Fast transport optimization on the circle

Julie Delon, Julien Salomon, Andrei Sobolevskii

FAST TRANSPORT OPTIMIZATION ON THE CIRCLE

JULIE DELON, JULIEN SALOMON, AND ANDREĬ SOBOLEVSKIĬ

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.

This theory is applied to a transportation problem arising in image processing: for two sets of point masses, both of which have the same total mass, find an optimal transport plan with respect to a given cost function c that satisfies the Monge condition. For the case of N real-valued point masses we present an $O(N \log \epsilon)$ algorithm that approximates the optimal cost within ϵ; when all masses are integer multiples of $1/M$, the algorithm gives an exact solution in $O(N \log M)$ operations.

1. Introduction

This work is motivated by a transport optimization problem arising in image processing [13]. A grayscale photograph may be characterized locally by the distribution of directions of brightness gradient. If two angular distributions of gradient for two different photographs match particularly well, this may indicate that the photographs feature the same object, even if it is slightly distorted or appears under different light conditions.

An effective measure of the quality of matching is provided by an optimal value of the *transport cost*. 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 $\mathbb{T}$, it is defined as the inf of the quantity

$$\hat{I}(\gamma) = \iint_{\mathbb{T}\times\mathbb{T}} \hat{c}(\hat{x}, \hat{y})\, \gamma(\mathrm{d}\hat{x} \times \mathrm{d}\hat{y}). \tag{1}$$

over the set of all *couplings* γ 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.*

Transport optimization on the circle appears in a number of applications and has long been studied from theoretic as well as algorithmic points of view (see, e.g., [10, 16, 1, 6]). In this paper we propose an efficient algorithm for minimizing (1) when the marginal measures are discrete, which is based on an analogy with the

Date: February 18, 2009.

2000 *Mathematics Subject Classification.* Primary 90C08; Secondary 68Q25, 90C25.

This work was started during the visit of JD and JS to UMI 2615 CNRS "Laboratoire J.-V. Poncelet," supported by ANR through grant ANR-07-BLAN-0235 OTARIE (`http://www.mccme.ru/~ansobol/otarie/`); AS thanks the Ministry of National Education of France for supporting his visit to the Observatoire de la Côte d'Azur, where part of this text was written, and to Russian Fund for Basic Research for the partial support via grant RFBR 07–01–92217-CNRSL-a.

weak KAM (Aubry–Mather) theory. The rest of this introduction contains an informal overview of our results; their relations to previous work are discussed in more detail in Section 5.

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 finite measures μ_0, μ_1, and redefine γ to be their coupling on $\mathbb{R}\times\mathbb{R}$. It is then convenient to replace the problem of minimizing the integral (1) with "minimization" of an integral

$$I(\gamma) = \iint_{\mathbb{R}\times\mathbb{R}} c(x,y)\,\gamma(\mathrm{d}x\times\mathrm{d}y). \tag{2}$$

Although the latter integral is infinite, it still makes sense to look for transport plans γ minimizing I with respect to *local* modifications, i.e., to require that for any compactly supported signed measure δ 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*:

$$c(x_1,y_1) + c(x_2,y_2) < c(x_1,y_2) + c(x_2,y_1) \tag{3}$$

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 $\mathrm{MK}_\lambda(\hat{\mu}_0,\hat{\mu}_1) = (\inf_\gamma \hat{I}(\gamma))^{1/\lambda}$ turns out to be a metric on the set of measures on the circle, referred to as the *Monge–Kantorovich distance* of order λ. The particular case $\lambda = 1$ is sometimes called the Kantorovich–Rubinshtein metric or, in image processing literature, the Earth Mover's distance [14].

The Monge condition (3) implies that whenever a transport plan reverses the mutual order of any two elements of mass, the transport cost can be strictly reduced by exchanging their destinations. It follows that a locally minimal transport plan must move elements of mass *monotonically*, preserving their spatial order.

The whole set of locally optimal transport plans for a given pair of marginals μ_0, μ_1 can be conveniently described using the following construction (fig. 1). Let F_0, F_1 be cumulative distribution functions of the measures μ_0, μ_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 (or, equivalently, to atoms of μ_0, μ_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 induces the relative measure (μ_0 or μ_1) on the Ou axis. 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 θ followed by an application of the correspondence F_1^{-1}, and still induces on the Ou axis the same measure μ_1 as F_1. A transport plan γ_θ 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 μ_0 and μ_1, and thus a locally optimal transport plan. Moreover, it is shown in Section 3 that all locally optimal transport plans can be obtained using this construction for different values of the parameter θ.

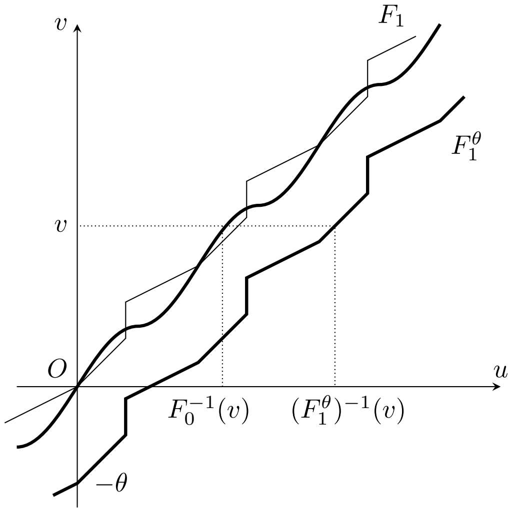

FIGURE 1. Construction of the locally optimal transport plan γ_θ.

