ELSEVIER Journal of Computational and Applied Mathematics 57 (1995) 345–361

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

The Kantorovich metric for probability measures on the circle

Carlos A. Cabrelli,*[,1] Ursula M. Molter[1]

Department of Pure Mathematics, University of Waterloo, Ont., Canada

Dedicated with affection and admiration to Prof. Luis A. Santaló

Received 11 November 1992

Abstract

In this paper we show that there exists an analytic expression for the Kantorovich distance between probability measures on the circle. Previously such an expression was only known for measures supported on the real line. In the case that the measures are discrete, this formula enables us to show that the Kantorovich distance can be computed in linear time. This is important for applications, in particular in pattern recognition where this distance is used for texture analysis. As another application we see that the analytic expression found allows us to solve a Minimal Matching Problem in linear time, for which so far only $n \log n$ algorithms were known.

Keywords: Probability metrics; Kantorovich metric; Texture metrics; Matching problems; Hutchinson distance

1. Introduction

Probability metrics have been studied extensively because of its importance in theory as well as in applications. Several metrics have been considered (see, for example, [26, 15, 14] and references therein).

One particular important metric is the Kantorovich metric:

$$d_Y(\mu, \nu) = \sup\left\{\int_Y f\,\mathrm{d}(\mu - \nu) \colon f\colon Y \to \mathbb{R},\ |f(x) - f(y)| \leqslant d(x,y)\ \forall x, y \in Y\right\},$$

where μ and ν are probability measures defined on the Borel sets of a suitable metric space (Y, d). When Y is compact, d_Y metrizes the weak $*$-topology of $M(Y)$. In the discrete case, when the measures are supported on a finite number of points, say $y_1, \ldots, y_n \in Y$, and $\mu = \sum_{i=1}^n \mu_i \delta_{y_i}$,

*Corresponding author. E-mail: ccabrell@mate.dm.uba.ar.

[1]Present address: Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria (Pabellón I), 1428 Capital Federal, Argentina.

SSDI 0377-0427(93)E0213-6

$\nu = \sum_{i=1}^{n} \nu_i \delta_{y_i}$, we have

$$d_Y(\mu, \nu) = \sup \left\{ \sum_{i=1}^{n} f(y_i)(\mu_i - \nu_i) \text{ with } |f(y_i) - f(y_j)| \leqslant d(y_i, y_j),\ 1 \leqslant i,j \leqslant n \right\}.$$

In this paper, we are going to consider the Kantorovich metric in the case when the underlying metric space Y is the unit circle T. The main result is Theorem 3.7 in which we show that an analytical expression for this distance can be obtained:

$$d_T(\mu, \nu) = \int_T |\alpha(t) - a_\alpha| \, dt,$$

where $\alpha(t) = \mu([0,t]) - \nu([0,t])$, $t \in T$ and a_α is a translation constant that depends on μ and ν, where μ and ν are measures supported on T.

We show that in the discrete case (i.e., μ and ν are supported on a finite set of points in T) this expression allows us to reduce the problem of calculating the Kantorovich distance between probability vectors (the Kantorovich Metric Problem, or KMP) to the weighted median problem for which several linear time algorithms exist [5, 10, 17, 19, 8]. As a consequence the KMP on the circle can be solved in linear time. We also show that the KMP on the circle generalizes the bipartite graph matching problem considered in [24], where an $n \log n$ algorithm to perform the minimal matching is derived. Using the KMP approach, we are able to obtain a linear-time algorithm.

The Kantorovich metric arises in very different contexts and under different names. In statistical applications it was known as the Wasserstein distance [23, 22, 18] and more recently it appeared with the development of fractal geometry and its applications to computer graphics under the name of Hutchinson distance [21, 6, 2–4] after being introduced in [9]. It also has proven to be very useful in Digital Image Processing. In particular, in some applications of texture analysis in pattern recognition Shen and Wong used this metric to enhance feature dissimilarity [20] (see also [12, 25]).

Furthermore, this metric admits a linear programming problem representation. Its dual problem is the well-known Kantorovich–Monge mass transfer problem [11, 13, 1, 16]. This duality has been established in 1958 by Kantorovich and Rubinstein (see, for example, [7]). A lot of attention has been given to this problem as well as to many of its generalizations since then. A complete account of these results can be found in the work of Rachev [14].

The outline of the paper is as follows: In Section 2 we study some properties of the Kantorovich metric on the line which we will need later. In Section 3 we analyze the case of the circle: First to each measure on the circle we associate a family of measures on the line. Using this association we relate the distance between measures on the circle to the distance between associated measures on the line. This relation then allows us to derive an explicit formula for the Kantorovich metric for measures on the circle (Theorem 3.7). Finally, in Section 4 we consider the discrete case and show an application to a Minimal Matching Problem.

2. Kantorovich metric on the line

In this section we study some properties of the Kantorovich metric for measures supported on the $[0, 1]$ interval. For this we first introduce some notation, then we define the Kantorovich metric on a metric space and finally we consider the particular case of the [0, 1] interval. We will need these properties to derive the results for measures on the circle in Section 3.

2.1. Notation

Let (Y, d) be a complete separable metric space. We use the notation $L(Y) = \{f: Y \to \mathbb{R}: |f(y_1) - f(y_2)| \leqslant d(y_1, y_2)$ for all $y_1, y_2 \in Y\}$, for the Lipschitz functions on Y with constant 1, and $M(Y)$ for the space of Borel probability measures on Y having bounded support.

We will denote by λ the Lebesgue measure on $\mathbb{R}$.

We identify the circle $K = \{z \in \mathbb{C}: |z| = 1\}$ with $T = [0, 1)$ as a fundamental domain for $\mathbb{R}/\mathbb{Z}$, via the transformation $t \to e^{i2\pi t}$. A natural metric on T is $\rho(s,t) = \min(|s-t|, 1-|s-t|)$, which corresponds to the minimum arclength on the circle. The functions on T can be identified with the periodic functions on $\mathbb{R}$ of period 1; using this identification we will use $f(t)$, with $t \in \mathbb{R}$, for functions on T.

