Expositiones Mathematicae 29 (2011) 387–398

Contents lists available at SciVerse ScienceDirect

Expositiones Mathematicae

journal homepage: www.elsevier.de/exmath

On the Kantorovich–Rubinstein theorem

D.A. Edwards

Mathematical Institute, 24–29 St Giles', Oxford OX 1 3LB, United Kingdom

A R T I C L E   I N F O

Article history:
Received 23 February 2011
Received in revised form
2 May 2011

MSC:
primary 49N15
secondary 46E27
90C46
28A35
28C05
46E15

Keywords:
Kantorovich–Rubinstein theorem
Wasserstein metric
Dobrushin formula
Lipschitz functions

A B S T R A C T

The Kantorovich–Rubinstein theorem provides a formula for the Wasserstein metric W_1 on the space of regular probability Borel measures on a compact metric space. Dudley and de Acosta generalized the theorem to measures on separable metric spaces. Kellerer, using his own work on Monge–Kantorovich duality, obtained a rapid proof for Radon measures on an arbitrary metric space. The object of the present expository article is to give an account of Kellerer's generalization of the Kantorovich–Rubinstein theorem, together with related matters. It transpires that a more elementary version of Monge–Kantorovich duality than that used by Kellerer suffices for present purposes. The fundamental relations that provide two characterizations of the Wasserstein metric are obtained directly, without the need for prior demonstration of density or duality theorems. The latter are proved, however, and used in the characterization of optimal measures and functions for the Kantorovich–Rubinstein linear programme. A formula of Dobrushin is proved.

1. Introduction

The present paper is an exposition of the elements of Kantorovich–Rubinstein theory. Little originality can be claimed for the proofs offered here, but our treatment is in several respects simpler and more direct than that in a number of standard accounts.

The theory has evolved from the original Kantorovich–Rubinstein theorem concerning optimal transport (see [12,11]). Let (X, d) be a metric space, let $\mathfrak{B}(X)$ denote the set of all Borel subsets of X, and suppose that μ, ν are regular probability Borel measures on X. We denote by $\mathcal{P}(\mu, \nu)$ the set of regular probability Borel measures π on the topological product $X \times X$ such that

$$\mu(E) = \pi(E \times X) \quad \text{and} \quad \nu(E) = \pi(X \times E) \quad \text{for all } E \in \mathfrak{B}(X). \tag{1.1}$$

E-mail address: edwardsd@maths.ox.ac.uk.

0723-0869/$ – see front matter
doi:10.1016/j.exmath.2011.06.005

For $f : X \to \mathbb{R}$, we define the expression $\|f\|_L$ by the equation

$$\|f\|_L = \sup\{|f(x) - f(y)|/d(x, y) : x, y \in X; x \neq y\}. \tag{1.2}$$

Then the Kantorovich–Rubinstein theorem of [12,11] states that, if the space (X, d) is compact, we have

$$\inf_{\pi \in \mathcal{P}(\mu,\nu)} \int_{X\times X} d(x, y)\pi(\mathrm{d}x\mathrm{d}y) = \sup\left\{\int_X f\,\mathrm{d}\mu - \int_X f\,\mathrm{d}\nu : \|f\|_L \leq 1\right\}. \tag{1.3}$$

Dudley [6], in work completed by de Acosta [1], generalized this theorem to separable metric spaces (see also [7]) and, in that setting, showed that the infimum is attained if the measures μ, ν are tight. In a slightly different direction, Kellerer [14, Theorem 1] generalized the theorem to Radon measures on an arbitrary metric space as follows (but see also Section 9). Let $\mathcal{P}_d(X)$ denote the set of probability Radon measures on a metric space (X, d) such that, for some and hence all $x_0 \in X$, we have $\int_X d(x, x_0)\mu(\mathrm{d}x) < \infty$, and let $\Pi(\mu, \nu)$ be the set of all probability Radon measures π on $X \times X$ satisfying the condition (1.1). Then Kellerer's result [14, Theorem 1], slightly modified, tells us that for an arbitrary metric space (X, d) Eq. (1.3) remains true if $\mu, \nu \in \mathcal{P}_d(X)$ and $\mathcal{P}(\mu, \nu)$ is replaced by $\Pi(\mu, \nu)$. Moreover, the infimum is attained.

Whereas Dudley and de Acosta [6,1,7] make use of Monge–Kantorovich duality for compact metric spaces, Kellerer [14] invokes a stronger version of Monge–Kantorovich duality, valid for arbitrary topological spaces. This allows Kellerer to abridge the Dudley–de Acosta argument to obtain a very direct proof of his theorem that obviates the need for an elaborate approximation from the case where the underlying space is compact. Note, however, that the Monge–Kantorovich duality theorem used by Dudley and de Acosta is much simpler to prove than the more general version used by Kellerer (in effect [13, Theorem 2.6]). In the present article, we show that it suffices for Kellerer's argument to invoke instead an elementary formulation of Monge–Kantorovich duality that has a relatively simple proof (see [8, Theorem 4.1]).

The paper is organized as follows. We work with measures on a non-empty metric space (X, d). After establishing our notation and giving various definitions in Section 2, we summarize in Section 3 what we shall need concerning Monge–Kantorovich duality. Then in Section 4, by means of arguments taken from [6,1,7,14], we prove the generalization of the Kantorovich–Rubinstein theorem to Radon measures on arbitrary metric spaces. As a consequence, we see at once that the expression $W(\mu, \nu)$ defined below by Eq. (2.2) is a metric for $\mathcal{P}_d(X)$. We treat also the Kantorovich–Rubinstein transshipment problem, in which the above condition (1.1) is replaced by the requirement that

$$\pi(E \times X) - \pi(X \times E) = \mu(E) - \nu(E) \quad \text{for all } E \in \mathfrak{B}(X). \tag{1.4}$$

An advantage of using Kellerer's improvement of the Dudley–de Acosta approach is that the results of Section 4 are obtained rapidly, without the need for a prior demonstration of density or duality theorems of the type we prove later in Sections 6 and 7 (compare [1,6,7,11,12,15,16]). In Section 5, we digress to show that in a metric space the formula (5.3) of Dobrushin [5] can be deduced from Lemma 4.2. The main theorem of Section 6 shows that the probability measures on X having finite support are dense in $\mathcal{P}_d(X)$ with respect to the metric W. This leads to the representation in Section 7 of $\mathrm{Lip}(X, d)$ modulo the constant functions as the Banach dual of a certain normed vector space of measures, a fact which allows us finally, in Section 8, to characterize the optimal measures and functions of Kantorovich–Rubinstein theory. In Section 9, we add a few concluding remarks.

In Sections 2–8, we confine attention to Radon measures, for the theory of which see, for instance, [2–4,10]. This is not a restriction if the metric space (X, d) is complete and separable since, by Ulam's theorem (see, for example, [7, Theorem 7.1.4]), every finite Borel measure on X is then a Radon measure.