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

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

It is shown in Section 4 that the Monge condition implies convexity of $C_{[F_0,F_1]}(\theta)$ and that its global minimum in θ coincides with the minimum value of the transport cost on the unit circle (1).

When the marginals μ_0, μ_1 are purely atomic with finite numbers n_0 and n_1 of atoms in each period, the function C becomes piecewise affine. In Section 4 we present an algorithm to approximate its minimum value to accuracy ϵ, 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 approximate solution turns out to be exact provided that $\epsilon < 1/M$. This gives an $O((n_0+n_1)\log M)$ exact transport optimization algorithm on the circle.

2. Preliminaries

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

2.1. **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 $y_1 < y_2$,

$$c(x_1,y_1) + c(x_2,y_2) - c(x_1,y_2) - c(x_2,y_1) < 0. \tag{4}$$

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

$$c(x+1,y+1) = c(x,y) \tag{5}$$

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

$$c(x,y) \geq P \quad \text{whenever } |x-y| \geq R(P). \tag{6}$$

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 (4) 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 (1), and, more generally, for any function of the form $c(x-y) + f(x) + g(y)$ with 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 written on the universal cover as $\hat{c}(x,y) = \inf_{k\in\mathbb{Z}} c(x,y+k)$.

Condition (4) is all that is needed in Section 3, which is concerned with general locally optimal transport plans on $\mathbb{R}$. Conditions (5), (6) come into play in Section 4, which deals with transport optimization on the circle.

2.2. **Distribution functions.** For a given locally finite measure μ on $\mathbb{R}$ define its *distribution function* F_μ by

$$F_\mu(0) = 0, \quad F_\mu(x) = \mu((0,x]) \text{ for } x > 0, \quad F_\mu(x) = -\mu((x,0]) \text{ for } x < 0. \tag{7}$$

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_μ by an additive constant (the normalization $F_\mu(0) = 0$ is arbitrary). When μ is periodic with unit mass in each period, the equality

$$F_\mu(x+1) = F_\mu(x) + 1 \tag{8}$$

holds for all x in $\mathbb{R}$.

The *inverse* of a distribution function F_μ is defined by

$$F_\mu^{-1}(y) = \inf\{x\colon y < F_\mu(x)\} = \sup\{x\colon y \geq F_\mu(x)\}. \tag{9}$$

Definitions (7) and (9) mean that F_μ, F_μ^{-1} are right-continuous. Discontinuities of F_μ correspond to atoms of μ and discontinuities of its inverse, to "vacua" of μ, i.e., to intervals of zero μ measure.

For a distribution function F_μ define its *complete graph* to be the continuous curve formed by the union of the graph of F_μ and the vertical segments corresponding to jumps of F_μ. 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.

2.3. **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 μ_0, μ_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 1. *A* (locally finite)[1] transport plan *with marginals* μ_0 *and* μ_1 *is a locally finite measure* γ *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;*

(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 γ. Condition (i) 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 2. *A* local modification *of the locally finite transport plan* γ *is a transport plan* γ' *such that* γ *and* γ' *have the same marginals and* $\gamma' - \gamma$ *is a compactly supported finite signed measure. A local modification is called* cost-reducing *if*

$$\iint c(x, y)\,(\gamma'(\mathrm{d}x \times \mathrm{d}y) - \gamma(\mathrm{d}x \times \mathrm{d}y)) < 0.$$

A locally finite transport plan γ *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.*

3. Conjugate transport plans and shifts

Let U_0, U_1 be two copies of $\mathbb{R}$ respectively equipped with positive periodic measures μ_0, μ_1 whose distribution functions F_0, F_1 satisfy (8), 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.

3.1. Normal plans and conjugation.

We introduce the following terminology:

Definition 3. *A locally finite transport plan* ν *on* $V_0 \times V_1$ *with uniform marginals is called* normal.

Definition 4. *For a normal transport plan* ν *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

$$\nu^{[F_0,F_1]}(A \times B) = \nu(F_0(A) \times F_1(B)). \tag{10}$$

Lemma 5. *For a normal transport plan* ν *its conjugate* $\nu^{[F_0,F_1]}$ *is a locally finite transport plan on* $U_0 \times U_1$ *with marginals* μ_0, μ_1.

Proof. Since distribution functions F_0, F_1 and their inverses preserve boundedness, condition (i) of Definition 1 is fulfilled. Definition 4, condition (ii) of Definition 1, and formula (7) 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 (ii) of Definition 1 on intervals and therefore on all Borel sets. □

[1]In what follows the words 'locally finite' defining a transport plan will often be dropped.