$(X, \cdot)$ will represent the unit interval on the line with the Euclidean distance.

For $\mu \in M(X)$, the distribution function of μ is the function F_μ: $X \to \mathbb{R}$, $F_\mu(x) = \mu([0, x])$.

2.2. The Kantorovich metric on $M(Y)$

If (Y, d) is a complete separable metric space, then the Kantorovich metric d_Y on $M(Y)$ is defined by

$$d_Y(\mu, \nu) = \sup\left\{\int_Y f(y)\,\mathrm{d}\mu(y) - \int_Y f(y)\,\mathrm{d}\nu(y):\ f \in L(Y)\right\}.$$

If we call

$$\langle \mu, f \rangle_Y = \int_Y f(y)\,\mathrm{d}\mu(y) \quad \text{and} \quad \eta = \mu - \nu,$$

then

$$\langle \eta, f \rangle_Y = \langle \mu, f \rangle_Y - \langle \nu, f \rangle_Y$$

and

$$d_Y(\mu, \nu) = \sup\{\langle \eta, f \rangle_Y, f \in L(Y)\}. \tag{2.1}$$

To see that d_Y is in fact a metric on $M(Y)$ see, for example, [9]. (The subscript Y will be omitted, where unnecessary.)

2.3. Properties of the Kantorovich metric on X

Let $\mu, \nu \in M(X)$, see $\eta = \mu - \nu$ and let $h: X \to \mathbb{R}$ be the function defined by $h(x) = F_\mu(x) - F_\nu(x)$ (h represents the mass distribution of η). We have the following theorem.

Theorem 2.1 (Vallender [23] and Barahona et al. [2]). *With the above notation we have*

$$d_X(\mu,\nu) = \sup\{\langle \eta, f\rangle_X, f \in L(X), f(0) = 0\}, \tag{2.2}$$

$$\langle \eta, f\rangle = -\int_X f'(x)h(x)\,\mathrm{d}x \quad \forall f \in L(X), \tag{2.3}$$

$$d_X(\mu,\nu) = \int_0^1 |h(x)|\,\mathrm{d}x. \tag{2.4}$$

Proof. Since $\eta(X) = 0$ we have $\langle \eta, f + a\rangle = \langle \eta, f\rangle$ $\forall a \in \mathbb{R}$, and then (2.2) follows.

For (2.3) let us consider the set $\mathscr{B} = \{(x,y) \in X \times X : x < y\}$ with the product measure $\varrho = \lambda \otimes \eta$.

Now, if $f \in L(X)$ then f is continuous on X and differentiable a.e. On the other hand, if we consider the function

$$g(x) = \begin{cases} f'(x) & \text{if it exists,} \\ 0 & \text{elsewhere,} \end{cases}$$

then g is λ-integrable on X, and $k(x,y) = g(x)$ is ϱ-integrable on $\mathscr{B}$. Now integrating k over $\mathscr{B}$ and applying Fubini's theorem, we obtain

$$\int_X f\,\mathrm{d}\eta = f(1)\eta(X) - \int_X f'(x)h(x)\,\mathrm{d}x.$$

Since $\eta(X) = 0$

$$\int_X f\,\mathrm{d}\eta = -\int_X f'(x)h(x)\,\mathrm{d}x.$$

In order to prove (2.4) we observe that

$$d_X(\mu,\nu) = \sup_{f \in L(X)} \int_X f\,\mathrm{d}\eta \leqslant \sup_{f \in L(X)} \left|\int_X f\,\mathrm{d}\eta\right| \leqslant \sup_{f \in L(X)} \int_X |f'(x)|\,|h(x)|\,\mathrm{d}x.$$

But since $|f'(x)| \leqslant 1$ a.e., $d_X(\mu,\nu)$ is therefore bounded as

$$d_X(\mu,\nu) \leqslant \int_X |h(x)|\,\mathrm{d}x.$$

To complete the proof of (2.4), we now exhibit a function $f^* \in L(X)$ such that $d_X(\mu,\nu) = \int_X f^*\,\mathrm{d}\eta$. To construct f^* consider first

$$g(x) = \begin{cases} -1 & \text{if } h(x) > 0, \\ 1 & \text{if } h(x) < 0, \\ 0 & \text{if } h(x) = 0. \end{cases}$$

Then $|g|$ is integrable on X. Now we define f^* as

$$f^*(x) = \int_0^x g(t)\,\mathrm{d}t.$$

Now if $0 \leqslant x \leqslant y \leqslant 1$ (the other case is symmetric) then

$$|f^*(x) - f^*(y)| = \left| \int_x^y g(t)\,\mathrm{d}t \right| \leqslant \int_x^y |g(t)|\,\mathrm{d}t \leqslant \int_x^y \mathrm{d}t = y - x = |y - x|$$

and therefore f^* is in $L(X)$. Now, by (2.3)

$$\int_X f\,\mathrm{d}\eta = -\int_X f'(x)h(x)\,\mathrm{d}x = \int_X |h(x)|\,\mathrm{d}x. \qquad \square$$

Let us call

$$L_0(X) = \{f \in L(X) \colon f(0) = 0\}. \tag{2.5}$$

In the proof of Eq. (2.4) of the theorem, we exhibited a function $f^* \in L_0$ such that $d_X(\mu, \nu) = \langle \eta, f^* \rangle$. We are interested now to find all the functions in L_0 that satisfy that equation.

