \( W_p : P_p(X) \times P_p(X) \to \mathbb{R} \) given through the assignment
\[
(\mu, \nu) \mapsto \left( \inf_{\pi \in \Lambda(\mu, \nu)} \int_{X \times X} d(x, y)^p \pi(dx, dy) \right)^{\frac{1}{p}}
\]
is called the \( p \)-th Wasserstein distance.

The \( p \)-th Wasserstein distance is well-defined for every real number \( p \in [1, \infty) \). To see this, use the triangle inequality and Minkowski’s inequality. If the metric space \( (X, d) \) is separable and a **Radon space**, that is, every Borel probability measure is a Radon measure, then it can be shown that the \( p \)-th Wasserstein distance defines a metric on \( P_p(X) \), cf. [AGS06, p. 151]. Examples of Radon spaces are separable complete metric spaces, cf. Proposition 18 in [Roy88, p. 411]. In the following we are interested in the case \( p = 1 \). We show for every metric space \( (X, d) \) that the first Wasserstein distance \( W_1 \) defines a metric on \( P_1(X) \), see Corollary 1.4. In order to establish Corollary 1.4 we first set up some notions and afterwards we prove the Kantorovich-Rubinstein duality Theorem, see Theorem 1.3. Let \( (X, d) \) denote a non-empty metric space and let \( f : X \to \mathbb{R} \) be a function. We define the quantity
\[
\text{Lip}(f) := \sup \left\{ \frac{|f(x) - f(y)|}{d(x, y)} : x, y \in X, x \neq y \right\} \in [0, +\infty].
\]
Moreover, we denote by \( \text{Lip}_1(X) \) the set that contains all functions \( f : X \to \mathbb{R} \) with \( \text{Lip}(f) \leq 1 \). If \( \mu \) is a measure contained in \( P_1(X) \) and \( f \) denotes a function contained in \( \text{Lip}_1(X) \), then the function \( f \) is Lebesgue integrable with respect to \( \mu \). Indeed, we compute
\[
|f(x)| \leq |f(x_0)| + d(x, x_0).
\]
For further references we put the prior statement in a Lemma.

**Lemma 1.2** Let \( (X, d) \) denote a non-empty metric space and let \( f \) be a function contained in \( \text{Lip}_1(X) \). Then the function \( f \) is Lebesgue integrable with respect to every measure contained in \( P_1(X) \).

The subsequent general version of the Kantorovich-Rubinstein duality Theorem is due to Hans Kellerer, cf. Theorem 1 in [Kel85].
Theorem 1.3 (Kantorovich-Rubinstein duality Theorem) Let $(X, d)$ be a metric space and let $\mu, \nu$ denote two Radon probability measures contained in $P_1(X)$. Then

$$W^1(\mu, \nu) = \sup \left\{ \int_X f \, d\mu - \int_X f \, d\nu : f \in \text{Lip}_1(X) \right\}. \tag{1.1}$$

Proof The following is a reformulation of the proof of [Edw11, Theorem 4.1]. First, we fix some notation. Let $B^\infty(X)$ denote the set that consists of all bounded Borel measurable functions $f : X \to \mathbb{R}$. For every two elements $f, g$ in $\mathbb{R}^X$ we define the function $f \oplus g : X \times X \to \mathbb{R}$ by the assignment $(x, y) \mapsto f(x) + g(y)$. Let $\mu$ and $\nu$ be two Radon probability measures contained in $P_1(X)$. Corollary 3.2 in [Edw11] tells us that

$$W^1(\mu, \nu) = \sup \left\{ \int_X f \, d\mu + \int_X g \, d\nu : f, g \in B^\infty(X), f \oplus g \leq d \right\}. \tag{1.2}$$

