No-collision Transportation Maps

Levon Nurbekyan · Alexander Iannantuono · Adam M. Oberman

Abstract Transportation maps between probability measures are critical objects in numerous areas of mathematics and applications such as PDE, fluid mechanics, geometry, machine learning, computer science, and economics. Given a pair of source and target measures, one searches for a map that has suitable properties and transports the source measure to the target one. Here, we study maps that possess the *no-collision* property; that is, particles simultaneously traveling from sources to targets in a unit time with uniform velocities do not collide. These maps are particularly relevant for applications in swarm control problems. We characterize these no-collision maps in terms of *half-space preserving* property and establish a direct connection between these maps and *binary-space-partitioning (BSP) tree* structures. Based on this characterization, we provide explicit BSP algorithms, of cost $O(n \log n)$, to construct no-collision maps. Moreover, interpreting these maps as approximations of optimal transportation maps, we find that they succeed in computing nearly optimal maps for q-Wasserstein metric ($q = 1, 2$). In some cases, our maps yield costs that are just a few percent off from being optimal.

Keywords no-collision transportation maps · optimal transportation · binary-space-partitioning trees · $k - d$ trees

L. Nurbekyan
Department of Mathematics, UCLA
E-mail: lnurbek@math.ucla.edu

A. Iannantuono
Department of Mathematics and Statistics, McGill
E-mail: alexander.iannantuono@mail.mcgill.ca

A. M. Oberman
Department of Mathematics and Statistics, McGill
E-mail: adam.oberman@mcgill.ca

1 Introduction

Given a probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ and a Borel measurable map $T : \mathbb{R}^d \to \mathbb{R}^d$ the push-forward of μ by T is a probability measure $T\sharp\mu$ that is given by

$$T\sharp\mu(B) = \mu(T^{-1}(B)), \ \forall B \text{ Borel}.$$

Furthermore, for probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and a Borel measurable mapping $T : \mathbb{R}^d \to \mathbb{R}^d$ we say that T transports μ to ν if

$$\nu = T\sharp\mu. \tag{1}$$

Heuristically, $T\sharp\mu$ is the measure obtained from μ by a rearrangement that sends (transports) a point x to $T(x)$. Therefore, one can think of (1) as a measure respecting correspondence of μ and ν where a point x in the support of μ is being matched with a point $T(x)$ in the support of ν. This perspective readily yields numerous applications of finding a suitable T for given μ, ν. However, T may not exist in general and, if exists, may not be unique. Therefore, a natural problem is to search for T with some useful properties.

Optimal transportation is a remarkably rich theory that addresses this problem [25, 26]. In this theory, one associates a cost, $c(x, y)$, for transporting a unit of mass from point x to y. Then, one searches for a transportation (correspondence) that minimizes the cost of transporting μ into ν; that is,

$$\min_{\nu = T\sharp\mu} \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x). \tag{2}$$

This setup of the transportation problem turns out to be incredibly fruitful far beyond the question of finding a T such that (1) holds. Indeed, applications of optimal transportation theory include PDE, fluid mechanics, geometry, machine learning, computer science, economics, and so on [25, 26, 20]. However, together with extremely useful theoretical properties (2) comes with a considerable computational cost. Hence, development of fast algorithms for solving or estimating (2) is vital for the applications.

In this note, we are motivated by particular applications of the transportation problem (1) in swarm control. Specifically, we aim at finding T that have *no-collision* property; that is, for every $x_1 \neq x_2 \in \text{supp}(\mu)$ one has that $(1 - \lambda)x_1 + \lambda T(x_1) \neq (1 - \lambda)x_2 + \lambda T(x_2)$ for all $\lambda \in (0, 1)$. In other words, if we transport mass (particles) along straight lines with constant speeds in a unit time, the point-masses will not collide on the way. Recall that $((1 - \lambda)\text{Id} + \lambda T)\sharp\mu$ is the *displacement interpolation* measure between μ and ν that is often used as a shape combination or interpolation technique in imaging. Thus, no-collision property can be interpreted as *no information loss* in that context.

Optimal transportation maps corresponding to $c(x, y) = h(x - y)$ with a strictly convex and even h do have no-collision property [25, 26, 3]. However, these maps are expensive to calculate and normally require global optimization. The Hungarian algorithm for assignment problem [24] is $O(n^3)$. The transportation plan computed by linear programming is potentially faster, but still requires optimization. The entropic regularization [5] is explicit and fast, but has an additional regularization parameter which affects both the quality (plan versus map) and the cost of the solution.

To the best of our knowledge, the only algorithms that achieve remarkable near-linear complexity are the ones in [1] (see also [5,6,10,2,20,23] for preceding work) and [15]. The algorithm in [1] achieves a complexity $\tilde{O}\left(\frac{n}{\epsilon^3}\left(\frac{C\log n}{\epsilon}\right)^d\right)$ for $c(x,y) = |x-y|^2$, where ϵ is the additive approximation error for the optimal cost. Bases of the algorithm are the entropic regularization of the optimal transportation problem [5] and low rank approximations of the Gaussian kernel matrices. However, the algorithm returns an *approximate transportation plan* instead of a map. Additionally, this plan is in an *implicit factored form.* Thus, calculation of the corresponding transportation map will impose quadratic complexity. Finally, the resulting map may not have the no-collision property due to an approximation error.

In [15], the authors achieve a complexity $O(n\log n)$ based on a clever back-and-forth H^1 gradient ascent scheme on the dual formulation of (2) and a fast calculation of c-transforms (generalized Legendre transforms) for $c(x,y) = \sum_{i=1}^d h_i(x_i - y_i)$. However, the algorithm requires a regular grid discretization.

We propose an explicit binary-space-partitioning (BSP) algorithm that constructs no-collision transportation maps in $O(n\log n)$ time in full generality without any optimization. We achieve this goal in two steps. First, we provide a complete characterization of no-collision maps in terms of *half-space preserving* property. More precisely, in Theorem 1, we show that a map has no-collision property iff for every pair of points in the source there exists a hyperplane that separates these points together with a parallel hyperplane separating the images of these points. In particular, any map that respects BSP tree structures on reference and target measures induced by the same partitioning directions is a no-collision map. This is a critical connection between BSP trees and no-collision maps that was not observed previously.

Second, we build transportation maps induced by BSP tree structures on reference and target measures. More precisely, given μ, ν, we partition $\text{supp}(\mu)$, $\text{supp}(\nu)$ and resulting subsets by a sequence of hyperplanes so that each resulting subset has equal masses on two sides of the partitioning hyperplane. Furthermore, to every point in $\text{supp}(\mu)$ and $\text{supp}(\nu)$, we assign a binary sequence that records in which side of partitioning hyperplanes the point fell during the construction. These binary sequences correspond to BSP tree structures on $\text{supp}(\mu)$, $\text{supp}(\nu)$ where nodes correspond to partitioning hyperplanes. Finally, the transportation map matches the leaves in $\text{supp}(\mu)$ and $\text{supp}(\nu)$.

Our analysis yields a flexible construction method since we allow for an arbitrary sequence of partitioning directions as long as constituent subsets shrink in diameter. A natural choice corresponds to successive partitions by hyperplanes orthogonal to elements of a basis. In Theorem 3, we prove that such partitions induce a well-defined transportation map. Moreover, in Theorem 4, we show that these maps are a.e. continuous.

Our method for constructing transportation maps depends on decomposing each measure recursively, using the half-spaces with the same direction vectors at each step of the recursion. In this case, the construction does not involve optimization: only a median search. Therefore, our run-time is $O(n\log n)$ by using a $O(n)$ median-search algorithm [4,19].

This method can also be used to decompose a single measure, that corresponds to a $k-d$ tree decompositions. These are particular instances of BSP trees and are

used for nearest neighbor search [12]. Typically, the setting is low dimensional, and the direction vectors are standard basis vectors. An important recent application of nearest neighbor search in high dimensions is image retrieval [16]. In this case, modern methods use a deep neural network to output a feature vector from an image. The cosine of the angle between feature vectors is used as a distance. Given a new image, similar images are found by nearest neighbor search using the distance on the feature vectors. However, $k - d$ trees are less effective for high dimensional data, since nearest neighbors can lie on different sides of the hyperplanes used to bisect the data. An alternative in high dimensions is to use randomized vectors [7, 8]. A further improvement is high dimensions is to use multiple randomized $k - d$ trees, [18], since multiple tree reduce the possibility of an unfortunate splitting.

BSP and $k - d$ trees are ubiquitous tools in computer science and image processing [12, 21, 27]. Nevertheless, to the best of our knowledge, current work is the first one to establish a direct connection between these structures and efficient constructions of transportation maps.

One can also optimize for the partitioning directions. For instance, one can search for directions that minimize transportation cost. Here, we do not address this problem which is a subject of our future research.

Given partitioning directions, BSP trees induce a total order on supp(μ) and supp(ν). Furthermore, the no-collision transportation map corresponding to this partitioning sequence is the unique *order-preserving* or *monotone* map between supp(μ) and supp(ν) that respects this order.

