



**HAL**  
open science

## Fast transport optimization for Monge costs on the circle

Julie Delon, Julien Salomon, Andrei Sobolevskii

► **To cite this version:**

Julie Delon, Julien Salomon, Andrei Sobolevskii. Fast transport optimization for Monge costs on the circle. *SIAM Journal on Applied Mathematics*, 2010, 70 (7), pp.2239. 10.1137/090772708 . hal-00362834v2

**HAL Id: hal-00362834**

**<https://hal.science/hal-00362834v2>**

Submitted on 8 Oct 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# FAST TRANSPORT OPTIMIZATION FOR MONGE COSTS ON THE CIRCLE\*

JULIE DELON<sup>†</sup>, JULIEN SALOMON<sup>‡</sup>, AND ANDREI SOBOLEVSKI<sup>§</sup>

**Abstract.** Consider the problem of optimally matching two measures on the circle, or equivalently two periodic measures on  $\mathbb{R}$ , and suppose the cost  $c(x, y)$  of matching two points  $x, y$  satisfies the Monge condition:  $c(x_1, y_1) + c(x_2, y_2) < c(x_1, y_2) + c(x_2, y_1)$  whenever  $x_1 < x_2$  and  $y_1 < y_2$ . We introduce a notion of *locally optimal* transport plan, motivated by the weak KAM (Aubry–Mather) theory, and show that all locally optimal transport plans are conjugate to shifts and that the cost of a locally optimal transport plan is a convex function of a shift parameter.

This theory is applied to a transportation problem arising in image processing: for two sets of point masses on the circle, both of which have the same total mass, find an optimal transport plan with respect to a given cost function  $c$  satisfying the Monge condition. In the circular case the sorting strategy fails to provide a unique candidate solution and a naive approach requires a quadratic number of operations. For the case of  $N$  real-valued point masses we present an  $O(N|\log \epsilon|)$  algorithm that approximates the optimal cost within  $\epsilon$ ; when all masses are integer multiples of  $1/M$ , the algorithm gives an exact solution in  $O(N \log M)$  operations.

**Key words.** Monge–Kantorovich problem, Monge cost, Aubry–Mather (weak KAM) theory.

**AMS subject classifications.** Primary 90C08; Secondary 68Q25, 90C25

**1. Introduction.** The transport optimization problem, introduced by G. Monge in 1781 and shown by L. Kantorovich in 1942 to be an instance of linear programming, is a convex optimization problem with strong geometric features. A typical example is minimization of mean-square displacement between two given finite marginal measures supported on convex compacta in Euclidean space: in this case a solution is defined by gradient of a convex function that satisfies a suitable Monge–Ampère equation. Various generalizations of this result and rich bibliographies can be found, e.g., in the survey [12] or the recent monograph [19].

Further constraints on the two marginals or their supports may furnish the problem with useful additional convex structure. One way of making this statement quantitative is to consider the algorithmic complexity of the corresponding numerical transport optimization schemes. In particular when the two measures live on segments of straight lines, the optimal map is monotone and may be found by sorting, which takes  $O(n \log n)$  operations when the data come in the form of discrete  $n$ -point histograms. If the input data are already sorted, this count falls to  $O(n)$ .

The optimal transport is well understood also when the marginals live on a compact Riemannian manifold [10]; the existence and characterization of optimal map in the case of a flat torus and quadratic cost have been established a decade ago [8]. However, the algorithmics of even the simplest setting of the unit circle is no longer trivial, because the support of the measures is now *oriented* rather than *ordered*. A naive approach would require solving the problem for each of  $n$  different alignments of two  $n$ -point histograms, thereby involving  $O(n^2)$  operations.

---

\*Supported by the French National Research Agency (project ANR-07-BLAN-0235 OTARIE, <http://www.mccme.ru/~ansobol/otarie/>). The hospitality of UMR 6202 CNRS “Laboratoire Caspiopée” (Observatoire de la Côte d’Azur) is gratefully acknowledged.

<sup>†</sup>LTCI CNRS, Télécom ParisTech.

<sup>‡</sup>Université Paris-Dauphine and CEREMADE.

<sup>§</sup>Institute for Information Transmission Problems and UMI 2615 CNRS “Laboratoire J.-V. Poncelet,” Moscow; partially supported by the Ministry of National Education of France and by the Russian Fund for Basic Research (project RFBR 07–01–92217-CNRS-a).

In this paper we present an efficient algorithm of transport optimization on the circle, which is based on a novel analogy with the Aubry–Mather (weak KAM) theory in Lagrangian dynamics (see, e.g., [3, 9, 14]). The key step is to lift the transport problem to the universal cover of the unit circle, rendering the marginals periodic and the cost of transport infinite. However, it still makes sense to look for those transport maps whose cost cannot be decreased by any *local* modification. Different locally optimal maps, which cannot be deformed into each other by any local rearrangement, form a family parameterized with an analogue of the rotation number in the Aubry–Mather theory. One can introduce a counterpart of the average Lagrangian, or Mather’s  $\alpha$  function in the Aubry–Mather theory, which turns out to be efficiently computable. As we show below, its minimization provides an efficient algorithm of transport optimization on the circle. The class of cost functions for which this theory works includes all costs with the Monge property, such as the quadratic cost or costs generated by natural Lagrangians with time-periodic potentials [5, 14].

Note that the problem of optimally matching circular distributions appears in a variety of applications. Important examples are provided by image processing and computer vision: image matching techniques for retrieval, classification, or stitching purposes [21, 6] are often based on matching or clustering “descriptors” of local features [15], which typically consist of one or multiple histograms of gradient orientation. Similar issues arise in object pose estimation and pattern recognition [15, 11]. Circular distributions also appear in a quite different context of analysis of color images, where hue is parameterized by polar angle. In all these applications, matching techniques must be robust to data quantization and noise and computationally effective, which is especially important with modern large image collections. These requirements are satisfied by the optimal value of a transport cost for a suitable cost function.

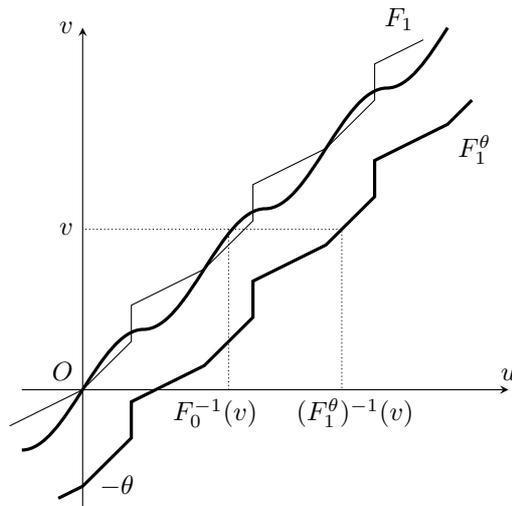
This paper is organized as follows. In §2 we give a specific but nontechnical overview of our results. After the basic definitions are given in §3, including that of locally optimal transport plans, in §4 we give an explicit description of the family of locally optimal transport plans: they are conjugate, in measure theoretic sense, to rotations of the unit circle (or equivalently to shifts of its universal cover). This result is in direct analogy with conjugacy to rotations in the one-dimensional Aubry–Mather theory [3]. As shown in §5, the average cost of a locally optimal transport plan is a convex function of the shift parameter. Moreover, the values of this function and its derivative are efficiently computable when the marginal measures are discrete, which enables us to present in §6 a fast algorithm for transport optimization on the circle. The same section contains results of a few numerical experiments. Finally a review of related work in the computer science literature is given in §7.

**2. Informal overview.** For two probability measures  $\hat{\mu}_0, \hat{\mu}_1$  on the unit circle  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$  and a given cost  $\hat{c}(\hat{x}, \hat{y})$  of transporting a unit mass from  $\hat{x}$  to  $\hat{y}$  in  $\mathbb{T}$ , the *transport cost* is defined as the inf of the quantity

$$\hat{I}(\gamma) = \iint_{\mathbb{T} \times \mathbb{T}} \hat{c}(\hat{x}, \hat{y}) \gamma(d\hat{x} \times d\hat{y}). \quad (2.1)$$

over the set of all *couplings*  $\gamma$  of the probability measures  $\hat{\mu}_0, \hat{\mu}_1$  (i.e., all measures on  $\mathbb{T} \times \mathbb{T}$  with marginals  $\hat{\mu}_0, \hat{\mu}_1$ ). These couplings are usually called *transport plans*.

Suppose that the cost function  $\hat{c}(\cdot, \cdot)$  on  $\mathbb{T} \times \mathbb{T}$  is determined via the relation  $\hat{c}(\hat{x}, \hat{y}) = \inf c(x, y)$  by a function  $c(\cdot, \cdot)$  on  $\mathbb{R} \times \mathbb{R}$  satisfying the condition  $c(x+1, y+1) = c(x, y)$  for all  $x, y$ ; here inf is taken over all  $x, y$  whose projections to the unit circle coincide with  $\hat{x}, \hat{y}$ . We lift the measures  $\hat{\mu}_0$  and  $\hat{\mu}_1$  to  $\mathbb{R}$ , obtaining periodic locally

FIG. 2.1. Construction of the locally optimal transport plan  $\gamma_\theta$ .

finite measures  $\mu_0$ ,  $\mu_1$ , and redefine  $\gamma$  to be their coupling on  $\mathbb{R} \times \mathbb{R}$ . It is then convenient to replace the problem of minimizing the integral (2.1) with “minimization” of an integral

$$I(\gamma) = \iint_{\mathbb{R} \times \mathbb{R}} c(x, y) \gamma(dx \times dy). \quad (2.2)$$

Although the latter integral is infinite, it still makes sense to look for transport plans  $\gamma$  minimizing  $I$  with respect to *local* modifications, i.e., to require that for any compactly supported signed measure  $\delta$  of zero mass and finite total variation, the difference  $I(\gamma + \delta) - I(\gamma)$ , which is defined by a finite integral, be nonnegative. These *locally optimal* transport plans are the main object of this paper.

Assume that the cost function  $c(x, y)$  satisfies the *Monge condition* (alternatively known as the continuous Monge property, see [1, 7]):

$$c(x_1, y_1) + c(x_2, y_2) < c(x_1, y_2) + c(x_2, y_1) \quad (2.3)$$

for all  $x_1 < x_2$  and  $y_1 < y_2$ . An example of such a cost function is  $|x - y|^\lambda$ , where  $\lambda > 1$ ; in this case the quantity  $\text{MK}_\lambda(\hat{\mu}_0, \hat{\mu}_1) = (\inf_\gamma \hat{I}(\gamma))^{1/\lambda}$  turns out to be a metric on the set of measures on the circle, referred to as the *Monge–Kantorovich distance* of order  $\lambda$ . The value  $\lambda = 1$  can still be treated in the same framework as the limiting case  $\lambda \rightarrow 1$ ; it is sometimes called the Kantorovich–Rubinshtein metric or, in image processing literature, the Earth Mover’s distance [18].

The Monge condition (2.3) implies that whenever under a transport plan the mutual order of any two elements of mass is reversed, the transport cost can be strictly reduced by exchanging their destinations. It follows that a locally minimal transport plan moves elements of mass *monotonically*, preserving their spatial order.