Corollary 3.2 in [Edw11] is an immediate consequence of Theorem 4.1 in [Edw10] whose proof is a sophisticated application of the Hahn-Banach Theorem, for the sake of space we decided to omit the proofs of these results in this text. Having Equation (1.2) on hand the derivation of Equation (1.1) is merely a clever computation. Let $\varepsilon > 0$ be a positive real number. Equation (1.2) tells us that there exists two bounded Borel measurable functions $f, g$ in $B^\infty(X)$ with $f \oplus g \leq d$ such that

$$W^1(\mu, \nu) - \varepsilon \leq \int_X f \, d\mu + \int_X g \, d\nu.$$

Now we construct a 1-Lipschitz function $k$ out of $f$ and $g$. Let the function $k : X \to \mathbb{R}$ be given by the assignment

$$x \mapsto \inf_{y \in X} (d(x, y) - g(y)) \cdot$$

As $g$ is bounded, the function $k$ is well-defined. Moreover, for every two points $x, x'$ in $X$ we compute

$$|k(x) - k(x')| = \left| \inf_{y \in X} (d(x, y) - g(y)) - \inf_{y \in X} (d(x', y) - g(y)) \right|$$

$$\leq \sup_{y \in X} |d(x, y) - d(x', y)| \leq d(x, x');$$

hence, the function $k$ is contained in the set $\text{Lip}_1(X)$. For each point $x$ in $X$ we have

$$f(x) \leq k(x) \leq d(x, x) - g(x) = -g(x);$$

3
thus, we obtain that \( f \leq k \) and \( g \leq -k \). Let \( \pi \in \Lambda(\mu, \nu) \) be a Radon coupling of \( \mu \) and \( \nu \). We compute

\[
\begin{align*}
W_1(\mu, \nu) - \varepsilon & \leq \int_X f \, d\mu + \int_X g \, d\nu \\
& \leq \sup \left\{ \int_X f \, d\mu - \int_X f \, d\nu : f \in \text{Lip}_1(X) \right\} \\
& = \sup \left\{ \int_{X \times X} (f \oplus -f) \, d\pi : f \in \text{Lip}_1(X) \right\} \\
& \leq \int_{X \times X} d(x, y) \, \pi(dx, dy);
\end{align*}
\]

thereby, the Theorem follows.

As a Corollary of the Kantorovich-Rubinstein duality Theorem, we obtain that the first Wasserstein distance defines a metric on \( P_1(X) \).

**Corollary 1.4** Let \( (X, d) \) denote a metric space. Then the first Wasserstein distance \( W_1 \) defines a metric on \( P_1(X) \).

**Proof** Let \( \mu \) and \( \nu \) be two measures contained in \( P_1(X) \). Suppose that \( W_1(\mu, \nu) = 0 \). In the following we show that \( \mu = \nu \). Let \( F \) be a closed subset of \( X \). For each integer \( k \geq 1 \) we define the function \( f_k : X \to \mathbb{R} \) through the assignment \( x \mapsto 1 \wedge (kd(x, F)) \). Note for all integers \( k \geq 1 \) that the function \( f_k \) is contained in the set \( \text{Lip}_1(X) \). As \( W_1(\mu, \nu) = 0 \), the Kantorovich-Rubinstein duality theorem implies that

\[
\frac{1}{k} \int_X f_k \, d\mu = \frac{1}{k} \int_X f_k \, d\nu
\]

for all integers \( k \geq 1 \). The sequence \( \{f_k\}_{k \geq 1} \) is increasing and converges pointwise to the indicator function of the open set \( F^c \); hence, on account of the monotone convergence Theorem and Equation (1.3) we obtain \( \mu(F^c) = \nu(F^c) \). Since the open subsets of \( X \) generate the Borel sigma algebra \( B(X) \), a well-known Corollary of Dynkin’s \( \pi\lambda \) Theorem tells us that \( \mu = \nu \), cf. Theorem A.1.5 in [Dur10]. We are left to show that the first Wasserstein distance \( W_1 \) is symmetric and satisfies the triangle inequality; but these properties are immediate consequences of the Kantorovich-Rubinstein duality theorem. The Corollary follows.