Lemma 6. *For any transport plan γ on $U_0 \times U_1$ with marginals μ_0 and μ_1 there exists a normal transport plan ν such that γ is conjugate to ν: $\gamma = \nu^{[F_0,F_1]}$.*

Proof. For non-atomic measures μ_0 and μ_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 (10). However if, e.g., μ_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 μ_0 be located in $(0,1]$ at points u_1, u_2, ... with respective masses m_1, m_2, ... Since $\gamma(\{u_i\} \times U_1) = \mu_0(\{u_i\}) = m_i > 0$, there exists a conditional probability measure $\rho(\cdot \mid u_i) = \gamma(\{u_i\} \times \cdot)/m_i$. For a set $A \subset (0,1]$ define a "residue" transport plan

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

where δ_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 γ whose projection to the first factor is atomic. Define a transport plan κ on $V_0 \times U_1$ by

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

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 κ in place of γ, we get a normal transport plan ν such that $\gamma(A \times B) = \nu(F_0(A) \times F_1(B))$. □

Since we are ultimately interested in transport optimization with marginals μ_0, μ_1 rather than with uniform marginals, two normal transport plans ν_1, ν_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 μ_0 or μ_1 have atoms, causing loss of information on the structure of ν in segments corresponding to these atoms. The proof of Lemma 6 gives a specific representative of this equivalence class of normal plans.

3.2. **Locally optimal normal transport plans are shifts.** Fix a cost function $c\colon U_0 \times U_1 \to \mathbb{R}$ that satisfies the Monge condition (4) and define

$$c_{[F_0,F_1]}(v_0, v_1) = c\big(F_0^{-1}(v_0), F_1^{-1}(v_1)\big). \tag{11}$$

For non-atomic measures μ_0, μ_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 (μ_0 or μ_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 Section 2.1, we will still call $c_{[F_0,F_1]}$ a cost function.

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

Lemma 7. *A transport plan γ on $U_0 \times U_1$ with marginals μ_0, μ_1 is c-locally optimal if and only if it is conjugate to a $c_{[F_0,F_1]}$-locally optimal normal transport plan ν. 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

$$\iint c_{[F_0,F_1]}(v_1, v_2)\,\big(\nu'(\mathrm{d}v_1 \times \mathrm{d}v_2) - \nu(\mathrm{d}v_1 \times \mathrm{d}v_2)\big)$$
$$= \iint c(u_1, u_2)\,\big(\gamma'(\mathrm{d}u_1 \times \mathrm{d}u_2) - \gamma(\mathrm{d}u_1 \times \mathrm{d}u_2)\big)$$

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). □

Transport optimization with marginals μ_0, μ_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 θ. More precisely, the following holds:

Theorem 8. *Let μ_0, μ_1 be two periodic positive measures defined respectively on U_0, U_1 with unit mass in each period and let $F_i \colon U_i \to V_i$, $i = 0, 1$, 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 ν_θ with* $\operatorname{supp} \nu_\theta = \{(v, w) \colon w = v + \theta\}$*, and conversely ν_θ is $c_{[F_0,F_1]}$-locally optimal for any real θ. All c-locally optimal transport plans on $U_0 \times U_1$ with marginals μ_0, μ_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 monotonicaly supported plan with uniform marginals can only be a shift. Then Lemma 7 is used to extend this result to transport plans on $U_0 \times U_1$.

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

$$r_\nu(v, w) = \nu((-\infty, v] \times (w, \infty)), \quad l_\nu(v, w) = \nu((v, \infty) \times (-\infty, w]). \tag{12}$$

To explain the notation r_v, l_v 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 9. *The function r_ν (resp. l_ν) 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 (12). To prove continuity observe that the second marginal of ν 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_ν, l_ν with respect to the second argument. A similar proof holds for the first argument. □

Lemma 10. *For any v there exist $w_\nu(v)$ and $m_\nu(v) \geq 0$ such that*

$$r_\nu(v, w_\nu(v)) = l_\nu(v, w_\nu(v)) = m_\nu(v). \tag{13}$$

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 (13). 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 12 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 $r_\nu(v_2, w) > r_\nu(v_1, w_\nu(v)) = l_\nu(v_1, w_\nu(v)) > l_\nu(v_2, w)$. □

Equalities (13) mean that the same amount of mass $m_\nu(v)$ goes under the plan ν 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 11. *For any v there exists a local modification ν_v of ν such that $w_{\nu_v}(v) = w_\nu(v)$ (with w_ν defined as in Lemma 10), $m_{\nu_v}(v) = 0$, and ν_v is either cost-reducing in the sense of Definition 2 or is equivalent to ν.*

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

$$w^- = \sup\{w' \colon l_\nu(v, w') = 0\}, \qquad w^+ = \inf\{w' \colon r_\nu(v, w') = 0\},$$
$$v^- = \sup\{v' \colon r_\nu(v', w) = 0\}, \qquad v^+ = \inf\{v' \colon l_\nu(v', w) = 0\}.$$

By local finiteness of the transport plan ν all these quantities are finite. Since $m > 0$, continuity of r_ν, l_ν implies that $w^- < w < w^+$ and $v^- < v < v^+$. Consider the measures

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

Equalities (13) 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 ρ^-, σ^-, σ^+, ρ^+ respectively, mapping them to the Lebesgue measure on $(0, m)$.

Under the plan ν, mass m is sent from (v^-, v) to (w, w^+) and from (v, v^+) to (w^-, w). We now construct a local modification ν_v of the transport plan ν 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 μ_0, μ_1 have atoms corresponding to the intervals under consideration.