The whole set of locally optimal transport plans for a given pair of marginals  $\mu_0$ ,  $\mu_1$  can be conveniently described using a construction represented in fig. 2.1. Let  $F_0$ ,  $F_1$  be cumulative distribution functions of the measures  $\mu_0$ ,  $\mu_1$  normalized so that  $F_0(0) = F_1(0) = 0$ . We shall regard graphs of  $F_0$ ,  $F_1$  as continuous curves

including, where necessary, the vertical segments corresponding to jumps of these functions (which are caused by atoms of  $\mu_0, \mu_1$ ). Each of these curves specifies a correspondence,  $F_0^{-1}$  or  $F_1^{-1}$ , between points of the vertical axis  $Ov$ , representing elements of mass, and points of the horizontal axis  $Ou$ , representing spatial locations, and maps the Lebesgue measure on  $Ov$  into  $\mu_0$  or  $\mu_1$  on  $Ou$ . This correspondence is monotone and defined everywhere except on an (at most countable) set of  $v$  values that correspond to vacua of the measure in the  $Ou$  axis.

Define now  $F_1^\theta(u) = F_1(u) - \theta$ . Then  $(F_1^\theta)^{-1}$  represents a *shift* of the  $Ov$  axis by  $\theta$  followed by an application of the correspondence  $F_1^{-1}$ , and still induces on the  $Ou$  axis the same measure  $\mu_1$  as  $F_1$ . A transport plan  $\gamma_\theta$  that takes an element of mass represented by  $v$  from  $F_0^{-1}(v)$  to  $(F_1^\theta)^{-1}(v)$  is, by construction, a monotone coupling of  $\mu_0$  and  $\mu_1$ , and thus a locally optimal transport plan. Moreover, it is shown in §4 that all locally optimal transport plans can be obtained using this construction as the parameter  $\theta$  runs over  $\mathbb{R}$ .

Finally define the *average cost*  $C_{[F_0, F_1]}(\theta)$  of the plan  $\gamma_\theta$  per unit period:

$$C_{[F_0, F_1]}(\theta) = \int_0^1 c(F_0^{-1}(v), (F_1^\theta)^{-1}(v)) dv.$$

It is shown in §5 that the Monge condition implies convexity of  $C_{[F_0, F_1]}(\theta)$  and that the global minimum of this function in  $\theta$  coincides with the minimum value of the transport cost on the unit circle (2.1).

When the marginals  $\mu_0, \mu_1$  are purely atomic with finite numbers  $n_0$  and  $n_1$  of atoms in each period, the function  $C$  becomes piecewise affine. In §6 we present an algorithm to approximate its minimum value to accuracy  $\epsilon$ , using a binary search that takes  $O((n_0 + n_1) \log(1/\epsilon))$  operations in the real number computing model. When masses of all atoms are rational numbers with the least common denominator  $M$ , this search returns an exact solution provided that  $\epsilon < 1/M$ . This gives an  $O((n_0 + n_1) \log M)$  exact transport optimization algorithm on the circle.

**3. Preliminaries.** Let  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$  be the unit circle, i.e., the segment  $[0, 1]$  with identified endpoints. By  $\pi: \mathbb{R} \rightarrow \mathbb{T}$  denote the projection that takes points of the universal cover  $\mathbb{R}$  to points of  $\mathbb{T}$ .

**3.1. The cost function.** A *cost function* is a real-valued function  $c(\cdot, \cdot)$  defined on the universal cover  $\mathbb{R}$  of the circle  $\mathbb{T}$ . We assume that it satisfies the *Monge condition*: for any  $x_1 < x_2$  and  $y_1 < y_2$ ,

$$c(x_1, y_1) + c(x_2, y_2) - c(x_1, y_2) - c(x_2, y_1) < 0. \quad (3.1)$$

Additionally  $c$  is assumed to be lower semicontinuous, to be invariant with respect to integer shifts, i.e.,

$$c(x + 1, y + 1) = c(x, y) \quad (3.2)$$

for all  $x, y$ , and to grow uniformly as  $|x - y| \rightarrow \infty$ : for any  $P$  there exists a finite  $R(P) \geq 0$  such that

$$c(x, y) \geq P \quad \text{whenever } |x - y| \geq R(P). \quad (3.3)$$

Note that the latter condition implies that the lower semicontinuous function  $c$  is bounded from below (and guarantees that the minima in a number of formulas below are attained).

Note that the Monge condition (3.1) holds for any twice continuously differentiable function  $c$  such that  $\partial^2 c(x, y)/\partial x \partial y < 0$ . If the cost function depends only on  $x - y$ , this reduces to a convexity condition:  $-\partial^2 c(x - y)/\partial x \partial y = c''(x - y) > 0$ . In particular, all the above conditions are satisfied for the function  $c(x, y) = |x - y|^\lambda$ , which appears in the definition of the Monge–Kantorovich distance with  $\lambda > 1$ , and, more generally, for any function of the form  $c(x - y) + f(x) + g(y)$  with strictly convex  $c$  and periodic  $f$  and  $g$ .

For a cost function  $c$  satisfying all the above conditions, the cost of transporting a unit mass from  $\hat{x}$  to  $\hat{y}$  on the circle is defined as  $\hat{c}(\hat{x}, \hat{y}) = \inf c(x, y)$ , where  $\hat{x}, \hat{y}$  are points of  $\mathbb{T}$  and  $\inf$  is taken over all  $x, y$  in  $\mathbb{R}$  such that  $\pi x = \hat{x}$  and  $\pi y = \hat{y}$ . Using the integer shift invariance, this definition can be lifted to the universal cover as  $\hat{c}(x, y) = \inf_{k \in \mathbb{Z}} c(x, y + k)$ .

Condition (3.1) is all that is needed in §4, which is concerned with locally optimal transport plans on  $\mathbb{R}$ . Conditions (3.2), (3.3) come into play in §5, which deals with transport optimization on the circle.

**3.2. Distribution functions.** For a given locally finite measure  $\mu$  on  $\mathbb{R}$  define its *distribution function*  $F_\mu$  by

$$F_\mu(0) = 0, \quad F_\mu(x) = \mu((0, x]) \text{ for } x > 0, \quad F_\mu(x) = -\mu((x, 0]) \text{ for } x < 0. \quad (3.4)$$

Then  $\mu((x_1, x_2]) = F_\mu(x_2) - F_\mu(x_1)$  whenever  $x_1 < x_2$ , and this identity also holds for any function that differs from  $F_\mu$  by an additive constant (the normalization  $F_\mu(0) = 0$  is arbitrary). When  $\mu$  is periodic with unit mass in each period, it follows that for all  $x$  in  $\mathbb{R}$

$$F_\mu(x + 1) = F_\mu(x) + 1. \quad (3.5)$$

The *inverse* of a distribution function  $F_\mu$  is defined by

$$F_\mu^{-1}(y) = \inf\{x: y < F_\mu(x)\} = \sup\{x: y \geq F_\mu(x)\}. \quad (3.6)$$

Definitions (3.4) and (3.6) mean that  $F_\mu, F_\mu^{-1}$  are right-continuous. Discontinuities of  $F_\mu$  correspond to atoms of  $\mu$  and discontinuities of its inverse, to “vacua” of  $\mu$ , i.e., to intervals of zero  $\mu$  measure.

For a distribution function  $F_\mu$  define its *complete graph* to be the continuous curve formed by the union of the graph of  $F_\mu$  and the vertical segments corresponding to jumps of  $F_\mu$ . Accordingly, by a slight abuse of notation let  $F_\mu(\{x\})$  denote the set  $[F_\mu(x - 0), F_\mu(x)]$  (warning:  $F_\mu(\{x\}) \supseteq \{F_\mu(x)\}$ ) and let  $F_\mu(A) = \bigcup_{x \in A} F_\mu(\{x\})$  for any set  $A$ .

**3.3. Local properties of transport plans.** Let  $\hat{\mu}_0, \hat{\mu}_1$  be two finite positive measures of unit total mass on  $\mathbb{T}$  and  $\mu_0, \mu_1$  their liftings to the universal cover  $\mathbb{R}$ , i.e., periodic measures such that  $\mu_i(A) = \hat{\mu}_i(\pi A)$ ,  $i = 0, 1$ , for any Borel set  $A$  that fits inside one period. Periodicity of measures here means that  $\mu(A + n) = \mu(A)$  for any integer  $n$  and any Borel  $A$ , where  $A + n = \{x + n: x \in A\}$ .

DEFINITION 3.1. A (locally finite)<sup>1</sup> transport plan *with marginals*  $\mu_0$  and  $\mu_1$  is a locally finite measure  $\gamma$  on  $\mathbb{R} \times \mathbb{R}$  such that

(i) for any  $x$  in  $\mathbb{R}$  the supports of measures  $\gamma((-\infty, x] \times \cdot)$  and  $\gamma(\cdot \times (-\infty, x])$  are bounded from above and the supports of measures  $\gamma((x, \infty) \times \cdot)$ ,  $\gamma(\cdot \times (x, \infty))$  are bounded from below;

<sup>1</sup>In what follows the words ‘locally finite’ defining a transport plan will often be dropped.

(ii)  $\gamma(A \times \mathbb{R}) = \mu_0(A)$  and  $\gamma(\mathbb{R} \times B) = \mu_1(B)$  for any Borel sets  $A, B$ .

The quantity  $\gamma(A \times B)$  is the amount of mass transferred from  $A$  to  $B$  under the transport plan  $\gamma$ . Condition (1) implies that the mass supported on any bounded interval gets redistributed over a bounded set (indeed, a bounded interval is the intersection of two half-lines), but is somewhat stronger.

DEFINITION 3.2. *A local modification of the locally finite transport plan  $\gamma$  is a transport plan  $\gamma'$  such that  $\gamma$  and  $\gamma'$  have the same marginals and  $\gamma' - \gamma$  is a compactly supported finite signed measure. A local modification is called cost-reducing if*

$$\iint c(x, y) (\gamma'(dx \times dy) - \gamma(dx \times dy)) < 0.$$

A locally finite transport plan  $\gamma$  is said to be locally optimal with respect to the cost function  $c$  or  $c$ -locally optimal if it has no cost-reducing local modifications.

**4. Conjugate transport plans and shifts.** Let  $U_0, U_1$  be two copies of  $\mathbb{R}$  equipped with positive periodic measures  $\mu_0, \mu_1$  whose distribution functions  $F_0, F_1$  satisfy (3.5), so that all intervals of unit length have unit mass. Let furthermore  $V_0, V_1$  be two other copies of  $\mathbb{R}$  equipped with the uniform (Lebesgue) measure.