Let \( f : (X, B(X)) \to (Y, B(Y)) \) denote a Borel measurable map between two metric spaces \( (X, d_X) \) and \( (Y, d_Y) \). We define the map

\[
f_* : P_1(X) \to P_1(Y)
\]

\[
\mu \mapsto \begin{cases} f_*\mu : B(Y) \to [0, 1] \\
B \mapsto \mu(f^{-1}(B)) \end{cases}
\]

4
It is well-known that the map $f_\#$ is well-defined and for every $\mu \in \mathcal{P}_1(X)$ the measure $f_\# \mu$ is called the *pushforward of $\mu$ by $f$ or the image measure of $\mu$ by $f$.*

**Lemma 1.5** Let $(X,d)$ and $(Y,d_Y)$ denote two metric spaces. If $\varphi : X \to Y$ is an isometric embedding, then the map $\varphi_\# : \mathcal{P}_1(X) \to \mathcal{P}_1(Y)$ is an isometric embedding.

**Proof** Let $\mu$ and $\nu$ be two measures contained in $\mathcal{P}_1(X)$. We have

$$W^1(\varphi_#\mu, \varphi_#\nu) = \sup \left\{ \int_Y f \, d\varphi_#\mu - \int_Y f \, d\varphi_#\nu : f \in \text{Lip}_1(Y) \right\}$$

$$= \sup \left\{ \int_X f \circ \varphi \, d\mu - \int_X f \circ \varphi \, d\nu : f \in \text{Lip}_1(Y) \right\};$$

consequently, if the map

$$\Phi : \text{Lip}_1(Y) \to \text{Lip}_1(X)$$

$$f \mapsto f \circ \varphi$$

is surjective, then it follows that

$$\sup \left\{ \int_X f \circ \varphi \, d\mu - \int_X f \circ \varphi \, d\nu : f \in \text{Lip}_1(Y) \right\}$$

$$= \sup \left\{ \int_X f \, d\mu - \int_X f \, d\nu : f \in \text{Lip}_1(X) \right\}$$

and we thus have $W^1(\varphi_#\mu, \varphi_#\nu) = W^1(\mu, \nu)$. Therefore, we are left to show that the map $\Phi$ is surjective. Let $f : X \to \mathbb{R}$ be a 1-Lipschitz function contained in $\text{Lip}_1(X)$. We define the function

$$\bar{f} : Y \to Y$$

$$y \mapsto \inf_{x \in X} (f(x) + d_Y(y, \varphi(x))).$$

We have,

$$f(x) = \inf_{x' \in X} f(x') + |f(x) - f(x')| \leq \inf_{x' \in X} f(x') + d_X(x, x')$$

$$= \inf_{x' \in X} f(x') + d_Y(\varphi(x), \varphi(x')) \leq f(x)$$

for all points $x$ in $X$. As a result, it follows that $f = \bar{f} \circ \varphi$. Moreover, we compute

$$|\bar{f}(y) - \bar{f}(y')| = \left| \inf_{x \in X} (f(x) + d(y, \varphi(x))) - \inf_{x \in X} (f(x) + d_Y(y', \varphi(x))) \right|$$

$$= \sup_{x \in X} |d(\varphi(x), y) - d(\varphi(x), y')| \leq d(y, y');$$

thus, the function $\bar{f}$ is contained in $\text{Lip}_1(Y)$. Hence we have asserted that the map $\Phi$ is a surjection. The Lemma follows. ■
In the following we introduce the concept of the support of a Radon probability measure and we show in Lemma 1.6 that the support of a Radon probability measure is separable. Let \((X, d)\) denote a non-empty metric space and let \(\mu\) be a Radon measure contained in \(P(X)\). The set \(\text{spt}(\mu)\) consisting of all points \(x\) in \(X\) such that \(\mu(U) > 0\) for all open neighborhoods of \(x\) is called the support of \(\mu\). The following Lemma shows that every Radon probability measure is concentrated on its support.

**Lemma 1.6** Let \((X, d)\) denote a metric space and let \(\mu\) be a measure contained in \(P(X)\). Then the support of \(\mu\) is closed and we have \(\mu(B) = \mu(B \cap \text{spt}(\mu))\) for all Borel measurable subsets \(B\) of \(X\). In particular, we have \(\mu(\text{spt}(\mu)) = 1\).