Observe first that the normal plan ν induces two transport plans τ_r, τ_l that map measures ρ^- to σ^+ and ρ^+ to σ^- correspondingly:

$$\tau_r(A \times B) = \nu(A \cap (-\infty, v) \times B \cap (w, +\infty)) = \nu(A \cap (v^-, v) \times B \cap (w, w^+)),$$
$$\tau_l(A \times B) = \nu(A \cap (v, +\infty) \times B \cap (-\infty, w)) = \nu(A \cap (v, v^+) \times B \cap (w^-, w)),$$

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 6, there exist two transport plans χ_r and χ_l mapping the Lebesgue measure on $(0, m)$ respectively to σ^+, σ^- and such that

$$\tau_r(A \times B) = \chi_r\big(r_w(A \cap (v^-, v)) \times B \cap (w, w^+)\big),$$
$$\tau_l(A \times B) = \chi_l\big(l_w(A \cap (v, v^+)) \times B \cap (w^-, w)\big).$$

Define now two transport plans $\bar\tau_l$, $\bar\tau_r$ that send mass elements to the same destinations but from interchanged origins:

$$\bar\tau_r(A \times B) = \chi_r\big(l_w(A \cap (v, v^+)) \times B \cap (w, w^+)\big),$$
$$\bar\tau_l(A \times B) = \chi_l\big(r_w(A \cap (v^-, v)) \times B \cap (w^-, w)\big).$$

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 ν_v has the same uniform marginals as ν, i.e., it is a local modification of ν. Observe furthermore that by the construction of ν_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 ν_v is either a cost-reducing modification of ν or equivalent to it. By the disintegration lemma (see, e.g., [2]) we can write $\chi_r(\mathrm{d}\alpha \times \mathrm{d}w') = \mathrm{d}\alpha\ \mathrm{d}G_r(w' \mid \alpha)$ and $\chi_l(\mathrm{d}\alpha \times \mathrm{d}w') = \mathrm{d}\alpha\ \mathrm{d}G_l(w' \mid \alpha)$, where $G_r(\,\cdot \mid \alpha)$ (resp. $G_l(\,\cdot \mid \alpha)$) are distribution functions of probability measures defined on $[w, w^+]$ (resp. $[w^-, w]$) for almost all $0 < \alpha < m$. Denote their respective inverses by $G_r^{-1}(\cdot \mid \alpha)$, $G_l^{-1}(\cdot \mid \alpha)$ and observe that $w^- \le G_l^{-1}(\beta' \mid \alpha) \le w \le G_r^{-1}(\beta'' \mid \alpha) \le w^+$ for any β', β''. Thus

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

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

$$\iint c(v', w')\ \tau_l(\mathrm{d}v' \times \mathrm{d}w') = \int_0^m \mathrm{d}\alpha \int_0^1 \mathrm{d}\beta\ c(l_w^{-1}(\alpha), G_l^{-1}(\beta \mid \alpha)),$$
$$\iint c(v', w')\ \bar\tau_r(\mathrm{d}v' \times \mathrm{d}w') = \int_0^m \mathrm{d}\alpha \int_0^1 \mathrm{d}\beta\ c(l_w^{-1}(\alpha), G_r^{-1}(\beta \mid \alpha)),$$
$$\iint c(v', w')\ \bar\tau_l(\mathrm{d}v' \times \mathrm{d}w') = \int_0^m \mathrm{d}\alpha \int_0^1 \mathrm{d}\beta\ c(r_w^{-1}(\alpha), G_l^{-1}(\beta \mid \alpha)).$$

The integral in Definition 2 now takes the form

$$\begin{aligned}\iint & c(v',w')\,(\nu_v(\mathrm{d}v'\times \mathrm{d}w')-\nu(\mathrm{d}v'\times \mathrm{d}w'))\\ &= \iint c(v',w')\,\big(-\tau_r(\mathrm{d}v'\times \mathrm{d}w')-\tau_l(\mathrm{d}v'\times \mathrm{d}w')\\ &\qquad\qquad + \bar\tau_r(\mathrm{d}v'\times \mathrm{d}w')+\bar\tau_l(\mathrm{d}v'\times \mathrm{d}w')\big)\\ &= \int_0^m \mathrm{d}\alpha \int_0^1 \mathrm{d}\beta\,\big(-c(r_w^{-1}(\alpha),G_r^{-1}(\beta\mid\alpha))-c(l_w^{-1}(\alpha),G_l^{-1}(\beta\mid\alpha))\\ &\qquad\qquad + c(l_w^{-1}(\alpha),G_r^{-1}(\beta\mid\alpha))+c(r_w^{-1}(\alpha),G_l^{-1}(\beta\mid\alpha))\big).\end{aligned}$$

As $r_w^{-1}(\alpha)\le v\le l_w^{-1}(\alpha)$ and $G_l^{-1}(\beta\mid\alpha)\le w\le G_r^{-1}(\beta\mid\alpha)$ for all α, β, the Monge condition (4) 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 ν_v is a cost-reducing local modification of ν; in the latter case ν_v is equivalent to ν. □

Lemma 12. *For any $v'<v''$ there exists a local modification $\nu_{v',v''}$ of ν such that $w_{\nu_{v',v''}}(v)=w_\nu(v)$ for $v'\le v\le v''$, $m_{\nu_{v',v''}}(v')=m_{\nu_{v',v''}}(v'')=0$, and in the strip $v'\le v\le v''$ 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 ν_i recursively to be the local modification of ν_{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 ν_i are either cost-reducing or equivalent to ν 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 ν_{j-1} is local to the interval $(v_{i'},v_{i''})$, where $v_{i'}=\max\{v_i\colon i<j, v_i<v_j\}$ and $v_{i''}=\min\{v_i\colon i<j, v_i>v_j\}$ (with max and min of empty set defined, as usual, as $-\infty$ and ∞). Thus there is a well-defined limit normal transport plan ν_∞ that is either a cost-reducing local modification or equivalent to ν and is such that, by continuity of the functions r_{ν_∞} and l_{ν_∞} 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_{ν_∞} in the second argument implies that

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

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 ν_∞ is supported thereon. □

Corollary 13. *For any normal transport plan ν there exists a real number θ_ν 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. □

We call the parameter θ_ν the *rotation number* of the normal transport plan ν.

