Fast Algorithms for Continuous Optimal Transport Between Histograms

Pankaj K. Agarwal
Department of Computer Science, Duke University, Durham, NC, USA

Sharath Raghvendra
Department of Computer Science, North Carolina State University, Raleigh, NC, USA

Keegan Yao
Department of Computer Science, Duke University, Durham, NC, USA

Abstract

We give the first (relative) $(1+\varepsilon)$-approximation algorithm for the continuous optimal transport (OT) problem between two axis-aligned histograms in the plane, each of which is represented as a piecewise-constant function over a rectangular subdivision. Our algorithm runs in nearly quadratic time in the complexity of the histograms in the worst case. In contrast to the discrete and semi-discrete optimal transport problems, which always admit linear-size OT plans, we demonstrate that there is a quadratic lower bound on the complexity of even a $(1+\varepsilon)$-approximate OT plan in the continuous setting in the worst case. This suggests that the runtime of our approximation algorithm nearly matches the worst-case lower-bound complexity of an explicit $(1+\varepsilon)$-approximate OT plan between two histograms. We additionally provide near-linear time algorithms with either weaker approximation guarantees or restrictions on the input histograms.

2012 ACM Subject Classification Theory of computation → Approximation algorithms analysis; Theory of computation → Computational geometry

Keywords and phrases Optimal transport, min-cost flow, minimum-weight matching, quadtrees, approximation algorithms, dynamic optimal transport algorithm

Digital Object Identifier 10.4230/LIPIcs.ESA.2026.123

Funding *Pankaj K. Agarwal*: NSF grants CCF-22-23870 and IIS-24-02823, and US-Israel Binational Science Foundation Grant 2022131.
Sharath Raghvendra: NSF grant CCF-2514753.
Keegan Yao: NSF grants CCF-22-23870 and IIS-24-02823, and US-Israel Binational Science Foundation Grant 2022131.

1 Introduction

Optimal transport (OT) is a widely used method for comparing two probability distributions, for deforming one distribution to another, and for transferring information between distributions; see e.g. [9, 11, 20, 21, 26, 45, 46, 59, 60, 64]. Unlike many other distance functions, which make pointwise comparisons on the heights of the probability mass functions and ignore the underlying geometry of the support sets, OT measures the total displacement of deforming one probability distribution into another. OT can be viewed as lifting the ground metric between points in a space to a metric between probability distributions over the points. Because of its widespread applications, OT has been extensively studied in many different fields; see the books [25, 32, 54, 58, 62] and survey papers [18, 38, 43, 52] and the references therein. The existing algorithms for computing an OT plan between two probability distributions, however, assume that at least one of them is a discrete distribution, i.e., it has a finite-size support set of points.

A natural finite-support representation of a continuous distribution is obtained by drawing an i.i.d. sample from it. However, using a sample of size $O(\varepsilon^{-d})$ yields only an *additive* approximation guarantee, incurring an error of $\varepsilon \cdot \mathrm{Diam}$ in the cost of the resulting optimal transport (OT) plan [25]. Such additive errors are inherent to sampling-based approaches and lead to a fundamental trade-off between accuracy and efficiency. While additive error may be acceptable when the OT cost is large, it becomes problematic when the true cost is small, in which case a *relative* approximation guarantee is more appropriate. If the distributions are either unknown or extremely complicated, sampling may be the only hope to compare distributions efficiently. However, if the distributions have geometric structure, there is hope to design algorithms that work directly with the continuous distributions and/or rely on adaptive sampling methods that lead to relative approximation error instead of additive error. In this paper, we study such an approach for the case when distributions are represented as 2D histograms, a natural representation of a probability distribution.

Problem statement. Let $\square^* \subset \mathbb{R}^2$ be an axis-aligned rectangle, and let Ξ_1, Ξ_2 be two (rectangular) subdivisions of $\square^*$ into n smaller axis-aligned rectangles. Let μ_1, μ_2 be two probability mass functions represented as *histograms* over Ξ_1, Ξ_2, respectively. That is, for $i = 1, 2$ and for any rectangle $R \in \Xi_i$, $\mu_i(x) = \mu_i(y)$ for all pairs $x, y \in \mathrm{int}(R)$. With some slight abuse of notation, we will also denote $\mu_i(U) := \int_U \mu_i(u)du$ as the measure of μ_i over U for any Lebesgue measurable subset $U \subseteq \square^*$ and $i = 1, 2$.

We define a *plan* to be any measure[1] σ over measurable subsets of $\square^* \times \square^*$ such that $\sigma(\square^* \times \square^*) \leq 1$. A plan is called a *transport plan* if $\sigma(A \times \square^*) = \mu_1(A)$ for all measurable sets $A \subseteq \square^*$ and $\sigma(\square^* \times B) = \mu_2(B)$ for all measurable sets $B \subseteq \square^*$. In particular, if σ is a transport plan then $\sigma(\square^* \times \square^*) = 1$. The cost of a plan σ is $w(\sigma) := \int_{\square^* \times \square^*} \|a - b\| d\sigma(a, b)$. The *continuous optimal transport problem* asks to compute $\sigma^* = \arg\min_\sigma w(\sigma)$, a minimum-cost transport plan[2].

We emphasize that while the input measures μ_1, μ_2 admit densities that can be Lebesgue integrated, the transport plan does not necessarily admit such a density. See e.g. [49, 54] for further treatment of the continuous optimal transport problem. The minimum such cost over all transport plans is known in the literature by the *1-Wasserstein distance* or the *earth mover's distance* between μ_1 and μ_2. For any $\alpha \geq 1$, a transport plan σ is called an *α-approximate OT plan* if $w(\sigma) \leq \alpha \cdot \min_{\sigma^*} w(\sigma^*)$ for any $\alpha \geq 1$. These definitions can be extended to discrete distributions in a straightforward manner; see [54].

Prior work. From an algorithmic standpoint, the OT problem has been extensively studied in the discrete setting where one assumes the distributions are over finite supports. In this regime, it is known that the OT problem can be formulated as a min-cost flow problem over a complete bipartite graph, and thus can be solved exactly in near-quadratic time using the recent min-cost flow algorithms [24]. An $(1 + \varepsilon)$-approximate OT plan in fixed dimensions can be computed in near-linear time [3, 31, 39] building upon a long line of work in geometric OT [1, 6, 22, 35, 56]. Subquadratic algorithms for computing an $(1 + \varepsilon)$-approximate OT plan in high dimensions were proposed in [10, 17]. Fast algorithms with provable additive

[1] Intuitively, a *measure* is a function σ mapping measurable subsets of $\square^* \times \square^*$ to the nonnegative reals such that $\sigma(\mathsf{X} \cup \mathsf{Y}) = \sigma(\mathsf{X}) + \sigma(\mathsf{Y}) - \sigma(\mathsf{X} \cap \mathsf{Y})$ for any measurable subsets $\mathsf{X}, \mathsf{Y} \subseteq \square^* \times \square^*$. For a more formal treatment of measure theory, we refer the reader to [63].

[2] While a cost-minimizing transport plan may not exist in the general case, it is known that the optimal transport plan between two histograms exists. For a more thorough proof of existence, see Theorem 4.1 of [62].

approximation guarantees using the so-called entropic regularization method [8, 27] are most widely used in machine learning applications. Many other efficient additive approximation algorithms (e.g. [36, 44]) and noteworthy heuristics (e.g. [30, 55]) are also studied in the machine learning community.

Partially alleviating the restriction that distributions have finite support yields the semi-discrete OT problem, where one of the distributions is discrete and the other is continuous. The key property of a semi-discrete OT plan is that a weight can be assigned to each point in the discrete distribution so that the OT plan transports the mass from each discrete point p_i to its Voronoi cell in the weighted Voronoi diagram (the weights ensure that the measure of the continuous distribution in the Voronoi cell of p_i is the same as the mass of p_i) [15, 19]. Using this observation, numerical algorithms have been developed to compute the weights of points in the discrete distribution [15, 23, 40, 41, 49, 50, 51]. Recently, combinatorial algorithms have been proposed for computing relative as well as additive approximations of OT plans that also exploit the Voronoi diagram property and that combine a cost-scaling method with the Hungarian algorithm [2, 3, 4].

Continuous OT represents the most challenging setting. Even how to represent an OT plan compactly is not understood except in some special cases: 1-dimensional OT [58, Chapter 2], continuous OT over tree metric spaces [48], and OT between two normal distributions [33, 37]. Compact representations of semi-discrete transport plans using weighted Voronoi diagrams [15] and Brenier maps [19] do not extend to the continuous setting as it requires defining the notion of a Voronoi diagram for a continuous set of points. As mentioned above, algorithms currently used in practice for continuous OT are sampling-based numerical algorithms [16, 28, 47, 53]. However, any naïve sampling strategy simultaneously fails to represent the continuous transport plan and provides an additive approximation in the best case due to aforementioned sampling lower bounds [25]. To our knowledge, the first combinatorial algorithm for the continuous OT problem between geometric supports was presented in [61]. It runs in $\widetilde{O}(n^3\Delta)$ time[3] when distributions are represented as histograms, where Δ denotes the spread of the vertex set (i.e. the ratio of farthest and closest pairwise distances of vertices), and computes a transport plan of cost at most $(1+\varepsilon)w(\sigma^*) + O\left(\frac{\varepsilon}{n}\right)$, where $w(\sigma^*)$ is the cost of the OT plan. Note that unlike the naïve sampling method, their additive approximation does not depend on the diameter of the support region.

Our contributions. Our main result is the first $(1+\varepsilon)$-approximation algorithm (without any additive-approximation) for computing an OT plan between two histograms in $\mathbb{R}^2$.

▶ **Theorem 1.** *Let $\square^*$ be an axis-aligned rectangle in $\mathbb{R}^2$ and let Ξ_1, Ξ_2 be two rectangular subdivisions of $\square^*$ with a total of n rectangles. Given two probability mass functions $\mu_1, \mu_2 \colon \square^* \to \mathbb{R}_{\geq 0}$ represented as histograms over Ξ_1 and Ξ_2, respectively, and a parameter $\varepsilon > 0$, a $(1+\varepsilon)$-approximate transport plan between μ_1 and μ_2 can be computed in $O(n^2\varepsilon^{-1}\log n \log\Delta + k\varepsilon^{-O(1)}\log^{O(1)}(n\Delta))$ time with probability at least $1/2$, where $k = O(n^2)$ is the number of rectangles in the overlay Ξ of Ξ_1 and Ξ_2 and Δ is the spread of vertices of Ξ.*

▶ Remark. Although $k = \Theta(n^2)$ in the worst case, it is often much smaller in practice. Furthermore, if the aspect ratio of all rectangles in the overlay Ξ is at most r, then the run time of our algorithm improves to $O(kr\varepsilon^{-1}\log r \log\Delta + k\varepsilon^{O(-1)}\log^{O(1)}\Delta)$ by a refined analysis. Ideally, we would like the algorithm to have such a run time if the rectangles in Ξ_1 and Ξ_2 have bounded aspect ratio, but such a result has remained elusive so far.

[3] $\widetilde{O}(\cdot)$ is used to suppress any poly-logarithmic and poly(ε^{-1}) terms in the rumtime for simplicity.