Maps that we obtain have several critical properties. Firstly, these are no-collision maps; that is, point-masses do not collide when we perform the transportation dynamically in a unit time. Secondly, these maps are scale-consistent. More precisely, the map on a coarser scale is the concatenation of maps on a finer scale: Proposition 1. Thirdly, these are transitive maps; that is, for a fixed sequence of partitioning directions, the map from μ to ν is the composition of the maps from μ to ρ and ρ to ν for any μ, ν, ρ: Proposition 2. The last two properties yield that the calculations of these maps are amenable to decentralization and parallelization techniques that can further enhance the speed and robustness of the calculations.

Since no-collision maps are closely related to optimal transportation maps, a natural question is to test the performance of these maps as estimators for the optimal maps. We perform tests in the two-dimensional case and successively choose horizontal and vertical partitioning directions. Interestingly, we find that these maps succeed in achieving nearly optimal costs for $c(x, y) = \|x - y\|_p^q$ for $p = 1, 2, \infty$ and $q = 1, 2$.

In our future research, we plan to investigate further these approximation results. Particularly interesting are the questions related to the characterization of distortion and approximation errors of these no-collision maps. Recall that approximate transportation maps are critical in creating nearest neighbor databases based on optimal transportation distances [14, 11, 13].

The paper is organized as follows. In Section 2, we fully characterize no-collision transportation maps in terms of the half-space-preserving property: Theorem 1. Furthermore, we prove our main existence and regularity results for half-space preserving maps, Theorems 3 and 4. Next, in Section 3, we discuss further critical properties of half-space-preserving transportation maps. In Section 4, we describe the implementation of our algorithm and present the numerical experiments. Finally, in the Appendix, we present the proofs of Theorems 3 and 4.

2 Half-space-preserving transportation maps

Here, we characterize the no-collision maps in terms of a half-space-preserving property. Based on this characterization, we introduce a new class of transportation maps that have the no-collision and other intriguing properties. In particular, the construction of these maps is very quick because it requires only finding medians.

Definition 1 A map $T : \Omega \subset \mathbb{R}^d \to \mathbb{R}^d$ is half-space-preserving if for every $x_1 \neq x_2 \in \Omega$ there exists a direction $v \in \mathbb{S}^{d-1}$ such that

$$x_2 \cdot v - x_1 \cdot v \geq 0, \quad T(x_2) \cdot v - T(x_1) \cdot v \geq 0,$$

and at least one of the inequalities is strict.

Geometrically, T is half-space-preserving if for every x_1, x_2 there exists a pair of parallel hyperplanes that separate x_1 from x_2 and $T(x_1)$ from $T(x_2)$ in such a way that x_i and $T(x_i)$ are on the same side of the corresponding hyperplane for $i = 1, 2$. It turns out that the half-space-preserving transportation maps are precisely the ones with the no-collision property.

Theorem 1 *A map $T : \Omega \subset \mathbb{R}^d \to \mathbb{R}^d$ is half-space-preserving if and only if it has the no-collision property.*

Proof Assume that T is a half-space-preserving map, and $x_1 \neq x_2 \in \Omega$. Thus, there exists $v \in \mathbb{S}^{d-1}$ is such that

$$x_2 \cdot v - x_1 \cdot v \geq 0, \quad T(x_2) \cdot v - T(x_1) \cdot v \geq 0,$$

and at least one of these inequalities is strict. Furthermore, assume by contradiction that there exists $\lambda \in (0, 1)$ such that

$$(1 - \lambda)(x_1 - x_2) + \lambda(T(x_1) - T(x_2)) = 0,$$

or equivalently

$$T(x_1) - T(x_2) = \frac{\lambda}{1 - \lambda}(x_2 - x_1).$$

But then we have that

$$T(x_2) \cdot v - T(x_1) \cdot v = -\frac{\lambda}{1 - \lambda}(x_2 \cdot v - x_1 \cdot v) \leq 0,$$

and so

$$x_2 \cdot v - x_1 \cdot v = T(x_2) \cdot v - T(x_1) \cdot v = 0,$$

that is a contradiction. Therefore, half-space-preserving maps possess the no-collision property.

For the converse implication, assume that T has the no-collision property, and, by a contradiction, assume that there exist $x_1 \neq x_2 \in \Omega$ for which the half-space-preserving condition does not hold. Hence, for every $v \in \mathbb{S}^{d-1}$ such that

$$x_2 \cdot v - x_1 \cdot v = 0$$

one has that

$$T(x_2) \cdot v - T(x_1) \cdot v = 0.$$

Therefore, there exists $\kappa \in \mathbb{R}$ such that

$$T(x_2) - T(x_1) = \kappa(x_2 - x_1).$$

Next, take $v = \frac{x_2 - x_1}{|x_2 - x_1|}$. Then, we have that

$$x_2 \cdot v - x_1 \cdot v > 0.$$

Furthermore, since x_1, x_2 do not satisfy the half-space-preserving condition

$$T(x_2) \cdot v - T(x_1) \cdot v = \kappa|x_2 - x_1| < 0,$$

and so $\kappa < 0$. But then, we have that

$$(1 - \lambda)(x_1 - x_2) + \lambda(T(x_1) - T(x_2)) = 0,$$

where

$$\lambda = \frac{\kappa}{\kappa - 1} \in (0, 1).$$

Hence, we arrive at a contradiction with the no-collision property.

Theorem 1 is the starting point for the construction of our transportation maps; that is, we incorporate the half-space-preserving property in the construction. More precisely, assume that $\mu \in \mathcal{P}(\mathbb{R}^d)$ is absolutely continuous and $\{v_k\}$ is an arbitrary sequence of directions. Denote by

$$\Omega_0 = \{\mathbb{R}^d\}.$$

Since μ is absolutely continuous there exists $h_1 \in \mathbb{R}$ such that

$$\mu(x \cdot v_1 \le h_1) = \mu(x \cdot v_1 > h_1).$$

Then we denote by

$$A_0 = \{x \cdot v_1 \le h_1\}, \quad A_1 = \{x \cdot v_1 > h_1\}, \quad \Omega_1 = \{A_0, A_1\}.$$

Furthermore, to every $x \in A$ we assign a binary sequence, $s(x)$, as follows

$$s(x) = \begin{cases} 0, & x \in A_0, \\ 1, & x \in A_1. \end{cases}$$

Then, we pick a set from Ω_1, say A_1, and find h_2 such that

$$\mu(A_1 \cap \{x \cdot v_2 \le h_2\}) = \mu(A_1 \cap \{x \cdot v_2 > h_2\}),$$

and denote by

$$A_{10} = A_1 \cap \{x \cdot v_2 \le h_2\}, \quad A_{11} = A_1 \cap \{x \cdot v_2 > h_2\}, \quad \Omega_2 = \{A_0, A_{10}, A_{11}\}.$$

Next, we update $s(x)$ for $x \in A_1$ as follows

$$s(x) = \begin{cases} 10, & x \in A_{10}, \\ 11, & x \in A_{11}. \end{cases}$$

The process continues by picking a subset and dividing it into two pieces of equal masses and updating the corresponding binary sequences. As a result, points in the support of μ receive a binary sequence $s(x)$. If directions $\{v_k\}$ are chosen appropriately, every two points in $\text{supp}(\mu)$ get separated at some step, and therefore the sequence $s(x)$ is a bijection between $\text{supp}(\mu)$ and $\{0, 1\}^\infty$. This procedure, with the same choice of directions and subsets, can be performed for any absolutely continuous measure ν. Suppose that the binary sequence of ν is $r(x)$. Then, the transportation map from μ to ν is

$$t = r^{-1} \circ s.$$

Although clear intuitively, a rigorous construction of this previous map is a technically challenging task. Denote by

$$\hat{s}_0(x) = 0, \quad x \in \mathbb{R}^d,$$

$$\hat{s}_1(x) = \begin{cases} 0 \cdot \frac{1}{3}, & x \in A_0, \\ 1 \cdot \frac{1}{3}, & x \in A_1, \end{cases}$$

$$\hat{s}_2(x) = \begin{cases} 0, & x \in A_0, \\ 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3^2}, & x \in A_{10}, \\ 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3^2}, & x \in A_{11}, \end{cases}$$

$$\vdots$$

So $\hat{s}_k(x),\ x \in A$ is the real number in the ternary system with digits $s(x)$.

Lemma 1 *The family $\{\hat{s}_k\}$ is a bounded, non-decreasing, and uniformly-convergent sequence of measurable step-functions with a limit*

$$\hat{s}(x) = \lim_{k \to \infty} \hat{s}_k(x), \quad x \in \mathbb{R}^d. \tag{3}$$

Moreover, if for some $k \in \mathbb{N}$ points $x \neq x'$ get separated by a hyperplane parallel to v_k, then

$$\hat{s}(x) \neq \hat{s}(x').$$

Proof For all x, k we have that

$$0 \leq \hat{s}_k(x) \leq \sum_{i=1}^\infty \frac{1}{3^i} = \frac{1}{2}.$$

By construction, the value $\hat{s}_k(x)$ at step $k+1$ can either stay the same or increase by a power of $\frac{1}{3}$. Hence,

$$0 \leq \hat{s}_{k+1}(x) - \hat{s}_k(x) \leq \frac{C}{k}, \ \forall x \in A,$$

for some universal constant C.