Definition 14. *A normal transport plan consisting of a uniform measure supported on the line* $\{(v,w) \colon w = v + \theta\}$ *is called a* shift *and denoted by* ν_θ.

Lemma 15. *For any* θ *the shift* ν_θ *is* $c_{[F_0,F_1]}$*-locally optimal.*

Proof. Let $\bar\nu$ be a local modification of ν_θ 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 12; it coincides with ν_θ 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 ν_θ, i.e., that ν_θ is a cost minimizer with respect to local modifications. □

Lemma 16. *Any* $c_{[F_0,F_1]}$*-locally optimal normal transport* ν *with rotation number* $\theta = \theta_\nu$ *is equivalent to the shift* ν_θ.

Proof. Let $v_i'' = -v_i' = i$ for $i = 1, 2, \dots$. All local modifications $\nu_i = \nu_{v_i',v_i''}$ of ν constructed as in Lemma 12 cannot be cost-reducing and are therefore equivalent to ν. On the other hand, this sequence stabilizes to the shift ν_θ on any bounded subset of $V_0 \times V_1$ as soon as this set is covered by the segment $(-i, i)$. Therefore ν_θ has the same conjugate as all ν_i and is equivalent to ν. □

Lemmas 15, 16, and 7 together imply Theorem 8.

4. Transport optimization for periodic measures

4.1. **The average cost of locally optimal transport.** Let now c be a cost function that satisfies the Monge condition (4), the integer shift invariance condition (5), the growth condition (6), and is bounded from below. Suppose that γ_θ is a locally optimal transport plan on $U_0 \times U_1$ with marginals μ_0, μ_1 conjugate to the shift ν_θ. Define $c_{[F_0,F_1]}$ as in (11) and let $F_1^\theta(u) = F_1(u) - \theta$ as illustrated in fig. 1.

Definition 17. *We call the quantity*

$$C_{[F_0,F_1]}(\theta) = \int_0^1 c_{[F_0,F_1]}(v', v' + \theta)\,\mathrm{d}v' = \int_0^1 c\big(F_0^{-1}(v'), (F_1^\theta)^{-1}(v')\big)\,\mathrm{d}v' \tag{14}$$

the average cost *(per period) of the transport plan* γ_θ.

Observe that it is indifferent whether to integrate here from 0 to 1 or from v to $v+1$ for any real v.

The following lemma provides a "bracket" for the global minimum of $C_{[F_0,F_1]}$ and estimates of its derivatives independent of μ_0, μ_1.

Lemma 18. *The average cost* $C_{[F_0,F_1]}$ *is a convex function that satisfies the inequalities*

$$\inf_{x,y} c(x,y) \le \underline{C}(\theta) \le C_{[F_0,F_1]}(\theta) \le \overline{C}(\theta) \tag{15}$$

with

$$\underline{C}(\theta) = \inf_{\substack{-1 \le u_1 \le 2 \\ \theta-1 \le u_2 \le \theta+2}} c(u_1,u_2), \quad \overline{C}(\theta) = \sup_{\substack{-1 \le u_1 \le 2 \\ \theta-1 \le u_2 \le \theta+2}} c(u_1,u_2). \tag{16}$$

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*

$$-\underline{L} \leq C'_{[F_0,F_1]}(\underline{\Theta} - 0) \leq 0 \leq C'_{[F_0,F_1]}(\overline{\Theta} + 0) \leq \overline{L}, \tag{17}$$

where $C'_{[F_0,F_1]}(\cdot)$ *is the derivative of* $C_{[F_0,F_1]}$. *These constants are independent on* μ_0, μ_1 *and are given explicitly by formulas* (18), (19) *and* (20) *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 $\underline{C}(\theta) = \operatorname{conv}\min(|\theta + 3|^\alpha, |\theta - 3|^\alpha)$, $\overline{C}(\theta) = \max(|\theta + 3|^\alpha, |\theta - 3|^\alpha)$, 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 $C_{[F_0,F_1]}\big(\frac{1}{2}(\theta' + \theta'')\big) \leq \frac{1}{2}\big(C_{[F_0,F_1]}(\theta') + C_{[F_0,F_1]}(\theta'')\big)$ for all θ', θ''. Let $\theta' < \theta''$, denote $\theta = \frac{1}{2}(\theta' + \theta'')$ and write

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

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

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

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

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 (15) on $C_{[F_0,F_1]}(\theta)$ follow from (14) 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 (6) 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 (16) are bounded by P from below. Therefore one can set

$$\underline{\Theta} = \inf\{\theta \colon \underline{C}(\theta) = \min_{\theta'} \overline{C}(\theta')\} > -\infty, \ \overline{\Theta} = \sup\{\theta \colon \underline{C}(\theta) = \min_{\theta'} \overline{C}(\theta')\} < \infty, \tag{18}$$

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 θ 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

$$\overline{L} = \inf_{\theta \geq \overline{\Theta}} \frac{\overline{C}(\theta) - \underline{C}(\overline{\Theta})}{\theta - \overline{\Theta}}. \tag{19}$$

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 (17) is given by a symmetrical argument; in particular

$$\underline{L} = \inf_{\theta \leq \underline{\Theta}} \frac{\overline{C}(\theta) - \underline{C}(\underline{\Theta})}{\underline{\Theta} - \theta}. \tag{20}$$ □

Definition 19. *A locally optimal transport plan* γ_{θ_0} *is called* globally optimal *if* $\theta_0 \in \arg\min_\theta C_{[F_0,F_1]}(\theta)$.