2. Definitions and notation

In what follows, (X, d) will always be a non-empty metric space. We shall denote by $\mathcal{M}(X)$ the space of real Radon measures on X and by $\mathcal{M}^+(X)$ the set of non-negative measures belonging to

$\mathcal{M}(X)$. We denote by $\mathcal{M}_d^+(X)$ the set of all $\mu \in \mathcal{M}^+(X)$ such that for some, and hence all, $x_0 \in X$, we have

$$\int_X d(x, x_0)\mu(\mathrm{d}x) < \infty,$$

and by $\mathcal{M}_d(X)$ the set of all $\mu \in \mathcal{M}(X)$ such that $|\mu| \in \mathcal{M}_d^+(X)$. (Thus we have $\mathcal{P}_d(X) = \{\mu \in \mathcal{M}_d^+(X) : \mu(X) = 1\}$.) Finally, we denote by $\mathcal{M}^0(X)$ the set of all $m \in \mathcal{M}(X)$ such that $m(X) = 0$, and by $\mathcal{M}_d^0(X)$ the set $\mathcal{M}^0(X) \cap \mathcal{M}_d(X)$.

In what follows, we shall frequently use the functional notation for integrals. For example, if $\mu \in \mathcal{M}^+(X)$ and $f \in \mathcal{L}^1(X, \mu)$, then $\mu(f)$ is shorthand for $\int_X f(x)\mu(\mathrm{d}x)$.

Given $f : X \to \mathbb{R}$, we have defined $\|f\|_L$ by the Eq. (1.2), we define $\mathrm{Lip}(X, d)$, the space of *Lipschitz functions* on X, as the set of all f such that $\|f\|_L < \infty$, and we denote by $\mathrm{Lip}_1(X, d)$ the set of f such that $\|f\|_L \le 1$. Note that $\|\cdot\|_L$ is a seminorm on $\mathrm{Lip}(X, d)$ and that $\|f\|_L = 0$ if and only if f is constant. It is obvious that $\mathrm{Lip}(X, d)$ is a vector subspace of $\mathbb{R}^X$; it is in fact a vector sublattice, as one sees by noting that $|\,|f(x)| - |f(y)|\,| \le |f(x) - f(y)|$. If $l \in \mathrm{Lip}_1(X, d)$ and $\rho \in \mathcal{M}_d^+(X)$ then $l \in \mathcal{L}^1(X, \rho)$. To see this, let $x_0 \in X$ and observe that

$$l(x_0) - d(x, x_0) \le l(x) \le l(x_0) + d(x, x_0). \tag{2.1}$$

The product space $Z = X \times X$ is metrizable, and we take its metric to be given by

$$\delta(z, z') = d(x, x') + d(y, y'),$$

where $x, x', y, y' \in X$ and $z = (x, y), z' = (x', y')$. The phrase 'uniformly continuous on Z' will always mean 'uniformly continuous with respect to δ'. Note that d is uniformly continuous on Z, because

$$|d(x, y) - d(x', y')| \le d(x, x') + d(y, y').$$

Given functions $f, g : X \to [-\infty, \infty)$, we shall denote by $f \oplus g$ the function on Z defined by the equation $(f \oplus g)(x, y) = f(x) + g(y)$. For example, if l is a real function on X, then we have $l \in \mathrm{Lip}_1(X, d)$ if and only if $l \oplus (-l) \le d$.

We define projections $p_1, p_2 : Z \to X$ by writing $p_1(x, y) = x$ and $p_2(x, y) = y$. The associated maps of measures $\mathcal{M}^+(Z) \to \mathcal{M}^+(X)$ will also be denoted by p_1 and p_2 (see, for example, [2, Proposition 1.15] or [3, Theorem 9.1.1]). Thus, when $\pi \in \mathcal{M}^+(Z)$ the measures $\mu = p_1(\pi)$ and $\nu = p_2(\pi)$ satisfy the condition (1.1).

Conversely, given $\mu, \nu \in \mathcal{M}^+(X)$ such that $\mu(X) = \nu(X)$, we write

$$\Pi(\mu, \nu) = \{\pi \in \mathcal{M}^+(Z) : p_1(\pi) = \mu, p_2(\pi) = \nu\}.$$

Then $\Pi(\mu, \nu) \neq \emptyset$, trivially if $\mu(X) = 0$, and because $\mu \otimes \nu/\mu(X) \in \Pi(\mu, \nu)$ if $\mu(X) > 0$. If f, g are real Borel functions in $\mathcal{L}^1(X, \mu)$ and $\mathcal{L}^1(X, \nu)$ respectively and $\pi \in \Pi(\mu, \nu)$, we have $f \oplus g \in \mathcal{L}^1(Z, \pi)$ and

$$\pi(f \oplus g) = \mu(f) + \nu(g).$$

Now assume, in addition, that $\mu, \nu \in \mathcal{M}_d^+(X)$. Then $d \in \mathcal{L}^1(Z, \pi)$. For let $x_0 \in X$. The inequalities $0 \le d(x, y) \le d(x, x_0) + d(x_0, y)$ can be written as $0 \le d \le a \oplus b$, where $a(x) = d(x, x_0)$ and $b(y) = d(x_0, y)$. But a, b are continuous, hence Borel, and $a \in \mathcal{L}^1(X, \mu), b \in \mathcal{L}^1(X, \nu)$. Hence $a \oplus b \in \mathcal{L}^1(Z, \pi)$ and therefore $d \in \mathcal{L}^1(Z, \pi)$. The set $\Pi(\mu, \nu)$ is a $\sigma(\mathcal{M}(Z), C_b(Z))$-compact subset of $\mathcal{M}^+(Z)$ (see [8, Theorem 2.3]). The map $\Pi(\mu, \nu) \ni \pi \mapsto \pi(d)$ is lower semicontinuous because $\pi(d) = \lim_{n\to\infty} \pi(d_n)$, where $d_n = \min(d, n)$. Hence $\pi \mapsto \pi(d)$ attains its infimum on $\Pi(\mu, \nu)$. The *Wasserstein metric* $W_1(\mu, \nu)$ for $\mathcal{P}_d(X)$, which we shall denote simply by $W(\mu, \nu)$, is defined by the equation

$$W(\mu, \nu) = \min\{\pi(d) : \pi \in \Pi(\mu, \nu)\}. \tag{2.2}$$

Clearly $0 \le W(\mu, \nu) < \infty$; we shall see below that W is indeed a metric for $\mathcal{P}_d(X)$.