Therefore let us introduce some notation: If $\gamma \colon X \to \mathbb{R}$ is an arbitrary measurable function, we will call

$$A^+(\gamma) = \{x \in X : \gamma(x) > 0\}, \qquad A^-(\gamma) = \{x \in X \colon \gamma(x) < 0\},$$

$$A^0(\gamma) = \{x \in X \colon \gamma(x) = 0\}, \qquad D(\gamma) = \{c \colon A^0(\gamma) \to \mathbb{R}, \text{ measurable: } |c(t)| \leqslant 1 \text{ a.e.}\}.$$

Corollary 2.2. *Using the preceding notation, let μ and ν be two measures in $M(X)$, and $h\colon X \to \mathbb{R}$ as before, then $f \in L_0(X)$ is optimal (i.e., $d_X(\mu, \nu) = \langle \mu - \nu, f \rangle_X$), if and only if there exists $c \in D(h)$*

$$f'(y) = \begin{cases} -1 & \lambda\text{-a.e. in } A^+(h), \\ 1 & \lambda\text{-a.e. in } A^-(h), \\ c(y) & \text{otherwise.} \end{cases} \tag{2.6}$$

Proof. *Sufficiency*: From (2.3) we have that $\langle \eta, f \rangle \leqslant \int_X |h(x)|\,\mathrm{d}x$, $\forall f \in L_0(X)$. Using that f satisfies (2.6), then the equality holds, and therefore f is optimal.

Necessity: Let us now assume that $g \in L_0(X)$ is optimal, but g does not satisfy the condition given in (2.6). Hence, there exists a set B of positive Lebesgue measure, such that either $B \subset A^+(h)$ and $g'(x) \neq -1$ $\forall x \in B$, or $B \subset A^-(h)$ and $g'(x) \neq 1$ $\forall x \in B$. Since both cases are completely symmetric, we assume $B \subset A^+(h)$.

Let us now define $r \in L_0(X)$ such that

$$r(x) = \int_0^x p(t)\,\mathrm{d}t, \quad \text{where } p(t) = \begin{cases} g'(t) & t \in X \backslash B, \\ -1 & t \in B. \end{cases}$$

Now using (2.3)

$$\begin{aligned}\langle \eta, g\rangle &= -\int_X g'(x)h(x)\,\mathrm{d}x \\ &= -\int_{X\backslash B} g'(x)h(x)\,\mathrm{d}x - \int_B g'(x)h(x)\,\mathrm{d}x \\ &< -\int_{X\backslash B} g'(x)h(x)\,\mathrm{d}x - \int_B (-1)h(x)\,\mathrm{d}x \\ &= \langle \eta, r\rangle_X \leqslant d_X(\mu, \nu)\end{aligned}$$

and hence g cannot be optimal. □

Note that in the case when the Lebesgue measure of $A^0(h)$ is equal to 0, there exists only one optimal solution.

2.4. Balanced mass distributions

The concept of *balanced* mass distributions will be essential to establish the link between the Kantorovich metric on the circle and on the line.

Using the previous notation, a function $\gamma\colon X \to \mathbb{R}$ is said to be *balanced* if

$$|\lambda(A^+(\gamma)) - \lambda(A^-(\gamma))| \leqslant \lambda(A^0(\gamma)). \tag{2.7}$$

The next lemma will show that, for a mass distribution α, being balanced is equivalent to the existence of optimal Lipschitz functions which take the same value on the endpoints (i.e., Lipschitz functions on T).

Lemma 2.3. *With the above notation, if $h = F_\mu - F_\nu$, for $\mu, \nu \in M(X)$, then h is balanced if and only if there exists $f^* \in L_0(X)$, optimal for the problem $\sup_{f\in L_0}\langle \eta, f\rangle$ such that $f^*(0) = f^*(1) = 0$.*

Proof. We first observe that if f^* is optimal, then by Corollary 2.2

$$f^*(1) = \int_0^1 (f^*)'(t)\,\mathrm{d}t = \lambda(A^-(h)) - \lambda(A^+(h)) + \int_{A^0(h)} c(t)\,\mathrm{d}t \tag{2.8}$$

for some $c \in D(h)$. Now, if h is balanced, from (2.8) it is clear that in order to obtain an optimal solution f^* satisfying $f^*(1) = 0$ it is enough to choose c to be a constant function. This constant is chosen to be $(\lambda(A^+(h)) - \lambda(A^-(h)))/\lambda(A^0(h))$ if $\lambda(A^0(h)) \neq 0$, and 0 if $\lambda(A^0(h)) = 0$.

For the converse we note that since $f^*(1) = 0$, from (2.8) we have

$$\lambda(A^+(h)) - \lambda(A^-(h)) = \int_{A^0(h)} c(t)\,\mathrm{d}t$$

and then, since $|c(t)| \leqslant 1$

$$|\lambda(A^+(h)) - \lambda(A^-(h))| \leqslant \int_{A^0(h)} |c(t)|\,dt \leqslant \lambda(A^0(h)). \qquad \square$$

3. The Kantorovich metric on the circle

In this section we are going to relate the distance between measures on the circle with the distance between measures on X obtained from the former by "cutting" the circle.

Given two measures on T, each "cut" of the circle determines two measures on X associated to that "cut". In Lemma 3.1 we show that the value of the distance between two measures on T is the infimum over all possible cuts of the values of the distances between the associated measures on X. Proposition 3.2 establishes that this infimum is attained if and only if there exists a "cut" such that the mass distribtion of the associated measures is balanced.

Proposition 3.6 shows that indeed such a "cut" always exists, and finally Theorem 3.7 provides an expression for $d_T(\mu, \nu)$.