We note that we require a reasonable way to represent a transport plan. The optimal plan may be quite complex even for histograms (see below). Extending the representation of OT plans in the discrete and semi-discrete settings, a natural representation of an OT plan in the continuous setting is in the form of a coupling between the regions of the supports of the two distributions. In our setting, we use the following representation. Let Ξ be the overlay of Ξ_1 and Ξ_2, the rectangular subdivisions of the two histograms, which we view as the common support of both μ_1 and μ_2. Let Ξ' be a further refinement of Ξ into axis-aligned rectangles. That is, Ξ' is obtained from Ξ by splitting each of its rectangles into subrectangles. We will represent a transport plan σ implicitly as a coupling $\sigma_{\Xi'}$ between pairs of boxes in Ξ', which we call a *box-to-box transport plan.* We then provide a parametric representation of the transport plan between every pair of boxes $R, R' \in \Xi'$ with $\sigma_{\Xi'}(R, R') > 0$, given the box-to-box transport plan $\sigma_{\Xi'}$. We call such a plan σ as *induced* by $\sigma_{\Xi'}$. See Section 2 for a formal definition of the box-to-box transport plans and its induced transport plan. We show that there exists a refined subdivision Ξ' and an $(1+\varepsilon)$-approximate OT plan σ that is induced by a box-to-box transport plan $\sigma_{\Xi'}$. We believe this representation of a continuous OT plan is of independent interest, and it is one of the technical contributions of the paper.

Note that for constructing a transport plan described as a coupling between pairs of boxes, a worst case complexity of at least $\Omega(k)$ is required since mass from every box may need to be routed at arbitrarily low or zero cost. One may hope that an $(1+\varepsilon)$-approximate OT plan with representation size $O(k)$ exists. We prove (in Section 5) that in fact, the complexity of an approximate transport plan may be $\Omega(n^2)$ in the worst case, even if $k = O(n)$.

► **Theorem 2.** *Given a positive integer $n > 20$ and a parameter $\varepsilon \leq \frac{1}{3}$, there exist an axis-aligned box $\square^*$, a subdivision Ξ of $\square^*$ with n axis-aligned boxes, and two probability distributions $\mu_1, \mu_2 \colon \square^* \to \mathbb{R}_{\geq 0}$ represented as histograms over Ξ such that any $(1+\varepsilon)$-approximate box-to-box transport plan σ^* must be induced by a refined subdivision of Ξ of size $\Omega((1-3\varepsilon)n^2)$.*

Our construction suggests that quadratic time is as good as one can hope to attain in the worst case for computing a $(1+\varepsilon)$-approximate OT plan for histograms if we wish to express a transport plan as a region-to-region coupling. This construction, however, does not preclude an efficient algorithm for estimating the 1-Wasserstein distance, i.e., the cost of an OT plan, between two histograms. We are not aware of a subquadratic time algorithm for this problem, and we leave this as an open problem.

Finally, we show that a weaker approximate OT plan can be computed implicitly in roughly $\widetilde{O}(k)$ time, where $\widetilde{O}(\cdot)$ ignores small factors in the runtime:

► **Theorem 3.** *Let $\square^*$ be an axis-aligned rectangle in $\mathbb{R}^2$ and let Ξ_1, Ξ_2 be two rectangular subdivisions of $\square^*$ with a total of n rectangles. Given two probability mass functions $\mu_1, \mu_2 \colon \square^* \to \mathbb{R}_{\geq 0}$ represented as histograms over Ξ_1 and Ξ_2, respectively, and a parameter $\delta > 0$, an $O(\log^2(\delta^{-1}))$-approximate transport plan between μ_1 and μ_2 can be computed implicitly in $O(k^{1+\delta} \log^{O(1)}(\Delta))$ time with probability at least $1/2$, where k is the number of rectangles in the overlay Ξ of Ξ_1 and Ξ_2 and Δ is the spread of the vertices of Ξ.*

► Remark. We remark that if δ is a sufficiently small constant, the bound on the size of the implicit representation in Theorem 3 becomes $\Omega(n^2)$. In fact, the size of the proposed implicit representation for the lower-bound construction in Theorem 2 is $\Omega(n^2)$ for $\varepsilon \leq \frac{1}{3}$. We are unaware of any implicit representation of subquadratic size for sufficiently small ε. An interesting open problem is whether a "meaningful" implicit representation of an $(1+\varepsilon)$-approximate OT plan between two histograms can be computed in subquadratic time.

Returning to Theorem 1, we prove a few key structural properties of the OT plan between two histograms (Section 2), which allows us to transport the mass locally and which also provide a lower bound on the cost of the OT plan. The structural properties immediately lead to a simple algorithm for computing a $(1+\varepsilon)$-approximate OT plan. However, the complexity of the resulting plan could be arbitrarily large. Next, we present a different algorithm (in Section 3) that nontrivially partitions the space using a randomly shifted quadtree, a technique widely used in geometric approximation algorithms [12, 13, 14, 42, 57] and OT algorithms [5, 31, 39, 56], transports some mass locally, discretizes the remaining mass, and uses an approximation algorithm for discrete OT problem to route the remaining mass. There is one critical difficulty with constructing a quadtree over a set of continuous objects, namely when to stop recursively subdividing the quadtree. If one subdivides a quadtree cell until it is contained in a single rectangle of the overlay Ξ (or intersects only one edge or a vertex of Ξ), then the size of the quadtree could be polynomial in the spread Δ of the vertex set V. We instead subdivide a cell if it contains at least one vertex of Ξ and it is "sufficiently large". A leaf cell containing no vertex of Ξ may intersect many edges of the subdivision Ξ. The construction of a transport plan at the leaf cells is therefore nontrivial. But in this case, all edges of Ξ passing through $\square$ are parallel and we reduce the problem of computing OT in such a leaf cell to a 1-dimensional OT problem (see Section 4). Constructing the quadtree in this manner and handling the leaf cells are the main technical contributions in this section.

To obtain the improved runtime of Theorem 3, we observe that the processing of leaf cells is the bottleneck in the quadratic-time algorithm, as each leaf cell is processed independently. We show (in Section 4) that all leaf cells at a fixed level of the quadtree can be processed by performing two line sweeps – one horizontal and one vertical – and by maintaining a set of 1D OT instances dynamically as red or blue mass is inserted onto or deleted from each instance. The dynamic 1D OT problem is solved by adapting the OT algorithm of Agarwal et al. [1], which computes an $O(\log^2(\delta^{-1}))$-approximate OT plan in $O(n^{1+\delta})$ time with high probability. Additionally, the dynamic 1-dimensional OT plan can be maintained in a persistent data structure so that leaf cells can be individually queried when required.

2 Structural Properties of Continuous OT Plan

In this section, we prove a few useful structural properties of a continuous OT plan and prove the existence of a box-to-box transport plan described in the introduction. Since we are aiming for an algorithm whose runtime is at least as large as the size of the overlay Ξ of Ξ_1 and Ξ_2, the rectangular subdivisions over which the histograms are defined, we *assume* Ξ *to be the common rectangular subdivision of both* μ_1 *and* μ_2. We begin with two key technical lemmas. The first lemma states that there exists an optimal plan which routes a maximal amount of mass from each point to itself and the second proves a lower bound on the cost of an OT plan if it is positive. These lemmas imply the existence of a near-optimal transport plan which transports rectangles to rectangles. Such a transport plan, which we will call a *box-to-box transport plan*, will be crucial in the design of efficient algorithms.

For any set $U \subseteq \square^*$, define $\mathsf{diag}(U \times U) := \{(u,u) : u \in U\}$. Define a transport plan σ between two histograms μ_1 and μ_2 over a rectangular subdivision Ξ to be *diagonalized* if $\sigma(\mathsf{diag}(U \times U)) = \min\{\mu_1(U), \mu_2(U)\}$ for every subset $U \subseteq R$ of every rectangle $R \in \Xi$.

▶ **Lemma 4.** *Let* $\square^*$ *be an axis-aligned rectangle in* $\mathbb{R}^2$*, let* Ξ *be a subdivision of* $\square^*$*, and let* $\mu_1, \mu_2 \colon \square^* \to \mathbb{R}_{\geq 0}$ *be two histograms over* Ξ*. There exists an optimal transport plan* σ *that is diagonalized.*

Figure 1 We show an example of two histograms on the left, and the same histograms over the overlay of the supports on the right.

Proof. Let σ^* be an arbitrary optimal transport plan between μ_1 and μ_2. We will deform σ^* into a diagonalized transport plan τ without increasing its cost.

Let τ_{diag} be the plan defined by $\tau_{\mathsf{diag}}(\mathsf{diag}(U \times U)) = \min\{\mu_1(U), \mu_2(U)\}$ for every subset U of a rectangle R of the overlay Ξ. More generally, for $U, U' \subseteq \square^*$, $\tau_{\mathsf{diag}}(U \times U') = \tau_{\mathsf{diag}}(\mathsf{diag}((U \cap U') \times (U \cap U')))$. We also define an additional plan τ_{off} and set $\tau := \tau_{\mathsf{off}} + \tau_{\mathsf{diag}}$. We argue that τ is a transport plan between μ_1 and μ_2 and $w(\tau) \leq w(\sigma^*)$.

For any subset U of a rectangle R in the overlay Ξ, define $U^\uparrow$ and $U^\downarrow$ to be the minimal sets such that $\sigma^*(U \times U^\uparrow) = \mu_1(U)$ and $\sigma^*(U^\downarrow \times U) = \mu_2(U)$. Let $A_U = (U^\downarrow \times U) \setminus \mathsf{diag}(U \times U)$, $B_U = (U^\downarrow \times U^\uparrow) \setminus \mathsf{diag}(U \times U)$ and $C_U = (U \times U^\uparrow) \setminus \mathsf{diag}(U \times U)$. Define $\delta(U) = \tau_{\mathsf{diag}}(\mathsf{diag}(U \times U)) - \sigma^*(\mathsf{diag}(U \times U))$ to be the additional amount that τ_{diag} transports along the diagonal of U over σ^*. Then we choose τ_{off} in a way such that

1. $\tau_{\mathsf{off}}(A_U) = \sigma^*(A_U) - \delta(U)$,
2. $\tau_{\mathsf{off}}(C_U) = \sigma^*(C_U) - \delta(U)$, and
3. $\tau_{\mathsf{off}}(B_U) = \sigma^*(B_U) + \delta(U)$

for every $U \subseteq R$ of every $R \in \Xi$. That is, we take τ_{off} to be σ^* and instead of transporting mass of $U^\downarrow$ into U or mass of U into $U^\uparrow$, we directly transport mass from $U^\downarrow$ to $U^\uparrow$. Since σ^* is a feasible transport plan, the choice of τ_{off} exists. That is, no rerouting of mass from $U_1^\downarrow$ to $U_1^\uparrow$ will prevent one from also rerouting mass from $U_2^\downarrow$ to $U_2^\uparrow$ for any two subsets U_1, U_2 of rectangles R_1, R_2. It follows from the definition of τ_{diag} that if $\mu_1(U) \leq \mu_2(U)$ (resp. $\mu_2(U) \leq \mu_1(U)$) then $\tau_{\mathsf{off}}(C_U) = 0$ (resp. $\tau_{\mathsf{off}}(A_U) = 0$). Furthermore, it can be verified that τ is indeed a transport plan between μ_1 and μ_2. We conclude that

$$
\begin{aligned}
&\int_{\square^* \times \square^*} \|u - v\| d\tau(u, v) - \int_{\square^* \times \square^*} \|u - v\| d\sigma^*(u, v) \\
&\leq \sum_{R \in \Xi} \int_{R^\downarrow \times R^\uparrow} \|u - v\| d(\tau_{\mathsf{off}} - \sigma^*)(u, v) - \sum_{R \in \Xi} \int_{R^\downarrow \times R} \|u - w\| d(\sigma^* - \tau_{\mathsf{off}})(u, w) \\
&\quad - \sum_{R \in \Xi} \int_{R \times R^\uparrow} \|w - v\| d(\sigma^* - \tau_{\mathsf{off}})(w, v) \\
&\leq 0
\end{aligned}
$$

by the triangle inequality. Note the integrals are well defined since $\tau_{\mathsf{off}} \geq \sigma^*$ over $R^\downarrow \times R^\uparrow$ and $\sigma^* \geq \tau_{\mathsf{off}}$ over $(R^\downarrow \times R) \cup (R \times R^\uparrow)$. ◀

By Lemma 4, for any rectangle $R \in \Xi$, one can immediately transport $\sigma(\mathsf{diag}(R \times R)) = \min\{\mu_1(R), \mu_2(R)\}$. It thus suffices to compute an OT plan between the residual demands $\hat{\mu}_i(R) := \mu_i(R) - \min\{\mu_1(R), \mu_2(R)\}$. With this initialization, we can additionally assume that for each rectangle R in Ξ, either $\mu_1(R) = 0$ or $\mu_2(R) = 0$. We now prove the second property, which states that the cost of an optimal transport plan is lower bounded by transporting mass from each rectangle R of Ξ at a distance approximately equal to its shortest sidelength $\underline{\ell}(R)$.