For $m \in \mathcal{M}_d^0(X)$ we let

$$\Gamma(m) = \{\gamma \in \mathcal{M}^+(Z) : p_1(\gamma), p_2(\gamma) \in \mathcal{M}_d^+(X); p_1(\gamma) - p_2(\gamma) = m\}$$

and we define the *Wasserstein functional* $\|m\|_W$ of m by the equation

$$\|m\|_W = \inf\{\gamma(d) : \gamma \in \Gamma(m)\};$$

note that $\Gamma(m) \supseteq \Pi(m^+, m^-) \neq \emptyset$, where $m = m^+ - m^-$ is the Jordan decomposition of m, and that $0 \le \|m\|_W < \infty$. Finally, for $m \in \mathcal{M}_d^0(X)$, we write

$$\|m\|_L^* = \sup\{m(l) : l \in \mathrm{Lip}_1(X, d)\}. \tag{2.3}$$

3. Monge–Kantorovich duality

We recall here a result from [8] that will be used below. Let U, V be non-empty completely regular spaces, let μ and ν be bounded non-negative Radon measures on U and V respectively such that $\mu(U) = \nu(V)$, and let T be the topological product $U \times V$. Let $\Lambda(\mu, \nu)$ denote the set of all bounded non-negative Radon measures λ on T such that $\mu(E) = \lambda(E \times V)$ and $\nu(F) = \lambda(U \times F)$ for all $E \in \mathfrak{B}(U)$ and $F \in \mathfrak{B}(V)$. For each lower semicontinuous function $c : T \to [0, \infty]$ let $\Theta(c)$ denote the set of pairs (u, v) of Borel functions $u : U \to [-\infty, \infty)$ and $v : V \to [-\infty, \infty)$ such that $u \in \mathcal{L}^1(\mu)$ and $v \in \mathcal{L}^1(\nu)$ and which satisfy

$$(u \oplus v)(x, y) \equiv u(x) + v(y) \le c(x, y) \quad \text{for all } (x, y) \in T.$$

Theorem 3.1. *Let $c : T \to [0, \infty]$ be a lower semicontinuous function. Then we have*

$$\min_{\lambda \in \Lambda(\mu,\nu)} \lambda(c) = \sup\{\mu(u) + \nu(v) : (u, v) \in \Theta(c)\}. \tag{3.1}$$

(Here, the case in which both terms are $+\infty$ is not excluded, but this will not occur in the application we make of this result below.)

For this, see [8, Theorem 4.1]. (As already remarked, Kellerer's Theorem 2.6 in [13] is stronger, but has a much more difficult proof.)

Given a topological space S, we shall denote by $\mathcal{B}^\infty(S)$ the set of all bounded real Borel functions on S.

Corollary 3.2. *Let $c : T \to [0, \infty]$ be a lower semicontinuous function. Then*

$$\min_{\lambda \in \Lambda(\mu,\nu)} \lambda(c) = \sup\{\mu(u) + \nu(v) : u \in \mathcal{B}^\infty(U), v \in \mathcal{B}^\infty(V); u \oplus v \le c\}. \tag{3.2}$$

Proof. The following proof is due to Kellerer [14]. By Theorem 3.1 and the monotone convergence theorem we have

$$\min_{\lambda \in \Lambda(\mu,\nu)} \lambda(c) = \sup\{\mu(u \wedge n) + \nu(v \wedge n) : (u, v) \in \Theta(c); n \in \mathbb{N}\}.$$

Hence, we can add to the right-hand side of Eq. (3.1) the requirement that the functions u, v are to be bounded above. But if $u, v \le n$ and $u \oplus v \le c$ then $u \le u \vee (-n)$, $v \le v \vee (-n)$, and $(u \vee (-n)) \oplus (v \vee (-n)) \le c$. It follows that, in Eq. (3.1), we can also stipulate that the functions u, v are to be bounded below. Hence we have Eq. (3.2). $\square$

4. The Kantorovich–Rubinstein theorem

Here is Kellerer's extension of the Kantorovich–Rubinstein theorem to Radon measures on an arbitrary metric space.

Theorem 4.1. *Let (X, d) be a metric space, let $\mu, \nu \in \mathcal{M}_d^+(X)$, and suppose that $\mu(X) = \nu(X)$. Then*

$$W(\mu, \nu) = \|\mu - \nu\|_L^*. \tag{4.1}$$

For comments on the proof which follows, see Section 1. First, note that we have, by Theorem 3.1 and Corollary 3.2,

$$W(\mu,\nu) = \sup\{\mu(f)+\nu(g) : f,g \in \mathcal{B}^\infty(X) : f\oplus g \le d\} \tag{4.2}$$
$$= \sup\{\mu(f)+\nu(g) : (f,g)\in \Theta(d)\}. \tag{4.3}$$

For $\mu,\nu \in \mathcal{M}_d^+(X)$ with $\mu(X)=\nu(X)$ let

$$R(\mu,\nu) = \sup\{\mu(l)-\nu(l) : l\in C_b(X), l\oplus(-l)\le d\}.$$

Lemma 4.2. *Let $\mu,\nu\in\mathcal{M}_d^+(X)$ and suppose that $\mu(X)=\nu(X)$. Then*

$$R(\mu,\nu) = W(\mu,\nu). \tag{4.4}$$

Proof. By Corollary 3.2,

$$\begin{aligned} R(\mu,\nu) &= \sup\{\mu(l)-\nu(l) : l\in C_b(X), l\oplus(-l)\le d\}\\ &\le \sup\{\mu(f)+\nu(g) : f,g\in\mathcal{B}^\infty(X) : f\oplus g\le d\} = W(\mu,\nu).\end{aligned}$$

Now, suppose that $\epsilon>0$. Then there exist $f,g\in\mathcal{B}^\infty(X)$ with $f\oplus g\le d$ such that

$$W(\mu,\nu)-\epsilon < \mu(f)+\nu(g).$$

We now make use of a standard construction, as follows. For $x\in X$ let

$$k(x) = \inf_{y\in X}(d(x,y)-g(y)).$$

Then $k(x)\in\mathbb{R}$ and

$$\begin{aligned} k(x)-k(x') &= \inf_{y\in X}\sup_{y'\in X}(d(x,y)-d(x',y')-g(y)+g(y'))\\ &\le \sup_{y'\in X}(d(x,y')-d(x',y')) \le d(x,x').\end{aligned}$$

Hence $k\in \mathrm{Lip}_1(X,d)$. Moreover

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

and hence k is bounded, and $f\oplus g\le k\oplus(-k)$. So, for $\pi\in\Pi(\mu,\nu)$, we have

$$\begin{aligned} W(\mu,\nu)-\epsilon &< \mu(f)+\nu(g) = \pi(f\oplus g)\le\pi(k\oplus(-k))\\ &= \mu(k)-\nu(k)\le R(\mu,\nu)\le W(\mu,\nu).\end{aligned}$$

This completes the proof. □

Proof of Theorem 4.1. If $l\in\mathrm{Lip}_1(X,d)$ then $(l,-l)\in\Theta(d)$ (see (2.1)). Hence, by Lemma 4.2, we have

$$\begin{aligned} W(\mu,\nu) &= \sup\{\mu(l)-\nu(l) : l\in C_b(X), \|l\|_L\le 1\}\\ &\le \sup\{\mu(l)-\nu(l) : l\in\mathrm{Lip}_1(X,d)\}\\ &\le \sup\{\mu(f)+\nu(g) : (f,g)\in\Theta(d)\} = W(\mu,\nu).\end{aligned}$$

By definition (2.3), this completes the proof. □

In the article [1] of de Acosta the results corresponding to Theorem 4.4 have a long and somewhat intricate proof. Theorem 4.1 allows us to make short work of this.