3.1. Identifications of measures on T with measures on X

If $\mu \in M(T)$, we define a function $G_\mu \colon \mathbb{R} \to \mathbb{R}$ as

$$G_\mu(x) = \mu([0,x]), \quad 0 \leqslant x < 1$$

and we extend the definition to $\mathbb{R}$ by the equation

$$G_\mu(x+1) = G_\mu(x) + 1. \tag{3.1}$$

Let us now call

$$M_D = \{\mu \in M(X) \colon \mu(\{0\}) = 0\} \tag{3.2}$$

and

$$M_I = \{\mu \in M(X) \colon \mu(\{1\}) = 0\}. \tag{3.3}$$

For each $s \in T$, we define natural identifications between $M(T)$ and these two sets in the following way: Let $s \in T$ be fixed. To every measure $\mu \in M(T)$ we associate the measures $\mu_s^D \in M_D$ and $\mu_s^I \in M_I$ determined by the distribution functions given by

$$\begin{aligned} &D_\mu^s \colon X \to \mathbb{R}, \qquad D_\mu^s(x) = G_\mu(x+s) - G_\mu(s), \\ &I_\mu^s \colon X \to \mathbb{R}, \qquad I_\mu^s(x) = G_\mu(x+s) - G_\mu(s-). \end{aligned} \tag{3.4}$$

Note that if $\mu(\{s\}) = 0$, then G_μ is continuous at s, and then $\mu_s^D = \mu_s^I$ and $D_\mu^s = I_\mu^s$.

Now we define a 1–1 correspondence between $L(T)$ and the subset $L_T \subset L(X)$ defined by

$$L_T = \{f \in L(X): f(0) = f(1)\}$$

in the following way: To every $f \in L(T)$ we associate the function $f_s \in L_T$ such that

$$f_s(x) = f(x+s) \quad \forall x \in X.$$

Note that the correspondences:

$$\begin{aligned} &M(T) \to M_D \subset M(X) \\ &\mu \mapsto \mu_s^D \\ &M(T) \to M_I \subset M(X) \\ &\mu \mapsto \mu_s^I \\ &L(T) \to L_T \subset L(X) \\ &f \mapsto f_s \end{aligned} \tag{3.5}$$

are 1–1, and different for each $s \in T$.

Informally speaking, μ_s^D and μ_s^I represent the measures on the line obtained by "cutting" the circle at s, and taking as fundamental domains the intervals $(s, s+1]$ and $[s, s+1)$ in $\mathbb{R}/\mathbb{Z}$, respectively.

We will refer to each $s \in T$ as a cut of the circle and to μ_s^D, μ_s^I and D_μ^s, I_μ^s as the right and left probability measures and distribution functions obtained by cutting the circle at s.

Note that with the preceding identifications, Lemma 2.3 states that if h is the difference of the distribution functions of two measures on $[0,1]$, then h is balanced if and only if the function f^* that realizes the Hutchinson distance between μ and ν satisfies $f^* \in L_T \cap L_0(X)$.

3.2. *Balanced distributions and measures on T*

It is easy to see that the Kantorovich metric on T is invariant under rotations, i.e.,

$$d_T(\mu, \nu) = d_T(\mu_t, \nu_t) \quad \text{for every } t \in T,\ \mu, \nu \in M(T),$$

where

$$\mu_t(A) = \mu(A - t) \quad \text{for any measurable set } A \subset T.$$

Therefore, in what follows we will consider without any loss of generality, the Kantorovich metric between measures $\mu, \nu \in M(T)$ such that $\mu(\{0\}) = \nu(\{0\}) = 0$, since we can always find a convenient rotation such that μ_t and ν_t satisfy the desired property.

Let then $\mu, \nu \in M(T)$, $\eta = \mu - \nu$, G_μ, G_ν as defined in (3.1) and $G(x) = G_\mu(x) - G_\nu(x)$. Note that G is a periodic function of period 1 with $G(0) = G(1) = 0$. If $s \in T$, let D_μ^s, D_ν^s, I_μ^s, I_ν^s be the distribution functions defined in (3.4) and μ_s^D, ν_s^D, μ_s^I, ν_s^I its associated measures on X.

We will first prove the following lemma.

Lemma 3.1. *Let* $\mu, \nu \in M(T)$, $\eta = \mu - \nu$, *then*

$$d_T(\mu,\nu) \leqslant \inf_{s\in T} d_X(\mu_s^D, \nu_s^D), \tag{3.6}$$

$$d_T(\mu,\nu) \leqslant \inf_{s\in T} d_X(\mu_s^I, \nu_s^I). \tag{3.7}$$

Proof. Using the identification given in (3.5) it is clear that for any fixed $s \in T$,

$$\langle \mu - \nu, f\rangle_T = \langle \mu_s^D - \nu_s^D, f_s\rangle_X \quad \forall f \in L(T).$$

Then

$$\begin{aligned} d_T(\mu,\nu) &\leqslant \sup_{f\in L(T)} \langle \mu - \nu, f\rangle_T \\ &= \sup_{f\in L(T)} \langle \mu_s^D - \nu_s^D, f_s\rangle_X \\ &\leqslant \sup_{f\in L(X)} \langle \mu_s^D - \nu_s^D, f\rangle_X \\ &= d_X(\mu_s^D, \nu_s^D) \quad \forall s \in T \end{aligned}$$

and then

$$d_T(\mu,\nu) \leqslant \inf_{s\in T} d_X(\mu_s^D, \nu_s^D).$$

Analogously

$$d_T(\mu,\nu) \leqslant \inf_{s\in T} d_X(\mu_s^I, \nu_s^I). \qquad \square$$

Proposition 3.2. *If for* $s \in T$, $D^s = D_\mu^s - D_\nu^s$ *is balanced, then*

$$d_T(\mu,\nu) = d_X(\mu_s^D, \nu_s^D) = \inf_{r\in T} d_X(\mu_r^D, \nu_r^D). \tag{3.8}$$

Analogously, if for $s \in T$, $I^s = I_\mu^s - I_\nu^s$ *is balanced we have*

$$d_T(\mu,\nu) = d_X(\mu_s^I, \nu_s^I) = \inf_{r\in T} d_X(\mu_r^I, \nu_r^I). \tag{3.9}$$

Proof. Let $s \in T$ be such that D^s is balanced, then from Lemma 2.3 we know that there exists $f \in L_T$ such that it is optimal, i.e.,

$$\langle \eta, f\rangle = d_X(\mu_s^D, \nu_s^D), \tag{3.10}$$

and since $f(0) = f(1)$,

$$\langle \eta, f\rangle \leqslant d_T(\mu,\nu). \tag{3.11}$$

That means that

$$d_X(\mu_s^D, \nu_s^D) \leqslant d_T(\mu, \nu)$$

and using (3.6) we have that

$$\inf_{r \in T} d_X(\mu_r^D, \nu_r^D) = d_X(\mu_s^D, \nu_s^D) = d_T(\mu, \nu).$$

The proof for I^s is analogous. □

3.3. Existence of an optimal cut

In this section we will prove the existence of $s \in T$, such that either D^s or I^s is balanced. For the proof we need some additional properties of right-continuous functions.