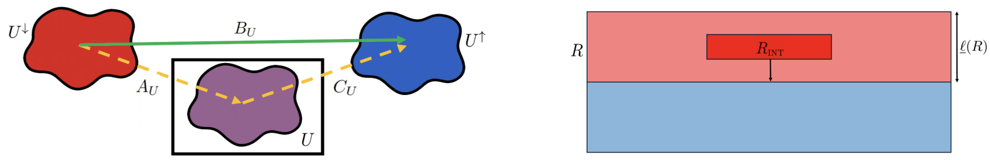

Figure 2 (Left) Illustraction of the proof of Lemma 4. (Right) Illustration of the proof of Lemma 5.

▶ **Lemma 5.** *Let $\square^*$ be an axis-aligned rectangle in $\mathbb{R}^2$, let Ξ be a subdivision of $\square^*$ into axis-aligned rectangles, and let $\mu_1, \mu_2 \colon \square^* \to \mathbb{R}_{\geq 0}$ be two histograms over Ξ such that for all $R \in \Xi$, $\mu_1(R) > 0$ implies $\mu_2(R) = 0$ and vice-versa. Then there exists a constant $C > 0$ such that $w(\sigma^*) \geq C \cdot \left(\sum_{R \in \Xi} \underline{\ell}(R) \cdot (\mu_1(R) + \mu_2(R))\right)$, where σ^* is an OT plan from μ_1 to μ_2 and $\underline{\ell}(R)$ is the sidelength of the shortest edge of a rectangle R.*

Proof. Let σ be an optimal transport plan between μ_1 and μ_2. Choose an arbitrary rectangle $R = [a, a + \ell_1(R)] \times [b, b + \ell_2(R)]$ in Ξ with bottom left corner (a, b) and sidelengths $\ell_1(R), \ell_2(R)$ in the x- and y-coordinate directions, respectively. Assume without loss of generality that $\mu_1(R) > 0$. Define the interior rectangle $R_{\text{INT}} = \left[a + \frac{\ell_1(R)}{3}, a + \frac{2\ell_1(R)}{3}\right] \times \left[b + \frac{\ell_2(R)}{3}, b + \frac{2\ell_2(R)}{3}\right]$ and let $U(R)$ be the minimal measurable set such that $\sigma(R_{\text{INT}} \times U(R)) = \mu_1(R_{\text{INT}})$. Since every rectangle R' in Ξ where $\mu_1(R') > 0$ satisfies $\mu_2(R') = 0$, we have that $U(R) \cap R = \varnothing$, and therefore $\inf_{(a,b) \in R_{\text{INT}} \times U(R)} \|a - b\| \geq \frac{1}{3} \min\{\ell_1(R), \ell_2(R)\} = \frac{\underline{\ell}(R)}{3}$. Moreover, $\mu_1(R_{\text{INT}}) = \frac{1}{9}\mu_1(R)$. The rectangle R was chosen arbitrarily. Therefore,

$$\begin{aligned} w(\mu_1, \mu_2) &\geq \sum_{R \in \Xi} \int_{R_{\text{INT}} \times U(R)} \|a - b\| d\sigma(a, v) \\ &\geq \sum_{R \in \Xi} \frac{\underline{\ell}(R)}{3} \cdot \left[\frac{\mu_1(R)}{9} + \frac{\mu_2(R)}{9}\right] = \frac{1}{27} \sum_{R \in \Xi} \underline{\ell}(R) \cdot (\mu_1(R) + \mu_2(R)) \,. \end{aligned}$$ ◀

box-to-box and induced transport plan. We now prove the existence of a box-to-box transport plan mentioned above. Let Ξ' be a refinement of Ξ into axis-aligned smaller rectangles, i.e., each rectangle of Ξ is further split into smaller rectangles. Then let $\sigma_{\Xi'} \colon \Xi' \times \Xi' \to \mathbb{R}_{\geq 0}$ be a function where $\sum_{R' \in \Xi'} \sigma_{\Xi'}(R, R') = \mu_1(R)$ and $\sum_{R' \in \Xi'} \sigma_{\Xi'}(R', R) = \mu_2(R)$ for all $R \in \Xi'$. We call $\sigma_{\Xi'}$ a *box-to-box transport plan* between μ_1 and μ_2 over Ξ'. See Figure 3 for an example.

We define a plan induced by $\sigma_{\Xi'}$ as follows. Let $R_1 = [x_1^-, x_1^+] \times [y_1^-, y_1^+]$ and $R_2 = [x_2^-, x_2^+] \times [y_2^-, y_2^+]$ be two rectangles in Ξ' such that $\sigma_{\Xi'}(R_1, R_2) > 0$. Let $\chi_i \colon [0,1]^2 \to R_i$ for $i = 1, 2$ be the natural parametrizations from the unit square to the rectangles R_1, R_2, i.e., $\chi_i(\lambda_1, \lambda_2) = (x_i^- + \lambda_1(x_i^+ - x_i^-), y_i^- + \lambda_2(y_i^+ - y_i^-))$ for $i = 1, 2$. With slight abuse of notation, we will extend χ_i to subsets of $[0,1]^2$ by $\chi_i(P) = \{\chi_i(\lambda_1, \lambda_2) : (\lambda_1, \lambda_2) \in P\}$. For any measurable $U_1 \subseteq R_1$ and $U_2 \subseteq R_2$, we define $\sigma(U_1 \times U_2) = \sigma_{\Xi'}(R_1, R_2) \cdot |\chi^{-1}(U_1) \cap \chi^{-1}(U_2)|$, where $|\cdot|$ denotes the Lebesgue measure. We call σ the transport plan *induced by* $\sigma_{\Xi'}$.

In the remainder of the paper, we use Lemmas 4 and 5 to show how to efficiently compute a refined subdivision Ξ' of bounded size and a box-to-box transport plan $\sigma_{\Xi'}$ such that the transport plan induced by $\sigma_{\Xi'}$ is a $(1 + \varepsilon)$-approximate transport plan.

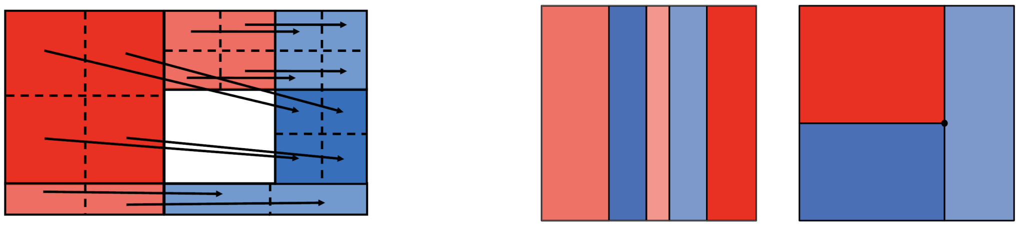

Figure 3 (Left) This figure shows a *box-to-box transport plan* between two histograms. The solid lines denote the subdivision Ξ, and the dashed lines denote its refinement. The box-to-box transport plan, depicted as the set of arrows, couples red-blue pairs of rectangles. (Right) An example of an e-leaf cell and a v-leaf cell.

3 General Algorithm

In this section, we present our main algorithm for computing an $(1+\varepsilon)$-approximate box-to-box transport plan between two histograms over the same rectangular subdivision Ξ in the plane. That is, it computes a refinement Ξ' of Ξ and a coupling between pairs of rectangles of Ξ'. We will assume, by Lemma 4, that $\mu_1(R) > 0$ implies $\mu_2(R) = 0$ ad vice-versa for every R in Ξ. At a very high level, the algorithm proceeds as follows. We construct a *randomly shifted quadtree* on the vertices of Ξ. We first greedily route some mass at small distances within each leaf cell. We then discretize the remaining mass, construct a spanner over the quadtree, and run the algorithm by Agarwal *et al.* [3] (see also [31]) to compute an $(1+\varepsilon)$-approximate transport plan between the discrete distributions. The discrete flow is dispersed arbitrarily to construct a continuous transport plan. The crux of the algorithm lies in the preprocessing step, that, given the subdivision Ξ, constructs a quadtree $\mathcal{Q}$ over the vertices of Ξ and processes all leaf cells of $\mathcal{Q}$. The rest of the section details the key steps in the algorithm.

Randomly shifted quadtree. We construct a randomly shifted quadtree $\mathcal{Q}$ on V, the set of vertices in the subdivision Ξ as in [12]. Each node v of $\mathcal{Q}$ is associated with a square $\square_v$ called its *cell*. We will not distinguish between a node and its cell. Without loss of generality, we assume $\square^* \subseteq [-\Delta, \Delta]^2$. Choose a vector $\vec{v}$ from $[-\Delta, \Delta]^2$ uniformly at random. We set $\square^\top = [-2\Delta, 2\Delta]^2 + \vec{v}$ to be the root of $\mathcal{Q}$ and construct $\mathcal{Q}$ recursively.

Let $\square = [x_1, x_2] \times [y_1, y_2]$ be a cell of $\mathcal{Q}$. Let $\Xi_\square = \Xi \cap \square$ denote the portion of Ξ restricted to $\square$, which is also a rectangular subdivision, and let $c(\square)$ (resp. $\ell(\square)$) denote the center (resp. sidelength) of $\square$. The cell $\square$ is a *leaf* cell if either

1. $V \cap \square = \varnothing$ ($\square$ is called an *e-leaf cell*; see Figure 3), or
2. $V \cap \square = \{p\}$ for some $p \in \square$, every rectangle in $\Xi_\square$ is incident to p and $\ell(\square) \leq \frac{\varepsilon^2}{27} \cdot \min_{R \in \Xi_\square} \underline{\ell}(R)$ ($\square$ is called a *v-leaf cell*; see Figure 3).

Otherwise, $\square$ is an *internal* (or *non-leaf*) *cell*. We subdivide $\square$ into its four children cells. For a point $u \in \square^*$, we use $\square_u$ to denote the leaf cell containing u. Let $\mathcal{L}$ (resp. $\mathcal{I}$) be the set of leaf (resp. internal) cells of $\mathcal{Q}$. For each cell $\square$ in $\mathcal{Q}$, place a set $X_\square$ of $4^{\lceil \log_4(\varepsilon^{-2}\log^2 \Delta) \rceil}$ uniform grid points in $\square$. Let $\mathrm{VD}(X_\square)$ be the Voronoi diagram of $X_\square$ within $\square$, and let $\mathrm{Vor}(x)$ be the Voronoi cell of $x \in X_\square$ in $\mathrm{VD}(X_\square)$. Set $X_{\mathcal{I}} = \bigcup_{\square \in \mathcal{I}} X_\square$.