**4.1. Normal plans and conjugation.** We introduce the following terminology:

DEFINITION 4.1. *A locally finite transport plan  $\nu$  on  $V_0 \times V_1$  with uniform marginals is called normal.*

DEFINITION 4.2. *For a normal transport plan  $\nu$  its conjugate transport plan  $\nu^{[F_0, F_1]}$  is a transport plan on  $U_0 \times U_1$  such that for any Borel sets  $A, B$*

$$\nu^{[F_0, F_1]}(A \times B) = \nu(F_0(A) \times F_1(B)). \quad (4.1)$$

LEMMA 4.3. *For a normal transport plan  $\nu$  its conjugate  $\nu^{[F_0, F_1]}$  is a locally finite transport plan on  $U_0 \times U_1$  with marginals  $\mu_0, \mu_1$ .*

*Proof.* Since distribution functions  $F_0, F_1$  and their inverses preserve boundedness, condition (1) of Definition 3.1 is fulfilled. Definition 4.2, condition (2) of Definition 3.1, and formula (3.4) together imply that

$$\begin{aligned} \nu^{[F_0, F_1]}((u_1, u_2] \times U_1) &= \nu(F_0((u_1, u_2]) \times F_1(U_1)) = \nu([F_0(u_1), F_0(u_2)] \times V_1) \\ &= F_0(u_2) - F_0(u_1) = \mu_0((u_1, u_2]). \end{aligned}$$

Similarly  $\nu^{[F_0, F_1]}(U_0 \times (u_1, u_2]) = \mu_1((u_1, u_2])$ . Thus  $\nu^{[F_0, F_1]}$  satisfies condition (2) of Definition 3.1 on intervals and therefore on all Borel sets.  $\square$

LEMMA 4.4. *For any transport plan  $\gamma$  on  $U_0 \times U_1$  with marginals  $\mu_0$  and  $\mu_1$  there exists a normal transport plan  $\nu$  such that  $\gamma$  is conjugate to  $\nu$ :  $\gamma = \nu^{[F_0, F_1]}$ .*

*Proof.* For non-atomic measures  $\mu_0$  and  $\mu_1$  the required transport plan is given by the formula  $\nu(A \times B) = \gamma(F_0^{-1}(A) \times F_1^{-1}(B))$ , which is dual to (4.1). However if, e.g.,  $\mu_0$  has an atom, then the function  $F_0^{-1}$  is constant over a certain interval and maps any subset  $A$  of this interval into one point of fixed positive measure in  $U_0$ , so information on the true Lebesgue measure of  $A$  is lost. In this case extra care has to be taken.

Recall that a locally finite measure has at most a countable set of atoms. Let atoms of  $\mu_0$  be located in  $(0, 1]$  at points  $u_1, u_2, \dots$  with masses  $m_1, m_2, \dots$ . Since  $\gamma(\{u_i\} \times U_1) = \mu_0(\{u_i\}) = m_i > 0$ , there exists a conditional probability measure  $\rho(\cdot | u_i) = \gamma(\{u_i\} \times \cdot) / m_i$ . For a set  $A \subset (0, 1]$  define a ‘‘residue’’ transport plan

$$\bar{\gamma}(A \times B) = \gamma(A \times B) - \sum_i m_i \delta_{u_i}(A) \rho(B | u_i),$$

where  $\delta_u$  is the Dirac unit mass measure on  $U_0$  concentrated at  $u$ , and extend  $\bar{\gamma}$  to general  $A$  using periodicity. We thus remove from  $\bar{\gamma}$  the part of  $\gamma$  whose projection to the first factor is atomic. Define a transport plan  $\kappa$  on  $V_0 \times U_1$  by

$$\kappa(C \times B) = \sum_i \lambda(C \cap F_0(\{u_i\})) \rho(B | u_i) + \bar{\gamma}(F_0^{-1}(C) \times B),$$

where  $C$  is a Borel set in  $V_0$  and  $\lambda(\cdot)$  denotes the Lebesgue measure in  $V_0$ . Clearly  $\kappa(F_0(A) \times B) = \gamma(A \times B)$ . Repeating this construction for the second factor, with  $\kappa$  in place of  $\gamma$ , we get a normal transport plan  $\nu$  such that  $\gamma(A \times B) = \nu(F_0(A) \times F_1(B))$ .  $\square$

Since we are ultimately interested in transport optimization with marginals  $\mu_0, \mu_1$  rather than with uniform marginals, two normal transport plans  $\nu_1, \nu_2$  will be called *equivalent* if they have the same conjugate. Two different normal transport plans can only be equivalent if one or both measures  $\mu_0$  or  $\mu_1$  have atoms, causing loss of information on the structure of  $\nu$  in segments corresponding to these atoms. The proof of Lemma 4.4 gives a specific representative of this equivalence class of normal plans.

**4.2. Locally optimal normal transport plans are shifts.** Fix a cost function  $c: U_0 \times U_1 \rightarrow \mathbb{R}$  that satisfies the Monge condition (3.1) and define

$$c_{[F_0, F_1]}(v_0, v_1) = c(F_0^{-1}(v_0), F_1^{-1}(v_1)). \quad (4.2)$$

For non-atomic measures  $\mu_0, \mu_1$ , it satisfies the Monge condition

$$c_{[F_0, F_1]}(v', w') + c_{[F_0, F_1]}(v'', w'') - c_{[F_0, F_1]}(v', w'') - c_{[F_0, F_1]}(v'', w') < 0$$

whenever  $v' < v''$  and  $w' < w''$ ; this inequality can only turn into equality if either  $v', v''$  or  $w', w''$  correspond to an atom of the respective marginal ( $\mu_0$  or  $\mu_1$ ) of  $\nu^{[F_0, F_1]}$ , i.e., if  $c_{[F_0, F_1]}$  is constant in either first or second argument. In spite of this slight violation of definition of §3.1, we will still call  $c_{[F_0, F_1]}$  a cost function.

Here and below, variables  $u, u', u_0, u_1, \dots$  are assumed to take values in  $U_0$  or  $U_1$  and variables  $v, v', v_0, v_1, \dots, w, w', \dots$ , in  $V_0$  or  $V_1$ .

LEMMA 4.5. *A transport plan  $\gamma$  on  $U_0 \times U_1$  with marginals  $\mu_0, \mu_1$  is  $c$ -locally optimal if and only if it is conjugate to a  $c_{[F_0, F_1]}$ -locally optimal normal transport plan  $\nu$ . In particular, all normal transport plans with the same locally optimal conjugate are locally optimal.*

*Proof.* Note that  $\nu' - \nu$  is compactly supported if and only if the difference of the respective conjugates  $\gamma' - \gamma$  is compactly supported. The rest of the proof follows from the identity

$$\begin{aligned} & \iint c_{[F_0, F_1]}(v_1, v_2) (\nu'(dv_1 \times dv_2) - \nu(dv_1 \times dv_2)) \\ &= \iint c(u_1, u_2) (\gamma'(du_1 \times du_2) - \gamma(du_1 \times du_2)) \end{aligned}$$

established by the change of variables  $v_1 = F_0(u_1), v_2 = F_1(u_2)$  (here jumps of the distribution functions are harmless because  $c_{[F_0, F_1]}$  is constant over respective ranges of its variables).  $\square$

Transport optimization with marginals  $\mu_0, \mu_1$  is thus reduced to a *conjugate problem* involving uniform marginals and the cost  $c_{[F_0, F_1]}$ . It turns out that any  $c_{[F_0, F_1]}$ -optimal normal transport plan must be supported on a graph of a monotone

function, and due to uniformity of marginals this function can only be a shift by a suitable real increment  $\theta$ . More precisely, the following holds:

**THEOREM 4.6.** *Let  $\mu_0, \mu_1$  be two periodic positive measures defined respectively on  $U_0, U_1$  with unit mass in each period and let  $F_i: U_i \rightarrow V_i, i = 0, 1$ , be their distribution functions. Then any  $c_{[F_0, F_1]}$ -locally optimal normal transport plan on  $V_0 \times V_1$  is equivalent to a normal transport plan  $\nu_\theta$  with  $\text{supp } \nu_\theta = \{(v, w): w = v + \theta\}$ , and conversely  $\nu_\theta$  is  $c_{[F_0, F_1]}$ -locally optimal for any real  $\theta$ . All  $c$ -locally optimal transport plans on  $U_0 \times U_1$  with marginals  $\mu_0, \mu_1$  are of the form  $\gamma_\theta = (\nu_\theta)^{[F_0, F_1]}$ .*

The proof, divided into a series of lemmas, is based on the classical argument: a nonoptimal transport plan can be modified by “swapping” pieces of mass to render its support monotone while decreasing its cost. This argument, carried out for plans with uniform marginals on  $V_0 \times V_1$ , is combined with the observation that a monotonically supported plan with uniform marginals can only be a shift. Then Lemma 4.5 is used to extend this result to transport plans on  $U_0 \times U_1$ .

Throughout the proof fix a normal transport plan  $\nu$  and define on  $V_0 \times V_1$  the functions

$$r_\nu(v, w) = \nu((-\infty, v] \times (w, \infty)), \quad l_\nu(v, w) = \nu((v, \infty) \times (-\infty, w]). \quad (4.3)$$

To explain the notation  $r_\nu, l_\nu$  observe that, e.g.,  $r_\nu(v, w)$  is the amount of mass that is located initially to the left of  $v$  and goes to the right of  $w$ .

**LEMMA 4.7.** *The function  $r_\nu$  (resp.  $l_\nu$ ) is continuous and monotonically increasing in its first (second) argument and is continuous and monotonically decreasing in its second (first) argument, while the other argument is kept fixed.*

*Proof.* Monotonicity is obvious from (4.3). To prove continuity observe that the second marginal of  $\nu$  is uniform, which together with positivity of all involved measures implies that in the decomposition

$$\nu(V_0 \times \cdot) = \nu((-\infty, v] \times \cdot) + \nu((v, \infty) \times \cdot),$$

both measures in the right-hand side cannot have atoms. This implies continuity of  $r_\nu, l_\nu$  with respect to the second argument. A similar proof holds for the first argument.  $\square$

**LEMMA 4.8.** *For any  $v$  there exist  $w_\nu(v)$  and  $m_\nu(v) \geq 0$  such that*

$$r_\nu(v, w_\nu(v)) = l_\nu(v, w_\nu(v)) = m_\nu(v). \quad (4.4)$$

*The correspondence  $v \mapsto w_\nu(v)$  is monotone:  $w_\nu(v_1) \leq w_\nu(v_2)$  for  $v_1 < v_2$ .*

*Proof.* Clearly  $r_\nu(v, -\infty) = \infty, r_\nu(v, \infty) = 0, l_\nu(v, -\infty) = 0, l_\nu(v, \infty) = \infty$ . The continuity of the functions  $r_\nu(v, \cdot), l_\nu(v, \cdot)$  in the second argument for a fixed  $v$  implies that their graphs intersect at some point  $(w_\nu(v), m_\nu(v))$ , which satisfies (4.4). Should the equality  $r_\nu(v, w) = l_\nu(v, w)$  hold on a segment  $[w', w'']$ , we set  $w_\nu(v)$  to its left endpoint  $w'$ ; this situation, however, will be ruled out by the corollary to Lemma 4.10 below. Monotonicity of  $w_\nu(v)$  follows from monotonicity of  $r_\nu(\cdot, w), l_\nu(\cdot, w)$  in the first argument for a fixed  $w$ : indeed, for  $v_2 > v_1$  the equality  $r_\nu(v_2, w) = l_\nu(v_2, w)$  is impossible for  $w < w_\nu(v_1)$  because for such  $w$  we have  $r_\nu(v_2, w) > r_\nu(v_1, w_\nu(v_1)) = l_\nu(v_1, w_\nu(v_1)) > l_\nu(v_2, w)$ .  $\square$

Equalities (4.4) mean that the same amount of mass  $m_\nu(v)$  goes under the plan  $\nu$  from the left of  $v$  to the right of  $w_\nu(v)$  and from the right of  $v$  to the left of  $w_\nu(v)$ .