Lemma 4.3. *The map $m\to\|m\|_L^*$ is a seminorm for the vector space $\mathcal{M}_d^0(X)$.*

Proof. Let $m\in\mathcal{M}_d^0(X)$. By Theorem 4.1 we have $\|m\|_L^* = W(m^+,m^-)$ and hence $0\le\|m\|_L^*<\infty$. Next, $-l\in\mathrm{Lip}_1(X,d)$ if and only if $l\in\mathrm{Lip}_1(X,d)$, and hence $\|-m\|_L^* = \|m\|_L^*$. More generally, we have $\|\alpha m\|_L^* = |\alpha|\|m\|_L^*$ for each real constant α. This is obvious if $\alpha \geq 0$, and for $\alpha < 0$, we have

$$\|\alpha m\|_L^* = \|(-\alpha)(-m)\|_L^* = (-\alpha)\| - m\|_L^* = |\alpha|\|m\|_L^*.$$

Finally, it is almost obvious that $\|\cdot\|_L^*$ is subadditive. □

For each $r > 0$ let $\mathcal{M}_d^+(X)_r$ denote the set of $\mu \in \mathcal{M}_d^+(X)$ such that $\mu(X) = r$.

Theorem 4.4. *For each $r > 0$ the map*

$$\mathcal{M}_d^+(X)_r \times \mathcal{M}_d^+(X)_r \ni (\mu, \nu) \mapsto W(\mu, \nu)$$

is a metric for $\mathcal{M}_d^+(X)_r$. Consequently the functional $\|\cdot\|_L^*$ is a norm for the vector space $\mathcal{M}_d^0(X)$. Moreover, if $\mu, \nu, \lambda \in \mathcal{M}_d^+(X)$ with $\mu(X) = \nu(X)$, then

$$W(\mu + \lambda, \nu + \lambda) = W(\mu, \nu). \tag{4.5}$$

Proof. The fact that $0 \leq W(\lambda, \mu) = W(\mu, \lambda) \leq W(\mu, \nu) + W(\nu, \lambda)$ is evident from Eq. (4.1) and the preceding lemma.

Now, suppose that $\mu, \nu \in \mathcal{M}_d^+(X)_r$ with $\mu(X) = \nu(X)$ and $W(\mu, \nu) = 0$. Let F be a closed non-empty subset of X and let $f(x) = d(x, F)$. Then $f \in \mathrm{Lip}_1(X, d)$. Let $f_n(x) = nf(x) \wedge 1$. Then $n^{-1}f_n \in \mathrm{Lip}_1(x, d)$ and hence, by Theorem 4.1, $\mu(f_n) = \nu(f_n)$ for all n. But $f_n \uparrow \mathbf{1}_G$ as $n \to \infty$, where $G = \complement F$. Hence $\mu(G) = \nu(G)$. Thus μ and ν agree on open sets and hence $\mu = \nu$. Thus $(\mu, \nu) \mapsto W(\mu, \nu)$ is a metric for $\mathcal{M}_d^+(X)_r$.

To prove that $\|\cdot\|_L^*$ is a norm it will suffice, by Lemma 4.3, to show that if $m \in \mathcal{M}_d^0(X)$ with $\|m\|_L^* = 0$ then $m = 0$. But the condition $\|m\|_L^* = 0$ is equivalent to saying that $W(m^+, m^-) = 0$, whence $m^+ = m^-$ and so $m = 0$.

Finally, Eq. (4.5) is now an immediate consequence of Theorem 4.1. □

Turning to the transshipment problem, we obtain another expression for $W(\mu, \nu)$, which will be important in Sections 6–8.

Theorem 4.5. *Let (X, d) be a metric space, let $\mu, \nu \in \mathcal{M}_d^+(X)$, and suppose that $\mu(X) = \nu(X)$. Then*

$$W(\mu, \nu) = \|\mu - \nu\|_W. \tag{4.6}$$

Consequently $\|m\|_W = \|m\|_L^*$ for all m in the vector space $\mathcal{M}_d^0(X)$, and hence $\|\cdot\|_W$ is a norm for $\mathcal{M}_d^0(X)$.

Proof. Our preceding results allow us to follow without further preliminaries the very last part of Appendix B in [1]. Since $\Pi(\mu, \nu) \subseteq \Gamma(\mu - \nu)$, we have

$$\|\mu - \nu\|_W \leq W(\mu, \nu).$$

Now, suppose $\epsilon > 0$ and choose $\gamma \in \Gamma(\mu - \nu)$ such that

$$\gamma(d) < \|\mu - \nu\|_W + \epsilon.$$

Let $m = \mu - \nu$, $\gamma_1 = p_1(\gamma)$, and $\gamma_2 = p_2(\gamma)$. Then

$$\gamma_1 - \gamma_2 = \mu - \nu = m^+ - m^-.$$

So there exist measures $\sigma, \tau \in \mathcal{M}_d^+(X)$ such that

$$\mu = m^+ + \sigma, \qquad \nu = m^- + \sigma, \qquad \gamma_1 = m^+ + \tau, \qquad \gamma_2 = m^- + \tau.$$

Now $W(\gamma_1, \gamma_2) = \min_{\pi \in \Pi(\gamma_1, \gamma_2)} \pi(d) \leq \gamma(d)$, and so, by Theorem 4.4,

$$W(\mu, \nu) = W(m^+, m^-) = W(\gamma_1, \gamma_2) \leq \gamma(d) < \|\mu - \nu\|_W + \epsilon,$$

hence $W(\mu, \nu) \leq \|\mu - \nu\|_W$, so in fact we have equality here.

By Theorems 4.1 and 4.4, the final assertion is now clear. □

5. The indicator metric

The *indicator metric* j for X is defined as follows.

$$j(x,y)=\begin{cases}0 & \text{if } x=y,\\ 1 & \text{if } x\neq y.\end{cases}$$

It is a bounded lower semicontinuous function on Z. The set $\mathrm{Lip}_1(X,j)$ contains an abundance of non-measurable functions (I owe this observation to Gordon Blower). To see this, note that $\mathrm{Lip}_1(X,j)$ consists of all functions $l : X \to \mathbb{R}$ such that $\operatorname{osc} l \le 1$, where

$$\operatorname{osc} l = \sup l - \inf l.$$

Now, let $j_n(x,y) = \min(nd(x,y),1)$ for $x,y \in X$ and $n \ge 1$. Then each j_n is a bounded metric for X, $j_n \le j_{n+1}$ for all n, and $j_n(x,y) \to j(x,y)$ as $n \to \infty$. Moreover, each metric j_n is topologically equivalent to d.