**Proof** For each point \(x\) in \(\text{spt}(\mu)^c\) there exists an open neighborhood \(U_x\) of \(x\) such that \(\mu(U_x) = 0\). We claim for all points \(x\) in \(\text{spt}(\mu)^c\) that \(U_x\) is a subset of \(\text{spt}(\mu)^c\). Let \(x\) be a point in \(\text{spt}(\mu)^c\). As \(U_x\) is open, every point \(x'\) in \(U_x\) has an open neighborhood \(U_{x'}\) contained in \(U_x\). Since we have \(0 \leq \mu(U_{x'}) \leq \mu(U_x) = 0\) for each point \(x'\) in \(U_x\), it follows that \(U_x\) is contained in \(\text{spt}(\mu)^c\). Hence the set \(\text{spt}(\mu)^c\) is open. Let now \(B\) be a Borel measurable subset of \(X\) and let \(K\) be a compact subset contained in \(B \cap \text{spt}(\mu)^c\). Note that \(\{U_x\}_{x \in K}\) is an open cover of \(K\); hence, there exists an integer \(n \geq 1\) and a finite subcover \(\{U_{x_k}\}_{1 \leq k \leq n}\) such that \(K = U_{x_1} \cup \cdots \cup U_{x_n}\). We have

\[
0 \leq \mu(K) \leq \sum_{k=1}^{n} \mu(U_{x_k}) = 0.
\]

Hence we may conclude that \(\mu(B \cap \text{spt}(\mu)^c) = 0\), since \(\mu\) is a Radon measure. The Lemma follows.

Recall that a subset \(A\) of a metric space is called separable if it contains a countable subset \(B \subset A\) such that \(A \subset \overline{B}\).

**Lemma 1.7** Let \((X, d)\) denote a metric space and let \(\mu\) be a measure contained in \(P(X)\). Then the support of \(\mu\) contains a countable subset \(A \subset \text{spt}(\mu)\) such that \(\text{spt}(\mu) \subset \overline{A}\), that is, the subspace \(\text{spt}(\mu) \subset X\) is separable.

**Proof** If the cardinality of the set \(\text{spt}(\mu)\) is less than or equal to one, then the statement is valid. Hence we may assume that \(\text{spt}(\mu)\) contains at least two points; thus, there exists an integer \(N \geq 1\) and two points \(x\) and \(y\) in \(\text{spt}(\mu)\) such that \(d(x, y) \geq \frac{1}{N}\). We may assume that \(N = 1\). We define for each integer \(k \geq 1\) the set

\[
\mathcal{C}_k := \{A \subset \text{spt}(\mu) : \text{for all } x, y \in A \text{ we have } d(x, y) \geq \frac{1}{k}\}
\]

of subsets of the set \(\text{spt}(\mu)\). For each integer \(k \geq 1\) we have that the set \(\mathcal{C}_k\) is non-empty and that \((\mathcal{C}_k, \subset)\) is a partially ordered set. Let \(k \geq 1\) be an integer...
and let $T \subset \mathcal{C}_k$ be a totally ordered subset. Note that the set
\[ U := \bigcup_{A \in T} A \]
is contained in $\mathcal{C}_k$. Moreover, the set $U$ is an upper bound of $T$. Now Zorn’s Lemma tells us that $\mathcal{C}_k$ contains a maximal element. Choose a maximal element in $\mathcal{C}_k$ and denote it by $A_k$. Now we distinguish two cases:

First, we suppose for each integer $k \geq 1$ that the set $A_k$ is countable.

Note that for all points $x$ in $\text{spt}(\mu)$ and all integers $k \geq 1$ we have that $d(x, A_k) < \frac{1}{2^k}$, since $A_k$ is maximal element of $\mathcal{C}_k$. We define the set
\[ A := \bigcup_{k=1}^{\infty} A_k. \]
The set $A$ is countable and we have that $A \subset \text{spt}(\mu) \subset \overline{A}$; hence, the support of $\mu$ is separable. Thereby, the Lemma follows in this first case.