We are now in position to use the Monge condition and show that this amount can be reduced to zero by modifying the transport plan locally without a cost increase.

LEMMA 4.9. *For any  $v$  there exists a local modification  $\nu_v$  of  $\nu$  such that  $w_{\nu_v}(v) = w_\nu(v)$  (with  $w_\nu$  defined as in Lemma 4.8),  $m_{\nu_v}(v) = 0$ , and  $\nu_v$  is either cost-reducing in the sense of Definition 3.2 or is equivalent to  $\nu$ .*

*Proof.* Let  $w = w_\nu(v)$  and  $m = m_\nu(v)$ . If  $m = 0$ , there is nothing to prove. Suppose that  $m > 0$  and define

$$\begin{aligned} w^- &= \sup\{w' : l_\nu(v, w') = 0\}, & w^+ &= \inf\{w' : r_\nu(v, w') = 0\}, \\ v^- &= \sup\{v' : r_\nu(v', w) = 0\}, & v^+ &= \inf\{v' : l_\nu(v', w) = 0\}. \end{aligned}$$

By local finiteness of the transport plan  $\nu$  all these quantities are finite. Since  $m > 0$ , continuity of  $r_\nu, l_\nu$  implies that the inequalities  $w^- < w < w^+$  and  $v^- < v < v^+$  are strict. Consider the measures

$$\begin{aligned} \rho^-(\cdot) &= \nu(\cdot \times (w, w^+)) \text{ on } (v^-, v), & \rho^+(\cdot) &= \nu(\cdot \times (w^-, w)) \text{ on } (v, v^+), \\ \sigma^-(\cdot) &= \nu((v, v^+) \times \cdot) \text{ on } (w^-, w), & \sigma^+(\cdot) &= \nu((v^-, v) \times \cdot) \text{ on } (w, w^+). \end{aligned}$$

Equalities (4.4) mean that all these measures have the same positive total mass  $m$ . Note that the Lebesgue measures of intervals  $(v^-, v)$ ,  $(v, v^+)$ ,  $(w^-, w)$ , and  $(w, w^+)$  may be greater than  $m$ , because some mass in these intervals may come from or go to elsewhere.

The functions  $r_w(\cdot) = r_\nu(\cdot, w)$ ,  $l_v(\cdot) = l_\nu(v, \cdot)$  are monotonically increasing and  $r_v(\cdot) = r_\nu(v, \cdot)$ ,  $l_w(\cdot) = l_\nu(\cdot, w)$  are monotonically decreasing, with their inverses  $r_w^{-1}$ ,  $l_v^{-1}$ ,  $r_v^{-1}$ ,  $l_w^{-1}$  defined everywhere except on an at most countable set of points. These functions may be regarded as a kind of distribution functions for the measures  $\rho^-$ ,  $\sigma^-$ ,  $\sigma^+$ ,  $\rho^+$  respectively, mapping them to the Lebesgue measure on  $(0, m)$ .

Under the plan  $\nu$ , mass  $m$  is sent from  $(v^-, v)$  to  $(w, w^+)$  and from  $(v, v^+)$  to  $(w^-, w)$ . We now construct a local modification  $\nu_v$  of the transport plan  $\nu$  that moves mass  $m$  from the interval  $(v^-, v)$  to  $(w^-, w)$  and from  $(v, v^+)$  to  $(w, w^+)$ , and show that it is cost-reducing unless measures  $\mu_0, \mu_1$  have atoms corresponding to the intervals under consideration.

Observe first that the normal plan  $\nu$  induces two transport plans  $\tau_r, \tau_l$  that map measures  $\rho^-$  to  $\sigma^+$  and  $\rho^+$  to  $\sigma^-$  correspondingly:

$$\begin{aligned} \tau_r(A \times B) &= \nu(A \cap (-\infty, v) \times B \cap (w, +\infty)) = \nu(A \cap (v^-, v) \times B \cap (w, w^+)), \\ \tau_l(A \times B) &= \nu(A \cap (v, +\infty) \times B \cap (-\infty, w)) = \nu(A \cap (v, v^+) \times B \cap (w^-, w)), \end{aligned}$$

where  $A \subset (v^-, v^+)$ ,  $B \subset (w^-, w^+)$  are two arbitrary Borel sets and the  $\cap$  operation takes precedence over  $\times$ . By an argument similar to the proof of Lemma 4.4, there exist two transport plans  $\chi_r$  and  $\chi_l$  mapping the Lebesgue measure on  $(0, m)$  respectively to  $\sigma^+$ ,  $\sigma^-$  and such that

$$\begin{aligned} \tau_r(A \times B) &= \chi_r(r_w(A \cap (v^-, v)) \times B \cap (w, w^+)), \\ \tau_l(A \times B) &= \chi_l(l_w(A \cap (v, v^+)) \times B \cap (w^-, w)). \end{aligned}$$

Define now two transport plans  $\bar{\tau}_l, \bar{\tau}_r$  that send mass elements to the same destinations but from interchanged origins:

$$\begin{aligned} \bar{\tau}_r(A \times B) &= \chi_r(l_w(A \cap (v, v^+)) \times B \cap (w, w^+)), \\ \bar{\tau}_l(A \times B) &= \chi_l(r_w(A \cap (v^-, v)) \times B \cap (w^-, w)). \end{aligned}$$

This enables us to define

$$\nu_v(A \times B) = \nu(A \times B) - \tau_r(A \times B) - \tau_l(A \times B) + \bar{\tau}_r(A \times B) + \bar{\tau}_l(A \times B).$$

Since  $\tau_r(A \times \mathbb{R}) = \bar{\tau}_r(A \times \mathbb{R}) = \rho^-(A)$  etc., the transport plan  $\nu_v$  has the same uniform marginals as  $\nu$ , i.e., it is a local modification of  $\nu$ . Observe furthermore that by the construction of  $\nu_v$  no mass is moved under this plan from the left-hand side of  $v$  to the right-hand side of  $w$  and inversely, i.e., that  $m_{\nu_v}(v) = 0$ .

It remains to show that  $\nu_v$  is either a cost-reducing modification of  $\nu$  or equivalent to it. By the disintegration lemma (see, e.g., [2]) we can write  $\chi_r(d\alpha \times dw') = d\alpha dG_r(w' | \alpha)$  and  $\chi_l(d\alpha \times dw') = d\alpha dG_l(w' | \alpha)$ , where  $G_r(\cdot | \alpha)$  (resp.  $G_l(\cdot | \alpha)$ ) are distribution functions of probability measures defined on  $[w, w^+]$  (resp.  $[w^-, w]$ ) for almost all  $0 < \alpha < m$ . Denote their respective inverses by  $G_r^{-1}(\cdot | \alpha)$ ,  $G_l^{-1}(\cdot | \alpha)$  and observe that  $w^- \leq G_l^{-1}(\beta' | \alpha) \leq w \leq G_r^{-1}(\beta'' | \alpha) \leq w^+$  for any  $\beta', \beta''$ . Thus

$$\begin{aligned} \iint c(v', w') \tau_r(dv' \times dw') &= \iint c(r_w^{-1}(\alpha), w') \chi_r(d\alpha \times dw') \\ &= \int_0^m d\alpha \int c(r_w^{-1}(\alpha), w') dG_r(w' | \alpha) \\ &= \int_0^m d\alpha \int_0^1 d\beta c(r_w^{-1}(\alpha), G_r^{-1}(\beta | \alpha)), \end{aligned}$$

where we write  $c$  instead of  $c_{[F_0, F_1]}$  to lighten notation, and similarly

$$\begin{aligned} \iint c(v', w') \tau_l(dv' \times dw') &= \int_0^m d\alpha \int_0^1 d\beta c(l_w^{-1}(\alpha), G_l^{-1}(\beta | \alpha)), \\ \iint c(v', w') \bar{\tau}_r(dv' \times dw') &= \int_0^m d\alpha \int_0^1 d\beta c(l_w^{-1}(\alpha), G_r^{-1}(\beta | \alpha)), \\ \iint c(v', w') \bar{\tau}_l(dv' \times dw') &= \int_0^m d\alpha \int_0^1 d\beta c(r_w^{-1}(\alpha), G_l^{-1}(\beta | \alpha)). \end{aligned}$$

The integral in Definition 3.2 now takes the form

$$\begin{aligned} &\iint c(v', w') (\nu_v(dv' \times dw') - \nu(dv' \times dw')) \\ &= \iint c(v', w') (-\tau_r(dv' \times dw') - \tau_l(dv' \times dw') \\ &\quad + \bar{\tau}_r(dv' \times dw') + \bar{\tau}_l(dv' \times dw')) \\ &= \int_0^m d\alpha \int_0^1 d\beta (-c(r_w^{-1}(\alpha), G_r^{-1}(\beta | \alpha)) - c(l_w^{-1}(\alpha), G_l^{-1}(\beta | \alpha)) \\ &\quad + c(l_w^{-1}(\alpha), G_r^{-1}(\beta | \alpha)) + c(r_w^{-1}(\alpha), G_l^{-1}(\beta | \alpha))). \end{aligned}$$

As  $r_w^{-1}(\alpha) \leq v \leq l_w^{-1}(\alpha)$  and  $G_l^{-1}(\beta | \alpha) \leq w \leq G_r^{-1}(\beta | \alpha)$  for all  $\alpha, \beta$ , the Monge condition (3.1) implies that either the value of this integral is negative or the function  $c$  (i.e.,  $c_{[F_0, F_1]}$ ) is constant in at least one of its arguments. In the former case the transport plan  $\nu_v$  is a cost-reducing local modification of  $\nu$ ; in the latter case  $\nu_v$  is equivalent to  $\nu$ .  $\square$

LEMMA 4.10. *For any  $v' < v''$  there exists a local modification  $\nu_{v', v''}$  of  $\nu$  such that  $w_{\nu_{v', v''}}(v) = w_\nu(v)$  for  $v' \leq v \leq v''$ ,  $m_{\nu_{v', v''}}(v') = m_{\nu_{v', v''}}(v'') = 0$ , and in*

the strip  $v' \leq v \leq v''$  the support of  $\nu_{v',v''}$  coincides with the complete graph of the monotone function  $w_\nu(\cdot)$ .