Let $\gamma: X \to \mathbb{R}$ be a right-continuous function. We then define $m_\gamma: \mathbb{R} \to [0,1]$, by

$$m_\gamma(t) = \lambda(\{x \in X: \gamma(x) \geqslant t\}), \tag{3.12}$$

and let

$$\begin{aligned} a_\gamma &= \sup\{t: m_\gamma(t) > \tfrac{1}{2}\}, \\ b_\gamma &= \inf\{t: m_\gamma(t) < \tfrac{1}{2}\}. \end{aligned} \tag{3.13}$$

The following properties are immediate:

(1) m_γ is a left-continuous, nonincreasing function.
(2) $m_\gamma(t) - m_\gamma(t+) = \lambda(\{x: \gamma(x) = t\})$.
(3) $a_\gamma \leqslant b_\gamma$.
(4) If $a_\gamma < b_\gamma$, then $\forall t \in (a, b]$, $m_\gamma(t) = \frac{1}{2}$ and if $m_\gamma(t) = \frac{1}{2} \Rightarrow t \in [a_\gamma, b_\gamma]$.
(5) $m_\gamma(a_\gamma) \geqslant \frac{1}{2}$.

We also have the following lemma.

Lemma 3.3. *For any neighborhood V_{a_γ} of a_γ, $\lambda(V_{a_\gamma} \cap \gamma(X)) > 0$, in particular, a_γ is in the closure of $\gamma(X)$; i.e., there exists a sequence $\{x_n\} \subset X$ such that $\gamma(x_n) \to a_\gamma$. The same holds true for b_γ.*

Proof. Consider $D_n = \{x \in X: |\gamma(x) - a_\gamma| < 1/2n\}$. We will show that $\lambda(D_n) > 0$ $\forall n \in \mathbb{N}$. For this assume that $\exists n \in \mathbb{N}$, such that $\lambda(D_n) = 0$. Then $m_\gamma(a_\gamma + 1/2n) = m_\gamma(a_\gamma - 1/2n) > \frac{1}{2}$ and this contradicts the fact that a_γ is a supremum. (The proof for b_γ is analogous.) □

The next lemma shows that a right-continuous function can always be transformed into a balanced function by simply adding a constant.

Lemma 3.4. *Let $\gamma: X \to \mathbb{R}$ be a right-continuous bounded function and let a_γ and b_γ defined as in (3.13). Then $\gamma - r$ is balanced if and only if $r \in [a_\gamma, b_\gamma]$.*

Proof. Let $r \in [a_\gamma, b_\gamma]$. We first notice that if $x, y, z \in [0,1]$, and $x + y + z = 1$, then

$$|x - y| \leqslant z \;\Leftrightarrow\; \begin{cases} 0 \leqslant x \leqslant \frac{1}{2}, \\ 0 \leqslant y \leqslant \frac{1}{2}. \end{cases} \tag{3.14}$$

Therefore, if we call $\delta = \gamma - a_\gamma$, in order to prove that δ is balanced, it is enough to see that $\lambda(A^+(\delta)) \leqslant \frac{1}{2}$ and $\lambda(A^-(\delta)) \leqslant \frac{1}{2}$.

For this we consider

$$A^+(\delta) = \{x\colon \gamma(x) > a_\gamma\} \tag{3.15}$$

$$= \bigcup_{n\in\mathbb{N}} \left\{x\colon \gamma(x) > a_\gamma + \frac{1}{n}\right\} \tag{3.16}$$

$$= \bigcup_{n\in\mathbb{N}} A^+\left(\delta - \frac{1}{n}\right). \tag{3.17}$$

Therefore,

$$\lambda(A^+(\delta)) = \lim_{n\to\infty} \lambda\left(A^+\left(\delta - \frac{1}{n}\right)\right).$$

But

$$\lambda\left(A^+\left(\delta - \frac{1}{n}\right)\right) \leqslant m_\gamma\left(a_\gamma + \frac{1}{n+1}\right) \leqslant \tfrac{1}{2}.$$

Then

$$\lambda(A^+(\delta)) \leqslant \tfrac{1}{2}.$$

In the same way

$$A^-(\delta) = \{x\colon \gamma(x) < a_\gamma\} \tag{3.18}$$

$$= \bigcup_{n\in\mathbb{N}} \left\{x\colon \gamma(x) < a_\gamma - \frac{1}{n}\right\} \tag{3.19}$$

$$= \bigcup_{n\in\mathbb{N}} A^-\left(\delta + \frac{1}{n}\right). \tag{3.20}$$

Therefore,

$$\lambda(A^-(\delta)) = \lim_{n\to\infty} \lambda\left(A^-\left(\delta + \frac{1}{n}\right)\right).$$

But

$$\lambda\left(A^-\left(\delta + \frac{1}{n}\right)\right) = 1 - m_\gamma\left(a_\gamma - \frac{1}{n+1}\right) < \tfrac{1}{2}.$$

Then

$$\lambda(A^-(\delta)) \leqslant \tfrac{1}{2}.$$

This shows that $\gamma - a_\gamma$ is balanced.