We can now reduce minimization of (1) on the unit circle to minimization of (2) on $\mathbb{R}$, which involves the cost function c rather than $\hat{c}$:

Theorem 20. *The canonical projection* $\pi\colon \mathbb{R} \to \mathbb{T}$ *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* (1).

Proof. A transport plan γ on $\mathbb{T}\times\mathbb{T}$ minimizes (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 (5), (6)).

Denote $S = \{(x,y)\colon c(x,y) = \hat{c}(x,y)\}$ and observe that the support of the globally optimal plan γ_{θ_0} lies within S: indeed, if it did not, there would exist a modification of γ_{θ_0} bringing some of the mass of each period to S and thus reducing the average cost. Therefore γ_{θ_0} is locally optimal with respect to the cost $\hat{c}(x,y)$ and its projection to $\mathbb{T}\times\mathbb{T}$ minimizes (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 $\mathbb{R}\times\mathbb{R}$. □

4.2. **Fast global transport optimization.** In a typical application, such as the image processing problem described in the introduction, measures μ_0 and μ_1 come in the form of *histograms*, i.e., discrete distributions supported on subsets $X = \{x_1, x_2, \dots, x_{n_0}\}$ and $Y = \{y_1, y_2, \dots, y_{n_1}\}$ of the unit circle. These two sets may coincide. In what follows we replace X and Y with their lifts to the universal cover and assume that the points are sorted and numbered in an increasing order:

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

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

Define $j(\theta)$ as the index of $\min\{y_j\colon F_1^\theta(y_j) > 0\}$ and denote $y_1^\theta = y_{j(\theta)}$, $y_2^\theta = y_{j(\theta)+1}$, …, $y_{n_1}^\theta = y_{j(\theta)+n_1-1}$. All the values

$$F_0(x_1), F_0(x_2), \dots, F_0(x_{n_0}), F_1^\theta(y_1^\theta), F_1^\theta(y_2^\theta), \dots, F_1^\theta(y_{n_1}^\theta) \tag{21}$$

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 $x_{(k)} = F_0^{-1}(v)$

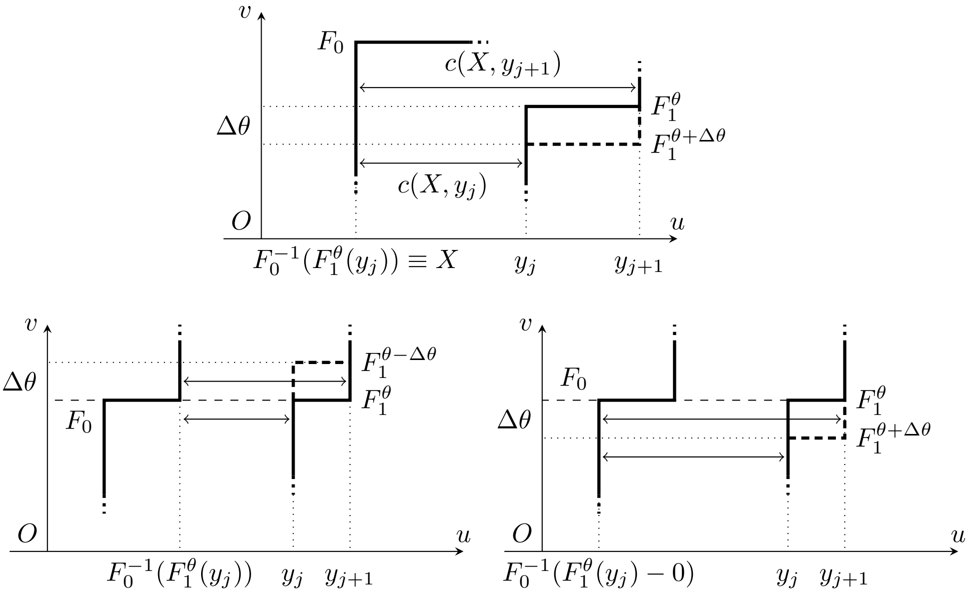

FIGURE 2. Derivation of expressions (23), (24) for $C'_{[F_0,F_1]}(\theta \pm 0)$. Thick lines show fragments of complete graphs of F_0, F_1^θ corresponding to jth terms in (23), (24). Top: θ not exceptional. Bottom: exceptional θ; thin dashed line marks the common value of F_0 and F_1^θ, left and right panes illustrate the cases $C'_{[F_0,F_1]}(\theta - 0)$ and $C'_{[F_0,F_1]}(\theta + 0)$. Note that $X = F_0^{-1}(F_1^\theta(y_j))$ can be alternatively written as $\inf\{x\colon F_0(x) > F_1^\theta(y_j)\}$.