Furthermore, suppose that $x \neq x'$ and they get separated at step k for the first time, and denote by l the length of $s(x) = s(x')$ right before step k. Then, we get that

$$|\hat{s}_n(x) - \hat{s}_n(x')| \geq \frac{1}{3^l} - \sum_{i=l+1}^\infty \frac{1}{3^i} = \frac{1}{2 \cdot 3^l},$$

for all $n \geq k$. Therefore, one has that

$$|\hat{s}(x) - \hat{s}(x')| \geq \frac{1}{2 \cdot 3^l} > 0.$$

For a generic absolutely continuous $\nu \in \mathcal{P}(\mathbb{R}^d)$ we denote by Δ_k the partition sets at step k. Furthermore, we denote by $\{\hat{r}_k\}$ and $\hat{r}$ the corresponding step-functions and the limiting function. We call a pair $(A, B) \in \Omega_k \times \Delta_k$ dual if $\hat{s}_k(x) = \hat{r}_k(y)$ for $x \in A, y \in B$. The following theorem is a technical result necessary for the proof of our main existence theorem.

Theorem 2 *Suppose that μ, ν are absolutely continuous probability measures, $\{v_k\}_{k=1}^{\infty}$ are some partitioning directions, and $\hat{s}, \hat{r}$ are functions defined by (3) for μ, ν, respectively. Furthermore, define a function*

$$\hat{t} = \hat{r}^{-1} \circ \hat{s} : \mathbb{R}^d \to \mathbb{R}^d,$$

where

$$\mathrm{Dom}(\hat{t}) \subset \{x \in \mathbb{R}^d \ s.t. \ \hat{r}^{-1}(\hat{s}(x)) \ is \ a \ singleton\}.$$

Then one has that

$$\hat{t}^{-1}(B) = \mathrm{Dom}(\hat{t}) \cap A, \ \forall B \in \bigcup_{k=1}^{\infty} \Delta_k, \tag{4}$$

where $A \in \bigcup_{k=1}^{\infty} \Omega_k$ is the dual set of B.

Consequently, if points in $\mathrm{Im}(\hat{t})$ get eventually separated then $\hat{t}$ is a half-space-preserving map.

Proof Suppose that $B \in \Delta_k$, and k is the smallest such number. Furthermore, assume that $A \in \Omega_k$ be the dual set corresponding to B. Finally, suppose that $x_0 \in A$, $y_0 \in B$. We first prove that

$$\begin{aligned} A = & \hat{s}^{-1}\left(\left[\hat{s}_k(x_0), \hat{s}_k(x_0) + \frac{1}{2 \cdot 3^l}\right]\right), \\ B = & \hat{r}^{-1}\left(\left[\hat{r}_k(y_0), \hat{r}_k(y_0) + \frac{1}{2 \cdot 3^l}\right]\right), \end{aligned} \tag{5}$$

where l is the length of the binary sequence assigned to the points of A and B at step k. Recall that by construction we have that

$$\hat{s}_k(x) = \hat{r}_k(y), \ \forall x \in A, \ y \in B,$$

and

$$\hat{s}_k(x) \leq \hat{s}_n(x) \leq \hat{s}_k(x) + \sum_{i=l+1}^{\infty} \frac{1}{3^i} = \hat{s}_k(x) + \frac{1}{2 \cdot 3^l}$$

for any $n \geq k$ and $x \in A$. Therefore,

$$\hat{s}_k(x) \leq \hat{s}(x) \leq \hat{s}_k(x) + \frac{1}{2 \cdot 3^l}, \ \forall x \in A,$$

and hence

$$A \subset \hat{s}^{-1}\left(\left[\hat{s}_k(x_0), \hat{s}_k(x_0) + \frac{1}{2 \cdot 3^l}\right]\right).$$

Now suppose that $x \notin A$. This means that $x \in A'$ for some $A' \in \Omega_k$, $A' \neq A$, and x, x_0 got separated at a step before k. Therefore, from the proof of Lemma 1 we have that

$$|\hat{s}(x') - \hat{s}(x_0)| \geq \frac{1}{2 \cdot 3^{l-1}},$$

and so

$$\hat{s}(x') \notin \left[\hat{s}_k(x_0), \hat{s}_k(x_0) + \frac{1}{2 \cdot 3^l} \right],$$

which means that

$$A \supset \hat{s}^{-1} \left(\left[\hat{s}_k(x_0), \hat{s}_k(x_0) + \frac{1}{2 \cdot 3^l} \right] \right).$$

Thus, (5) is proven. Furthermore, denote by

$$I = \left[\hat{s}_k(x_0), \hat{s}_k(x_0) + \frac{1}{2 \cdot 3^l} \right] = \left[\hat{r}_k(y_0), \hat{r}_k(y_0) + \frac{1}{2 \cdot 3^l} \right].$$

Then we have shown that

$$A = \hat{s}^{-1}(I), \quad B = \hat{r}^{-1}(I).$$

Next, we have that $x \in \hat{t}^{-1}(B)$ if and only if $x \in \mathrm{Dom}(\hat{t})$ and $\hat{t}(x) = \hat{r}^{-1}(\hat{s}(x)) \in B = \hat{r}^{-1}(I)$. Latter is equivalent to $\hat{s}(x) \in I$ or $x \in \hat{s}^{-1}(I) = A$. Hence we arrive at (4).

Finally, assume that $x_1 \neq x_2 \in \mathrm{Dom}(\hat{t})$. If $\hat{t}(x_1) = \hat{t}(x_2)$ then we can take $v = \frac{x_2 - x_1}{|x_2 - x_1|}$ and obtain

$$x_2 \cdot v - x_1 \cdot v = |x_2 - x_1| > 0, \quad \hat{t}(x_2) \cdot v - \hat{t}(x_1) \cdot v = 0.$$

If $\hat{t}(x_1) \neq \hat{t}(x_2)$ we have that they get separated at some step. Suppose that one step before separating they both belong to some B, and we cut B in direction v. Then, we get that B gets partitioned into B', B'', where

$$B' = B \cap \{y \cdot v \leq \eta\}, \quad B'' = B \cap \{y \cdot v > \eta\},$$

for some $\eta \in \mathbb{R}$. Without loss of generality assume that

$$\hat{t}(x_1) \in B', \quad \hat{t}(x_2) \in B''.$$

Denote by A, A', A'' the dual sets of B, B', B'', respectively. Then we have that

$$B' = B \cap \{x \cdot v \leq h\}, \quad A'' = B \cap \{x \cdot v > h\}.$$

Furthermore, from (4) we have that

$$x_1 \in A', \quad x_2 \in A''.$$

Thus,

$$x_2 \cdot v - x_1 \cdot v > 0, \quad \hat{t}(x_2) \cdot v - \hat{t}(x_1) \cdot v > 0.$$

Now we are in the position to state and prove our main existence result for no-collision transportation maps.

Theorem 3 *Suppose that*

$$\mu = f dx, \quad \nu = g dx,$$

are probability measures with compact supports such that

$$\mu\left(\partial(\mathrm{supp}(\mu))\right) = \nu\left(\partial(\mathrm{supp}(\nu))\right) = 0.$$

Furthermore, assume that

$$c \le f(x) \le C, \; \mu \; a.e., \quad c \le g(x) \le C, \; \nu \; a.e.,$$

for some constants $c, C > 0$.

Finally, suppose that $\{v_k\}_{k=1}^{\infty} \subset \{e_i\}_{i=1}^{d}$, where latter is an arbitrary basis in $\mathbb{R}^d$. Moreover, suppose that each set appearing during the partition gets partitioned in each of $\{e_i\}_{i=1}^{d}$ directions infinitely many times.

Then, we have that:

1. points inside $\mathrm{int}(\mathrm{supp}(\mu))$ and $\mathrm{int}(\mathrm{supp}(\nu))$ get separated,

2. there exists $F_0 \in \mathcal{B}(\mathbb{R}^d)$ such that $\mu(F_0) = 0$, and for every $x \in \mathrm{int}(\mathrm{supp}(\mu)) \setminus F_0$ there exits a unique $y \in \mathrm{int}(\mathrm{supp}(\nu))$ such that $\hat{s}(x) = \hat{r}(y)$,

3. the map $\hat{t} : \mathrm{int}(\mathrm{supp}(\mu)) \setminus F_0 \to \mathrm{int}(\mathrm{supp}(\nu))$, $x \mapsto \hat{r}^{-1}(\hat{s}(x))$ is Borel measurable and $\hat{t}\sharp\mu = \nu$.

Next, we prove a regularity result for these maps.

Theorem 4 *Under the assumptions of Theorem 3 one has that $\hat{t}$ is a.e. continuous.*

We postpone the proofs of these related and rather technical theorems to the Appendix.

3 Further properties of half-space-preserving transportation maps

Here, we discuss further critical properties of half-space-preserving maps constructed in Theorem 3.

Proposition 1 (Synthesis) *Suppose that $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, $\{v_k\}_{k=1}^{\infty}$ are some partitioning directions, and $k \in \mathbb{N}$. Furthermore, assume that for all dual pairs $(A, B) \in \Omega_k \times \Delta_k$ there exists a Borel measurable half-space-preserving map $t_{A,B} : A \setminus F_A \to B$ such that*

$$t_{A,B}\sharp(\mu|_{A \setminus F_A}) = \nu_B,$$

where F_A are some Borel sets such that $\mu(F_A) = 0$. Then the map $t : \mathbb{R}^d \setminus \bigcup_{A \in \Omega_k} F_A \to \mathbb{R}^d$ given by

$$t = \sum_{A \in \Omega_k} \mathbf{1}_A t_{A,B},$$

is a half-space-preserving map such that

$$t\sharp\mu = \nu.$$

This previous proposition asserts that the calculation of half-space-preserving transportation maps on partition sets are mutually independent. Thus, the global-map calculation is decentralized and stable under local perturbations.