*Proof.* Let  $\{v_i\}$  be a dense countable subset of  $[v', v'']$  including its endpoints. Set  $\nu_0 = \nu$  and define  $\nu_i$  recursively to be the local modification of  $\nu_{i-1}$  given by the previous lemma and such that  $w_{\nu_i}(v_i) = w_\nu(v_i)$  and  $m_{\nu_i}(v_i) = 0$ . Then all  $\nu_i$  are either cost-reducing or equivalent to  $\nu$  and  $w_{\nu_j}(v_i) = w_\nu(v_i)$ ,  $m_{\nu_j}(v_i) = 0$  for all  $j > i$ . Indeed, denote  $w_i = w_{\nu_i}(v_i)$  and observe that if e.g.  $v_j > v_i$ , then, as  $m_{\nu_i}(v_i) = r_{\nu_i}(v_i, w_i) = 0$ , mass from  $(-\infty, v_i]$  does not appear to the right of  $w_i$  and so does not contribute to the balance of mass around  $w_j$ . Therefore for any  $j$  the possible modification of  $\nu_{j-1}$  is local to the interval  $(v_{i'}, v_{i''})$ , where  $v_{i'} = \max\{v_i : i < j, v_i < v_j\}$  and  $v_{i''} = \min\{v_i : i < j, v_i > v_j\}$  (with max and min of empty set defined as  $v'$  and  $v''$ ). Thus there is a well-defined limit normal transport plan  $\nu_\infty$  that is either a cost-reducing local modification or equivalent to  $\nu$  and is such that, by continuity of the functions  $r_{\nu_\infty}$  and  $l_{\nu_\infty}$  in the first argument,  $m_{\nu_\infty}(v)$  vanishes everywhere on  $[v', v'']$ .

Consider now the function  $w_{\nu_\infty}(\cdot)$ , which coincides with  $w_\nu(\cdot)$  on a dense subset of  $[v', v'']$ , so that their complete graphs coincide. For any quadrant of the form  $(-\infty, v_0) \times (w_0, \infty)$  such that  $w_0 > w_{\nu_\infty}(v_0)$ , monotonicity of  $r_{\nu_\infty}$  in the second argument implies that

$$0 \leq \nu_\infty((-\infty, v_0) \times (w_0, \infty)) = r_{\nu_\infty}(v_0, w_0) \leq r_{\nu_\infty}(v_0, w_{\nu_\infty}(v_0)) = 0,$$

i.e.,  $\nu_\infty((-\infty, v_0) \times (w_0, \infty)) = 0$ . Similarly  $\nu_\infty((v_0, \infty) \times (-\infty, w_0)) = 0$  for any quadrant with  $w_0 < w_{\nu_\infty}(v_0)$ . The union of all such quadrants is the complement of the complete graph of the function  $v \mapsto w_\nu(v)$ ; this implies that  $\nu_\infty$  is supported thereon.  $\square$

**COROLLARY 4.11.** *For any normal transport plan  $\nu$  there exists a real number  $\theta_\nu$  such that  $w_\nu(v) = v + \theta_\nu$ .*

*Proof.* It is enough to show that  $w_\nu(v') - v' = w_\nu(v'') - v''$  for all  $v', v''$ . Let  $v' < v''$  and  $\nu_{v',v''}$  be the local modification constructed in the previous lemma. Since it has uniform marginals and monotone support, we have  $w_\nu(v'') - w_\nu(v') = \nu_{v',v''}((v', v'') \times (w_\nu(v'), w_\nu(v''))) = v'' - v'$ , which completes the proof.  $\square$

We call the parameter  $\theta_\nu$  the *rotation number* of the normal transport plan  $\nu$ .

**DEFINITION 4.12.** *A normal transport plan consisting of a uniform measure supported on the line  $\{(v, w) : w = v + \theta\}$  is called a shift and denoted by  $\nu_\theta$ .*

**LEMMA 4.13.** *For any  $\theta$  the shift  $\nu_\theta$  is  $c_{[F_0, F_1]}$ -locally optimal.*

*Proof.* Let  $\bar{\nu}$  be a local modification of  $\nu_\theta$  such that the signed measure  $\nu_\theta - \bar{\nu}$  is supported in  $(v', v'') \times (w', w'')$ . Let  $\bar{\nu}_{v',v''}$  be a local modification of  $\bar{\nu}$  constructed in Lemma 4.10; it coincides with  $\nu_\theta$  over  $v' < v < v''$ , and hence everywhere. Since it is either cost-reducing or equivalent to  $\bar{\nu}$ , it follows that  $\bar{\nu}$  cannot be cost-reducing with respect to  $\nu_\theta$ , i.e., that  $\nu_\theta$  is a cost minimizer with respect to local modifications.  $\square$

**LEMMA 4.14.** *Any  $c_{[F_0, F_1]}$ -locally optimal normal transport  $\nu$  with rotation number  $\theta = \theta_\nu$  is equivalent to the shift  $\nu_\theta$ .*

*Proof.* Let  $v_i'' = -v_i' = i$  for  $i = 1, 2, \dots$ . All local modifications  $\nu_i = \nu_{v_i', v_i''}$  of  $\nu$  constructed as in Lemma 4.10 cannot be cost-reducing and are therefore equivalent to  $\nu$ . On the other hand, this sequence stabilizes to the shift  $\nu_\theta$  on any bounded subset of  $V_0 \times V_1$  as soon as this set is covered by  $(-i, i) \times (-i, i)$ . Therefore  $\nu_\theta$  has the same conjugate as all  $\nu_i$  and is equivalent to  $\nu$ .  $\square$

Lemmas 4.13, 4.14, and 4.5 together imply Theorem 4.6.

**5. Transport optimization for periodic measures.** Let now  $c$  be a cost function that satisfies the Monge condition (3.1), the integer shift invariance condition (3.2), the growth condition (3.3), and is bounded from below. Suppose that  $\gamma_\theta$  is a locally optimal transport plan on  $U_0 \times U_1$  with marginals  $\mu_0, \mu_1$  conjugate to the shift  $\nu_\theta$ . Define  $c_{[F_0, F_1]}$  as in (4.2) and let  $F_1^\theta(u) = F_1(u) - \theta$  as illustrated in fig. 2.1.

DEFINITION 5.1. *We call the quantity*

$$C_{[F_0, F_1]}(\theta) = \int_0^1 c_{[F_0, F_1]}(v', v' + \theta) dv' = \int_0^1 c(F_0^{-1}(v'), (F_1^\theta)^{-1}(v')) dv' \quad (5.1)$$

the average cost (per period) of the transport plan  $\gamma_\theta$ . Observe that it is indifferent whether to integrate here from 0 to 1 or from  $v$  to  $v + 1$  for any real  $v$ . Examples of average cost functions  $C_{[F_0, F_1]}$  for different marginals  $\mu_0, \mu_1$  and different cost functions  $c$  are shown in fig. 5.1.

The following technical lemma provides a “bracket” for the global minimum of  $C_{[F_0, F_1]}$  and estimates of its derivatives independent of  $\mu_0, \mu_1$ .

LEMMA 5.2. *The average cost  $C_{[F_0, F_1]}$  is a convex function that satisfies the inequalities*

$$\inf_{x, y} c(x, y) \leq \underline{C}(\theta) \leq C_{[F_0, F_1]}(\theta) \leq \overline{C}(\theta) \quad (5.2)$$

with

$$\underline{C}(\theta) = \inf_{\substack{-1 \leq u_1 \leq 2 \\ \theta - 1 \leq u_2 \leq \theta + 2}} c(u_1, u_2), \quad \overline{C}(\theta) = \sup_{\substack{-1 \leq u_1 \leq 2 \\ \theta - 1 \leq u_2 \leq \theta + 2}} c(u_1, u_2). \quad (5.3)$$

There exist constants  $\underline{\Theta} < \overline{\Theta}$  and  $\underline{L}, \overline{L} > 0$  such that the global minimum of  $C$  is achieved on the interval  $[\underline{\Theta}, \overline{\Theta}]$  and

$$-\underline{L} \leq C'_{[F_0, F_1]}(\underline{\Theta} - 0) \leq 0 \leq C'_{[F_0, F_1]}(\overline{\Theta} + 0) \leq \overline{L}, \quad (5.4)$$

where  $C'_{[F_0, F_1]}(\cdot)$  is the derivative of  $C_{[F_0, F_1]}$ . These constants are independent on  $\mu_0, \mu_1$  and are given explicitly by formulas (5.5), (5.6) and (5.7) below.

The bounds given in the present lemma are rather loose. E.g., for  $c(x, y) = |x - y|^\alpha$  with  $\alpha > 1$ , they are  $\underline{C}(\theta) = \text{conv min}(|\theta + 3|^\alpha, |\theta - 3|^\alpha)$ ,  $\overline{C}(\theta) = \max(|\theta + 3|^\alpha, |\theta - 3|^\alpha)$ , and  $-\underline{\Theta} = \overline{\Theta} = 6$ . For symmetric costs like this one it is often possible to replace  $[\underline{\Theta}, \overline{\Theta}]$  by the interval  $[-1, 1]$  which may be tighter.

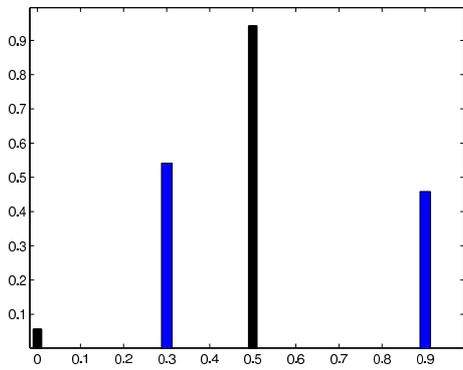
*Proof.* To prove convexity of  $C_{[F_0, F_1]}$  it is sufficient to show that  $C_{[F_0, F_1]}(\frac{1}{2}(\theta' + \theta'')) \leq \frac{1}{2}(C_{[F_0, F_1]}(\theta') + C_{[F_0, F_1]}(\theta''))$  for all  $\theta', \theta''$ . Let  $\theta' < \theta''$ , denote  $\theta = \frac{1}{2}(\theta' + \theta'')$  and write

$$C_{[F_0, F_1]}(\theta) = \int_0^1 c_{[F_0, F_1]}(v, v + \theta) dv = \int_{\theta - \theta'}^{\theta - \theta' + 1} c_{[F_0, F_1]}(v', v' + \theta) dv',$$

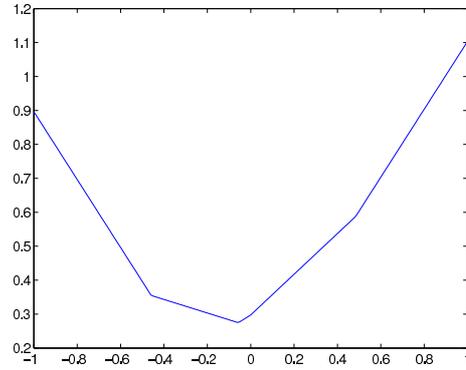
$$C_{[F_0, F_1]}(\theta') = \int_{\theta - \theta'}^{\theta - \theta' + 1} c_{[F_0, F_1]}(v', v' + \theta') dv', \quad C_{[F_0, F_1]}(\theta'') = \int_0^1 c_{[F_0, F_1]}(v, v + \theta'') dv.$$