We define a tree distance $\mathsf{d}_{\mathcal{Q}} \colon \square^* \times \square^* \to \mathbb{R}_{\geq 0}$, based on $\mathcal{Q}$, between pairs of points in $\square^*$. Let u, v be two points in $\square^*$. Let $\square$ be the lowest common ancestor of $u, v \in \square^*$ in $\mathcal{Q}$. If $\square \in \mathcal{L}$, then define $\mathsf{d}_{\mathcal{Q}}(u, v) = \|u - v\|$. Otherwise, let $x_u \in X_\square$ (resp. $x_v \in X_\square$) denote

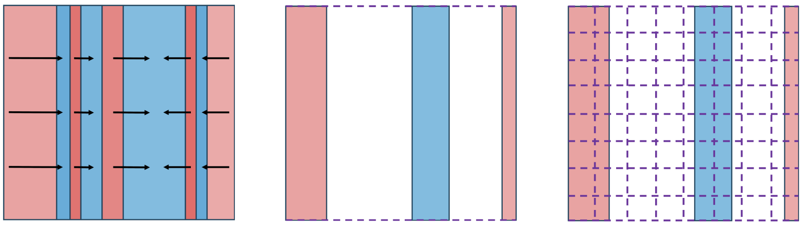

Figure 4 For each e-leaf, we first compute a 1D OT plan which equivalently defines a box-to-box plan between the distributions restricted to the e-leaf cell. We then transport mass which incurs small cost in the 1D plan. Finally, we cover the leaf cell with small squares and collapse the remaining mass to the center of its corresponding square. With the resulting discrete distributions, we compute a discrete OT plan and disperse the plan throughout rectangles.

the grid point such that $u \in \mathrm{Vor}(x_u)$ (resp. $v \in \mathrm{Vor}(x_v)$), and $\mathsf{d}_{\mathcal{Q}}(u, v)$ is defined to be the lengths of the shortest paths in the quadtree connecting u to x_u and v to x_v plus the distance to connect x_u and x_v, i.e.,

$$\mathsf{d}_{\mathcal{Q}}(u, v) = \|x_u - x_v\| + \sqrt{2}\left(\frac{\varepsilon}{\log \Delta}\ell(\square) - \ell(\square_u) - \ell(\square_v)\right) + \|u - c(\square_u)\| + \|v - c(\square_v)\|.$$

The second term is the length of the paths from $c(\square_u)$ to x_u and x_v to $c(\square_v)$ following parent-child edges in the quadtree. We conclude this discussion with the following lemma whose proof follows from earlier results [56].

▶ **Lemma 6.** *The quadtree $\mathcal{Q}$ has $O(k \log \Delta)$ many cells and $\|u - v\| \le \mathbb{E}_{\mathcal{Q}}[\mathsf{d}_{\mathcal{Q}}(u, v)] \le (1 + \varepsilon) \cdot \|u - v\|$ for all $u, v \in \square^*$.*

Local routing of leaf cells. The most interesting part of the algorithm happens at the leaf cells, when the distance between points in the cell is not approximated by a discrete path between cell centers. Unlike quadtrees over discrete point sets, where the size of a leaf cell is small enough compared to the distance from a point to its nearest neighbor, the sidelength of a leaf cell $\square$ may be much larger than $\min_{R \in \Xi_\square} \underline{\ell}(R)$, the smallest sidelength of any rectangle in $\square$. Therefore we need the ability to assign mass at an arbitrarily small cost within each leaf cell in the event that the OT cost is very small. We do this by computing an OT plan within $\Xi_\square$ and committing the mass that is transported a very small distance in the resulting plan. This step also computes a refinement $\Xi'_\square$ of $\Xi_\square$ by splitting rectangles of $\Xi_\square$ into smaller subrectangles. We now describe the process at each leaf in detail.

Let $\square$ and $\Xi_\square$ be as defined above. For each rectangle R in $\Xi_\square$ and for $i = 1, 2$, let $\mu_{i,\square}$ be the restriction of the probability mass μ_i within $\square$. Note that $\sum_{R \in \Xi_\square} \mu_{1,\square}(R)$ need not be equal to $\sum_{R \in \Xi_\square} \mu_{2,\square}(R)$. We compute an *unbalanced* box-to-box OT plan[4] $\sigma_{\Xi'_\square}$, where $\Xi'_\square$ is a refinement of $\Xi_\square$. Then $\bigcup_\square \Xi'_\square$ forms a refinement of Ξ, which will be used to define the desired box-to-box plan between μ_1 and μ_2.

The construction of $\sigma_{\Xi'_\square}$ and which of $\sigma_{\Xi'_\square}$ is committed depends on whether $\square$ is an e-leaf or a v-leaf.

1. $\square$ is an e-leaf. In this case, all edges of $\Xi_\square$ are parallel and their endpoints lie on $\partial\square$. Suppose $\square = [x_1, x_2] \times [y_1, y_2]$. Without loss of generality, assume that the edges are parallel to the y-axis. Let $\Xi^*_\square$ be the set of x-projections of boxes in $\Xi_\square$, which

[4] In unbalanced OT, where one measure has more mass than the other (assume without loss of generality μ_1 has more mass), a transport plan is a plan σ where $\sigma(U \times \square^*) \le \mu_1(U)$ and $\sigma(\square^* \times U) = \mu_2(U)$ for all measurable $U \subseteq \square^*$.

is the same as the intersections of each box in $\Xi_\square$ with any horizontal line crossing $\square$ (since all rectangles of $\Xi_\square$ span the y-interval of $\square$). For each interval R^* of $\Xi^*_\square$, which is the x-projection of a rectangle $R \in \Xi_\square$, we assign $\mu_1(R^*) = \mu_1(R)$ (resp. $\mu_2(R^*) = \mu_2(R)$). Since $\sum_{R^* \in \Xi^*_\square} \mu_1(R^*)$ is not necessarily equal to $\sum_{R^* \in \Xi^*_\square} \mu_2(R^*)$, we have a 1D *unbalanced* continuous OT problem at hand. We compute a α-approximate 1D OT plan $\hat{\sigma}_\square$ between μ_1 and μ_2, for some $\alpha \geq 1$, as described in Section 4. By construction, the 1D OT plan $\hat{\sigma}_\square$ will transport mass from intervals to intervals. This 1D plan $\hat{\sigma}_\square$ can be immediately lifted to 2D to obtain a box-to-box (unbalanced) α-approximate OT plan between $\mu_{1,\square}$ and $\mu_{2,\square}$. Among the mass transported by the unbalanced OT instance, we commit the transportation of mass that is transported within distance $\frac{\varepsilon^2}{\log \Delta} \cdot \ell(\square)$ to the box-to-box OT plan $\sigma_{\Xi'_\square}$. For each pair of intervals I_1, I_2 such that $\hat{\sigma}_\square(I_1, I_2) > 0$, let $I'_1 \subseteq I_1$ and $I'_2 \subseteq I_2$ be maximal subintervals such that $\frac{|I'_1|}{|I_1|} = \frac{|I'_2|}{|I_2|}$ and $|x - y| \leq \frac{\varepsilon^2}{\log \Delta} \cdot \ell(\square)$ for all $x \in I'_1, y \in I'_2$. Then for the rectangles $R'_1 = I'_1 \times [y_1, y_2]$ and $R'_2 = I'_2 \times [y_1, y_2]$, we define $\sigma_{\Xi'_\square}(R'_1, R'_2) = \frac{|I'_1|}{|I_1|} \cdot \hat{\sigma}_\square(I_1, I_2)$. This choice of committed mass for the box-to-box transport plan therefore defines a refinement $\Xi'_\square$ of $\Xi_\square$.

2. <u>$\square$ is a v-leaf</u>. In this case $\Xi'_\square = \Xi_\square$ and we transport within $\square$ as much mass as we can. We do so in a greedy manner. Let $\Xi^i_\square = \{R \in \Xi_\square : \mu_i(R) > 0\}$ for $i = 1, 2$. Repeat the following until either $\Xi^1_\square$ or $\Xi^2_\square$ is empty. Choose $R_1 \in \Xi^1_\square$ and $R_2 \in \Xi^2_\square$ arbitrarily. Assign $\sigma_{\Xi'_\square}(R_1, R_2) = \min\{\mu_1(R_1), \mu_2(R_2)\}$ and subtract $\min\{\mu_1(R_1), \mu_2(R_2)\}$ from both $\mu_1(R_1)$ and $\mu_2(R_2)$. If $\mu_1(R_1) = 0$ (resp. $\mu_2(R_2) = 0$), then remove R_1 (resp. R_2) from $\Xi^1_\square$ (resp. $\Xi^2_\square$).

For each leaf cell $\square$ of the quadtree $\mathcal{Q}$ and each rectangle R of $\Xi'_\square$, we define the remaining available mass at R as $\widetilde{\mu}_1(R) := \mu_1(R) - \sum_{R' \in \Xi'_\square} \sigma_{\Xi'_\square}(R, R')$ and $\widetilde{\mu}_2(R) := \mu_2(R) - \sum_{R' \in \Xi'_\square} \sigma_{\Xi'_\square}(R', R)$. We observe that an OT plan that routes the remaining masses $\widetilde{\mu}_1, \widetilde{\mu}_2$ can be combined with the current computed transport plan σ, defined as the combination of local transport plans within each leaf cell, to construct a $(1+\varepsilon)$-approximate OT plan.

▶ **Lemma 7.** *Let σ be the partial transport plan committed in the local routing at the leaf cells. Additionally let $\widetilde{\sigma}$ be a $(1+\varepsilon)$-approximate OT plan from $\widetilde{\mu}_1$ to $\widetilde{\mu}_2$ with respect to $\mathsf{d}_\mathcal{Q}$. Then the plan $\tau = \widetilde{\sigma} + \sigma$ is a transport plan from μ_1 to μ_2 and*

$$\int_{\square^* \times \square^*} \mathsf{d}_\mathcal{Q}(u, v) d\tau(u, v) \leq \max\left\{(1+\varepsilon)\,\alpha, (1+\varepsilon)^2\right\} \mathit{OPT},$$

where OPT is the cost of the optimal transport plan from μ_1 to μ_2 with respect to $\mathsf{d}_\mathcal{Q}$ and α is the approximation factor of the 1-dimensional OT solver used to compute σ.

The proof of Lemma 7 shares many of the same technical ideas as the proof of Lemma 4, so we omit the details from this version of the paper. What remains is to compute the transport plan $\widetilde{\sigma}$ approximately.

Coupling the remaining mass. We recall that for each leaf cell $\square$, the set $X_\square$ consists of $4^{\lceil \log_4(\varepsilon^{-2} \log^2 \Delta) \rceil}$ uniform grid points in $\square$. Let $X_\mathcal{L} = \bigcup_{\square \in \mathcal{L}} X_\square$. We discretize the distributions over the support set $X_\mathcal{L}$ to some demand function λ, solve an instance of the discrete OT problem, and then convert back to continuous OT. The main challenge is how to discretize the mass without incurring too much error. We make the following discretization $\lambda \colon X_\mathcal{L} \to \mathbb{R}$ of the residual histograms in each type of leaf cell.