Proposition 2 (Transitivity) *Suppose that $\mu_1, \mu_2, \cdots, \mu_n \in \mathcal{P}(\mathbb{R}^d)$, and $\{v_k\}_{k=1}^\infty$ are some partitioning directions. Furthermore, assume that*

$$t_k : \mathbb{R}^d \setminus F_k \to \mathbb{R}^d, \ 1 \le k \le n-1,$$

are half-space-preserving maps generated by $\{v_k\}_{k=1}^\infty$ such that

$$\mu_k(F_k) = 0, \quad t_k \sharp \mu_k = \mu_{k+1}, \ 1 \le k \le n-1.$$

Then, the mapping

$$t = t_{n-1} \circ t_{n-2} \circ \cdots \circ t_1 : \mathbb{R}^d \setminus F_1 \to \mathbb{R}^d,$$

is a half-space-preserving mapping corresponding to directions $\{v_k\}_{k=1}^\infty$, and

$$t \sharp \mu_1 = \mu_n.$$

4 Numerical Results

Our algorithm proceeds by successively partitioning the source and target measures. These partitions generate BSP tree decompositions of these measures: see Figures 2 and 3 for uniform and Gaussian measures illustrated in Figure 1.

The algorithm has a run-time of $O(n \log n)$ [4, 19]: the points are divided into halves or cells, along a horizontal cut, and then each half is divided into vertical halves. The algorithm is applied recursively until each cell has one point in it or until a stopping criterion is reached. Thus, the algorithm does not involve explicit optimization.

We compare transportation costs of maps generated by our algorithm with optimal ones for various transportation metrics. As a reference point, we also compute transportation costs of lexicographical-order-preserving or Knothe-Rosenblatt maps [17, 22]: these are fair candidates with the same run-time and without any optimization of the transportation cost.

We conducted our experiments on an Intel i7 2.9 GHz CPU with 8 GB of DDR3 RAM for $n = 2^{2N}$, $N = 2, 3, \ldots, 6$ number of points in the source and target measures.

To compare transportation costs generated by our algorithm, we implemented

(i) the Hungarian algorithm for the assignment problem that has a run-time of order n^3,

(ii) the linear programming (LP) algorithm for the Kantorovich relaxation of the optimal transportation problem,

(iii) our algorithm with horizontal-vertical partitioning (HV),

(iv) the Sinkhorn (SH) algorithm [5] for the entropic regularization (ER) of the optimal transportation problem as implemented in [9],

(v) the lexicographical sorting algorithm (LEX) to generate the Knothe-Rosenblatt map,
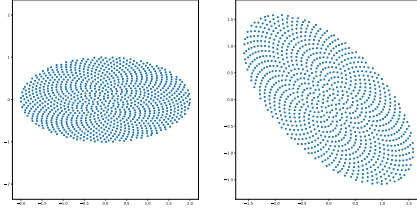
(i) Quasi-equally spaced source points within an ellipse
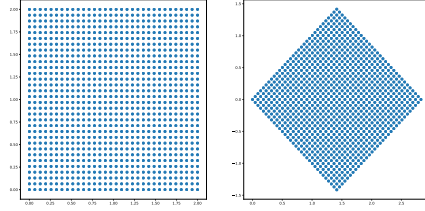
(ii) Equally spaced source points within a grid
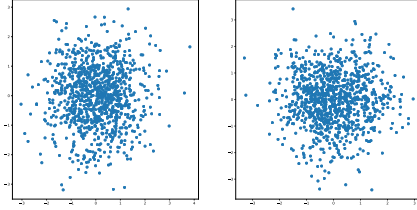
(iii) Normally distributed

Fig. 1 Visualization of computational examples.

The Hungarian and LP algorithms yielded the same cost with former breaking down for $N = 3$, so we report only the LP cost. In our experiments, we consider transportation costs of the form $c(x, y) = \|x - y\|_p^q$ for $p = 1, 2, \infty$ and $q = 1, 2$.

In the first set of examples, we consider empirical measures of quasi-uniformly spaced points in an ellipse, uniformly spaced points on a grid, and normally distributed points with zero mean and an identity covariance matrix, Figure 1, as source measures. As target measures, we consider their rotated versions by $45°$ counter-clockwise.

We summarize the results in Tables 1, 2, and 3. Note that only the Hungarian, HV, and LEX algorithms return a map. In practice, however, the LP solution yields a plan that is nearly a map. The SH solution is a plan that can have large support, depending on the ER regularization parameter. We found that the algorithm was faster with a regularization parameter 0.01 that led to quite wide plans. We report
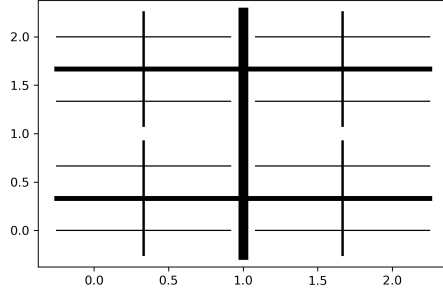
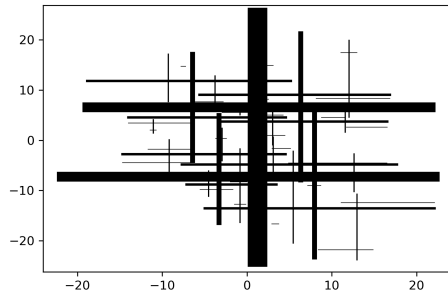**Fig. 2** Bisection on equally spaced points within a grid (source)

Fig. 3 Bisection on points sampled from $\mathcal{N}(0,1)$ (target)

the costs of maps or plans as appropriate. The conversion of plans to maps imposes extra costs associated with the projection that we also include.

From Tables 1, 2, and 3 we find that the accuracy of HV costs are comparable to those of SH. However, HV algorithm yields a map directly. Furthermore, HV maps perform much better than Knothe-Rosenblatt maps.

In the second set of examples, we map the uniform distribution on the unit square to the standard Gaussian, Table 4, and the latter to a rotated Gaussian with a 3 : 1 aspect ration, Table 5. We again observe that HV maps consistently estimate the optimal cost within a factor of 2 and in some cases, achieve remarkable accuracy of a few percent.

Finally, in Figure 4, we use HV maps to construct barycenters or displacement interpolations of various shapes. The results clearly indicate that HV maps are shape-sensitive and encode geometric information.

Appendix

Proof (Theorem 3) We assume that $\{e_i\}_{i=1}^d$ is the standard basis in $\mathbb{R}^d$ because the proof for a general basis is identical up to a multiplication by a suitable volume element.
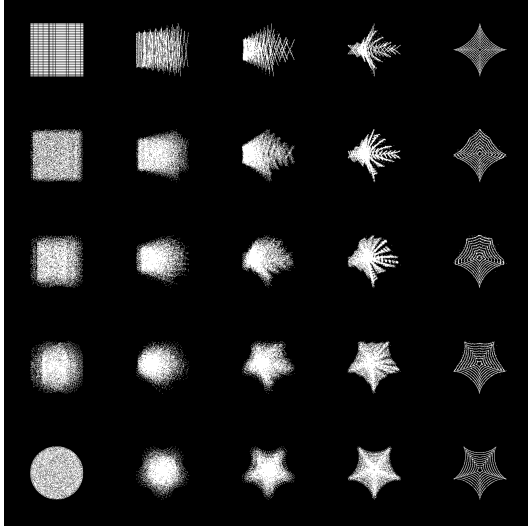
Fig. 4 Barycenters of four starting shapes (shown in corners of image) at various weights

Cost	Method		n							
			64		256		1024		4096	
$\|\cdot\|_2^2$	OT		0.32		0.26		0.24		0.25	
	HV	(ratio)	0.38	$(1.18\times)$	0.30	$(1.16\times)$	0.28	$(1.15\times)$	0.27	$(1.08\times)$
	SH	(ratio)	0.34	$(1.07\times)$	0.30	$(1.15\times)$	0.29	$(1.18\times)$	0.29	$(1.12\times)$
	LEX	(ratio)	0.85	$(2.66\times)$	0.78	$(3.02\times)$	0.76	$(3.09\times)$	0.74	$(2.92\times)$
$\|\cdot\|_2$	OT		0.49		0.44		0.43		0.42	
	HV	(ratio)	0.58	$(1.17\times)$	0.51	$(1.17\times)$	0.50	$(1.18\times)$	0.50	$(1.18\times)$
	SH	(ratio)	0.52	$(1.05\times)$	0.48	$(1.08\times)$	0.47	$(1.10\times)$	0.46	$(1.11\times)$
	LEX	(ratio)	0.85	$(1.72\times)$	0.81	$(1.84\times)$	0.80	$(1.88\times)$	0.79	$(1.89\times)$
$\|\cdot\|_1^2$	OT		0.49		0.39		0.36		0.38	
	HV	(ratio)	0.64	$(1.30\times)$	0.48	$(1.23\times)$	0.43	$(1.20\times)$	0.42	$(1.11\times)$
	SH	(ratio)	0.51	$(1.04\times)$	0.42	$(1.09\times)$	0.40	$(1.11\times)$	0.40	$(1.05\times)$
	LEX	(ratio)	1.40	$(2.86\times)$	1.28	$(3.33\times)$	1.24	$(3.45\times)$	1.22	$(3.24\times)$
$\|\cdot\|_\infty^2$	OT		0.23		0.19		0.18		0.18	
	HV	(ratio)	0.30	$(1.28\times)$	0.25	$(1.31\times)$	0.24	$(1.32\times)$	0.23	$(1.26\times)$
	SH	(ratio)	0.26	$(1.11\times)$	0.23	$(1.20\times)$	0.22	$(1.24\times)$	0.22	$(1.20\times)$
	LEX	(ratio)	0.70	$(2.97\times)$	0.64	$(3.34\times)$	0.62	$(3.45\times)$	0.61	$(3.31\times)$