Let us now take an arbitrary r, $a_\gamma \leqslant r \leqslant b_\gamma$. Because of the properties of m_γ stated in Section 3.3 we have that $\forall t \in (a_\gamma, b_\gamma], m_\gamma(t) = \frac{1}{2}$. Then $m_\gamma(a_\gamma +) - m_\gamma(b_\gamma) = 0$, which implies

$$\lambda(\{x\colon a_\gamma < \gamma(x) \leqslant b_\gamma\}) = 0. \tag{3.21}$$

Now for $a_\gamma < r \leqslant b_\gamma$ we have

$$\lambda(\{x\colon \gamma(x) > r\}) = \lambda(\{x\colon \gamma(x) > a_\gamma\}) = \lambda(A^+(\delta)) \leqslant \tfrac{1}{2}.$$

Using Eq. (3.21),

$$\lambda(\{x\colon \gamma(x) = r\}) = 0,$$

and then

$$\lambda(\{x\colon \gamma(x) > r\}) = \lambda(\{x\colon \gamma(x) \geqslant r\}) = m_\gamma(r) = \tfrac{1}{2},$$

and therefore,

$$\lambda(\{x\colon \gamma(x) < r\}) = \tfrac{1}{2}.$$

For the converse, assume $r \notin [a, b]$. If $r < a$ then

$$\begin{aligned}\lambda(\{x \in X\colon \gamma(x) > r\}) &= \lambda\left(\bigcup_n \left\{x \in X\colon \gamma(x) > r + \frac{1}{n}\right\}\right)\\ &= \lim_{n \to +\infty} \lambda\left(\left\{x \in X\colon \gamma(x) > r + \frac{1}{n}\right\}\right).\end{aligned}$$

But $c_n \equiv \lambda(\{x \in X\colon\ \gamma(x) > r + 1/n\})$ is an increasing sequence. Therefore, if n is such that $r + 1/n < a$, then $c_n > \frac{1}{2}$ (since $m_\gamma(x) > \frac{1}{2}$ for $x < a_\gamma$). Hence

$$\lambda(\{x \in X\colon \gamma(x) > r\}) > \tfrac{1}{2},$$

and using (3.14) $\gamma - r$ cannot be balanced.

For the case $b < r$ the proof is analogous. □

We have an immediate corollary.

Corollary 3.5. *If $\gamma(0) = \gamma(1)$, and β is the extension of γ to $\mathbb{R}$ as a periodic function of period* 1, *and we define $\gamma_s\colon X \to \mathbb{R}$ by $\gamma_s(t) \equiv \beta(s + t)$, then $\forall r \in [a_\gamma, b_\gamma]$, $\gamma_s - r$ is balanced.*

The proof is straightforward.

Remark. Considering γ as a uniformly distributed random variable on the $[0, 1]$ interval, the preceding lemma showed that the interval $[a_\gamma, b_\gamma]$ contains exactly the "medians" of the distribution of γ.

The next proposition shows the existence of an optimal cut.

Proposition 3.6. *If μ and ν are two measures on T, G_μ, G_ν are as defined in* (3.1) *and $G(x) = G_\mu(x) - G_\nu(x)$, consider $\alpha: X \to \mathbb{R}$ to be the restriction of G to X, i.e., $\alpha(x) = G_\mu(x) - G_\nu(x)$, $x \in [0,1]$, then there exists $s \in T$ such that either D^s or I^s is balanced, where $D^s = D^s_\mu - D^s_\nu$ and $I^s = I^s_\mu - I^s_\nu$ as defined in* (3.4).

Proof. Since α is right-continuous, we can apply Lemma 3.3 to α. Therefore, there exists a sequence $\{s_n\} \subset X$ such that $\alpha(s_n) \to a_\alpha$, where a_α is as in the lemma. Let now s be a limit point of $\{s_n\}$, then there exists either

(i) a decreasing subsequence $s_{n_k} \to s$, or
(ii) an increasing subsequence $s_{n_k} \to s$.

In the first case $\alpha(s_{n_k}) \to \alpha(s+) = \alpha(s) = a_\alpha$, and hence $D^s(x) = G(x+s) - G(s) = G(x+s) - a_\alpha$. By Lemma 3.4 α is balanced, and by the corollary, D^s is balanced.

In the second case $\alpha(s_{n_k}) \to \alpha(s-) = a_\alpha$, and hence $I^s(x) = G(x+s) - G(s-) = G(x+s) - a_\alpha$, and therefore I^s is balanced. □

We can now prove the following theorem.

Theorem 3.7. *Let $\mu, \nu \in M(T)$, then*

$$d_T(\mu,\nu) = \int_0^1 |\alpha(x) - a_\alpha| \, dx,$$

where $\alpha: X \to \mathbb{R}$ is defined by

$$\alpha(x) = \mu([0,x]) - \nu([0,x]), \quad 0 \leqslant x < 1,$$
$$\alpha(1) = 0$$

and

$$a_\alpha = \sup\{t \in T : \lambda(\{x \in X : \alpha(x) \geqslant t\}) > \tfrac{1}{2}\}.$$

Proof. Since from the definition of α, α is the restriction of G to X with G as defined above, by the preceding Proposition 3.6 we know that there exists $s \in T$ such that either D^s or I^s is balanced.

Assume first that s is such that D^s is balanced. Then by Proposition 3.2

$$d_T(\mu,\nu) = d_X(\mu^D_s, \nu^D_s).$$

Now, using Eq. (2.4)

$$d_X(\mu^D_s, \nu^D_s) = \int_0^1 |D^s(x)| \, dx = \int_0^1 |\alpha(x) - a_\alpha| \, dx.$$

If, on the other hand, s is such that I^s is balanced, then by the same proposition

$$d_T(\mu,\nu) = d_X(\mu^I_s, \nu^I_s).$$

Now, using Eq. (2.4)

$$d_X(\mu_s^I, \nu_s^I) = \int_0^1 |I^s(x)|\,\mathrm{d}x = \int_0^1 |\alpha(x) - a_\alpha|\,\mathrm{d}x.$$