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 $C_{[F_0,F_1]}$:

$$C_{[F_0,F_1]}(\theta) = \sum_{1 \le k \le n_0+n_1} c(x_{(k)}, y_{(k)})\,(v_{(k)} - v_{(k-1)}). \tag{22}$$

Observe that, as the parameter θ increases by $\Delta\theta$, those $v_{(k)}$ that correspond to values F_1^θ decrease by the same increment. Let $F_1^\theta(y_{j_0})$ be such a value. As it appears in (22) 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. 2, top).

Moreover, there are exceptional values of θ for which two of the values in (21) coincide and their ordering in the sequence $(v_{(k)})$ changes. For such values of θ the derivative $C'_{[F_0,F_1]}$ has different right and left limits, as illustrated in fig. 2, bottom:

$$C'_{[F_0,F_1]}(\theta - 0) = \sum_{1 \le j \le n_1} \bigl(c(F_0^{-1}(F_1^\theta(y_j)), y_{j+1}) - c(F_0^{-1}(F_1^\theta(y_j)), y_j)\bigr), \tag{23}$$

$$C'_{[F_0,F_1]}(\theta + 0) = \sum_{1 \le j \le n_1} \bigl(c(F_0^{-1}(F_1^\theta(y_j) - 0), y_{j+1}) - c(F_0^{-1}(F_1^\theta(y_j) - 0), y_j)\bigr). \tag{24}$$

If θ 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 linear; moreover, from the Monge condition (4) 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 21. *Values of C and its left and right derivatives can be computed for any θ using at most $O(n_0+n_1)$ comparisons and evaluations of $c(x,y)$.*

Proof. Sorting the n_0+n_1 values (21) 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 (22) 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 θ is exceptional or not. All this can again be done in $O(n_0+n_1)$ operations. □

Now 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 algorihtm below, are defined by explicit formulas in Lemma 18 and do not depend on measures μ_0, μ_1. The minimum of $C_{[F_0,F_1]}(\theta)$ can be found to accuracy ϵ using the following binary search technique:

(A1) Initially set $\underline{\theta}:=\underline{\Theta}$ and $\overline{\theta}:=\overline{\Theta}$, where $\underline{\Theta}$, $\overline{\Theta}$ are defined in Lemma 18.
(A2) Set $\theta:=\frac{1}{2}(\underline{\theta}+\overline{\theta})$.
(A3) Compute $C'_{[F_0,F_1]}(\theta-0)$, $C'_{[F_0,F_1]}(\theta+0)$.
(A4) If $C'_{[F_0,F_1]}(\theta-0)\le 0\le C'_{[F_0,F_1]}(\theta+0)$, then θ is the required minimum; stop.
(A5) If $\overline{\theta}-\underline{\theta}<\epsilon/L$, then:
   (a) compute $C_{[F_0,F_1]}(\underline{\theta})$, $C_{[F_0,F_1]}(\overline{\theta})$;
   (b) set θ equal to the solution of

$$C_{[F_0,F_1]}(\underline{\theta})+C'_{[F_0,F_1]}(\underline{\theta}+0)(\theta-\underline{\theta})=C_{[F_0,F_1]}(\overline{\theta})+C'_{[F_0,F_1]}(\overline{\theta}-0)(\theta-\overline{\theta}); \tag{25}$$

   (c) stop.
(A6) Otherwise set $\underline{\theta}:=\theta$ if $C'_{[F_0,F_1]}(\theta+0)<0$, or $\overline{\theta}:=\theta$ if $C'_{[F_0,F_1]}(\theta-0)>0$.
(A7) Go to step (A2).

It follows from inequalities (17) of Lemma 18 that the minimizing value of θ belongs to the segment $[\underline{\Theta},\overline{\Theta}]$. Therefore at all steps

$$C'_{[F_0,F_1]}(\underline{\theta}+0)\le 0\le C'_{[F_0,F_1]}(\overline{\theta}-0) \tag{26}$$

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

Step (A5) requires some comments. By convexity, $-\underline{L}\le C'_{[F_0,F_1]}(\theta\pm 0)\le\overline{L}$ for all $\underline{\Theta}\le\theta\le\overline{\Theta}$, i.e., $|C'_{[F_0,F_1]}(\theta\pm 0)|\le L$ at all steps. When $\overline{\theta}-\underline{\theta}<\epsilon/L$, this bound ensures that for any θ' in $[\underline{\theta},\overline{\theta}]$ 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 θ in that interval, then it is located precisely at the solution of (25) and must be a minimum of C because of (26), so the final value of θ is the exact solution; otherwise it is an approximation with guaranteed accuracy.

The final value of θ 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 $[\underline{\theta},\overline{\theta}]$ of length ϵ can contain at most one exceptional value of θ.

Since at each iteration the interval $[\underline{\theta}, \overline{\theta}]$ is halved, step (A5) will be achieved in $O(\log_2((\overline{\Theta} - \underline{\Theta})/(\epsilon/L)))$ iterations. By Lemma 21 each step (A3) (and (A5b)) takes $O(n_0 + n_1)$ operations. Together these observations establish

Theorem 22. *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 θ is within ϵ/L from the global minimum, and $C_{[F_0,F_1]}(\theta) \le \min_\theta C_{[F_0,F_1]}(\theta) + \epsilon$. When all masses m'_i, m''_j 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.*

5. Relations to the previous work

The problem of finding an optimal coupling between two measures is very well understood when both marginals are finite measures supported on convex compacts in the n-dimensional Euclidean space (see, e.g., [9] or the comprehensive recent monograph [15] and references therein). In particular, when $c(x, y) = |x - y|^2$, there is a unique optimal transport plan which is essentially a map T from the support of μ_0 to that of μ_1 such that $\mu_1(A) = \mu_0(T^{-1}A)$ for all A. Moreover this map is the (sub)gradient of a suitable convex function, which satisfies a certain Monge–Ampère equation in a weak sense.