Table 1 Average costs of set of quasi-equally spaced points within an ellipse to the same set of points rotated $45°$ counter-clockwise measured in various cost functions

1. Suppose that $x \in \text{int}\,(\text{supp}(\mu))$, and $x' \in \mathbb{R}^d$, $x \neq x'$, but they never get separated by a hyperplane. Therefore, whenever a common subset containing x, x' gets partitioned they always stay in the same side. Suppose that $\{A_k\}$ is the sequence of subsets that they both belong during the cutting process. By construction, we have that

$$A_{k+1} \subset A_k, \quad \mu(A_{k+1}) = \frac{\mu(A_k)}{2}.$$

Since $\{v_k\}_{k=1}^\infty \subset \{e_1, e_2, \cdots, e_d\}$ we have that

$$A_k = (\alpha_1^k, \beta_1^k] \times (\alpha_2^k, \beta_2^k] \cdots \times (\alpha_d^k, \beta_d^k],$$

Cost	Method		64	n 256		1024		4096	
$\|\cdot\|_2^2$	OT		0.31		0.30		0.30		0.31	
	HV	(ratio)	0.31	(1.01×)	0.31	(1.01×)	0.30	(1.01×)	0.30	(0.98×)
	SH	(ratio)	0.35	(1.12×)	0.34	(1.14×)	0.34	(1.14×)	0.34	(1.11×)
	LEX	(ratio)	0.42	(1.36×)	0.40	(1.33×)	0.40	(1.32×)	0.39	(1.27×)
$\|\cdot\|_2$	OT		0.54		0.54		0.54		0.54	
	HV	(ratio)	0.55	(1.01×)	0.54	(1.01×)	0.54	(1.01×)	0.54	(1.01×)
	SH	(ratio)	0.57	(1.05×)	0.57	(1.05×)	0.57	(1.05×)	0.57	(1.05×)
	LEX	(ratio)	0.60	(1.11×)	0.59	(1.09×)	0.59	(1.09×)	0.59	(1.08×)
$\|\cdot\|_1^2$	OT		0.51		0.51		0.50		0.50	
	HV	(ratio)	0.52	(1.02×)	0.52	(1.02×)	0.51	(1.02×)	0.51	(1.01×)
	SH	(ratio)	0.53	(1.04×)	0.53	(1.05×)	0.53	(1.05×)	0.53	(1.05×)
	LEX	(ratio)	0.69	(1.35×)	0.67	(1.32×)	0.65	(1.30×)	0.65	(1.29×)
$\|\cdot\|_\infty^2$	OT		0.26		0.25		0.25		0.25	
	HV	(ratio)	0.26	(1.03×)	0.26	(1.02×)	0.26	(1.02×)	0.26	(1.02×)
	SH	(ratio)	0.28	(1.09×)	0.28	(1.10×)	0.28	(1.10×)	0.28	(1.10×)
	LEX	(ratio)	0.35	(1.35×)	0.33	(1.32×)	0.33	(1.30×)	0.32	(1.28×)

Table 2 Average costs of set of points equally spaced within the unit grid to the same set of points rotated 45° counter-clockwise measured in various cost functions

Cost	Method		64	n 256		1024		4096	
$\|\cdot\|_2^2$	OT		0.23		0.10		0.04		0.01	
	HV	(ratio)	0.49	(2.15×)	0.22	(2.20×)	0.09	(2.49×)	0.04	(3.08×)
	SH	(ratio)	0.24	(1.08×)	0.13	(1.30×)	0.08	(2.12×)	0.06	(4.55×)
	LEX	(ratio)	1.14	(5.05×)	1.14	(11.23×)	1.20	(32.91×)	1.18	(91.63×)
$\|\cdot\|_2$	OT		0.39		0.19		0.15		0.07	
	HV	(ratio)	0.49	(1.25×)	0.32	(1.63×)	0.24	(1.64×)	0.13	(1.69×)
	SH	(ratio)	0.41	(1.06×)	0.22	(1.15×)	0.19	(1.28×)	0.13	(1.76×)
	LEX	(ratio)	0.86	(2.19×)	0.86	(4.44×)	0.97	(6.68×)	0.95	(12.85×)
$\|\cdot\|_1^2$	OT		0.41		0.14		0.06		0.02	
	HV	(ratio)	0.74	(1.78×)	0.27	(1.94×)	0.16	(2.49×)	0.05	(3.06×)
	SH	(ratio)	0.43	(1.05×)	0.16	(1.19×)	0.10	(1.61×)	0.06	(3.80×)
	LEX	(ratio)	1.60	(3.86×)	1.98	(14.39×)	1.92	(30.63×)	1.86	(117.01×)
$\|\cdot\|_\infty^2$	OT		0.17		0.09		0.03		0.01	
	HV	(ratio)	0.31	(1.81×)	0.21	(2.33×)	0.08	(3.28×)	0.04	(3.95×)
	SH	(ratio)	0.19	(1.13×)	0.12	(1.37×)	0.07	(2.64×)	0.06	(5.77×)
	LEX	(ratio)	0.97	(5.66×)	0.96	(10.92×)	0.97	(38.32×)	0.96	(99.29×)

Table 3 Average costs of sets of points sampled from $\mathcal{N}(0,1)$ to the same set of points rotated 45° counter-clockwise measured in various cost functions

where $-\infty \le \alpha_i^k < \beta_i^k \le +\infty$. Denote by

$$x = (x_1, x_2, \cdots, x_d), \quad x' = (x'_1, x'_2, \cdots, x'_d).$$

Without loss of generality, assume that

$$x_i \ne x'_i, \ 1 \le i \le l, \quad x_i = x'_i, \ i > l.$$

Since $\{A_k\}$ are rectangles we have that

$$R = \bigcap_k A_k,$$

Cost	Method		n 64		256		1024		4096	
$\|\cdot\|_2^2$	OT		1.56		1.76		1.53		1.55	
	HV	(ratio)	1.58	$(1.01\times)$	1.76	$(1.00\times)$	1.53	$(1.00\times)$	1.54	$(0.99\times)$
	SH	(ratio)	1.60	$(1.02\times)$	1.80	$(1.02\times)$	1.57	$(1.03\times)$	1.59	$(1.02\times)$
	LEX	(ratio)	2.86	$(1.83\times)$	3.02	$(1.72\times)$	2.65	$(1.73\times)$	2.68	$(1.73\times)$
$\|\cdot\|_2$	OT		0.96		1.07		1.05		1.09	
	HV	(ratio)	0.98	$(1.01\times)$	1.08	$(1.01\times)$	1.05	$(1.00\times)$	1.09	$(1.00\times)$
	SH	(ratio)	0.99	$(1.03\times)$	1.10	$(1.03\times)$	1.08	$(1.03\times)$	1.12	$(1.03\times)$
	LEX	(ratio)	1.42	$(1.47\times)$	1.46	$(1.36\times)$	1.42	$(1.35\times)$	1.45	$(1.33\times)$
$\|\cdot\|_1^2$	OT		3.23		2.42		2.53		2.52	
	HV	(ratio)	3.31	$(1.03\times)$	2.46	$(1.02\times)$	2.56	$(1.01\times)$	2.53	$(1.01\times)$
	SH	(ratio)	3.25	$(1.01\times)$	2.45	$(1.01\times)$	2.56	$(1.01\times)$	2.54	$(1.01\times)$
	LEX	(ratio)	5.44	$(1.69\times)$	4.26	$(1.76\times)$	4.43	$(1.75\times)$	4.41	$(1.75\times)$
$\|\cdot\|_\infty^2$	OT		0.97		1.34		1.25		1.28	
	HV	(ratio)	1.01	$(1.04\times)$	1.35	$(1.01\times)$	1.25	$(1.00\times)$	1.29	$(1.00\times)$
	SH	(ratio)	0.99	$(1.02\times)$	1.36	$(1.02\times)$	1.28	$(1.02\times)$	1.31	$(1.02\times)$
	LEX	(ratio)	2.15	$(2.22\times)$	2.27	$(1.70\times)$	2.16	$(1.73\times)$	2.23	$(1.74\times)$

Table 4 Average costs of set of points equally spaced within a grid to a set of points sampled from $\mathcal{N}(0,1)$ measured in various cost functions