Now let $\mathfrak{d}$ be a bounded metric for X that is topologically equivalent to d. Note that $\mathcal{M}^+_{\mathfrak{d}}(X) = \mathcal{M}^+(X)$. Suppose we are given two measures $\mu,\nu \in \mathcal{M}^+(X)$ such that $\mu(X)=\nu(X)$. The set $\Pi(\mu,\nu)$ remains unchanged if we replace the metric d by $\mathfrak{d}$. So, if we define $q(\mathfrak{d})$ by the equation

$$q(\mathfrak{d}) = \sup\{\mu(l)-\nu(l) : l \in C_b(X)\cap \mathrm{Lip}_1(X,\mathfrak{d})\},$$

then, by Lemma 4.2, we have $\min_{\pi\in\Pi(\mu,\nu)} \pi(\mathfrak{d}) = q(\mathfrak{d})$. In particular,

$$\min_{\pi\in\Pi(\mu,\nu)} \pi(j_n) = q(j_n)$$

for all n. Next, $\mathrm{Lip}_1(X,j_n) \subseteq \mathrm{Lip}_1(X,j)$, and hence $q(j_n) \le q(j)$, where

$$q(j) = \sup\{\mu(l)-\nu(l) : l \in C_b(X)\cap \mathrm{Lip}_1(X,j)\}.$$

Noting also that for $l \in C_b(X)\cap\mathrm{Lip}_1(X,j)$, we have $l\oplus(-l) \le j$ and hence $\mu(l)-\nu(l) = \pi(l\oplus(-l)) \le \pi(j)$, we see that

$$\min_{\pi\in\Pi(\mu,\nu)} \pi(j_n) = q(j_n) \le q(j) \le \min_{\pi\in\Pi(\mu,\nu)} \pi(j). \tag{5.1}$$

Lemma 5.1. *Let Ω be a compact Hausdorff space and let (u_n) be an increasing sequence in $C(\Omega)$ with pointwise limit $u : \Omega \to (-\infty,\infty]$. Then*

$$\lim_{n\to\infty}\min_{\omega\in\Omega} u_n(\omega) = \sup_n \min_{\omega\in\Omega} u_n(\omega) = \min_{\omega\in\Omega} u(\omega).$$

Proof. This is a close relative of Dini's theorem. For a proof see, for example, [8, Lemma 4.2]. □

Now, for $h \in C_b(Z)$ and $\pi \in \Pi(\mu,\nu)$ write $\hat{h}(\pi) = \pi(h)$. Then $\hat{h} : \Pi(\mu,\nu) \to \mathbb{R}$ is continuous with respect to the $\sigma(\mathcal{M}(Z), C_b(Z))$-topology, and we recalled in Section 2 that the latter makes $\Pi(\mu,\nu)$ a compact Hausdorff space. By the monotone convergence theorem $\lim_{n\to\infty}\hat{j}_n(\pi) = \hat{j}(\pi)$. Hence, by Lemma 5.1,

$$\min_{\pi\in\Pi(\mu,\nu)} \pi(j_n) = \min_{\pi\in\Pi(\mu,\nu)} \hat{j}_n(\pi) \to \min_{\pi\in\Pi(\mu,\nu)} \hat{j}(\pi) = \min_{\pi\in\Pi(\mu,\nu)} \pi(j)$$

as $n \to \infty$. Applying this to the inequalities (5.1), we deduce that

$$\min_{\pi\in\Pi(\mu,\nu)} \pi(j) = q(j). \tag{5.2}$$

Now let $D(\mu,\nu)$ be defined by

$$D(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \pi(j) = \min_{\pi\in\Pi(\mu,\nu)} \pi(Z\setminus\Delta),$$

where Δ denotes the diagonal in $Z = X\times X$. By Eq. (5.2), we have the following result.

Theorem 5.2. *Let $\mu,\nu \in \mathcal{M}^+(X)$ with $\mu(X)=\nu(X)$ Then*

$$D(\mu,\nu) = \sup\{\mu(l)-\nu(l) : l \in C_b(X);\ 0 \le l \le 1\}.$$

This leads to the formula of Dobrushin [5]:

Corollary 5.3. *Under the same conditions, we have*

$$D(\mu,\nu) = \sup\{\mu(E) - \nu(E) : E \in \mathfrak{B}(X)\}. \tag{5.3}$$

Proof. By Theorem 5.2,

$$\sup\{\mu(\phi) - \nu(\phi) : \phi \in \mathcal{B}^\infty(X);\ 0 \le \phi \le 1\} \ge D(\mu,\nu).$$

On the other hand, when $\phi \in \mathcal{B}^\infty(X)$ with $0 \le \phi \le 1$, we have $\phi \oplus (-\phi) \le j$, and so, for $\pi \in \Pi(\mu,\nu)$, we have

$$\mu(\phi) - \nu(\phi) = \pi(\phi \oplus (-\phi)) \le \pi(j)$$

and therefore

$$\sup\{\mu(\phi) - \nu(\phi) : \phi \in \mathcal{B}^\infty(X);\ 0 \le \phi \le 1\} \le D(\mu,\nu).$$

So we must, in fact, have equality here. Now let $X = X_+ \cup X_-$ be a Hahn decomposition of X for the measure $m = \mu - \nu$ and let

$$\psi(x) = \begin{cases} 1 & \text{if } x \in X_+, \\ 0 & \text{if } x \in X_-. \end{cases}$$

Then $\psi \in \mathcal{B}^\infty(X)$, $0 \le \psi \le 1$, and it is evident that

$$m(\psi) = \sup\{m(\phi) : \phi \in \mathcal{B}^\infty(X);\ 0 \le \phi \le 1\} = D(\mu,\nu).$$

On the other hand,

$$\begin{aligned} m(\psi) &= m(X_+) = m^+(X) = \sup\{m(E) : E \in \mathfrak{B}(X)\} \\ &= \sup\{\mu(E) - \nu(E) : E \in \mathfrak{B}(X)\}. \end{aligned}$$

This completes the proof of Dobrushin's formula (5.3). □

This is not the shortest proof of Dobrushin's formula, because it can also be proved directly, without using Theorem 5.2, by evaluating $\max\{\pi(\Delta) : \pi \in \Pi(\mu,\nu)\}$—see [18] for some indications. Our object here has simply been to derive the formula by an application of Lemma 4.2.

6. Two density theorems

Let us denote by $\mathcal{M}_f^+(X)$ the set of all $\mu \in \mathcal{M}^+(X)$ such that supp μ is a finite set. Thus, the elements of $\mathcal{M}_f^+(X)$ are precisely the measures that are of the form $\sum_{\alpha\in A} t_\alpha \varepsilon_\alpha$, where A is a finite subset of X and $t_\alpha \in \mathbb{R}_+$ for all $\alpha \in A$.