1. In the e-leaf cells $\square$, we further refine $\Xi'_\square$. Each rectangle R in $\Xi'_\square$ is replaced with its set of intersections $\{R \cap \mathrm{Vor}(x) : x \in X_\square\}$ in $\Xi'_\square$. Since $X_\square$ is a set of grid points, each $\mathrm{Vor}(x) \cap \square$ is a square and therefore $R \cap \mathrm{Vor}(x)$ is indeed a rectangle (if nonempty). Additionally note $\widetilde{\mu}_i(\mathrm{Vor}(x) \cap \square)$ is equal to the sum of $\widetilde{\mu}_i(R)$ over all rectangles R of $\Xi'_\square$ contained in $\mathrm{Vor}(x)$ for $i = 1, 2$. Then we define the discrete demand $\lambda(x) = \widetilde{\mu}_1(\mathrm{Vor}(x) \cap \square) - \widetilde{\mu}_2(\mathrm{Vor}(x) \cap \square)$ for each grid point $x \in X_\square$.
2. In the v-leaf cells $\square$, where $|V \cap \square| = 1$, we note by construction that $\ell(\square) \le \varepsilon \cdot \mathsf{OPT}$. Therefore, we simply define $\lambda(c(\square)) = \widetilde{\mu}_1(\square) - \widetilde{\mu}_2(\square)$, where again $c(\square)$ is the center of the cell $\square$.

Given the demand function λ, we can define the discrete distributions $\mu_1^\circ, \mu_2^\circ \colon X_{\mathcal{L}} \to \mathbb{R}_{\ge 0}$ as $\mu_1^\circ(x) = \max\{\lambda(x), 0\}$ and $\mu_2^\circ(x) = \max\{-\lambda(x), 0\}$ for all $x \in X_{\mathcal{L}}$.

After constructing the distributions μ_1°, μ_2°, we compute an $(1+\varepsilon)$-approximate discrete OT plan with respect to distance $\mathsf{d}_{\mathcal{Q}}$ using now standard techniques [3, 31]. For every $x_1, x_2 \in X$ such that $\sigma_{\mathtt{disc}}(x_1, x_2) > 0$, we compute a box-to-box plan between rectangles $R_1 \subseteq \mathrm{Vor}(x_1)$ and $R_2 \subseteq \mathrm{Vor}(x_2)$ in a greedy manner similar to the local coupling of v-leaf cells on item 2. This concludes the construction of $\Xi' = \bigcup_\square \Xi'_\square$ and the box-to-box transport plan $\widetilde{\sigma}_{\Xi'}$. Following the choice of parametrization from the box-to-box transport plan, we simultaneously obtain the transport plan $\widetilde{\sigma}$ induced by the box-to-box transport plan $\widetilde{\sigma}_{\Xi'}$. Then we prove that the computed $\widetilde{\sigma}$ is an $(1+\varepsilon)$-approximate OT plan between $\widetilde{\mu}_1$ and $\widetilde{\mu}_2$, satisfying the condition of Lemma 7.

▶ **Lemma 8.** *The transport plan $\widetilde{\sigma}$ is an $(1+\varepsilon)$-approximate OT plan between $\widetilde{\mu}_1$ and $\widetilde{\mu}_2$ with respect to $\mathsf{d}_{\mathcal{Q}}$, i.e., $\int_{\square^* \times \square^*} \mathsf{d}_{\mathcal{Q}}(u,v)\, d\widetilde{\sigma}(u,v) \le (1+\varepsilon)\mathsf{OPT}$, where OPT is the cost of the optimal transport plan between $\widetilde{\mu}_1$ and $\widetilde{\mu}_2$ with respect to $\mathsf{d}_{\mathcal{Q}}$.*

Combining Lemmas 6, 7 and 8, we conclude that $w(\sigma) \le (1+\varepsilon)\alpha\mathsf{OPT}$. Let $T_\alpha(n, k, \Delta)$ be the total time spent in computing α-approximate OT plans at all leaves of $\mathcal{Q}$ (see Section 4). Here n is the number of distinct x- and y-coordinate values of points in V, k is the number of rectangles in Ξ, and Δ is the spread of the points in V. By Lemma 6, $\mathcal{Q}$ has $O(k \log \Delta)$ many cells. Moreover, the discrete point set constructed by the algorithm at each leaf cell has $O(\varepsilon^{-6} \log^2 \Delta)$ points. Therefore, the expected time spent in computing the discrete transport plan is $k\varepsilon^{-O(1)} \log^{O(1)} \Delta$. Hence, we obtain the following.

▶ **Lemma 9.** *The cost of the transport plan σ computed by the algorithm is at most $(1+\varepsilon)\alpha\mathsf{OPT}$ with probability at least $1/2$, and the expected runtime of the algorithm is $O(T_\alpha(n, k, \Delta) + k\varepsilon^{-O(1)} \log^{O(1)} \Delta)$.*

4 Processing Leaf Cells

In this section, we describe two algorithms for computing an α-approximate OT plan on the e-leaf cells of $\mathcal{Q}$. The first algorithm computes a $(1+\varepsilon)$-approximate OT plan by processing each leaf independently. The second algorithm is based on the observeation that a rectangle R of Ξ may appear in many e-leaf cells and thus $\Xi_\square$ and $\Xi_{\square'}$ for two e-leaf cells $\square$ and $\square'$ may overlap significantly. Therefore, it bundles the processing of these cells using a dynamic data structure and computes an implicit representation of their transport plans using a persistent data structure [29].

A $(1+\varepsilon)$-approximation algorithm. Let $\mathcal{I}$ be a partition of an interval I_0 into subintervals (corresponding to $\Xi^*_\square$ for an e-leaf $\square$), and let $\hat{\mu}_1, \hat{\mu}_2 \colon I_0 \to \mathbb{R}_{\ge 0}$ be the two piecewise-constant mass functions such that the mass over each subinterval I of $\mathcal{I}$ is constant and either $\hat{\mu}_1(I) = 0$

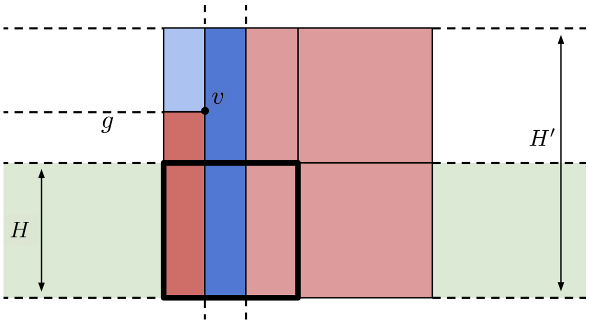

Figure 5 Illustration of the proof of Lemma 10.

or $\hat{\mu}_2(I) = 0$. We place $\frac{1}{\varepsilon}$ points uniformly within each subinterval I and assign $\varepsilon\hat{\mu}_1(I)$ and $\varepsilon\hat{\mu}_2(I)$ mass to each of them (note that one of them will be 0). Let $X \subset I_0$ be the resulting set of points, and let $\widetilde{\mu}_1, \widetilde{\mu}_2 \colon X \to \mathbb{R}_{\geq 0}$ be their mass functions. We can compute an optimal transport plan $\widetilde{\sigma}$ of $\widetilde{\mu}_1, \widetilde{\mu}_2$ in $O(|X| \log |X|)$ time using the algorithm by Aggrawal et al. [7]. We can easily convert $\widetilde{\sigma}$ into a transport plan $\hat{\sigma}$ of $\hat{\mu}_1, \hat{\mu}_2$. We omit the straightforward details from here. Lemma 5 implies that $\hat{\sigma}$ is a $(1+\varepsilon)$-approximate OT plan of $\hat{\mu}_1, \hat{\mu}_2$.

Repeating this procedure for all e-leaf cells of $\mathcal{Q}$, we compute $(1+\varepsilon)$-approximate OT plans for all of them. The total time spent is $\sum_{\square \in \mathcal{L}} O(\varepsilon^{-1} |\Xi_\square| \log |\Xi_\square|) = O(\varepsilon^{-1} \log n) \cdot \sum_{\square \in \mathcal{L}} |\Xi_\square|$. Therefore, it suffices to bound $\sum_{\square \in \mathcal{L}} |\Xi_\square|$.

▶ **Lemma 10.** $\sum_{\square \in \mathcal{L}} |\Xi_\square| = O(n^2 \log \Delta)$.

Proof. We note there are at most $k \leq n^2$ many v-leaf cells since it is assumed that each v-leaf cell contains a vertex of the overlay Ξ. Moreover, $|\Xi_\square| \leq 4$ for every v-leaf cell. Therefore it suffices to compute the total complexity over all e-leaf cells. We bound the total complexity of these e-leaf cells that lie at a fixed level and whose edges are parallel to the y-axis. An analogous argument bounds the complexity of the remaining e-leaf cells.

Fix a level ℓ and consider the set $\mathcal{L}_\ell$ of all e-leaves at level ℓ whose edges are parallel to the y-axis. They lie on a uniform grid $\mathbb{G}_\ell$ of size $L = \frac{\Delta}{2^{\ell-1}}$. Consider the subset $\mathcal{L}_H \subseteq \mathcal{L}_\ell$ of these e-leaves that lie in a fixed row, which is a horizontal strip H of width L. An edge of $\Xi_\square$ for some e-leaf cell lying in H is a portion of a vertical edge of Ξ that completely crosses H. Since the vertical edges of Ξ lie on $O(n)$ vertical segments, $O(n)$ vertical edges of Ξ completely cross H. Therefore, the total complexity of e-leaf cells in $\mathcal{L}_H$ is $O(n)$.

Next, we prove that there are $O(n)$ rows in the grid $\mathbb{G}_\ell$ that contain an e-leaf cell with y-parallel edges. This will imply that $\sum_{\square \in \mathcal{L}_\ell} |\Xi_\square| = O(n^2)$. Let $\square \in \mathcal{L}_H$ be an e-leaf cell lying in the horizontal strip H of $\mathbb{G}_\ell$. We claim that the parent cell of $\square$, denoted by $\square'$, contains a vertex v of the subdivision Ξ because otherwise $\square'$ is also an e-leaf cell. Note that v lies on some horizontal edge g of Ξ, and g lies in the horizontal strip H' spanned by $\square'$. H' is a row in the uniform grid $\mathbb{G}_{\ell-1}$ corresponding to the level $\ell - 1$ and it contains two rows of $\mathbb{G}_\ell$. Since the horizontal edges of the subdivision Ξ have $O(n)$ distinct y-coordinate values, there are $O(n)$ rows of $\mathbb{G}_{\ell-1}$ that contain a horizontal edge of the subdivision Ξ. Consequently, there are $O(n)$ rows of $\mathbb{G}_\ell$ that contain e-leaf cells with y-parallel edges. Hence, $\sum_{\square \in \mathcal{L}_\ell} |\Xi_\square| = O(n^2)$.

Summing this bound over all $O(\log \Delta)$ levels of the tree and over all e-leaf cells with x-parallel edges, and adding the complexity of the v-leaf cells, we obtain that $\sum_{\square \in \mathcal{L}} |\Xi_\square| = O(n^2 \log \Delta)$. ◀

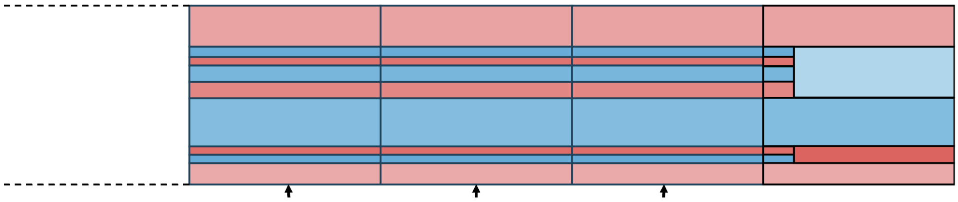

Figure 6 In this example, the highlighted e-leaf cells all share an optimal box-to-box transport plan. It is sufficient to only compute changes in the 1D OT plan over a sequence of leaf cells.