Cost	Method		n 64		256		1024		4096	
$\|\cdot\|_2^2$	OT		0.97		1.02		1.23		1.14	
	HV	(ratio)	1.49	$(1.54\times)$	1.23	$(1.21\times)$	1.32	$(1.07\times)$	1.16	$(1.02\times)$
	SH	(ratio)	0.98	$(1.02\times)$	1.05	$(1.03\times)$	1.26	$(1.03\times)$	1.16	$(1.02\times)$
	LEX	(ratio)	6.52	$(6.72\times)$	6.59	$(6.46\times)$	7.00	$(5.72\times)$	7.13	$(6.27\times)$
$\|\cdot\|_2$	OT		1.01		0.77		0.78		0.84	
	HV	(ratio)	1.15	$(1.13\times)$	0.83	$(1.08\times)$	0.81	$(1.04\times)$	0.85	$(1.02\times)$
	SH	(ratio)	1.04	$(1.02\times)$	0.80	$(1.03\times)$	0.81	$(1.04\times)$	0.87	$(1.04\times)$
	LEX	(ratio)	2.25	$(2.22\times)$	2.18	$(2.83\times)$	2.27	$(2.92\times)$	2.31	$(2.75\times)$
$\|\cdot\|_1^2$	OT		1.27		1.47		1.22		1.09	
	HV	(ratio)	1.74	$(1.38\times)$	1.75	$(1.19\times)$	1.37	$(1.13\times)$	1.15	$(1.05\times)$
	SH	(ratio)	1.28	$(1.01\times)$	1.49	$(1.01\times)$	1.25	$(1.03\times)$	1.10	$(1.01\times)$
	LEX	(ratio)	9.42	$(7.44\times)$	10.37	$(7.07\times)$	11.27	$(9.27\times)$	11.13	$(10.21\times)$
$\|\cdot\|_\infty^2$	OT		1.39		1.23		1.06		1.00	
	HV	(ratio)	1.79	$(1.29\times)$	1.49	$(1.21\times)$	1.19	$(1.12\times)$	1.05	$(1.05\times)$
	SH	(ratio)	1.41	$(1.01\times)$	1.26	$(1.02\times)$	1.08	$(1.03\times)$	1.03	$(1.02\times)$
	LEX	(ratio)	6.58	$(4.75\times)$	5.97	$(4.85\times)$	5.89	$(5.57\times)$	5.97	$(5.94\times)$

Table 5 Average costs of sets of points sampled from $\mathcal{N}(0,1)$ to a new set of points sampled from $\mathcal{N}(0,1)$ (then scaled to a 3:1 aspect ratio and rotated $90°$ counter-clockwise) measured in various cost functions

is also a rectangle, and we denote by

$$R = [\alpha_1, \beta_1] \times [\alpha_2, \beta_2] \times \cdots [\alpha_d, \beta_d],$$

and we have that

$$\lim_{k \to \infty} \alpha_i^k = \alpha_i, \quad \lim_{k \to \infty} \beta_i^k = \beta_i, \ 1 \le i \le d.$$

Since $x, x' \in R$ we have that

$$\beta_i - \alpha_i \ge |x_i - x_i'| > 0, \ 1 \le i \le l.$$

Furthermore, we have that

$$\mu(R) = \lim_{k \to \infty} \mu(A_k) = 0.$$

If $\mathcal{L}^d(R) > 0$ then we have that $\text{int}(R) \neq \emptyset$ and $\text{int}(R) \cap \text{supp}(\mu) = \emptyset$ which contradicts to the fact that $x \in \text{int}(\text{supp}(\mu))$. Therefore, we have that $\mathcal{L}^d(R) = 0$ which means that $\alpha_i = \beta_i$ for some $i > l$. Without loss of generality assume that

$$\alpha_i < \beta_i, \ 1 \leq i \leq q, \quad \alpha_i = \beta_i, \ i > q.$$

We have that $q \geq l$. Moreover, $\alpha_i = \beta_i = x_i = x_i'$, and $-\infty < \alpha_i^k < \beta_i^k < \infty$ for all $i > q$ and k large enough. Additionally, if $-\infty < \alpha_i < \beta_i < \infty$ for some $1 \leq i \leq q$ then $-\infty < \alpha_i^k < \beta_i^k < \infty$ for k large enough. In what follows we assume that k is so large that this previous statements hold.

Furthermore, assume that $M > 0$ is such that

$$\text{supp}(\mu) \subset [-M, M]^d.$$

Since $\mu = f dx$, by construction we have that

$$\int_{A_k \setminus A_{k+1}} f dx = \int_{A_{k+1}} f dx, \ \forall k.$$

Therefore, using $c \leq f \leq C$, μ a.e. we get that

$$\frac{\mathcal{L}^d(A_k \setminus A_{k+1} \cap \text{supp}(\mu))}{\mathcal{L}^d(A_{k+1} \cap \text{supp}(\mu))} \geq \frac{c}{C} > 0, \ \forall k. \tag{6}$$

Now, suppose that for some k the set A_k gets partitioned in the direction e_1. There are three possibilities: a) $-\infty < \alpha_1 < \beta_1 < \infty$, b) $-\infty < \alpha_1 < \beta_1 = \infty$, c) $-\infty = \alpha_1 < \beta_1 < \infty$.
a) $-\infty < \alpha_1 < \beta_1 < \infty$. In this case, we have that $-\infty < \alpha_1^k < \beta_1^k < \infty$ since k is large enough. Therefore, either

$$A_{k+1} = (\alpha_1^k, \gamma] \times (\alpha_2^k, \beta_2^k] \cdots \times (\alpha_d^k, \beta_d^k],$$
$$A_k \setminus A_{k+1} = (\gamma, \beta_1^k] \times (\alpha_2^k, \beta_2^k] \cdots \times (\alpha_d^k, \beta_d^k],$$

or

$$A_{k+1} = (\gamma, \beta_1^k] \times (\alpha_2^k, \beta_2^k] \cdots \times (\alpha_d^k, \beta_d^k],$$
$$A_k \setminus A_{k+1} = (\alpha_1^k, \gamma] \times (\alpha_2^k, \beta_2^k] \cdots \times (\alpha_d^k, \beta_d^k],$$

for some $\alpha_1^k < \gamma < \beta_1^k$. Suppose that we are in the former case.

Since $x \in \text{int}(\text{supp}(\mu))$ we have that there exists a $\sigma > 0$ such that

$$\times_{i=1}^d [x_i - \sigma, x_i + \sigma] \subset \text{supp}(\mu).$$

We have that

$$A_k \setminus A_{k+1} \cap \text{supp}(\mu) \subset A_k \setminus A_{k+1} \cap [-M, M]^d,$$

and therefore

$$\mathcal{L}^d(A_k \setminus A_{k+1} \cap \text{supp}(\mu))$$

$$\leq \mathcal{L}^d(A_k \setminus A_{k+1} \cap [-M, M]^d) \leq (\beta_1^k - \gamma) \prod_{i=2}^d \min\{\beta_i^k - \alpha_i^k, 2M\}$$

$$\leq (\beta_1^k - \beta_1)(2M)^{q-1} \prod_{i>q} (\beta_i^k - \alpha_i^k),$$

where we used the fact that $\beta_1 \leq \gamma < \beta_1^k$ and

$$\lim_{k \to \infty} \alpha_i^k = \lim_{k \to \infty} \beta_i^k = x_i, \ i > q.$$

On the other hand, we have that

$$A_{k+1} \cap \text{supp}(\mu) \supset A_{k+1} \cap \times_{i=1}^d [x_i - \sigma, x_i + \sigma],$$

therefore

$$\mathcal{L}^d(A_{k+1} \cap \text{supp}(\mu))$$

$$\geq \mathcal{L}^d(A_{k+1} \cap \times_{i=1}^d [x_i - \sigma, x_i + \sigma]) \geq \min\{\gamma - \alpha_1^k, \sigma\} \prod_{i=2}^d \min\{\beta_i^k - \alpha_i^k, \sigma\}$$

$$\geq \prod_{i=1}^q \min\{\beta_i - \alpha_i, \sigma\} \prod_{i>q} (\beta_i^k - \alpha_i^k).$$

Hence, we obtain that

$$\frac{\mathcal{L}^d(A_k \setminus A_{k+1} \cap \text{supp}(\mu))}{\mathcal{L}^d(A_{k+1} \cap \text{supp}(\mu))} \leq (\beta_1^k - \beta_1) \frac{(2M)^{q-1}}{\prod_{i=1}^q \min\{\beta_i - \alpha_i, \sigma\}}.$$

Similarly, if x, x' fall in the upper (in e_1 direction) half of A_k we get that

$$\frac{\mathcal{L}^d(A_k \setminus A_{k+1} \cap \text{supp}(\mu))}{\mathcal{L}^d(A_{k+1} \cap \text{supp}(\mu))} \leq (\alpha_1 - \alpha_1^k) \frac{(2M)^{q-1}}{\prod_{i=1}^q \min\{\beta_i - \alpha_i, \sigma\}}.$$

b) $-\infty < \alpha_1 < \beta_1 = \infty$. In this case we have that $-\infty < \alpha_1^k < \beta_1^k = \infty$, and

$$A_{k+1} = (\gamma, \infty] \times (\alpha_2^k, \beta_2^k] \cdots \times (\alpha_d^k, \beta_d^k],$$
$$A_k \setminus A_{k+1} = (\alpha_1^k, \gamma] \times (\alpha_2^k, \beta_2^k] \cdots \times (\alpha_d^k, \beta_d^k],$$

for some $\alpha_1^k < \gamma < \infty$. As before, we have that

$$\mathcal{L}^d(A_k \setminus A_{k+1} \cap \text{supp}(\mu))$$
$$\leq (\gamma - \alpha_1^k)(2M)^{q-1} \prod_{i>q} (\beta_i^k - \alpha_i^k) \leq (\alpha_1 - \alpha_1^k)(2M)^{q-1} \prod_{i>q} (\beta_i^k - \alpha_i^k).$$