Making the change of variables  $v' = v + \theta - \theta'$  and taking into account that  $\theta - \theta' + \theta = 2\theta - \theta' = \theta''$ , we get

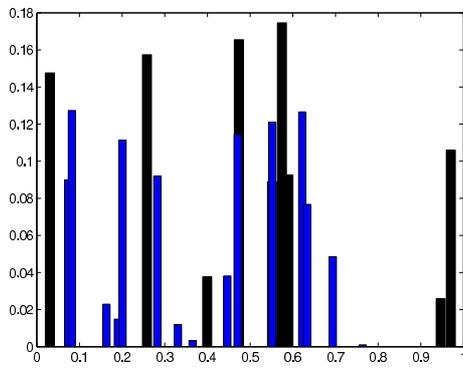
$$\begin{aligned} & 2C_{[F_0, F_1]}(\theta) - C_{[F_0, F_1]}(\theta') - C_{[F_0, F_1]}(\theta'') \\ &= \int_0^1 (c_{[F_0, F_1]}(v, v + \theta) + c_{[F_0, F_1]}(v + \theta - \theta', v + \theta'') \\ &\quad - c_{[F_0, F_1]}(v + \theta - \theta', v + \theta) - c_{[F_0, F_1]}(v, v + \theta'')) dv. \end{aligned}$$



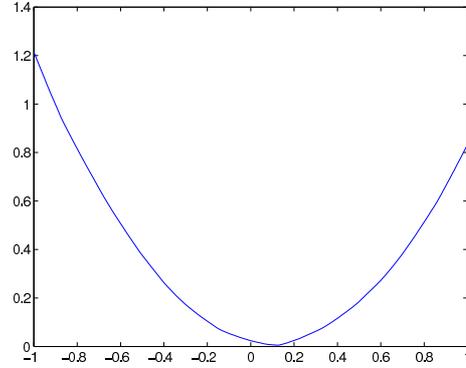
(a) Black bars:  $\mu'_0$ , light bars:  $\mu'_1$



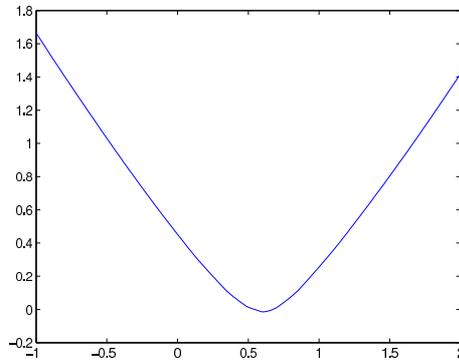
(b)  $C_{[F'_0, F'_1]}(\cdot)$  for  $c(x, y) = |x - y|$



(c) Black bars:  $\mu''_0$ , light bars:  $\mu''_1$



(d)  $C_{[F''_0, F''_1]}(\cdot)$  for  $c(x, y) = |x - y|^2$



(e)  $C_{[F''_0, F''_1]}(\cdot)$  for a non-symmetric cost  $c(x, y) = |0.5 + x - y|^{1.2} + 0.1 \cos(2\pi x + 1) - 0.3 \sin(2\pi y - 0.5)$

FIG. 5.1. Average cost functions  $C_{[F_0, F_1]}$  for some atomic marginals and costs. The cost  $|x - y|$  is regarded as  $\lim |x - y|^\lambda$ ,  $\lambda > 1$ , as  $\lambda \rightarrow 1$ .

Since  $v + \theta - \theta' > v$  and  $v + \theta' > v + \theta$ , the Monge condition for  $c$  implies that the integrand here is negative on a set of nonzero measure, yielding the desired inequality for the function  $C_{[F_0, F_1]}$ . Note that convexity of  $C_{[F_0, F_1]}$  implies its continuity because  $C_{[F_0, F_1]}$  is finite everywhere.

Bounds (5.2) on  $C_{[F_0, F_1]}(\theta)$  follow from (5.1) with  $v = 0$  because  $v' - 1 \leq F_0^{-1}(v') \leq v' + 1$ ,  $v' + \theta - 1 \leq (F_1^\theta)^{-1}(v') \leq v' + \theta + 1$ , and  $0 \leq v' \leq 1$ . Furthermore, the growth condition (3.3) implies that  $\underline{C}(\theta) \geq P$  as soon as  $|\theta| > R(P) + 3$ . Indeed, in this case  $|u_2 - u_1| \geq |\theta| - 3 \geq R(P)$  and right-hand sides of formulas (5.3) are bounded by  $P$  from below. Therefore one can set

$$\underline{\Theta} = \inf\{\theta: \underline{C}(\theta) = \min_{\theta'} \overline{C}(\theta')\} > -\infty, \quad \overline{\Theta} = \sup\{\theta: \underline{C}(\theta) = \min_{\theta'} \overline{C}(\theta')\} < \infty, \quad (5.5)$$

where min is attained because  $C$  is continuous.

The set  $\arg \min_{\theta'} C_{[F_0, F_1]}(\theta')$  lies on the segment  $[\underline{\Theta}, \overline{\Theta}]$ . (Indeed, if e.g.  $\theta < \underline{\Theta}$ , then  $C_{[F_0, F_1]}(\theta) \geq \underline{C}(\theta) > \min_{\theta'} \overline{C}(\theta') \geq \min_{\theta'} C_{[F_0, F_1]}(\theta')$ , so  $\theta$  cannot belong to  $\arg \min_{\theta'} C_{[F_0, F_1]}(\theta')$ ; a similar conclusion holds if  $\theta > \overline{\Theta}$ .) It follows that  $C'_{[F_0, F_1]}(\underline{\Theta} - 0) \leq 0 \leq C'_{[F_0, F_1]}(\overline{\Theta} + 0)$ .

By convexity  $C'_{[F_0, F_1]}(\overline{\Theta} + 0) \leq (C_{[F_0, F_1]}(\theta) - C_{[F_0, F_1]}(\overline{\Theta})) / (\theta - \overline{\Theta})$  for all  $\theta \geq \overline{\Theta}$ . The right-hand side of the latter inequality can be estimated from above by

$$\overline{L} = \inf_{\theta \geq \overline{\Theta}} \frac{\overline{C}(\theta) - \underline{C}(\overline{\Theta})}{\theta - \overline{\Theta}}. \quad (5.6)$$

The ratio in the right-hand side takes finite values, so  $\overline{L}$  is finite. This establishes the inequality  $C'_{[F_0, F_1]}(\overline{\Theta} + 0) \leq \overline{L}$ . The rest of (5.4) is given by a symmetrical argument; in particular

$$\underline{L} = \inf_{\theta \leq \underline{\Theta}} \frac{\overline{C}(\theta) - \underline{C}(\underline{\Theta})}{\underline{\Theta} - \theta}. \quad (5.7)$$

□

**DEFINITION 5.3.** *A locally optimal transport plan  $\gamma_{\theta_0}$  is called globally optimal if  $\theta_0 \in \arg \min_{\theta} C_{[F_0, F_1]}(\theta)$ .*

We can now reduce minimization of (2.1) on the unit circle to minimization of (2.2) on  $\mathbb{R}$ , which involves the cost function  $c$  rather than  $\hat{c}$ :

**THEOREM 5.4.** *The canonical projection  $\pi: \mathbb{R} \rightarrow \mathbb{T}$  establishes a bijection between globally optimal transport plans on  $\mathbb{R} \times \mathbb{R}$  and transport plans on  $\mathbb{T} \times \mathbb{T}$  that minimize (2.1).*

*Proof.* A transport plan  $\gamma$  on  $\mathbb{T} \times \mathbb{T}$  minimizes (2.1) if it is a projection of a transport plan on  $\mathbb{R} \times \mathbb{R}$  that locally minimizes the transport cost defined by the cost function  $\hat{c}(x, y) = \min_{k \in \mathbb{Z}} c(x, y + k)$  (see introduction; min here is attained because of the integer shift invariance and growth conditions (3.2), (3.3)).

Denote  $S = \{(x, y): c(x, y) = \hat{c}(x, y)\}$  and observe that the support of the globally optimal plan  $\gamma_{\theta_0}$  lies within  $S$ : indeed, if it did not, there would exist a (nonlocal but periodic) modification of  $\gamma_{\theta_0}$  bringing some of the mass of each period to  $S$  and thus reducing the average cost. Therefore  $\gamma_{\theta_0}$  is locally optimal with respect to the cost  $\hat{c}(x, y)$  and its projection to  $\mathbb{T} \times \mathbb{T}$  minimizes (2.1).

Conversely, a minimizing transport plan on  $\mathbb{T} \times \mathbb{T}$  can be lifted to  $\mathbb{R} \times \mathbb{R}$  in such a way that its support lies inside  $S$  (translations of arbitrary pieces of support by integer increments along  $x$  and  $y$  axes are allowed because they leave  $\hat{c}(x, y)$  invariant). Therefore its average cost per period cannot be less than that of a globally optimal transport plan on  $\mathbb{R} \times \mathbb{R}$ . □

**6. Fast global transport optimization.** In a typical application, such as the image processing problem described in the introduction, measures  $\mu_0$  and  $\mu_1$  come in the form of *histograms*, i.e., discrete distributions supported on subsets  $\hat{X} = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{n_0}\}$  and  $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{n_1}\}$  of the unit circle. These two sets may coincide. In what follows we replace  $\hat{X}$  and  $\hat{Y}$  with their lifts to the universal cover  $X$  and  $Y$  and assume that the points of the latter pair of sets are sorted and numbered in an increasing order:

$$\begin{aligned} \dots < x_{-1} < x_0 = x_{n_0} - 1 \leq 0 < x_1 < \dots < x_{n_0} \leq 1 < x_{n_0+1} = x_1 + 1 < \dots, \\ \dots < y_{-1} < y_0 = y_{n_1} - 1 \leq 0 < y_1 < \dots < y_{n_1} \leq 1 < y_{n_1+1} = y_1 + 1 < \dots \end{aligned}$$

Denote masses of these points by  $\mu_0(\{x_i\}) = m_i^{(0)}$ ,  $\mu_1(\{y_j\}) = m_j^{(1)}$ ; these are assumed to be arbitrary positive real numbers satisfying  $\sum_{1 \leq i \leq n_0} m_i^{(0)} = \sum_{1 \leq j \leq n_1} m_j^{(1)} = 1$ .