▶ **Lemma 11.** *If the subdivision Ξ consists of boxes where the ratio of longest to shortest edge is $r \geq 1$, then $\sum_{\square \in \mathcal{L}} |\Xi_\square| = O(kr \log \Delta)$.*

Putting everything together, we obtain the following:

▶ **Lemma 12.** *The $(1+\varepsilon)$-approximate OT plans at all leaves can be computed in a total of $O(n^2\varepsilon^{-1} \log n \log \Delta)$ time. If the rectangles in Ξ are all r-fat, i.e. the ratio of longest to shortest sidelength of every rectangle in Ξ is at most r, the running time improves to $O(kr\varepsilon^{-1} \log r \log \Delta)$.*

Using Lemma 12 as a bound on $T_{1+\varepsilon}(n, k, \Delta)$ in Lemma 9, we obtain Theorem 1, the first main result of the paper.

A more efficient algorithm. We now describe a more efficient algorithm for processing all the e-leaf cells by representing their transport plan implicitly and handling them in batches. In particular, for any constant $\delta > 0$, we compute an $O(\log^2(\delta^{-1}))$-approximate OT plan in time $O(k^{1+\delta} \log^{O(1)} \Delta)$ with high probability using a Monte Carlo algorithm.

We fix a level i of $\mathcal{Q}$, and let $\mathcal{L}_i$ be the set of all e-leaf cells at level i that contain vertical (y-parallel) edges. The e-leaf cells of level i containing horizontal edges or of different levels are processed in a similar manner. Each cell of $\mathcal{L}_i$ is a grid cell of a uniform grid $\mathbb{G}_i$ of size $\ell \times \ell$ where $\ell = 2^i \cdot \min_{\square' \in \mathcal{Q}} \ell(\square')$. Let Σ be the set of vertical strips (columns) of $\mathbb{G}_i$ that contain $\mathcal{L}_i$, let L be the (vertical) boundary lines of Σ, and let $\mathcal{I}$ be the set of x-projections (x-intervals) of strips in Σ; each interval in $\mathcal{I}$ has length ℓ. We compute the transport plans of $\mathcal{L}_i$ by sweeping a horizontal line in the $+y$-direction, and we represent them using a persistent data structure, as described below.

Let $\mathcal{R}$ be the set of rectangles of Ξ that appear in at least one cell of $\mathcal{L}_i$. If a rectangle $R \in \mathcal{R}$ intersects a line of L, we split R at the lines of L, only keep the leftmost and rightmost rectangles, and discard the middle portions (the ones that are bounded by L on both sides) as they do not appear in any cell of $\mathcal{L}_i$ (because these portions completely contain a cell at level i and this cell cannot intersect any edge of $\Xi_\square$). Abusing the notation a little, let $\mathcal{R}$ denote the resulting set of rectangles. Each rectangle now lies within a single strip of Σ, and $|\mathcal{R}| \leq 2k$. For each rectangle $R = I_x \times I_y \in \mathcal{R}$, let $p_R \in \mathbb{R}$ be the midpoint of the x-interval I_x. We assign the mass $\widetilde{\mu}_1(p_R)$ (resp. $\widetilde{\mu}_2(p_R)$) to be $\hat{\mu}_1(R) \cdot \ell \cdot |I_x|$ (resp. $\hat{\mu}_2(R) \cdot \ell \cdot |I_x|$), i.e., the mass over a subrectangle of R of width $|I_x|$ and length ℓ; again one of the masses is 0.

We sweep a horizontal line H from $y = -\infty$ to $y = +\infty$, and maintain the set of points $P = \{p_R : R \cap H \neq \varnothing\}$ contained in the rectangles of $\mathcal{R}$ that intersect H. When H passes through the bottom edge of a rectangle R, p_R is inserted into P, and when H reaches the top edge of R, p_R is deleted from P. For each interval $I \in \mathcal{I}$, we maintain a transport plan σ_I of $\{\widetilde{\mu}_1(p) : p \in P \cap I\}$ and $\{\widetilde{\mu}_2(p) : p \in P \cap I\}$ dynamically as points are inserted or deleted from $P_I = P \cap I$, as described below. Each plan σ_I will be maintained using a tree, and we will use persistence to record all versions of the transport plans. In particular, let $y_1, \ldots, y_n$ be the y-coordinates of horizontal edges of rectangles in $\mathcal{R}$. The persistent data structure enables us to record transport plans for every interval (y_i, y_{i+1}). Let $\square = [x^-_\square, x^+_\square] \times [y^-_\square, y^+_\square]$ be a leaf in $\mathcal{L}_i$, and let $I_\square = [x^-_\square, x^+_\square]$. Suppose $y^+_\square \in [y_i, y_{i+1}]$. Then the transport plan $\sigma_{I_\square}$ corresponding to the interval (y_i, y_{i+1}) is the desired transport plan for $\square$.

We maintain a transport plan σ_I for P_I using the tree-based greedy algorithm as described in [1, 6, 22, 34]. For simplicity, first assume $\Delta = n^{O(1)}$, and we present a $O(\log n)$-approximation algorithm. In this case, we construct a 1D quadtree (also called dyadic tree) T_I on the interval I – recursively partition I into two halves until an interval contains at most one point. Since $\Delta = n^{O(1)}$, the height of T_I is $O(\log n)$. The greedy algorithm proceeds in a top-down manner. If a node $v \in T_I$ has points with both positive $\widetilde{\mu}_1$ and $\widetilde{\mu}_2$, it transports locally as much as possible and pushes the excess mass to its parent. We omit the details from here. Since $\mathcal{Q}$ is a randomly shifted quadtree, I is also a randomly shifted interval, so a standard analysis shows that the algorithm computes a $O(\log n)$-approximate OT plan for P_I. When a point is inserted or deleted, σ_I is updated only along a path of T_I. Hence, we can use the path-copying method to make T_I persistent, which requires $O(\log n)$ space, the height of T_I, at each update. Hence, the total space needed to represent all transport plans is $O(k \log n)$, and the total time spent is also $O(k \log n)$. Repeating this procedure for e-leaf cells of level i with horizontal edges and of all other levels, the total time and space required is $O(k \log n \log \Delta)$.

Recall that the general algorithm not only requires the full transport plan for each leaf, but also the mass that is locally transported, i.e. the mass that is transported within distance $\frac{\ell}{\log \Delta}$ (suffices to choose $\varepsilon = 1$) and the leftover mass. This can also be computed by keeping track of the mass that is transported locally within T_I (i.e. mass that is transported at nodes below certain height). We omit the details from this version.

As shown in [1] (see also [3]), the approximation factor can be improved by increasing the fanout of each node of T_I to roughly k^δ. This improves the approximation factor to $O(\log \frac{1}{\delta})$ but the update time now becomes $O(k^\delta \log n)$. Hence, we can compute a $O(\log \delta^{-1})$-approximate OT plan in $O(k^{1+\delta} \log^2 n)$ time. Finally, we can handle arbitrary spread using the idea of moats as in [1], which rules out certain random shifts and performs a random shift at each level of the tree. Omitting all the details, which can be found in [1], we conclude the following.

▶ **Lemma 13.** *For any parameter $\delta > 0$, an $O(\log^2(\delta^{-1}))$-approximate OT plan at all leaf cells of $\mathcal{Q}$ can be computed in $O(k^{1+\delta} \log^{O(1)} \Delta)$ time with high probability.*

Combining Lemmas 9 and 13, we obtain Theorem 3.

5 Lower-Bound Construction of a Transport Plan

We have proven algorithmically in Section 3 that there exists a $(1+\varepsilon)$-approximate OT plan which maps boxes to boxes, where the complexity of representing this plan is roughly $\widetilde{O}(n^2)$. In this section, we provide a construction of two histograms where any $(1+\varepsilon)$-approximate box-to-box transport plan of this form incurs quadratic complexity, even if the complexity of the subdivision is linear.

We consider the following example, shown in Figure 7. Intuitively, we construct a sequence of long and skinny strips at a distance of roughly 1 apart. A lower bound on the cost of a transport plan would be to couple each rectangle in the histogram with its nearest rectangle. However, due to the equal spacing of rectangles, every red rectangle has two choices of closest

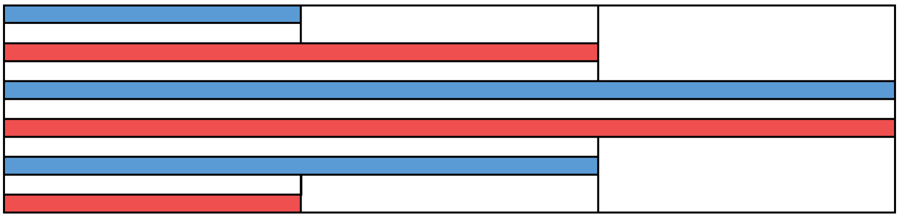

Figure 7 This figure gives a rough sketch of the histograms μ_1 and μ_2 in the lower bound construction. We emphasize that the complexity of the overlay is $k = O(n)$.

blue rectangle. The only assignment that allows one to couple each box with its nearest box of the opposite color is to divide the boxes into vertical strips and alternate the direction of the assignment.

Formally, we describe the two histograms μ_1 and μ_2 by their point-wise probability density functions. With some slight abuse of notation, we also use μ_1 and μ_2 to denote the corresponding probability density functions. Let $n \geq 1$ be a fixed integer. For each $0 < i \leq n$, define

$$\mu_1(x) = \begin{cases} \frac{2n}{n+1}, & i \text{ odd}, x \in [0, (n-i+1) \cdot n^4] \times \left[i - \frac{1}{2n^6} - \frac{1}{2}, i + \frac{1}{2n^6} - \frac{1}{2}\right], \\ \frac{2n}{n+1}, & i \text{ even}, x \in [0, (n-i+1) \cdot n^4] \times \left[-i - \frac{1}{2n^6} + \frac{1}{2}, -i + \frac{1}{2n^6} + \frac{1}{2}\right], \\ 0, & \text{otherwise}, \end{cases}$$

and similarly

$$\mu_2(x) = \begin{cases} \frac{2n}{n+1}, & i \text{ even}, x \in [0, (n-i+1) \cdot n^4] \times \left[i - \frac{1}{2n^6} - \frac{1}{2}, i + \frac{1}{2n^6} - \frac{1}{2}\right], \\ \frac{2n}{n+1}, & i \text{ odd}, x \in [0, (n-i+1) \cdot n^4] \times \left[-i - \frac{1}{2n^6} + \frac{1}{2}, -i + \frac{1}{2n^6} + \frac{1}{2}\right], \\ 0, & \text{otherwise}. \end{cases}$$

This completes the construction of μ_1 and μ_2. See Figure 7 for an illustration of this example. Then Lemma 14 on this choice of μ_1 and μ_2 implies Theorem 2.