This completes the proof. □

Corollary 3.8. *Under the same hypothesis of the preceding theorem,*

$$d_T(\mu, \nu) = \min\left(\inf_{s\in T}\int_0^1 |\alpha(x) - \alpha(s)|\,\mathrm{d}x,\ \inf_{s\in T}\int_0^1 |\alpha(x) - \alpha(s-)|\,\mathrm{d}x\right).$$

Proof. Recalling that G defined above is a periodic function, we have

$$d_X(\mu_s^D, \nu_s^D) = \int |D^s(x)|\,\mathrm{d}x = \int_0^1 |G(x+s) - G(s)|\,\mathrm{d}x$$
$$= \int_0^1 |G(x) - G(s)|\,\mathrm{d}x = \int_0^1 |\alpha(x) - \alpha(s)|\,\mathrm{d}x.$$

Similarly,

$$d_X(\mu_s^I, \nu_s^I) = \int_0^1 |\alpha(x) - \alpha(s-)|\,\mathrm{d}x.$$

Now using Propositions 3.2 and 3.6 the result follows. □

Corollary 3.9. *If μ and $\nu \in M(T)$ are continuous (i.e., $\mu(\{s\}) = \nu(\{s\}) = 0\ \forall s \in T$) then*

$$d_T(\mu, \nu) = \inf_{s\in T}\int_0^1 |\alpha(x) - \alpha(s)|\,\mathrm{d}x.$$

Proof. If μ and ν are continuous, then α is continuous and therefore $D^s = I^s$ for every $s \in T$. □

4. Discrete case

We want to show in this section that the computation of the distance for the discrete case has linear complexity.

We therefore consider measures supported on a finite number of points. The points of the support do not need to be equidistributed; and when analyzing the distance between two measures μ and ν, we only consider those points such that either μ or ν are $\neq 0$.

Let us first consider the linear case of measures supported on X. In this case, the problem of the calculation of the Kantorovich metric can be described in the following way:

Let $0 \leqslant x_1 < x_2 < \cdots < x_n \leqslant 1$ and $\mu = (\mu_1, \dots, \mu_n)$, $\nu = (\nu_1, \dots, \nu_n)$ be such that $0 \leqslant \mu_i, \nu_i \leqslant 1$ and $\sum \mu_i = \sum \nu_i = 1$.

The Kantorovich metric is then

$$d_X(\mu, \nu) = \sup \left\{ \sum_i (\mu_i - \nu_i) f_i \colon |f_{i+1} - f_i| \leqslant x_{i+1} - x_i, i = 1, \dots, n-1 \right\}.$$

Using (2.4) we obtain

$$d_X(\mu, \nu) = \sum_{i=1}^{n-1} d_i |F_i|, \tag{4.1}$$

where $d_i = x_{i+1} - x_i$ and $F_i = \sum_{k=1}^{i} \mu_k - \nu_k$.

Let us consider now the case of measures on the circle T.

Let $0 \leqslant s_1 < s_2 < \cdots < s_n < 1$ be n points in T, and $\mu = (\mu_1, \dots, \mu_n)$, $\nu = (\nu_1, \dots, \nu_n)$ be such that $0 \leqslant \mu_i, \nu_i \leqslant 1$ and $\sum \mu_i = \sum \nu_i = 1$.

The Kantorovich metric on T is then

$$d_T(\mu, \nu) = \sup \left\{ \sum_{i=1}^{n} (\mu_i - \nu_i) f_i \right\},$$

where the supremum is taken over $f = (f_1, \dots, f_n)$ such that

$$|f_{i+1} - f_i| \leqslant \rho(s_{i+1}, s_i), \quad i = 1, \dots, n-1,$$

$$|f_1 - f_n| \leqslant \rho(s_1, s_n)$$

and ρ is the minimum arclength metric defined in Section 2.

Then, using Corollary 3.8 it is easy to see that

$$d_T(\mu, \nu) = \min_{1 \leqslant s \leqslant n} \sum_{k=1}^{n} d_k |\alpha_k - \alpha_s|, \tag{4.2}$$

where $\alpha_j = \sum_{i=1}^{j} \mu_i - \nu_i$, $1 \leqslant j \leqslant n$ and $d_j = \rho(s_{j+1}, s_j)$, $1 \leqslant j \leqslant n-1$, $d_n = \rho(s_1, s_n)$.

The problem of finding the minimum of the formula above is known as the weighted median problem. It is known that this problem has linear time complexity (see [8] and references therein).

Therefore, the problem of calculating the Kantorovich distance on the line and on the circle for finite, not necessarily equally distributed weights turns out to be solvable in linear time.

4.1. An application to a Minimal Matching Problem

In this section we are going to show an application of the discrete Kantorovich metric to solve a Minimal Matching Problem. Let us first describe the problem.

Let (T, ρ) be as before the circle with the minimal arclength distance. Let

$$A = \{a_1, \dots, a_n\}, \qquad B = \{b_1, \dots, b_n\}$$

be two sorted sets of points on the circle, $0 \leqslant a_i, b_i < 1$. Let π be a permutation of $\{1, \dots, n\}$. The matching associated to π (or simply the matching π) is the pairing $(a_i, b_{\pi(i)})$, $i = 1, \dots, n$. The cost of the matching is the sum of the distances between the points in each pair, i.e.,

$$C(\pi) = \sum_{i=1}^{n} \rho(a_i, b_{\pi(i)}).$$

The Minimal Matching Problem is to find the minimum of $C(\pi)$ over all the permutations. In the case that the points are on the line it is known (see [20, 24]) that the minimum is obtained when π is the identity (i.e., the matching is (a_1, b_1), $(a_2, b_2), \dots, (a_n, b_n)$). In the circular case an $n \log n$ algorithm was obtained in [24].

Let us now interpret the Minimal Matching Problem in the context of the Kantorovich Metric Problem: Assume that μ and ν are the uniform measures supported on A, B, respectively (i.e., $\mu(a_i) = 1 = \nu(b_i)$, $i = 1, \dots, n$). (Note: For simplicity of the notation we do not normalize the measures, since it is irrelevant to the analysis.)