**6.1. Computation of the average cost and its derivative.** Define  $j(\theta)$  as the index of  $\min\{y_j: F_1^\theta(y_j) > 0\}$  and denote  $y_1^\theta = y_{j(\theta)}$ ,  $y_2^\theta = y_{j(\theta)+1}$ ,  $\dots$ ,  $y_{n_1}^\theta = y_{j(\theta)+n_1-1}$ . All the values

$$F_0(x_1), F_0(x_2), \dots, F_0(x_{n_0}), F_1^\theta(y_1^\theta), F_1^\theta(y_2^\theta), \dots, F_1^\theta(y_{n_1}^\theta) \quad (6.1)$$

belong to the segment  $(0, 1]$ . We now sort these values into an increasing sequence, denote its elements by  $v_{(1)} \leq v_{(2)} \leq \dots \leq v_{(n_0+n_1)}$  and set  $v_{(0)} = 0$ . Note that for each  $v$  such that  $v_{(k-1)} < v < v_{(k)}$  with  $1 \leq k \leq n_0 + n_1$  the values  $x_{(k)} = F_0^{-1}(v)$  and  $y_{(k)} = (F_1^\theta)^{-1}(v)$  are uniquely defined and belong to  $X, Y$ . It is now easy to write an expression for the function  $C_{[F_0, F_1]}$ :

$$C_{[F_0, F_1]}(\theta) = \sum_{1 \leq k \leq n_0+n_1} c(x_{(k)}, y_{(k)}) (v_{(k)} - v_{(k-1)}). \quad (6.2)$$

Observe that, as the parameter  $\theta$  increases by  $\Delta\theta$ , those  $v_{(k)}$  that correspond to values  $F_1^\theta$  decrease by the same increment. Let  $F_1^\theta(y_{j_0})$  be such a value. As it appears in (6.2) twice, first as  $v_{(k)}$  and then as  $-v_{(k-1)}$  in the next term of the sum, it will make two contributions to the derivative  $C'_{[F_0, F_1]}(\theta)$ :  $-c(F_0^{-1}(F_1^\theta(y_{j_0})), y_{j_0})$  and  $c(F_0^{-1}(F_1^\theta(y_{j_0})), y_{j_0+1})$  (see fig. 6.1, top).

Moreover, there are exceptional values of  $\theta$  for which two of the values in (6.1) coincide and their ordering in the sequence  $(v_{(k)})$  changes. For such values of  $\theta$  the derivative  $C'_{[F_0, F_1]}$  has different right and left limits, as illustrated in fig. 6.1, bottom:

$$C'_{[F_0, F_1]}(\theta - 0) = \sum_{1 \leq j \leq n_1} (c(F_0^{-1}(F_1^\theta(y_j)), y_{j+1}) - c(F_0^{-1}(F_1^\theta(y_j)), y_j)), \quad (6.3)$$

$$C'_{[F_0, F_1]}(\theta + 0) = \sum_{1 \leq j \leq n_1} (c(F_0^{-1}(F_1^\theta(y_j) - 0), y_{j+1}) - c(F_0^{-1}(F_1^\theta(y_j) - 0), y_j)). \quad (6.4)$$

If  $\theta$  is not exceptional, the value of  $C'_{[F_0, F_1]}(\theta)$  is given by the first of these formulas.

The function  $C_{[F_0, F_1]}$  is therefore piecewise affine (see in particular fig. 5.1, where this function is plotted for atomic marginals  $\mu_0, \mu_1$ ). Moreover, from the Monge condition (3.1) it follows that  $C'_{[F_0, F_1]}(\theta - 0) < C'_{[F_0, F_1]}(\theta + 0)$  at exceptional points, giving an alternative proof of convexity of  $C_{[F_0, F_1]}(\theta)$  in the discrete case.

**LEMMA 6.1.** *Values of  $C$  and its left and right derivatives can be computed for any  $\theta$  using at most  $O(n_0 + n_1)$  comparisons and evaluations of  $c(x, y)$ .*

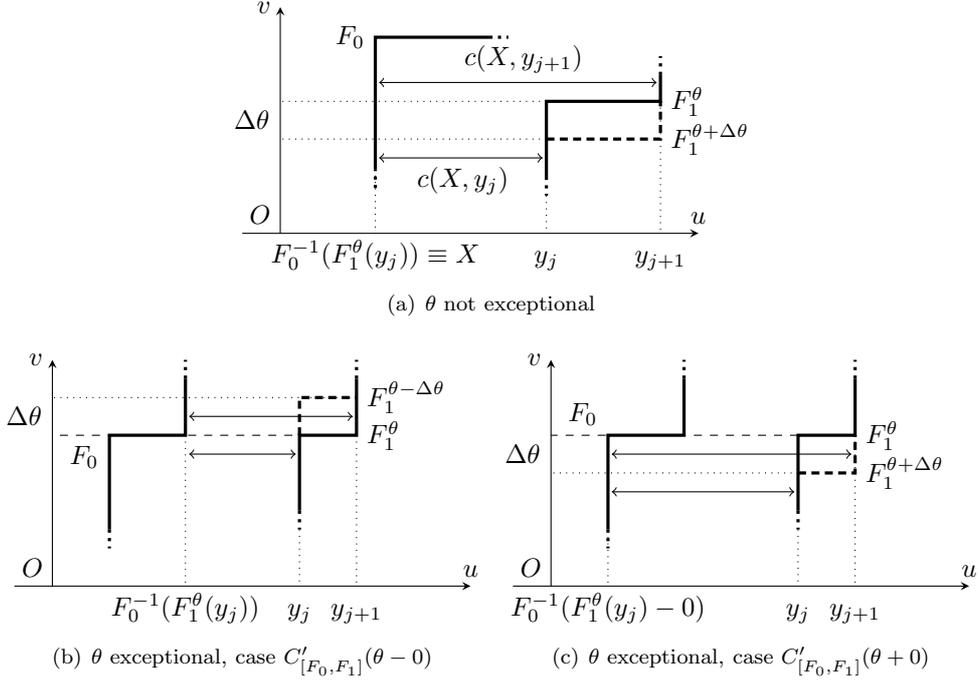


FIG. 6.1. Derivation of expressions (6.3), (6.4) for  $C'_{[F_0, F_1]}(\theta \pm 0)$ . Thick lines show fragments of complete graphs of  $F_0$ ,  $F_1^\theta$  corresponding to  $j$ th terms in (6.3), (6.4), thin dashed line (bottom) marks the common value of  $F_0$  and  $F_1^\theta$ . Note that  $X = F_0^{-1}(F_1^\theta(y_j))$  is equivalently expressed as  $\inf\{x: F_0(x) > F_1^\theta(y_j)\}$ .

*Proof.* Sorting the  $n_0 + n_1$  values (6.1) into an increasing sequence requires  $n_0 + n_1 - 1$  comparisons (one starts with comparing  $F_0(x_1)$  and  $F_1^\theta(y_1^\theta)$  to determine  $v_{(1)}$ , and after this each of the remaining values is considered once until there remains only one value, which is assigned to  $v_{(n_0+n_1)}$  with no further comparison). At the same time, pointers to  $x_{(k)}$  and  $y_{(k)}$  should be stored. After this preliminary stage, to find the values for  $C'_{[F_0, F_1]}$  and its one-sided derivatives it suffices to evaluate each of the  $n_0 + n_1$  terms in (6.2) and to take into account the corresponding contribution of plus or minus  $c(x_{(k)}, y_{(k)})$  to the value of  $C'_{[F_0, F_1]}(\theta)$ , paying attention to whether the value of  $\theta$  is exceptional or not. All this can again be done in  $O(n_0 + n_1)$  operations.  $\square$

**6.2. Transport optimization algorithm.** Fix  $\epsilon > 0$  and set  $L = \max\{\underline{L}, \overline{L}\}$ . Recall that  $\underline{L}, \overline{L}$ , as well as the parameters  $\underline{\Theta}, \overline{\Theta}$  that are used in the algorithm below, are defined by explicit formulas in Lemma 5.2 and do not depend on measures  $\mu_0, \mu_1$ . The minimum of  $C'_{[F_0, F_1]}(\theta)$  can be found to accuracy  $\epsilon$  using the following binary search technique:

1. Initially set  $\underline{\theta} := \underline{\Theta}$  and  $\overline{\theta} := \overline{\Theta}$ , where  $\underline{\Theta}, \overline{\Theta}$  are defined in Lemma 5.2.
2. Set  $\theta := \frac{1}{2}(\underline{\theta} + \overline{\theta})$ .
3. Compute  $C'_{[F_0, F_1]}(\theta - 0)$ ,  $C'_{[F_0, F_1]}(\theta + 0)$ .
4. If  $C'_{[F_0, F_1]}(\theta - 0) \leq 0 \leq C'_{[F_0, F_1]}(\theta + 0)$ , then  $\theta$  is the required minimum; stop.

5. If  $\bar{\theta} - \underline{\theta} < \epsilon/L$ , then compute  $C_{[F_0, F_1]}(\underline{\theta})$ ,  $C_{[F_0, F_1]}(\bar{\theta})$ , solve the linear equation

$$C_{[F_0, F_1]}(\underline{\theta}) + C'_{[F_0, F_1]}(\underline{\theta} + 0)(\theta - \underline{\theta}) = C_{[F_0, F_1]}(\bar{\theta}) + C'_{[F_0, F_1]}(\bar{\theta} - 0)(\theta - \bar{\theta}) \quad (6.5)$$

for  $\theta$ , and stop.

6. Otherwise set  $\underline{\theta} := \theta$  if  $C'_{[F_0, F_1]}(\theta + 0) < 0$ , or  $\bar{\theta} := \theta$  if  $C'_{[F_0, F_1]}(\theta - 0) > 0$ .

7. Go to step 2.

It follows from inequalities (5.4) of Lemma 5.2 that the minimizing value of  $\theta$  belongs to the segment  $[\underline{\theta}, \bar{\theta}]$ . Therefore at all steps

$$C'_{[F_0, F_1]}(\underline{\theta} + 0) \leq 0 \leq C'_{[F_0, F_1]}(\bar{\theta} - 0) \quad (6.6)$$

and the segment  $[\underline{\theta}, \bar{\theta}]$  contains the minimum of  $C$ .

Step 5 requires some comments. By convexity,  $-\underline{L} \leq C'_{[F_0, F_1]}(\theta \pm 0) \leq \bar{L}$  for all  $\underline{\theta} \leq \theta \leq \bar{\theta}$ , i.e.,  $|C'_{[F_0, F_1]}(\theta \pm 0)| \leq L$  at all steps. When  $\bar{\theta} - \underline{\theta} < \epsilon/L$ , this bound ensures that for any  $\theta'$  in  $[\underline{\theta}, \bar{\theta}]$  the minimal value of  $C$  is within  $\epsilon/L \cdot L = \epsilon$  from  $C_{[F_0, F_1]}(\theta')$ . If there is a single exceptional value of  $\theta$  in that interval, then it is located precisely at the solution of (6.5) and must be a minimum of  $C$  because of (6.6), so the final value of  $\theta$  is the exact solution; otherwise it is an approximation with guaranteed accuracy.

The final value of  $\theta$  will certainly be exact when masses of all atoms are rational numbers having the least common denominator  $M$  and  $\epsilon < 1/M$ . Indeed, in this case any interval  $[\underline{\theta}, \bar{\theta}]$  of length  $\epsilon$  can contain at most one exceptional value of  $\theta$ .

Since at each iteration the interval  $[\underline{\theta}, \bar{\theta}]$  is halved, step 5 will be achieved in  $O(\log_2((\bar{\theta} - \underline{\theta})/(\epsilon/L)))$  iterations. By Lemma 6.1 each instance of step 3 (and equation (6.5)) takes  $O(n_0 + n_1)$  operations. Thus we obtain the following result.