Similarly, we have that

$$\mathcal{L}^d(A_{k+1} \cap \text{supp}(\mu)) \geq \prod_{i=1}^q \min\{\beta_i - \alpha_i, \sigma\} \prod_{i>q} (\beta_i^k - \alpha_i^k),$$

and thus

$$\frac{\mathcal{L}^d(A_k \setminus A_{k+1} \cap \text{supp}(\mu))}{\mathcal{L}^d(A_{k+1} \cap \text{supp}(\mu))} \le (\alpha_1 - \alpha_1^k)\frac{(2M)^{q-1}}{\prod_{i=1}^q \min\{\beta_i - \alpha_i, \sigma\}}.$$

c) $-\infty = \alpha_1 < \beta_1 < \infty$. In this case, we have that $-\infty = \alpha_1^k < \beta_1^k < \infty$, and

$$\frac{\mathcal{L}^d(A_k \setminus A_{k+1} \cap \text{supp}(\mu))}{\mathcal{L}^d(A_{k+1} \cap \text{supp}(\mu))} \le (\beta_1^k - \beta_1)\frac{(2M)^{q-1}}{\prod_{i=1}^q \min\{\beta_i - \alpha_i, \sigma\}}.$$

Summarizing, we get whenever A_k gets partitioned in e_1 we have that

$$\frac{\mathcal{L}^d(A_k \setminus A_{k+1} \cap \text{supp}(\mu))}{\mathcal{L}^d(A_{k+1} \cap \text{supp}(\mu))} \le o(1).$$

We get similar estimates for partitions in any of the directions $\{e_i\}_{i=1}^q$. Thus, if we take a subsequence $\{A_{k_m}\}$ that get partitioned in one of these directions we get that

$$\lim_{m \to \infty} \frac{\mathcal{L}^d(A_{k_m} \setminus A_{k_m+1} \cap \text{supp}(\mu))}{\mathcal{L}^d(A_{k_m+1} \cap \text{supp}(\mu))} = 0,$$

which contradicts to (6). Thus, the first item is proven.

2. Firstly, we will show that for every $x \in \text{int}(\text{supp}(\mu))$ there exists $y \in \text{supp}(\nu)$ such that

$$\hat{s}(x) = \hat{r}(y).$$

Assume that $\{A_k\}$ is the sequence of partition sets that contain x. Again, we have that

$$A_{k+1} \subset A_k, \quad \mu(A_{k+1}) = \frac{\mu(A_k)}{2},$$

and

$$A_k = (\alpha_1^k, \beta_1^k] \times (\alpha_2^k, \beta_2^k] \cdots \times (\alpha_d^k, \beta_d^k],$$

for some $-\infty \le \alpha_i^k < \beta_i^k \le +\infty$. Moreover, from the previous item we obtain that

$$\bigcap_k A_k = \{x\},$$

because $x \in \text{int}(\text{supp}(\mu))$, and the intersection cannot contain any other point. Therefore, we have that $-\infty < \alpha_i^k < \beta_i^k < +\infty$, and

$$\alpha_i^k \nearrow x_i, \quad \beta_i^k \searrow x_i, \text{ as } k \to \infty,$$

where $x = (x_1, x_2, \cdots, x_d)$.

Denote by $\{B_k\}$ the dual sequence of $\{A_k\}$ that partition ν. Again, we have that

$$B_k = (\gamma_1^k, \delta_1^k] \times (\gamma_2^k, \delta_2^k] \cdots \times (\gamma_d^k, \delta_d^k],$$

for some $-\infty < \gamma_i^k < \delta_i^k < +\infty$. Thus, our first task is to show that

$$\bigcap_k B_k \cap \text{supp}(\nu) \ne \emptyset.$$

Since $\alpha_i^k < x_i$ we have that $\{\alpha_i^k\}_k$ is not an eventually constant sequence. Therefore, $\{\gamma_i^k\}_k$ is also not an eventually constant sequence. Besides, $\{\gamma_i^k\}_k$ and

$\{\delta_i^k\}_k$ are, respectively, nondecreasing and nonincreasing sequences. Therefore, we have that

$$W_x = \bigcap_k B_k = [\gamma_1, \delta_1] \times [\gamma_2, \delta_2] \cdots \times [\gamma_d, \delta_d],$$

where

$$\gamma_i = \sup_k \gamma_i^k, \quad \delta_i = \inf_k \delta_i^k, \ 1 \le i \le d.$$

In fact, we have that

$$W_x = \bigcap_k \mathrm{cl}(B_k).$$

Since $\nu(B_k) > 0$ we have that $\mathrm{cl}(B_k) \cap \mathrm{supp}(\nu) \ne \emptyset$. Thus,

$$\{\mathrm{cl}(B_k) \cap \mathrm{supp}(\nu)\}_k$$

is a nested family of nonempty compact sets. Therefore, we have that

$$W_x \cap \mathrm{supp}(\nu) = \bigcap_k \mathrm{cl}(B_k) \cap \mathrm{supp}(\nu) \ne \emptyset.$$

If $W_x \cap \mathrm{int}(\mathrm{supp}(\nu)) \ne \emptyset$ then by item 1, we get that

$$W_x = \{y\},$$

for some $y \in \mathrm{int}(\mathrm{supp}(\nu))$. Hence, to complete the proof of item 1, we need to show that there exists a $F_0 \in \mathcal{B}(\mathbb{R}^d)$ such that $\mu(F_0) = 0$, and

$$W_x \cap \mathrm{int}(\mathrm{supp}(\nu)) \ne \emptyset, \quad \forall x \notin \mathrm{int}(\mathrm{supp}(\mu)) \setminus F_0.$$

For every k denote by

$$\Delta_k' = \{B \in \Delta_k \text{ s,t. } B \cap \partial(\mathrm{supp}(\nu)) \ne \emptyset\},$$

and

$$\Delta_k'' = \Delta_k \setminus \Delta_k'.$$

Since

$$\bigcup_{\Delta_k} B = \mathbb{R}^d,$$

we obtain that

$$\partial(\mathrm{supp}(\nu)) \subset \bigcup_{\Delta_k'} B = H_k.$$

Furthermore, for every $B \in \Delta_k''$ we have that $\nu(B) > 0$, and therefore $B \cap \mathrm{supp}(\nu) \ne \emptyset$. On the other hand, $B \cap \partial(\mathrm{supp}(\nu)) = \emptyset$, and B is connected. Hence, $B \subset \mathrm{int}(\mathrm{supp}(\nu))$, and

$$G_k = \bigcup_{\Delta_k''} B \subset \mathrm{int}(\mathrm{supp}(\nu)).$$

Note that

$$H_k \supset H_{k+1}, \quad G_k \subset G_{k+1}, \ \forall k.$$

By item 1 we have that for every $y \in \text{int}(\text{supp}(\nu))$ there exists a partition rectangle B such that $y \in B \subset \text{int}(\text{supp}(\nu))$. Hence, $B \in \Delta_k''$ for some k, and $y \in G_k$. Therefore, we obtain that

$$\bigcup_k G_k = \text{int}(\text{supp}(\nu)).$$

Consequently,

$$\bigcap_k H_k = \mathbb{R}^d \setminus \text{int}(\text{supp}(\nu)) \supset \partial(\text{supp}(\nu)).$$

Since $\nu(\text{int}(\text{supp}(\nu))) = 1$ we get that

$$\nu\left(\bigcup_k G_k\right) = 1, \quad \nu\left(\bigcap_k H_k\right) = 0.$$

Denote by Ω_k' and Ω_k'' the families dual to Δ_k' and Δ_k''. Furthermore, denote by

$$F_k = \bigcup_{\Omega_k'} A, \quad E_k = \bigcup_{\Omega_k''} A.$$

By construction, we have that

$$F_k \supset F_{k+1}, \quad E_k \subset E_{k+1}, \ \forall k.$$

Moreover,

$$\mu(F_k) = \nu(H_k), \quad \mu(E_k) = \nu(G_k).$$

Denote by

$$F_0 = \bigcap_k F_k.$$

Then, we have that $F_0 \in \mathcal{B}(\mathbb{R}^d)$, and

$$\mu(F_0) = \lim_{k \to \infty} \mu(F_k) = \lim_{k \to \infty} \nu(H_k) = \nu\left(\bigcap_k H_k\right) = 0.$$

Finally, note that if $W_x \cap \text{int}(\text{supp}(\nu)) = \emptyset$ then $x \in F_0$.