Second, we suppose now that there exists an integer $k \geq 1$ such that the set $A_k$ is uncountable. In the following we show that this is impossible. For every point $x$ contained in $A_k$ we define the set $B_x := \{ y \in X : d(x, y) < \frac{1}{2^k} \}$. Note that if $x$ and $y$ are two distinct points of $A_k$, then we have that the sets $B_x$ and $B_y$ are disjoint. On account of the inclusion $A_k \subset \text{spt}(\mu)$ we have that $\mu(B_x) > 0$ for all points $x$ in $A_k$, as $A_k$ is a subset of $\text{spt}(\mu)$. For each integer $l \geq 1$ we define the set $B_l := \{ x \in A_k : \mu(B_x) > \frac{1}{l} \}$. It follows for each integer $l \geq 1$ that the set $B_l$ is finite, since $\mu$ is a probability measure and the sets $B_x$ and $B_y$ are disjoint if $x$ and $y$ are two distinct points in $A_k$. Furthermore, we have $A_k = \bigcup_{l=1}^{\infty} B_l$; as a result, the set $A_k$ is countable, as it is the countable union of countable sets. But this is impossible, since we assume that $A_k$ is uncountable. Thus, there is no integer $k \geq 1$ such that the set $A_k$ is uncountable. Hence we have shown that the second case does not occur, and thereby the Lemma follows.

\[ \square \]

**Theorem 1.8** Let $(X, d)$ denote a complete non-empty metric space. Then the metric space $(\mathcal{P}_1(X), W^1)$ is complete. Furthermore, if the metric space $(X, d)$ is bounded, then the metric space $(\mathcal{P}_1(X), W^1)$ is bounded.

**Proof** Let $\{\mu_k\}_{k \geq 1}$ be a Cauchy sequence in $(\mathcal{P}_1(X), W^1)$. We define the set
\[ \mathcal{X} := \bigcup_{k \geq 1} \text{spt}(\mu_k). \]

Lemma 1.7 tells us that each subspace $\text{spt}(\mu_k)$ is separable, hence the subspace $\mathcal{X}$ is separable, as it is the closure of a countable union of separable subspaces of $X$. Let $i : \mathcal{X} \hookrightarrow X$ denote the inclusion from $\mathcal{X}$ to $X$. We claim that each measure $\mu_k$ is contained in $i_*(\mathcal{P}_1(\mathcal{X}))$. Note that $\mathcal{B}(\mathcal{X}) = \{ B \cap \mathcal{X} :$
Thus, we have for each integer \( k \geq 1 \) that the map \( \nu_k : \mathcal{B}(\mathcal{X}) \to \mathbb{R} \) given by the assignment \( \mathcal{B} \cap \mathcal{X} \mapsto \mu_k(\mathcal{B} \cap \mathcal{X}) \) defines a measure on \( (\mathcal{X}, \mathcal{B}(\mathcal{X})) \). It is immediate that each measure \( \nu_k \) is contained in \( P_1(\mathcal{X}) \). Due to Lemma 1.6 we have \( i_* \nu_k(\mathcal{B}) = \mu_k(\mathcal{B} \cap \mathcal{X}) = \mu_k(B \cap \text{spt}(\mu_k)) = \mu_k(\mathcal{B} \cap \text{spt}(\mu)) = \mu_k(B) \) for each integer \( k \geq 1 \) and each Borel measurable subset \( B \) of \( X \). Consequently, each measure \( \mu_k \) is contained in \( i_*(P_1(\mathcal{X})) \). Lemma 1.5 asserts that the map \( i_* \) is an isometric embedding; hence, the sequence \( \{ \nu_k(\mu_k) \}_{k \geq 1} \) is a Cauchy sequence in \( (P_1(\mathcal{X}), W^1) \) As the metric space \( (X, d) \) is complete and separable, Theorem 6.18 in [Vil09], whose proof is rather involved and therefore omitted in this text, tells us that there exists a measure \( \nu \) contained in \( P_1(\mathcal{X}) \) such that \( \nu_k \to \nu \) with \( k \to +\infty \). Therefore, we have that \( \mu_k \to i_* \nu \) with \( k \to +\infty \). Consequently, we have shown that the metric space \( (P_1(\mathcal{X}), W^1) \) is complete if the metric space \( (X, d) \) is complete. Suppose now that \( (X, d) \) is bounded. Let \( \mu \) and \( \nu \) be two measures contained in \( P_1(\mathcal{X}) \). We compute