Theorem 6.1. *Let $\mu \in \mathcal{M}_d^+(X)$ and suppose that $\epsilon > 0$. Then there exists $\sigma \in \mathcal{M}_f^+(X)$ such that $\mu(X) = \sigma(X)$ and $W(\mu,\sigma) < \epsilon$.*

Proof. For the case of compact or separable X, this result has been given a number of proofs, some decidedly difficult. Here we adapt the argument of de Acosta [1], who credits B. Simon with the main idea. We first recall a couple of properties of Radon measures. Suppose that $\mu \in \mathcal{M}^+(X)$ and $A \in \mathfrak{B}(X)$. Let $\mu_A : \mathfrak{B}(X) \to \mathbb{R}$ be defined by the formula $\mu_A(B) = \mu(A \cap B)$. Then it is trivial to show that $\mu_A \in \mathcal{M}^+(X)$. Next, suppose that S, T are Hausdorff spaces, that $h : S \to T$ is a continuous map, and that $\sigma \in \mathcal{M}^+(S)$. Then the image measure $h(\sigma) = \sigma \circ h^{-1}$ belongs to $\mathcal{M}^+(T)$ (see, for instance, [2, Proposition 1.15] or [3, Theorem 9.1.1]). Moreover, for $f \in \mathcal{L}^1(T, h(\sigma))$, we have $f \circ h \in \mathcal{L}^1(S,\sigma)$ and $\int_S f \circ h(s)\sigma(\mathrm{d}s) = \int_T f(t)\tau(\mathrm{d}t)$, where $\tau = h(\sigma)$.

It will suffice to prove the theorem for the case in which $\mu \in \mathcal{P}_d(X)$. Suppose that $\epsilon > 0$, and let $a_0 \in X$. Then $d(\cdot, a_0) \in \mathcal{L}^1(\mu)$ and hence, because the measure μ is tight, we can find a non-empty compact set K such that

$$\int_{\complement K} d(x,a_0)\mu(\mathrm{d}x) < \epsilon.$$

We can partition K into a finite union $\bigcup_{r=1}^{n} A_r$ of Borel sets A_r each of diameter at most ϵ. For each r, choose $a_r \in A_r$ and let $h_r(x) = (x, a_r) \in Z$ for all $x \in X$ and $r = 0, 1, \ldots, n$. For each $E \in \mathfrak{B}(Z)$, let

$$\pi(E) = \sum_{r=1}^{n} \mu(A_r \cap h_r^{-1}(E)) + \mu(O \cap h_0^{-1}(E)),$$

where $O = \complement K$. Then, for instance,

$$\mu(A_r \cap h_r^{-1}(E)) = \mu_{A_r}(h_r^{-1}(E)) = h_r(\mu_{A_r})(E),$$

so it is clear that $\pi \in \mathcal{M}^+(Z)$.

Next we compute $p_1(\pi)$ and $p_2(\pi)$. For $A \in \mathfrak{B}(X)$, we have

$$p_1(\pi)(A) = \pi(A \times X) = \sum_{r=1}^{n} \mu(A_r \cap A) + \mu(O \cap A) = \mu(A).$$

Thus $p_1(\pi) = \mu$. And, for $B \in \mathfrak{B}(X)$, we have

$$h_r^{-1}(X \times B) = \{x \in X : (x, a_r) \in X \times B\} = \begin{cases} X & \text{if } a_r \in B, \\ \emptyset & \text{if } a_r \notin B, \end{cases}$$

and hence

$$p_2(\pi)(B) = \pi(X \times B) = \sum_{r=1}^{n} \mu(A_r)\varepsilon_{a_r}(B) + \mu(O)\varepsilon_{a_0}(B).$$

Thus $p_2(\pi) = \sum_{r=1}^{n} \mu(A_r)\varepsilon_{a_r} + \mu(O)\varepsilon_{a_0} \in \mathcal{M}_f^+(X)$; so $\pi \in \Pi(\mu, \sigma)$, where $\sigma = p_2(\pi)$.

Finally, we estimate $\int_Z d(x, y)\pi(\mathrm{d}x\mathrm{d}y)$. For each r we have

$$\int_Z d(x, y)h_r(\mu_{A_r})(\mathrm{d}x\mathrm{d}y) = \int_X d \circ h_r(x)\mu_{A_r}(\mathrm{d}x) = \int_X d(x, a_r)\mu_{A_r}(\mathrm{d}x).$$

Hence

$$\begin{aligned} W(\mu, \sigma) &\le \int_Z d(x, y)\pi(\mathrm{d}x\mathrm{d}y) = \sum_{r=1}^{n} \int_X d(x, a_r)\mu_{A_r}(\mathrm{d}x) + \int_X d(x, a_0)\mu_O(\mathrm{d}x) \\ &= \sum_{r=1}^{n} \int_{A_r} d(x, a_r)\mu(\mathrm{d}x) + \int_O d(x, a_0)\mu(\mathrm{d}x) < \epsilon\mu(K) + \epsilon \le 2\epsilon, \end{aligned}$$

so the proof is complete. □

Given $x, y \in X$, we denote by ε_{xy} the measure $\varepsilon_x - \varepsilon_y$. A *simple* element of $\mathcal{M}^0(X)$ is one that is a finite linear combination of measures of the form ε_{xy}. We denote by $\mathcal{M}_s^0(X)$ the set of all simple elements of $\mathcal{M}^0(X)$. Let us also denote by $\mathcal{M}_f^0(X)$ the set of $\mu \in \mathcal{M}^0(X)$ such that $\operatorname{supp}\mu$ is a finite set. (Recall that, for a signed Radon measure, $\operatorname{supp}\mu = \operatorname{supp}\mu^+ \cup \operatorname{supp}\mu^-$.)

Lemma 6.2. *An element of $\mathcal{M}^0(X)$ is simple if and only if it has finite support. Thus $\mathcal{M}_s^0(X) = \mathcal{M}_f^0(X) \subseteq \mathcal{M}_d^0(X)$.*

Proof. The simple elements of $\mathcal{M}^0(X)$ are obviously measures with finite support. Conversely, suppose that $\mu \in \mathcal{M}_f^0(X)$. Then, for some finite subset A of X and real numbers t_a, we have $\mu = \sum_{a \in A} t_a\varepsilon_a$, with $\sum_{a \in A} t_a = 0$. Now let $b \in X$. Then $\mu = \sum_{a \in A} t_a(\varepsilon_a - \varepsilon_b) = \sum_{a \in A} t_a\varepsilon_{ab} \in \mathcal{M}_s^0(X)$. The inclusion $\mathcal{M}_f^0(X) \subseteq \mathcal{M}_d^0(X)$ is obvious. □

Theorem 6.3. *The set $\mathcal{M}_s^0(X)$ of simple elements of $\mathcal{M}^0(X)$ is a dense vector subspace of $\mathcal{M}_d^0(X)$ with respect to the norm $\|\cdot\|_W$.*

Proof. Let $m \in \mathcal{M}_d^0(X)$ and suppose that $\epsilon > 0$. By Theorem 6.1 we can choose measures $\sigma, \tau \in \mathcal{M}_f^+(X)$ such that $\sigma(X) = m^+(X)$, $\tau(X) = m^-(X)$ and

$$W(m^+, \sigma) < \epsilon, \qquad W(m^-, \tau) < \epsilon,$$

and let $\rho = \sigma - \tau$. Then, by Lemma 6.2, ρ is a simple element of $\mathcal{M}^0(X)$ and

$$\begin{aligned} \|m - \rho\|_W &= \|(m^+ - \sigma) + (\tau - m^-)\|_W \\ &\leq \|m^+ - \sigma\|_W + \|\tau - m^-\|_W \\ &= W(m^+, \sigma) + W(\tau, m^-) < 2\epsilon. \quad \square \end{aligned}$$

7. The dual of $\mathcal{M}_d^0(X)$

Here we study $\mathcal{M}_d^0(X)$ as a vector space endowed with the norm $\|\cdot\|_W$.