3. From items 1, 2 we have that the map $\hat{t} : \text{int}(\text{supp}(\mu)) \setminus F_0 \to \text{int}(\text{supp}(\nu))$ given by

$$\hat{t}(x) = \hat{r}^{-1}(\hat{s}(x)),$$

is well defined. Our first task is to show that $\hat{t}$ is Borel measurable. For that, we need to show that $\hat{t}^{-1}(G) \in \mathcal{B}(\mathbb{R}^d)$ for any open set $G \subset \mathbb{R}^d$. Since $\text{Im}(\hat{t}) \subset \text{int}(\text{supp}(\nu))$ we have that

$$\hat{t}^{-1}(G) = \hat{t}^{-1}\left(G \cap \text{int}(\text{supp}(\nu))\right).$$

Therefore, we may assume that $G \subset \text{int}(\text{supp}(\nu))$. Furthermore, denote by

$$\Delta_k'' = \{B \in \Delta_k \text{ s.t. } B \subset G\}, \quad \Delta_k' = \Delta_k \setminus \Delta_k''.$$

Next, define

$$G_k = \bigcup_{\Delta_k''} B.$$

Then we have that
$$G_k \subset G_{k+1}, \quad G_k \subset G, \ \forall k.$$
From item 1, we have that for every $y \in G$ there exists a partition set B such that $y \in B \subset G$. Hence, we get that
$$\bigcup_k G_k = G.$$
This means that
$$\hat{t}^{-1}(G) = \bigcup_k \hat{t}^{-1}(G_k) = \bigcup_k \bigcup_{\Delta_k''} \hat{t}^{-1}(B).$$
On the other hand, from equation (4) in Theorem 2 we have that
$$\hat{t}^{-1}(G) = \bigcup_k \bigcup_{\Omega_k''} (A \cap \mathrm{int}(\mathrm{supp}(\mu)) \setminus F_0),$$
where Ω_k', Ω_k'' are the dual families of Δ_k', Δ_k''. Therefore, we have that $\hat{t}^{-1}(G) \in \mathcal{B}(\mathbb{R}^d)$. Furthermore, denote by
$$F_k = \bigcup_{\Omega_k''} (A \setminus F_0), \ \forall k.$$
By construction, we have that
$$F_k \subset F_{k+1}, \ \forall k,$$
and hence
$$\mu\left(\hat{t}^{-1}(G)\right) = \lim_{k \to \infty} \mu(F_k).$$
On the other hand, from construction and equations $\mu(F_0) = 0$, $\mu(\mathrm{int}(\mathrm{supp}(\mu))) = 1$, we have that
$$\mu(F_k) = \sum_{\Omega_k''} \mu(A) = \sum_{\Delta_k''} \nu(B) = \nu(G_k).$$
Therefore, we obtain that
$$\mu\left(\hat{t}^{-1}(G)\right) = \lim_{k \to \infty} \mu(F_k) = \lim_{k \to \infty} \nu(G_k) = \nu(G).$$
Thus, $\hat{t}$ is Borel measurable and $\hat{t}\sharp\mu = \nu$. Finally, from item 1 we have that points in $\mathrm{Im}(\hat{t}) \subset \mathrm{int}(\mathrm{supp}(\nu))$ get eventually separated. Therefore, by Theorem 2 we obtain that $\hat{t}$ is half-space-preserving.

Proof (Theorem 4) Denote by H the union of all the hyperplanes that partition μ. Then we have that $\mathcal{L}^d(H) = \mu(H) = 0$. We prove that $\hat{t}$ is continuous on $\mathrm{Dom}(\hat{t}) \setminus H$.

Suppose $x \in \mathrm{Dom}(\hat{t}) \setminus H$, and $y = \hat{t}(x)$. Furthermore, denote by $\{A_k\}$ and $\{B_k\}$ the rectangles that contain x and y, respectively. We have that
$$\bigcap_k A_k = \{x\}, \quad \bigcap_k B_k = \{y\}.$$
Moreover, since $x \notin H$ we have that $x \in \mathrm{int}(A_k)$ for all k. Therefore, by construction, we have that $y \in \mathrm{int}(B_k)$ for all k. Thus, we obtain that
$$\bigcap_k \mathrm{int}(A_k) = \{x\}, \quad \bigcap_k \mathrm{int}(B_k) = \{y\},$$
which yields the continuity.

References

1. Altschuler, J., Bach, F., Rudi, A., Weed, J.: Approximating the quadratic transportation metric in near-linear time. arXiv preprint arXiv:1810.10046 (2018)
2. Altschuler, J., Weed, J., Rigollet, P.: Near-linear time approximation algorithms for optimal transport via sinkhorn iteration. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, pp. 1961–1971. Curran Associates Inc., USA (2017). URL http://dl.acm.org/citation.cfm?id=3294771.3294958
3. Ambrosio, L., Gigli, N., Savaré, G.: Gradient flows in metric spaces and in the space of probability measures, second edn. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel (2008)
4. Blum, M., Floyd, R.W., Pratt, V., Rivest, R.L., Tarjan, R.E.: Time bounds for selection. Journal of Computer and System Sciences **7**(4), 448 – 461 (1973). DOI https://doi.org/10.1016/S0022-0000(73)80033-9. URL http://www.sciencedirect.com/science/article/pii/S0022000073800339
5. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. In: Advances in neural information processing systems, pp. 2292–2300 (2013)
6. Cuturi, M., Doucet, A.: Fast computation of wasserstein barycenters. In: E.P. Xing, T. Jebara (eds.) Proceedings of the 31st International Conference on Machine Learning, *Proceedings of Machine Learning Research*, vol. 32, pp. 685–693. PMLR, Bejing, China (2014). URL http://proceedings.mlr.press/v32/cuturi14.html
7. Dasgupta, S., Freund, Y.: Random projection trees and low dimensional manifolds. Citeseer (2008)
8. Dasgupta, S., Sinha, K.: Randomized partition trees for exact nearest neighbor search. In: Conference on Learning Theory, pp. 317–337 (2013)
9. Flamary, R., Courty, N.: POT Python optimal transport library (2017). URL https://github.com/rflamary/POT
10. Genevay, A., Cuturi, M., Peyré, G., Bach, F.: Stochastic optimization for large-scale optimal transport. In: Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16, pp. 3440–3448. Curran Associates Inc., USA (2016). URL http://dl.acm.org/citation.cfm?id=3157382.3157482
11. Indyk, P.: Algorithms for dynamic geometric problems over data streams. In: Proceedings of the Thirty-sixth Annual ACM Symposium on Theory of Computing, STOC '04, pp. 373–380. ACM, New York, NY, USA (2004). DOI 10.1145/1007352.1007413. URL http://doi.acm.org/10.1145/1007352.1007413
12. Indyk, P.: Nearest neighbors in high-dimensional spaces. Handbook of Discrete and Computational Geometry. Citeseer (2004)
13. Indyk, P.: A near linear time constant factor approximation for euclidean bichromatic matching (cost). In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07, pp. 39–42. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (2007). URL http://dl.acm.org/citation.cfm?id=1283383.1283388
14. Indyk, P., Thaper, N.: Fast image retrieval via embeddings
15. Jacobs, M., Léger, F.: A fast approach to optimal transport: the back-and-forth method. Preprint (2019). URL https://arxiv.org/abs/1905.12154. ArXiv:1905.12154 [math.OC]
16. Johnson, J., Douze, M., Jégou, H.: Billion-scale similarity search with gpus. IEEE Transactions on Big Data (2019)
17. Knothe, H.: Contributions to the theory of convex bodies. Michigan Math. J. **4**, 39–52 (1957). URL http://projecteuclid.org/euclid.mmj/1028990175
18. Li, K., Malik, J.: Fast k-nearest neighbour search via prioritized dci. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70, pp. 2081–2090. JMLR.org (2017)
19. Musser, D.R.: Introspective sorting and selection algorithms. Software: Practice and Experience **27**(8), 983–993 (1997). DOI 10.1002/(SICI)1097-024X(199708)27:8⟨983::AID-SPE117⟩3.0.CO;2-\#. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291097-024X%28199708%2927%3A8%3C983%3A%3AAID-SPE117%3E3.0.CO%3B2-%23
20. Peyré, G., Cuturi, M.: Computational Optimal Transport: With Applications to Data Science. now (2019). URL https://ieeexplore.ieee.org/document/8641476
21. Radha, H., Vetterli, M., Leonardi, R.: Image compression using binary space partitioning trees. IEEE Transactions on Image Processing **5**(12), 1610–1624 (1996). DOI 10.1109/83.544569

22. Rosenblatt, M.: Remarks on a multivariate transformation. Ann. Math. Statistics **23**, 470–472 (1952). DOI 10.1214/aoms/1177729394. URL `https://doi.org/10.1214/aoms/1177729394`

23. Schmitzer, B.: Stabilized sparse scaling algorithms for entropy regularized transport problems. SIAM Journal on Scientific Computing **41**(3), A1443–A1481 (2019)

24. Schrijver, A.: Combinatorial optimization: polyhedra and efficiency, vol. 24. Springer Science & Business Media (2003)

25. Villani, C.: Topics in optimal transportation, *Graduate Studies in Mathematics*, vol. 58. American Mathematical Society, Providence, RI (2003). DOI 10.1007/b12016. URL `https://doi.org/10.1007/b12016`

26. Villani, C.: Optimal transport, *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, vol. 338. Springer-Verlag, Berlin (2009). DOI 10.1007/978-3-540-71050-9. URL `https://doi.org/10.1007/978-3-540-71050-9`. Old and new

27. Wu, X.: Image coding by adaptive tree-structured segmentation. IEEE Transactions on Information Theory **38**(6), 1755–1767 (1992). DOI 10.1109/18.165448