From Corollary 3.8 and using that $\alpha(s)$ in this case only takes a finite number of values we have

$$d_T(\mu, \nu) = \inf_s \int_T |\alpha(x) - \alpha(s)| \, \mathrm{d}x. \tag{4.3}$$

In [24] it is shown that

$$\inf_\pi C(\pi) = \inf_s \int_T |\alpha(x) - \alpha(s)| \, \mathrm{d}x.$$

Therefore, using (4.2) we see that the Minimal Matching Problem in the circle can be reduced to the Weighted Median Problem and then be solved in linear time.

5. Conclusions

In this paper we study the Kantorovich metric for probability measures on the circle and we obtain an explicit analytical expression for it. This expression is related to the one already known for measures on the line. In applications of Digital Image Processing the Kantorovich metric on the circle has been found to be very useful. In particular, in pattern recognition for texture analysis and discrimination, this metric has been applied for comparing circular features as orientability and gradient directions. This fact makes it very relevant to have a linear-time algorithm for its calculation.

We also show that a combinatorial Minimal Matching Problem for which only an $n \log n$ algorithm was known can be interpreted in the Kantorovich metric context and therefore be solved in linear time.

Acknowledgements

The authors thank F. Barahona for many stimulating discussions, valuable insights on this subject and bringing many references to our attention, and R. Baeza-Yates for pointing out [8]. We also want to thank Z. Rachev for his encouragement to publish this paper. We gratefully acknowledge support from the Faculty of Mathematics of the University of Waterloo.

References

[1] E.J. Anderson and P. Nash, *Linear Programming in Infinite-Dimensional Spaces* (Wiley, Chichester, 1987).
[2] F. Barahona, C.A. Cabrelli and U.M. Molter, Computing the Hutchinson distance by network flows methods, *Random Comput. Dynamics* **1**(1) (1992) 117–129.
[3] F. Barahona, C.A. Cabrelli and U.M. Molter, Matching probability measures on the line under translation, *Random Comput. Dynamics*, to appear.
[4] M.F. Barnsley, *Fractals Everywhere* (Academic Press, San Diego, CA, 1988).
[5] C. Bleich and M.L. Overton, A linear-time algorithm for the weighted median problem, Tech. Report 75, Dept. of Computer Science, New York Univ., 1983.
[6] J. Brandt, C.A. Cabrelli and U.M. Molter, An algorithm for the computation of the Hutchinson distance, *Inform. Process. Lett.* **40** (1991) 113–117.
[7] R.M. Dudley, Probabilities and metrics: convergence of laws on metric spaces, with a view to statistical testing, Aarhus University Mathematics Institute Lecture Notes Series, Vol. 45, Aarhus University, Aarhus, 1976.
[8] C. Gurwitz, Weighted median algorithms for L_1 approximation, *BIT* **30** (1990) 301–310.
[9] J. Hutchinson, Fractals and self-similarity, *Indiana Univ. J. Math.* **30** (1981) 713–747.
[10] D.B. Johnson and T. Mizoguchi, Selecting the kth element in $X + Y$ and $X_1 + X_2 + \cdots + X_m$, *SIAM J. Comput.* **7** (1978) 147–153.
[11] V.L. Levin, General Monge–Kantorovich problem and its applications in measure theory and mathematical economics, in: Lev.J. Leifman, Ed., *Functional Analysis, Optimization and Mathematical Economics* (Oxford University Press, Oxford, 1990) 141–176.
[12] H.M. Raafat and A.K.C. Wong, A texture information-directed region growing algorithm for image segmentation and region classification, *Comput. Vision Graphics Image Process.* **43** (1988) 1–21.
[13] S.T. Rachev, The Monge–Kantorovich mass transference problem and its stochastic applications, *Theory Probab. Appl.* **29** (1984) 647–686.
[14] S.T. Rachev, *Probability Metrics and the Stability of Stochastic Models* (Wiley, New York, 1991).
[15] S.T. Rachev and R.M. Shortt, Classification problem for probability metrics, *Contemp. Math.* **94** (1989) 221–262.
[16] S.T. Rachev and R.M. Shortt, Duality theorems for Kantorovich–Rubinstein and Wasserstein functionals, *Dissertationes Math.* **CCXCIX** (1990).
[17] A. Reiser, A linear selection algorithm for sets of elements with weights, *Inform. Process. Lett.* **7** (1978) 159–162.
[18] L. Rüschendorf, The Wasserstein distance and approximation theorems, *Z. Wahrsch. Verw. Gebiete* **70** (1985) 117–129.
[19] M.I. Shamos, Geometry and statistics, in: J.F. Traub, Ed., *Algorithms and Complexity: New Directions and Recent Results* (Academic Press, New York, 1976) 251–280.
[20] H.C. Shen and A.K.C. Wong, Generalized texture representation and metric, *Comput. Vision Graphics Image Process.* **23** (1983) 187–206.
[21] J. Stark, A neural network to compute the Hutchinson metric in fractal image processing, *IEEE Trans. Neural Networks* **2**(1) (1991) 156–158.
[22] A. Szulga, On the Wasserstein metric, in: *Proc. Transactions of the 8th Prague Conf. on Information Theory, Statistical Decision Functions, and Random Processes, Vol. B* (Akademia, Prague, 1978) 267–273.
[23] S.S. Vallender, Calculation of the Wasserstein distance between probability distributions on the line, *Theory Probab. Appl.* **18** (1973) 784–786.
[24] M. Werman, S. Peleg, R. Melter and T.Y. Kong, Bipartite graph matching for points on a line or a circle, *J. Algorithms* **7** (1986) 277–284.
[25] A.K.C. Wong, H.C. Shen and P.W. Wong, Search-effective multi-class texture classification, *Internat. J. Pattern Recognition Artificial Intell.* **4**(4) (1990) 527–552.
[26] V.M. Zolotarev, Probability metrics, *Theory Probab. Appl.* **28** (1983) 278–302.