Everything becomes especially simple in the one-dimensional case, when the optimal coupling is monotone and can thus be established by sorting the elements of the two measures left-to-right. In fact sorting works for any cost function satisfying the Monge condition (alternatively known as the continuous Monge property, see [1, 5]). If the data come in the form of discrete n-point histograms whose points are already sorted, then the solution can be constructed in $O(n)$ operations; if sorting is needed, the operation count rises to $O(n \log n)$.

When marginals live on a compact Riemannian manifold rather than in Euclidean space, existence of the optimal coupling is also well-known [8]. For the specific case of $\mathbb{T}^n$ and a quadratic cost it has been established almost a decade ago by Cordero-Erausquin [6]. Again, the optimal coupling here is realized by a gradient of a suitable function. This characterization becomes particularly transparent if one lifts the problem to the universal cover $\mathbb{R}^n$ of the torus $\mathbb{T}^n$, where this function is convex and its gradient is monotone and satisfies an integer shift invariance condition similar to (8).

However in the simplest one-dimensional case the fundamental set $\mathbb{T}$ is no longer *ordered* but merely *oriented*. Hence, for the sorting method to work, one has to artificially "cut" $\mathbb{T}$ at some point to make it into the ordered convex set $(0, 1]$. A brute-force approach to finding a minimizer of (1) for n-point histograms requires to cut one of the circular histograms at each of the n possible locations and to compute the corresponding values of transport cost, thereby involving $O(n^2)$ operations.

Faster algorithms for the transportation problem on the circle have been proposed in a number of works. Karp and Li [10] 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 [16]; clearly, the balanced problem can always be treated as a particular case of the unbalanced one. In both these works $O(n \log n)$ algorithms are obtained for the case where all points have unit mass and the cost function on the universal cover is given by the Euclidean distance $|x - y|$.

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 $1/M$.

The results of [1] were extended in a different direction by McCann [12], 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.

Finally, we note that an important device of the present work is the notion of locally optimal transport plan in an unbounded domain, i.e., a plan whose (infinite) cost cannot be improved by any local modification. Different locally optimal transport plans cannot be deformed into each other by local rearrangement. There is a common pattern between this notion and action minimizing measures in the weak KAM (Aubry–Mather) theory (see, e.g., [7, 11]). In particular, a cost function satisfying conditions (4)–(6) can be generated by a natural Lagrangian with a time-periodic potential [11, 4], the conjugacy to shifts of the universal cover (or rotations of the circle) established in Section 3 is a counterpart of conjugacy to (irrational) rotations in the one-dimensional weak KAM theory [3], whereas the average cost $C_{[F_0,F_1]}(\theta)$ is similar to the averaged Lagrangian or Mather's α function (see, e.g., [7]).

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*, pages 583–592, 1992.

[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(3):381–422, 1983.

[4] P. Bernard and B. Buffoni. Optimal mass transportation and Mather theory. *Journal of the European Mathematical Society*, 9:85–121, 2007.

[5] R. E. Burkard, B. Klinz, and R. Rudolf. Perspectives of Monge properties in optimization. *Discrete Appl. Math.*, 70(2):95–161, 1996.

[6] D. Cordero-Erausquin. Sur le transport de mesures périodiques. *Comptes Rendus de l'Académie des Sciences. Series I Math.*, 329(3):199–202, August 1999.

[7] A. Fathi. *Weak KAM theorem in Lagrangian dynamics.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.

[8] M. Feldman and R. J. McCann. Monge's transport problem on a Riemannian manifold. *Trans. Amer. Math. Soc.*, 354(4), 2002.

[9] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996.

[10] R. M. Karp and S. Y. R. Li. Two special cases of the assignment problem. *Discrete Mathematics*, 13:129–142, 1975.

[11] O. Knill. *Jürgen Moser, selected chapters in the calculus of variations.* Birkhäuser Verlag, 2003.
[12] R. McCann. Exact solutions to the transportation problem on the line. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 455(1984):1341–1380, 1999.
[13] J. Rabin, J. Delon, and Y. Gousseau. Circular Earth Mover's distance for the comparison of local features. In *Proc. ICPR*. IEEE Computer Society, 2008.
[14] Y. Rubner, C. Tomasi, and L. Guibas. The Earth Mover's Distance as a Metric for Image Retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.
[15] C. Villani. *Optimal transport: Old and new.* Grundlehren der mathematischen Wissenschaften, vol. 338. Springer-Verlag, 2009.
[16] M. Werman, S. Peleg, R. Melter, and T. Kong. Bipartite graph matching for points on a line or a circle. *Journal of Algorithms*, 7(2):277–284, 1986.

LTCI CNRS, Télécom ParisTech

Université Paris-Dauphine/CEREMADE

Physics department, M.V. Lomonosov Moscow State University

UMI 2615 CNRS "Laboratoire J.-V. Poncelet"
E-mail address: sobolevski@phys.msu.ru