\[
W^1(\mu, \nu) \leq \int_{X \times X} d(x, y) (\mu \otimes \nu)(dx, dy) \leq \sup_{x, y \in X} d(x, y).
\]

Thus, the metric space \( (P_1(\mathcal{X}), W^1) \) is bounded and the Theorem follows. ■

Let \( (X, d) \) denote a metric space. We define

\[
P_Q(\mathcal{X}) := \left\{ \sum_{i=1}^{n} \alpha_i \delta_{x_i} : n \geq 1, \sum_{i=1}^{n} \alpha_i = 1, \alpha_i \in \mathbb{Q}_{\geq 0}, x_i \in \mathcal{X} \right\} \quad \text{and} \quad P_\mathcal{R}(\mathcal{X}) := \left\{ \sum_{i=1}^{n} \alpha_i \delta_{x_i} : n \geq 1, \sum_{i=1}^{n} \alpha_i = 1, \alpha_i \in \mathbb{R}_{\geq 0}, x_i \in \mathcal{X} \right\}.
\]

The set \( P_Q(\mathcal{X}) \) is called the set of atomic probability measures with non-negative rational masses and the set \( P_\mathcal{R}(\mathcal{X}) \) is called the set of atomic probability measures with non-negative real masses. It turns out that \( P_Q(\mathcal{X}) \) is \( W^1 \)-dense in \( P_1(\mathcal{X}) \). This is the content of the following Proposition.

**Proposition 1.9** Let \( (X, d) \) be a non-empty metric space and let \( \mu \) be a Radon probability measure contained in \( P_1(\mathcal{X}) \). If \( \varepsilon > 0 \) is a positive real number, then there exists a measure \( \nu_\varepsilon \) contained in \( P_Q(\mathcal{X}) \) with \( \text{spt}(\nu_\varepsilon) \subset \text{spt}(\mu) \) such that \( W^1(\mu, \nu_\varepsilon) < \varepsilon \).

**Proof** The following is a conglomerate of the proof of Theorem 6.1 in [Edw11] and the proof of Theorem 6.18 in [Vil09]. Suppose that \( \varepsilon > 0 \) is a positive real number. Recall that \( \mu(\text{spt}(\mu)) = 1 \). Consequently, we have in particular that the set \( \text{spt}(\mu) \) is non-empty. Let \( x_0 \) be a point contained in \( \text{spt}(\mu) \). Since the measure \( \mu \) is a Radon measure, there exists a non-empty compact subset \( K \subset \text{spt}(\mu) \) with the property that

\[
\int_{K^c} d(x, x_0) \, d\mu < \frac{\varepsilon}{4},
\]
As the subset $K$ is compact, there exists an integer $n \geq 1$ and disjoint non-empty Borel measurable subsets $B_1, \ldots, B_n$ of $K$ with $\text{diam}(B_k) < \frac{\varepsilon}{4}$ for all integers $1 \leq k \leq n$ such that

$$K = \bigcup_{k \in \{1, \ldots, n\}} B_k.$$ 

Choose for every integer $1 \leq k \leq n$ an element $x_k$ contained in $B_k$. For each integer $0 \leq k \leq n$ we define the map $h_k : X \to X \times X$

$$x \mapsto (x, x_k).$$

Abbreviate $B_0 := K^c$. We set

$$\pi : \B(X \times X) \to \mathbb{R}
B \mapsto\sum_{k=0}^{n} \mu((h_k)^{-1}(B) \cap B_k).$$