▶ **Lemma 14.** *For any refined subdivision Ξ' of Ξ and $\varepsilon \leq \frac{1}{3}$, if $\sigma_{\Xi'}$ is a $(1+\varepsilon)$-approximate box-to-box transport plan between μ_1 and μ_2, then $|\Xi'| = \Omega((1-3\varepsilon)n^2)$.*

6 Conclusion

In this paper, we presented the first relative approximation algorithm for computing an optimal transport plan between two continuous probability distributions, each represented as a histogram over a two-dimensional rectangular subdivision. We also presented a lower-bound construction that suggests our algorithm being close to optimal. We conclude by mentioning a few open problems:

- Can an $(1+\varepsilon)$-approximate value of the cost of an OT plan between a pair of two-dimensional histograms be computed in near-linear time?
- Can our algorithm be extended to more general continuous probability distributions? For example, what if each probability distribution is represented as a piecewise linear function over a triangulation in $\mathbb{R}^2$, or as a histogram in $\mathbb{R}^d$?

References

1 Pankaj K Agarwal, Kyle Fox, Debmalya Panigrahi, Kasturi R Varadarajan, and Allen Xiao. Faster algorithms for the geometric transportation problem. In *Proceedings of the 33rd Annual Symposium on Computational Geometry*, pages 7:1–7:16, 2017. doi:10.4230/LIPIcs.SOCG.2017.7.

2 Pankaj K Agarwal, Sharath Raghvendra, Pouyan Shirzadian, and Keegan Yao. A combinatorial algorithm for the semi-discrete optimal transport problem. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, pages 25857–25887, 2024.

3 Pankaj K Agarwal, Sharath Raghvendra, Pouyan Shirzadian, and Keegan Yao. Fast and accurate approximations of the optimal transport in semi-discrete and discrete settings. In *Proceedings of the 35th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 4514–4529, 2024. doi:10.1137/1.9781611977912.159.

4 Pankaj K Agarwal, Sharath Raghvendra, Pouyan Shirzadian, and Keegan Yao. Efficient algorithms for robust and partial semi-discrete optimal transport. In *Proceedings of the 39th Annual Conference on Neural Information Processing Systems*, 2025.

5 Pankaj K Agarwal, Sharath Raghvendra, Pouyan Shirzadian, and Keegan Yao. Efficient approximation algorithm for computing Wasserstein barycenter under Euclidean metric. In *Proceedings of the 36th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 4809–4826, 2025. doi:10.1137/1.9781611978322.163.

6 Pankaj K. Agarwal and Kasturi R. Varadarajan. A near-linear constant-factor approximation for Euclidean bipartite matching? In *Proceedings of the 20th Annual Symposium on Computational Geometry*, pages 247–252, 2004. doi:10.1145/997817.997856.

7 Alok Aggarwal, Amotz Bar-Noy, Samir Khuller, Dina Kravets, and Baruch Schieber. Efficient minimum cost matching using quadrangle inequality. In *Proceedings of the 33rd Annual Symposium on Foundations of Computer Science*, pages 583–592, 1992. doi:10.1109/SFCS.1992.267793.

8 Jason Altschuler, Jonathan Weed, and Philippe Rigollet. Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 1961–1971, 2017.

9 David Alvarez-Melis and Tommi Jaakkola. Gromov-Wasserstein alignment of word embedding spaces. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1881–1890, 2018. doi:10.18653/V1/D18-1214.

10 Alexandr Andoni and Hengjie Zhang. Sub-quadratic $(1+\varepsilon)$-approximate Euclidean spanners, with applications. In *Proceedings of the 64th Annual IEEE Symposium on Foundations of Computer Science*, pages 98–112, 2023.

11 Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning*, pages 214–223, 2017.

12 Sanjeev Arora. Polynomial time approximation schemes for Euclidean TSP and other geometric problems. In *Proceedings of the 37th Annual Symposium on Foundations of Computer Science*, pages 2–11, 1996. doi:10.1109/SFCS.1996.548458.

13 Sanjeev Arora. Approximation schemes for NP-hard geometric optimization problems: A survey. *Mathematical Programming*, 97(1):43–69, 2003. doi:10.1007/S10107-003-0438-Y.

14 Sanjeev Arora, Prabhakar Raghavan, and Satish Rao. Approximation schemes for Euclidean k-medians and related problems. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 106–113, 1998. doi:10.1145/276698.276718.

15 Franz Aurenhammer, Friedrich Hoffmann, and Boris Aronov. Minkowski-type theorems and least-squares clustering. *Algorithmica*, 20(1):61–76, 1998. doi:10.1007/PL00009187.

16 Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000. doi:10.1007/S002110050002.

17 Lorenzo Beretta, Vincent Cohen-Addad, Rajesh Jayaram, and Erik Waingarten. Approximating high-dimensional earth mover's distance as fast as closest pair. In *Proceedings of the 66th Annual Symposium on Foundations of Computer Science*, pages 241–266, 2025. doi:10.1109/FOCS63196.2025.00017.

18 Nicolas Bonneel and Julie Digne. A survey of optimal transport for computer graphics and computer vision. *Computer Graphics Forum*, 42(2):439–460, 2023. doi:10.1111/CGF.14778.

19 Yann Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *Comptes rendus de l'Académie des Sciences Paris Série I Math.*, 305:805–808, 1987.

20 Yann Brenier. Optimal transport, convection, magnetic relaxation and generalized boussinesq equations. *Journal of Nonlinear Science*, 19(5):547–570, 2009. `doi:10.1007/S00332-009-9044-3`.

21 Wanxing Chang, Ye Shi, and Jingya Wang. Csot: curriculum and structure-aware optimal transport for learning with noisy labels. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pages 8528–8541, 2023.

22 Moses S Charikar. Similarity estimation techniques from rounding algorithms. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 380–388, 2002. `doi:10.1145/509907.509965`.

23 Rick Chartrand, Brendt Wohlberg, Kevin Vixie, and Erik Bollt. A gradient descent solution to the Monge-Kantorovich problem. *Applied Mathmematical Sciences*, 3(22):1071–1080, 2009.

24 Li Chen, Rasmus Kyng, Yang Liu, Richard Peng, Maximilian Probst Gutenberg, and Sushant Sachdeva. Maximum flow and minimum-cost flow in almost-linear time. *Journal of the ACM*, 72(3):1–103, 2025. `doi:10.1145/3728631`.

25 Sinho Chewi, Jonathan Niles-Weed, and Philippe Rigollet. *Statistical Optimal Transport*. Springer, 2025.

26 Ching-Yao Chuang, R Devon Hjelm, Xin Wang, Vibhav Vineet, Neel Joshi, Antonio Torralba, Stefanie Jegelka, and Yale Song. Robust contrastive learning against noisy views. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16670–16681, 2022.

27 Marco Cuturi. Sinkhorn distances: lightspeed computation of optimal transport. In *Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 2*, pages 2292–2300, 2013. URL: `https://proceedings.neurips.cc/paper/2013/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html`.

28 Fernando De Goes, Katherine Breeden, Victor Ostromoukhov, and Mathieu Desbrun. Blue noise through optimal transport. *ACM Transactions on Graphics*, 31(6):1–11, 2012. `doi:10.1145/2366145.2366190`.

29 James R Driscoll, Neil Sarnak, Daniel Dominic Sleator, and Robert Endre Tarjan. Making data structures persistent. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*, pages 109–121, 1986. `doi:10.1145/12130.12142`.

30 Aden Forrow, Jan-Christian Hütter, Mor Nitzan, Philippe Rigollet, Geoffrey Schiebinger, and Jonathan Weed. Statistical optimal transport via factored couplings. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, pages 2454–2465, 2019. URL: `http://proceedings.mlr.press/v89/forrow19a.html`.

31 Emily Fox and Jiashuai Lu. A deterministic near-linear time approximation scheme for geometric transportation. In *Proceedings of the 64th Annual IEEE Symposium on Foundations of Computer Science*, pages 1301–1315, 2023. `doi:10.1109/FOCS57990.2023.00078`.

32 Alfred Galichon. *Optimal Transport Methods in Economics*. Princeton University Press, 2016.

33 Matthias Gelbrich. On a formula for the l_2 Wasserstein metric between measures on Euclidean and hilbert spaces. *Mathematische Nachrichten*, 147(1):185–203, 1990.

34 Gramoz Goranci, Peter Kiss, Neel Patel, Martin P Seybold, Eva Szilagyi, and Da Wei Zheng. Fully dynamic Euclidean bi-chromatic matching in sublinear update time. In *Proceedings of the 42nd International Conference on Machine Learning*, 2025.

35 Piotr Indyk and Nitin Thaper. Fast image retrieval via embeddings. In *3rd International Workshop on Statistical and Computational Theories of Vision*, 2003.

36 Arun Jambulapati, Aaron Sidford, and Kevin Tian. A direct $\widetilde{O}(1/\varepsilon)$ iteration parallel algorithm for optimal transport. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 11359–11370, 2019.

37 Hicham Janati, Boris Muzellec, Gabriel Peyré, and Marco Cuturi. Entropic optimal transport between unbalanced gaussian measures has a closed form. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, pages 10468–10479, 2020.

38 Abdelwahed Khamis, Russell Tsuchida, Mohamed Tarek, Vivien Rolland, and Lars Petersson. Scalable optimal transport methods in machine learning: A contemporary survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

39 Andrey Boris Khesin, Aleksandar Nikolov, and Dmitry Paramonov. Preconditioning for the geometric transportation problem. In *Proceedings of the 35th Annual Symposium on Computational Geometry*, pages 15–1, 2019.

40 Jun Kitagawa. An iterative scheme for solving the optimal transportation problem. *Calculus of Variations and Partial Differential Equations*, 51(1):243–263, 2014.

41 Jun Kitagawa, Quentin Mérigot, and Boris Thibert. Convergence of a Newton algorithm for semi-discrete optimal transport. *Journal of the European Mathematical Society*, 21(9):2603–2651, 2019.

42 Stavros G Kolliopoulos and Satish Rao. A nearly linear-time approximation scheme for the euclidean k-median problem. *SIAM Journal on Computing*, 37(3):757–782, 2007. `doi:10.1137/S0097539702404055`.

43 Soheil Kolouri, Se Rim Park, Matthew Thorpe, Dejan Slepcev, and Gustavo K Rohde. Optimal mass transport: Signal processing and machine-learning applications. *IEEE Signal Processing Magazine*, 34(4):43–59, 2017. `doi:10.1109/MSP.2017.2695801`.

44 Nathaniel Lahn, Sharath Raghvendra, and Kaiyi Zhang. A combinatorial algorithm for approximating the optimal transport in the parallel and MPC settings. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pages 21675–21686, 2023.

45 Zhengfeng Lai, Chao Wang, Sen-ching Cheung, and Chen-Nee Chuah. Sar: Self-adaptive refinement on pseudo labels for multiclass-imbalanced semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4091–4100, 2022.

46 Marc Lambert, Sinho Chewi, Francis Bach, Silvère Bonnabel, and Philippe Rigollet. Variational inference via wasserstein gradient flows. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, pages 14434–14447, 2022.

47 Hugo Lavenant, Sebastian Claici, Edward Chien, and Justin Solomon. Dynamical optimal transport on discrete surfaces. *ACM Transactions on Graphics (TOG)*, 37(6):1–16, 2018.

48 Tam Le, Makoto Yamada, Kenji Fukumizu, and Marco Cuturi. Tree-sliced variants of wasserstein distances. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 12304–12315, 2019.

49 Bruno Lévy and Erica L Schwindt. Notions of optimal transport theory and how to implement them on a computer. *Computers & Graphics*, 72:135–148, 2018. `doi:10.1016/J.CAG.2018.01.009`.

50 Quentin Mérigot. A multiscale approach to optimal transport. *Computer graphics forum*, 30(5):1583–1592, 2011. `doi:10.1111/J.1467-8659.2011.02032.X`.