Lemma 7.1. *For all $x, y \in X$, we have $\|\varepsilon_{xy}\|_W \leq d(x, y)$.*

Proof. We can assume that $x \neq y$. Then $\varepsilon_x \otimes \varepsilon_y = \varepsilon_{xy}^+ \otimes \varepsilon_{xy}^- \in \Gamma(\varepsilon_{xy})$. Hence

$$\|\varepsilon_{xy}\|_W \leq \int_{X\times X} d(s, t)(\varepsilon_x \otimes \varepsilon_y)(\mathrm{d}s\mathrm{d}t) = d(x, y). \quad \square$$

Now, suppose that $f \in \mathrm{Lip}(X, d)$. Then for each $m \in \mathcal{M}_d^0(X)$, we define $\hat{f}(m)$ by the equation

$$\hat{f}(m) = \int_X f\, \mathrm{d}m.$$

Lemma 7.2. *For $f \in \mathrm{Lip}(X, d)$ the map $\hat{f} : \mathcal{M}_d^0(X) \to \mathbb{R}$ is a continuous linear functional, and $\|\hat{f}\| = \|f\|_L$.*

Proof. For $\gamma \in \Gamma(m)$, we have $\hat{f}(m) = \int_Z (f(x) - f(y))\gamma(\mathrm{d}x\mathrm{d}y)$. Hence

$$|\hat{f}(m)| \leq \int_Z |f(x) - f(y)|\gamma(\mathrm{d}x\mathrm{d}y) \leq \|f\|_L \int_Z d(x, y)\gamma(\mathrm{d}x\mathrm{d}y) = \|f\|_L \gamma(d).$$

Taking the infimum over $\gamma \in \Gamma(m)$, we obtain $|\hat{f}(m)| \leq \|f\|_L\|m\|_W$, and thus $\|\hat{f}\| \leq \|f\|_L$.

On the other hand, we have, by Lemma 7.1,

$$\begin{aligned} \|\hat{f}\| &= \sup\{|\hat{f}(m)| : m \in \mathcal{M}_d^0(X);\ \|m\|_W \leq 1\} \\ &\geq \sup\{|\hat{f}(\varepsilon_{xy})|/d(x, y) : x, y \in X;\ x \neq y\} \\ &= \sup\{(f(x) - f(y))/d(x, y) : x, y \in X;\ x \neq y\} = \|f\|_L. \end{aligned}$$

Hence in fact $\|\hat{f}\| = \|f\|_L$. $\square$

Now, let $\phi \in \mathcal{M}_d^0(X)^*$. Fix some $a \in X$ and let $u(x) = \phi(\varepsilon_{xa})$ for all x. Then, for all x, y, we have $\phi(\varepsilon_{xy}) = \phi(\varepsilon_{xa} - \varepsilon_{ya}) = u(x) - u(y)$. Therefore

$$|u(x) - u(y)| = |\phi(\varepsilon_{xy})| \leq \|\phi\|\, \|\varepsilon_{xy}\|_W \leq \|\phi\| d(x, y).$$

This shows that $u \in \mathrm{Lip}(X, d)$. Moreover, $\hat{u}(\varepsilon_{xy}) = \int_X u(s)\varepsilon_{xy}(\mathrm{d}s) = u(x) - u(y)$. Therefore, we see that $\hat{u}$ and ϕ are continuous linear functionals on $\mathcal{M}_d^0(X)$ that agree on $\mathcal{M}_s^0(X)$. So, by Theorem 6.3, we have $\phi = \hat{u}$. We have proved the following theorem.

Theorem 7.3. *The map $f \mapsto \hat{f}$ is a linear surjection of $\mathrm{Lip}(X, d)$ onto the Banach dual of $(\mathcal{M}_d^0(X), \|\cdot\|_W)$; the kernel of the map $f \mapsto \hat{f}$ is the set of all constant functions on X. Moreover $\|\hat{f}\| = \|f\|_L$ for all $f \in \mathrm{Lip}(X, d)$. Thus, the dual of $(\mathcal{M}_d^0(X), \|\cdot\|_W)$ is $\mathrm{Lip}(X, d)$ modulo the constant functions.*

Corollary 7.4. *Suppose that $m \in \mathcal{M}_d^0(X)$. Then there exists $f \in \mathrm{Lip}(X, d)$ with $\|f\|_L = 1$ such that $\hat{f}(m) = \|m\|_W$.*

Proof. This follows from Theorem 7.3, by the Hahn–Banach theorem. □

We can also obtain as a corollary the Kirszbraun–McShane–Whitney theorem:

Corollary 7.5. *Suppose that $\emptyset \neq Y \subseteq X$, let $L = \mathrm{Lip}(X, d)$ and $L' = \mathrm{Lip}(Y, d)$, and let $f \in L'$. Then f can be extended to an element F of L such that $\|F\|_L = \|f\|_{L'}$.*

Proof. By uniform continuity, we can suppose that f has been extended to $\overline{Y}$. This means that we can assume that Y is a closed subset of X. That in turn means that we can regard $\mathcal{M}_d^0(Y)$ as a vector subspace of $\mathcal{M}_d^0(X)$. Let $\bar{f}(m) = \int_Y f \, \mathrm{d}m$ for all $m \in \mathcal{M}_d^0(Y)$. Then $\bar{f}$ is a continuous linear functional on $\mathcal{M}_d^0(Y)$ with norm $\|f\|_{L'}$. By the Hahn–Banach theorem, $\bar{f}$ can be extended to a continuous linear functional $\phi : \mathcal{M}_d^0(X) \to \mathbb{R}$ without change of norm. But now, by Theorem 7.3, we have $\phi = \widehat{G}$ for some $G \in L$ and

$$\|G\|_L = \|\widehat{G}\| = \|\bar{f}\| = \|f\|_{L'}.$$

Since $\widehat{G}$ and $\bar{f}$ agree on $\mathcal{M}_d^0(Y)$, we must have $G(y) - f(y) = c$ for some constant c and all $y \in Y$. Let $F = G - c$. Then $F \in L$, F extends f, and $\|F\|_L = \|G\|_L = \|f\|_{L'}$. □

8. Optimal measures and functions

Choose and fix $m \in \mathcal{M}_d^0(X)$. By an *optimal measure* on Z for m, we shall understand a measure $\rho \in \Gamma(m)$ such that $\rho(d) = \inf\{\gamma(d) : \gamma \in \Gamma(m)\}$. The existence of such a ρ follows from Theorem 4.5, together with the fact that $W(m^+, m^-) = \min\{\pi(d) : \pi \in \Pi(m^+, m^-)\}$. By an *optimal function* for m, we shall mean a function $g \in \mathrm{Lip}_1(X, d)$ such that $m(g) = \sup\{m(l) : l \in \mathrm{Lip}_1(X, d)\}$. The existence of such a function follows from Corollary 7.4, together with Theorem 4.5.