**THEOREM 6.2.** *The above binary search algorithm takes  $O((n_0 + n_1) \log(1/\epsilon))$  comparisons and evaluations of  $c(x, y)$  to terminate. The final value of  $\theta$  is within  $\epsilon/L$  from the global minimum, and  $C_{[F_0, F_1]}(\theta) \leq \min_{\theta} C_{[F_0, F_1]}(\theta) + \epsilon$ . When all masses  $m_i^{(0)}$ ,  $m_j^{(1)}$  are rational with the least common denominator  $M$ , initializing the algorithm with  $\epsilon = 1/2M$  leads to an exact solution in  $O((n_0 + n_1) \log M)$  operations.*

**6.3. Experiments.** We tested experimentally the estimates of Theorem 6.2 for time complexity as a function of parameters of the problem. The average computing time of the algorithm for different values of  $n_0 + n_1$  and  $\epsilon$  is illustrated in fig. 6.2. These results have been obtained using the following procedure. For each value of  $n_0$  and  $n_1$ , points  $\{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{n_0}\}$  and  $\{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{n_1}\}$ , which constitute the support of distributions  $\mu_0$  and  $\mu_1$ , are drawn independently from the uniform distribution on  $[0, 1]$  and sorted. The masses  $m_i^{(0)}$ ,  $i = 1 \dots n_0$  and  $m_j^{(1)}$ ,  $j = 1 \dots n_1$  are then drawn from the uniform distribution and normalized such that  $\sum_{1 \leq i \leq n_0} m_i^{(0)} = \sum_{1 \leq j \leq n_1} m_j^{(1)} = 1$ . Finally the transport cost is minimized for  $c(x, y) = |x - y|$ . The code used to produce this figure is available online at the web site of the OTARIE project <http://www.mccme.ru/~ansobol/otarie/software.html>.

In the first experiment the value of  $\epsilon$  was set to  $10^{-10}$ , the algorithm was run 10 times for each pair  $(n_0, n_1)$  with  $1 \leq n_0, n_1 \leq 100$ , and the computing times were averaged. In the second experiment  $(n_0, n_1)$  was fixed at  $(10, 10)$  and the average computing time was similarly computed for different values of  $\epsilon$ . The averaged computing times for the two experiments are plotted in fig. 6.2. Observe the manifest linear dependence of computing time on  $n_0 + n_1$  and  $\log \epsilon$ .

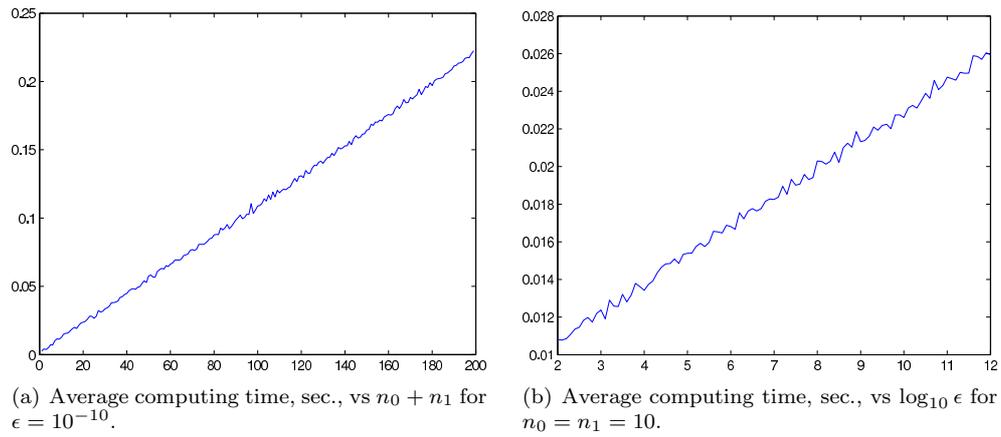


FIG. 6.2. Average computing time of the algorithm for different values of  $(n_0 + n_1)$  and  $\log_{10} \epsilon$ . The experiment was performed on a PC with a 3.00 GHz processor.



(a) Top: F. Maliavin, *Whirlwind* (1916).



(b) Right: P. Puvis de Chavanne, *Jeunes filles au bord de la mer* (1879).



(c) Puvis' hues transported to Maliavin's canvas.

FIG. 6.3. Optimal matching of the hue component of color.

The next figure is a concrete, if not entirely serious, illustration of optimal matching in the case of distributions on the “color circle.”

Recall that in the HSL (Hue, Saturation, and Lightness) color model, the color space is represented in cylindrical coordinates. The polar angle corresponds to the *hue*, or the degree to which a color can be described as similar to or different from other colors (as opposed to difference in saturation or lightness between shades of the same color).

We chose two famous paintings, one Russian and one French, whose highly different coloring is characteristic of the two painters, the Expressionist Filipp Maliavin (1869–1940) and the Symbolist Pierre Puvis de Chavannes (1824–1898). An optimal matching of the hue distributions according to the linear cost  $c(x, y) = |x - y|^2$  was used to substitute hues of the first painting with the corresponding hues of the second one while preserving the original values of saturation and brightness. In spite of the drastic change in coloring, the optimality of matching ensures that warm and cold colors retain their quality and the overall change of aspect does not feel arbitrary or artificial.

**7. Related algorithmic work.** Fast algorithms for the transportation problem on the circle, with the Euclidean distance  $|x - y|$  as a cost, have been proposed in a number of works. Karp and Li [13] consider an *unbalanced* matching, where the total mass of the two histograms are not equal and elements of the smaller mass have to be optimally matched to a subset of elements of the larger mass. A balanced optimal matching problem has later been considered independently by Werman et al [20]; clearly, the balanced problem can always be treated as a particular case of the unbalanced one. In both of these works  $O(n \log n)$  algorithms are obtained for the case where all points have unit mass.

Aggarwal et al [1] present an algorithm improving Karp and Li’s results for an unbalanced transportation problem on the circle with general integer weights and the same cost function  $|x - y|$ . They also consider a general cost function  $c(x, y)$  that satisfies the Monge condition and an additional condition of *bitonicity*: for each  $x$ , the function  $c(x, y)$  is nonincreasing in  $y$  for  $y < y_0(x)$  and nondecreasing in  $y$  for  $y > y_0(x)$ . Note that this rules out the circular case. The second algorithm of [1] is designed for bitonic Monge costs and runs in  $O(n \log M)$  time for an unbalanced transportation problem with integer weights on the line, where  $M$  is the total weight of the matched mass and  $n$  is the number of points in the larger histogram.

The algorithm proposed in the present article only applies to the *balanced* problem for a Monge cost. However it does not involve bitonicity and is therefore applicable on the circle, where it achieves the same  $O(n \log M)$  time as the second algorithm of [1] if all weights are integer multiples of  $1/M$ . Although our theory is developed for the case of costs satisfying a strict inequality in the Monge condition, it can be checked that the discrete algorithm works for the case  $c(x, y) = |x - y|$ , which can be treated as a limit of  $|x - y|^\lambda$ ,  $\lambda > 1$ , as  $\lambda \rightarrow 1$  [17].

Finally we note that results of [1] were extended in a different direction by McCann [16], who provides, again in the balanced setting, a generalization of their first algorithm to the case of a general cost of the *concave* type on the open line. This case is opposite to Monge costs and requires completely different tools. Indeed, for a strictly concave cost such as  $c(x, y) = \sqrt{|x - y|}$  the notion of locally optimal transport plan on the universal cover does not make sense: concave costs favor long-haul transport over local rearrangements, destroying local finiteness.

## REFERENCES

- [1] A. AGGARWAL, A. BAR-NOY, S. KHULLER, D. KRAVETS, AND B. SCHIEBER, *Efficient minimum cost matching using quadrangle inequality*, in Foundations of Computer Science, 1992. Proceedings of 33rd Annual Symposium, 1992, pp. 583–592.
- [2] L. AMBROSIO, N. GIGLI, AND G. SAVARÉ, *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*, Birkhauser, 2005.
- [3] S. AUBRY AND P. Y. LE DAERON, *The discrete Frenkel–Kontorova model and its extensions I. Exact results for the ground-states*, Physica D: Nonlinear Phenomena, 8 (1983), pp. 381–422.
- [4] R. VENKATESH BABU, P. PÉREZ, AND P. BOUTHEMY, *Robust tracking with motion estimation and local Kernel-based color modeling*, Image and Vision Computing, 25 (2007), pp. 1205–1216.
- [5] P. BERNARD AND B. BUFFONI, *Optimal mass transportation and Mather theory*, Journal of the European Mathematical Society, 9 (2007), pp. 85–121.
- [6] M. BROWN, R. SZELISKI, AND S. WINDER, *Multi-image matching using multi-scale oriented patches*, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005, vol. 1, 2005.
- [7] R. E. BURKARD, B. KLINZ, AND R. RUDOLF, *Perspectives of Monge properties in optimization*, Discrete Appl. Math., 70 (1996), pp. 95–161.
- [8] D. CORDERO-ERAUSQUIN, *Sur le transport de mesures périodiques*, Comptes Rendus de l’Académie des Sciences - Series I - Mathematics, 329 (1999), pp. 199–202.
- [9] A. FATHI, *Weak KAM Theorem in Lagrangian Dynamics*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, February 2009.
- [10] M. FELDMAN AND R. J. MCCANN, *Monge’s transport problem on a Riemannian manifold*, Trans. Amer. Math. Soc., 354 (2002).
- [11] W. GANGBO AND R. J. MCCANN, *Shape recognition via Wasserstein distance*, Quarterly of Applied Mathematics, 58 (2000), pp. 705–738.
- [12] W. GANGBO AND R. J. MCCANN, *The geometry of optimal transportation*, Acta Math., 177 (1996), pp. 113–161.
- [13] R. M. KARP AND S. Y. R. LI, *Two special cases of the assignment problem*, Discrete Mathematics, 13 (1975), pp. 129–142.
- [14] O. KNILL, *Jürgen Moser, selected chapters in the calculus of variations*, Birkhäuser Verlag, 2003.
- [15] D. G. LOWE, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, 60 (2004), pp. 91–110.
- [16] R. J. MCCANN, *Exact solutions to the transportation problem on the line*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 455 (1999), pp. 1341–1380.
- [17] J. RABIN, J. DELON, AND Y. GOUSSEAU, *Transportation distances on the circle*, arXiv:0906.5499 (2009).
- [18] Y. RUBNER, C. TOMASI, AND L.J. GUIBAS, *The Earth Mover’s Distance as a Metric for Image Retrieval*, International Journal of Computer Vision, 40 (2000), pp. 99–121.
- [19] C. VILLANI, *Optimal transport: Old and new*, vol. 338 of Grundlehren der mathematischen Wissenschaften, Springer-Verlag, Dec 2009.
- [20] M. WERMAN, S. PELEG, R. MELTER, AND TY KONG, *Bipartite graph matching for points on a line or a circle*, Journal of Algorithms, 7 (1986), pp. 277–284.
- [21] J. ZHANG, M. MARSZALEK, S. LAZEBNIK, AND C. SCHMID, *Local features and kernels for classification of texture and object categories: A comprehensive study*, International Journal of Computer Vision, 73 (2007), pp. 213–238.