We claim that $\pi$ is contained in $\mathcal{M}^+(X \times X)$, that is, $\pi$ is a non-negative Radon measure on $(X \times X, \mathcal{B}(X \times X))$. To see this, note that each map $h_k$ is continuous; thus, as $\mu$ is a non-negative Radon measure on $(X, \mathcal{B}(X))$ we obtain by unraveling the definitions that each measure $\mu((h_k)^{-1}(\cdot) \cap B_k)$ is a non-negative Radon measure on $(X \times X, \mathcal{B}(X \times X))$ and thus the measure $\pi$ is contained in $\mathcal{M}^+(X \times X)$, as it is a finite sum of non-negative Radon measures. We define the map $\nu : \B(X) \to \mathbb{R}$ through the assignment

$$B \mapsto\sum_{k=0}^{n} \mu(B_k) \delta_{x_k}(B).$$

Note that the map $\nu$ is a measure contained in $\mathcal{P}(X)$ and we have $\text{spt}(\nu) \subset \text{spt}(\mu)$. We compute

$$\pi(X \times B) = \sum_{k=0}^{n} \mu(B_k) \delta_{x_k}(B) = \nu(B)$$

and

$$\pi(B \times X) = \sum_{k=0}^{n} \mu(B_k \cap B) = \mu(B)$$

for all Borel measurable sets $B$ contained in $\mathcal{B}(X)$. Equation (1.6) tells us that the measure $\pi$ is a Radon coupling of the Radon probability measures $\mu$ and $\nu$. Therefore, we deduce

$$W^1(\mu, \nu) \leq \int_{X \times X} d(x, y) \, d\pi(dx, dy)
= \sum_{k=0}^{n} \int_{B_k} d(x, x_k) \, d\mu < \frac{\varepsilon}{4} + \mu(K) \frac{\varepsilon}{4} \leq \frac{\varepsilon}{2}.$$
Thus, we are left to construct a measure \( \nu_\varepsilon \) contained in \( P \) such that \( \nu_\varepsilon \subseteq spt(\nu) \) such that \( W^1(\nu, \nu_\varepsilon) \leq \frac{\varepsilon}{2} \). Suppose that \( \tau := \sum_{k=0}^{n} \alpha_k \delta_{x_k} \) is an element of \( P \). We define the two non-negative measures

\[
\begin{align*}
(\nu - \tau)_+ &:= \sum_{k=0}^{n} 0 \vee (\mu(B_k) - \alpha_k) \delta_{x_k} \\
(\nu - \tau)_- &:= \sum_{k=0}^{n} 0 \vee (\alpha_k - \mu(B_k)) \delta_{x_k}.
\end{align*}
\]

Note that \( \nu - \tau = (\nu - \tau)_+ - (\nu - \tau)_- \). Since \( \nu \) and \( \tau \) are probability measures, we obtain that \( a := (\nu - \tau)_+ + (\nu - \tau)_- \). It is immediate that \( a > 0 \).

Let \( x^* \) be a point contained in \( X \). We have

\[
W^1(\nu, \tau) = \sup \left\{ \int_X f \, d\nu - \int_X f \, d\tau : f \in Lip_1(X) \right\}
\]

\[
= \sup \left\{ \int_X f \, d(\nu - \tau)_+ - \int_X f \, d(\nu - \tau)_- : f \in Lip_1(X) \right\}
\]

\[
= aW^1 \left( \frac{(\nu - \tau)_+}{a}, \frac{(\nu - \tau)_-}{a} \right)
\]

\[
\leq \frac{1}{a} \int_{X \times X} d(x, y) (\nu - \tau)_+ \otimes (\nu - \tau)_- (dx, dy)
\]

\[
\leq \int_X d(x, x^*) (\nu - \tau)_+ (dx) + \int_X d(x, x^*) (\nu - \tau)_- (dx)
\]

\[
\leq \left( \max_{k \in \{0, \ldots, n\}} d(x_k, x^*) \right) \sum_{k=0}^{n} |\mu(B_k) - \alpha_k|.
\]