Theorem 8.1. *Let $m \in \mathcal{M}_d^0(X)$, $\gamma \in \Gamma(m)$, and $g \in \mathrm{Lip}_1(X, d)$. Then the following statements are equivalent:*

(i) *γ is an optimal measure, and g an optimal function, for m;*
(ii) *$g(x) - g(y) = d(x, y)$ for all $(x, y) \in \mathrm{supp}\,\gamma$.*

Proof. Suppose that γ and g are both optimal for m. Then, writing $\gamma_1 = p_1(\gamma)$ and $\gamma_2 = p_2(\gamma)$, we have

$$\begin{aligned}\|m\|_W &= \gamma(d) \geq \int_Z (g(x) - g(y))\gamma(\mathrm{d}x\mathrm{d}y) \\ &= \gamma_1(g) - \gamma_2(g) = m(g) = \hat{g}(m) = \|m\|_W.\end{aligned}$$

Hence we have equality throughout, and so

$$\int_Z (d(x, y) - g(x) + g(y))\gamma(\mathrm{d}x\mathrm{d}y) = 0,$$

with the integrand non-negative, and so statement (ii) follows.

Suppose, conversely, that we are given that γ and g satisfy statement (ii). Then

$$\begin{aligned}\|m\|_W &\leq \gamma(d) = \int_Z (g(x) - g(y))\gamma(\mathrm{d}x\mathrm{d}y) \\ &= \gamma_1(g) - \gamma_2(g) = m(g) = \hat{g}(m) \leq \|m\|_W.\end{aligned}$$

We again must have equality throughout, and statement (i) follows. □

9. Concluding remarks

Variants of Theorem 4.1 are proved by Fernique [9] and Bogachev [3] by approximation from the case where the initial measures have finite supports. See also [17,18]. For Polish spaces, a radically different treatment of Monge–Kantorovich duality has been given by Villani [20], which yields formula (4.1) as an immediate consequence—see [20, Remark 6.5]. A second generalization by Kellerer of the Kantorovich–Rubinstein theorem to arbitrary metric spaces is given by [14, Theorem 2], where a formulation for τ-additive measures is obtained. This, however, lies outside the scope of the present article.

We obtained Dobrushin's formula (5.3) above as a corollary of a result for the indicator metric in the style of the Kantorovich–Rubinstein theorem, namely Theorem 5.2. On the other hand, Villani [19], using Monge–Kantorovich duality, offers for Polish spaces a Kantorovich–Rubinstein formula for *arbitrary* lower semicontinuous metrics, noting that the indicator metric is just a special case. Unfortunately, his argument appears to be incomplete, since the measurability of certain functions in play is not established.

The Kantorovich–Rubinstein theorem has given rise to an extensive literature, and only the most elementary treatment has been attempted in the present article. A survey of many variants and generalizations of the theory may be found in [15, Chapter 4]. Villani [20, Chapter 6] treats the Wasserstein metrics W_p ($1 \leq p < \infty$) and their applications at some length, and also provides a very valuable guide to the literature.

References

[1] A. de Acosta, Invariance principles in probability for triangular arrays of B-valued random vectors and some applications, Annals of Probability 10 (2) (1982) 346–373.
[2] C. Berg, J.P.R. Christensen, P. Ressel, Harmonic Analysis on Semigroups. Theory of Positive Definite and Related Functions, in: Graduate Texts in Mathematics, vol. 100, Springer, Berlin, 1984.
[3] V.I. Bogachev, Measure Theory, vol. II, Springer, Berlin, 2007.
[4] N. Bourbaki, Elements of Mathematics: Integration, vol. II, Springer, Berlin, 2004.
[5] R.L. Dobrushin, Prescribing a system of random variables by conditional distributions, Theory of Probability and its Applications 15 (1970) 458–486.
[6] R.M. Dudley, Probabilities and Metrics, in: Lecture Notes Series, vol. 45, Matematisk Institut, Aarhus University, Aarhus, 1976.
[7] R.M. Dudley, Real Analysis and Probability, Wadsworth and Brook Cole, Pacific Grove, CA, 1989.
[8] D.A. Edwards, A simple proof in Monge–Kantorovich duality theory, Studia Mathematica 200 (1) (2010) 67–77.
[9] X. Fernique, Sur le théorème de Kantorovitch–Rubinstein dans les espces polonais, in: Seminar on Probability XV, Strasbourg 1979/1980, in: Lecture Notes in Mathematics, vol. 850, Springer, Berlin, 1981, pp. 6–10.
[10] D.H. Fremlin, Measure Theory, vol. IV, Torres Fremlin, Colchester, 2003.
[11] L.V. Kantorovich, G.P. Akilov, Functional Analysis, 2nd ed., Pergamon Press, Oxford, 1982.
[12] L.V. Kantorovich, G. Rubinstein, On a space of completely additive functions, Vestnik Leningradskogo Universiteta 13 (7) (1958) 52–59.
[13] H.G. Kellerer, Duality theorems for marginal problems, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 67 (1984) 399–432.
[14] H.G. Kellerer, Duality theorems and probability metrics, in: Proceedings of the Seventh Conference on Probability Theory, in: BraŞov. 1982, VNU Press, Utrecht, 1985, pp. 211–220.
[15] S.T. Rachev, L. Rüschendorf, Mass Transportation Problems, vol. I, Springer, Berlin, 1998.
[16] S.T. Rachev, R.M. Shortt, Duality theorems for Kantorovich–Rubinstein and Wasserstein functionals, Dissertationes Mathematicae 299 (1990) 25 pp.
[17] A. Szulga, On the Wasserstein metric, Prague, 1978, in: Transactions of the Eighth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, vol. B, Reidel, Dordrecht, Boston, Mass, 1978, pp. 267–273.
[18] A. Szulga, On minimal metrics in the space of random variables, Teoriya Veroyatnosteĭ i ee Primeneniya 27 (2) (1982) 401–405.
[19] C. Villani, Topics in Optimal Transportation, in: Graduate Studies in Mathematics, vol. 58, American Math. Soc., Providence, RI, 2003.
[20] C. Villani, Optimal Transport, Old and New, in: Grundlehren der Mathematischen Wissenschaften, vol. 338, Springer, Berlin, 2009.