It is possible to choose non-negative rational numbers \( \alpha_0, \ldots, \alpha_n \) in \( \mathbb{Q}_{\geq 0} \) with \( \sum_{k=0}^{n} \alpha_k = 1 \) such that

\[
\left( \max_{k \in \{0, \ldots, n\}} d(x_k, x^*) \right) \sum_{k=0}^{n} |\mu(B_k) - \alpha_k| \leq \frac{\varepsilon}{2}.
\]

Hence, the measure \( \nu_\varepsilon := \sum_{k=0}^{n} \alpha_k \delta_{x_k} \) satisfies \( W^1(\nu, \nu_\varepsilon) \leq \frac{\varepsilon}{2} \) and \( spt(\nu_\varepsilon) \subset spt(\nu) \). The Proposition follows. \( \blacksquare \)

To conclude this section we derive an explicit formula for the first Wasserstein distance on \( P(X) \).

**Proposition 1.10** Let \( (X, d) \) denote a metric space. Let \( n \geq 1 \) denote an integer and let \( x_1, y_1, \ldots, x_n, y_n \) be points in \( X \). Then we have that

\[
W^1 \left( \frac{1}{n} (x_1 + \cdots + x_n), \frac{1}{n} (y_1 + \cdots + y_n) \right) = \frac{1}{n} \min_{\tau \in S_n} \sum_{i=1}^{n} d(x_i, y_{\tau(i)}). \quad (1.7)
\]
Proof. An outline of the following proof may be found in the introduction of [Vil03]. Abbreviate $d := W^1 \left( \frac{1}{n} (x_1 + \cdots + x_n), \frac{1}{n} (y_1 + \cdots + y_n) \right)$. By definition,

$$d = \inf_{\pi \in \Lambda(\mu, \nu)} \int_{X \times X} d(x, y) \, d\pi(x, y)$$

Note that $\pi \in \Lambda(\mu, \nu)$ if and only if

$$\sum_{i,j=1}^n \pi_{ij} \delta_{(x_i, y_j)} = 1$$

(1.8)

Hence, there is a one-to-one correspondence between the set $\Lambda(\mu, \nu)$ and the set of all $n \times n$ doubly stochastic matrices

$$B_n := \left\{ (\pi_{ij})_{i,j} : \pi_{ij} \geq 0, \text{ and } (\pi_{ij})_{i,j} \text{ satisfies (1.8)} \right\}.$$

Thus,

$$d = \inf \left\{ \frac{1}{n} \sum_{i,j=1}^n \pi_{ij} d(x_i, y_j) : (\pi_{ij})_{i,j} \in B_n \right\}. \quad (1.9)$$

Let $\{\pi^{(k)}\}_{k \geq 1} \subset B_n$ denote a sequence of doubly stochastic matrices such that

$$\lim_{k \to +\infty} \frac{1}{n} \sum_{i,j=1}^n \pi_{ij}^{(k)} d(x_i, y_j) = d.$$

Identify the symmetric group $S_n$ with the set of all $n \times n$ permutation matrices, i.e. identify $\tau \in S_n$ with the matrix $P_{\tau} := (\delta_{\tau(i), j})_{i,j}$. The Birkhoff-von Neumann theorem tells us that every doubly stochastic matrix is a convex combination of permutation matrices; thus, we obtain for every $k \geq 1$ that

$$d' \leq \frac{1}{n} \sum_{i,j=1}^n \pi_{ij}^{(k)} d(x_i, y_j), \quad (1.10)$$

where $d' := \min \left\{ \frac{1}{n} \sum_{i,j=1}^n \delta_{\tau(i), j} d(x_i, y_j) : \tau \in S_n \right\}$. Note that $d \leq d'$. Now combine (1.9) and (1.10) to deduce that $d' \leq d$; therefore, the Proposition follows. \qed
Bibliography