51 Quentin Mérigot, Jocelyn Meyron, and Boris Thibert. An algorithm for optimal transport between a simplex soup and a point cloud. *SIAM Journal on Imaging Sciences*, 11(2):1363–1389, 2018. `doi:10.1137/17M1137486`.

52 Eduardo Fernandes Montesuma, Fred Maurice Ngole Mboula, and Antoine Souloumiac. Recent advances in optimal transport for machine learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

53 Vladimir I Oliker and Laird D Prussner. On the numerical solution of the equation $\frac{\partial^2 z}{\partial x^2}\frac{\partial^2 z}{\partial y^2} - \left(\frac{\partial^2 z}{\partial x \partial y}\right) = f$ and its discretizations, i. *Numerische Mathematik*, 54(3):271–293, 1989.

54 Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

55 Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *Proceedings of the 3rd International Conference on Scale Space and Variational Methods in Computer Vision*, pages 435–446, 2012.

56 Sharath Raghvendra and Pankaj K Agarwal. A near-linear time ε-approximation algorithm for geometric bipartite matching. *Journal of the ACM*, 67(3):1–19, 2020. doi:10.1145/3393694.

57 Satish B Rao and Warren D Smith. Approximating geometrical graphs via "spanners" and "banyans". In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 540–550, 1998.

58 Filippo Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, volume 87. Birkhäuser, 2015.

59 Justin Solomon, Fernando De Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (ToG)*, 34(4):1–11, 2015. doi:10.1145/2766963.

60 Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, pages 1–34, 2024.

61 Marc van Kreveld, Frank Staals, Amir Vaxman, and Jordi Vermeulen. Approximating the earth mover's distance between sets of geometric objects. *arXiv preprint arXiv:2104.08136*, 2021. arXiv:2104.08136.

62 Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2008.

63 Richard L Wheeden and Antoni Zygmund. *Measure and Integral: An Introduction to Real Analysis.* CRC Press, 2015.

64 Mikhail Yurochkin, Sebastian Claici, Edward Chien, Farzaneh Mirzazadeh, and Justin Solomon. Hierarchical optimal transport for document representation. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, pages 1601–1611, 2019.

A Missing Proofs from the Main Text

▶ **Lemma 8.** *The transport plan $\widetilde{\sigma}$ is an $(1+\varepsilon)$-approximate OT plan between $\widetilde{\mu}_1$ and $\widetilde{\mu}_2$ with respect to d_Q, i.e., $\int_{\square^*\times\square^*} \mathsf{d}_Q(u,v)d\widetilde{\sigma}(u,v) \leq (1+\varepsilon)\mathsf{OPT}$, where OPT is the cost of the optimal transport plan between $\widetilde{\mu}_1$ and $\widetilde{\mu}_2$ with respect to d_Q.*

Proof. We note that the algorithm in [31] computes a $(1+\varepsilon)$-approximate OT plan between μ_1° and μ_2°. It therefore suffices to show that the choice of discretization incurs an $\varepsilon \cdot \mathsf{OPT}$ additive error. We perform an analysis on each type of leaf cell.

1. In the e-leaf cells $\square$, where $|V \cap \square| = 0$, we note that by construction, mass is locally transported in the box-to-box plan $\sigma_{\Xi'_\square}$ if and only if it is transported at a distance at most $\frac{\varepsilon^2}{\log\Delta} \cdot \ell(\square)$. Moreover, $\widetilde{\mu}_i$ is the leftover mass from μ_i not transported by $\sigma_{\Xi'_\square}$. Therefore, if $\widetilde{\mu}_1(U_1) > 0$ for $U_1 \subseteq \square$, any $U_2 \subseteq \square^*$ where $\widetilde{\sigma}(U_1 \times U_2) > 0$ can be assumed to have cost $\mathsf{d}_Q(u,v) \geq \frac{\varepsilon^2}{\log\Delta} \cdot \ell(\square)$ for all $(u,v) \in U_1 \times U_2$. We conclude that snapping the leftover mass in $\square$ to a sufficiently small grid of points in the leaf cell to construct μ_i° admits an $\varepsilon \cdot \mathsf{OPT}$ additive error.
2. In the v-leaf cells $\square$, where $|V \cap \square| = 1$, we note that $\ell(\square) \leq \varepsilon \cdot \mathsf{OPT}$ is explicitly maintained by construction (via application of Lemma 4). ◀

▶ **Lemma 11.** *If the subdivision Ξ consists of boxes where the ratio of longest to shortest edge is $r \geq 1$, then $\sum_{\square\in\mathcal{L}} |\Xi_\square| = O(kr\log\Delta)$.*

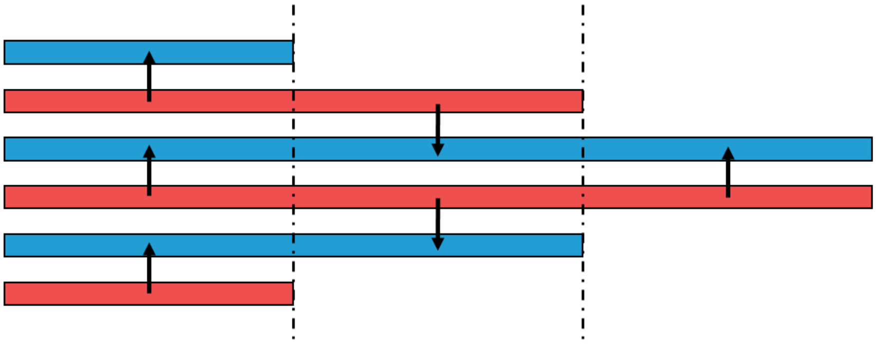

Figure 8 Here we showcase the lower bound example described in Section 5, as well as the box-to-box transport plan $\sigma_{\Xi'}$ described in the proof of Lemma 14, shown in black. The distribution μ_1 is shown in red, and the distribution μ_2 is shown in blue. Each long horizontal rectangle is given in the input decomposition Ξ, and the refined subdivision will divide each rectangle along the vertical dotted lines. When traversing from vertical strip i to $i+1$, the direction of the box-to-box transport plan flips. Unlike the image here, the length of each long horizontal box is polynomially long and polynomially skinny.

Proof. It suffices to prove the complexity of e-leaf cells in the quadtree since by construction $|\Xi_\square| \le 4$ for any v-leaf cell $\square$.

For any e-leaf cell $\square$ with sidelength $\ell > 0$ and any rectangle R intersecting $\square$, we know that $\overline{\ell}(R) \ge \ell$ since the rectangle R is long with respect to $\square$ by definition of an e-leaf cell. Therefore by the fatness assumption, we note that $\underline{\ell}(R) \ge \frac{\ell}{r}$. By a simple packing argument in 1-dimension, it follows that at most $r+1$ many rectangles can intersect the e-leaf cell $\square$. We conclude that $|\Xi_\square| = O(\max\{r, 1\})$ for every $\square \in \mathcal{L}$.

Putting everything together, we conclude that there are at most $O(k \log \Delta)$ many leaves by Lemma 6 and each leaf cell has complexity at most $O(r)$. Therefore, $\sum_{\square \in \mathcal{L}} |\Xi_\square| = O(kr \log \Delta)$. ◀

▶ **Lemma 14.** *For any refined subdivision Ξ' of Ξ and $\varepsilon \le \frac{1}{3}$, if $\sigma_{\Xi'}$ is a $(1+\varepsilon)$-approximate box-to-box transport plan between μ_1 and μ_2, then $|\Xi'| = \Omega((1-3\varepsilon)n^2)$.*

Proof. We assume for technicality that $\varepsilon \ge \frac{12}{n+1}$; if instead $\varepsilon < \frac{12}{n+1}$, then any $(1+\varepsilon)$-approximate transport plan is also a $(1+\frac{12}{n+1})$-approximate transport plan and $(1-3\cdot\frac{12}{n+1})n^2 = \Omega((1-3\varepsilon)n^2)$. We note there exists a box-to-box transport plan which has a cost of 1. The subdivision Ξ' consists of the following sets. Define

$$R_{i,j} = \left[(i-1)\cdot n^4, i\cdot n^4\right] \times \left[j - \frac{1}{2n^6} + \frac{1}{2}, j + \frac{1}{2n^6} + \frac{1}{2}\right].$$

Then the partition Ξ' divides the boxes of Ξ with positive mass into the boxes $R_{i,j}$ constructed above, i.e., $\{R_{i,j} : -n \le j \le n, 0 < i \le |j|\} \subseteq \Xi'$ and the remainder of $\square^*$ can be subdivided arbitrarily to construct Ξ'. That is, every long and skinny horizontal box with positive mass is divided into subregions of length n^4 in the x-coordinate direction. The box-to-box transport plan is shown in Figure 8. Note that $|\Xi'| = O(n^2)$. Moreover, the minimum distance between any $p, q \in \square^*$ such that $\mu_1(p) > 0$ and $\mu_2(q) > 0$ is $1 - \frac{1}{n^6}$ by construction. We will define the *vertical strip* as $\mathcal{V}_i = \{R_{i',j'} \in \Xi' : i' = i\}$, i.e. $\mathcal{V}_i$ is the set of rectangles $R_{i,j}$ for some j.

Let σ be an arbitrary $(1+\varepsilon)$-approximate optimal transport plan. Then since the optimal transport plan has a cost of at most 1, we note that $w(\sigma) \le 1+\varepsilon$. Given a parameter $L > 1$, let $U_L \subseteq \square^* \times \square^*$ denote the set of pairs of points (u, v) maximizing $\sigma(U_L) > 0$ such that $\|u - v\| > L$ for all $(u, v) \in U_L$. Then observe that

$$
\begin{aligned}
1+\varepsilon &\geq w(\sigma) \\
&= \int_{U_L} \|u-v\| d\sigma(u,v) + \int_{\square^* \times \square^* \setminus U_L} \|u-v\| d\sigma(u,v) \\
&\geq L\sigma(U_L) + \left(1 - \frac{1}{n^6}\right) \sigma(\square^* \times \square^* \setminus U_L) \\
&= L\sigma(U_L) + \left(1 - \frac{1}{n^6}\right) (1 - \sigma(U_L)) \\
&= \left(L - 1 + \frac{1}{n^6}\right) \cdot \sigma(U_L) + \left(1 - \frac{1}{n^6}\right).
\end{aligned}
$$

From this, we observe $\sigma(U_L) \leq \left(\varepsilon + \frac{1}{n^6}\right) \cdot \left(L - 1 + \frac{1}{n^6}\right)^{-1} \leq \frac{2\varepsilon}{L-1}$ for $\varepsilon > \frac{6}{n-1}$ and any $L > 1$.

Suppose $L = n^3$. We note the length of each box in Ξ' (or equivalently the length of each step in the staircase) is n^4. Moreover, the mass of any $R_{i,j}$ satisfies $\max\{\mu_1(R_{i,j}), \mu_2(R_{i,j})\} = \left(2\sum_{i=1}^n i\right)^{-1} = \frac{1}{n(n+1)}$ by construction (there are exactly 2 long rectangles of each length ranging from 1 to n times n^4 in the construction of Ξ, and each one is divided into intervals of length n^4 in Ξ'). Therefore,

$$
\sigma(U_L) \leq \frac{2\varepsilon}{L-1} = \frac{2\varepsilon}{n^3 - 1} \leq \frac{1}{n \cdot n(n+1)} = \frac{1}{n} \max\{\mu_1(R_{i,j}), \mu_2(R_{i,j})\}
$$

for any $R_{i,j}$ and any sufficiently large n, assuming $\varepsilon \leq \frac{1}{2}$. We conclude that for every box in Ξ', at most a $\frac{1}{n}$ fraction of its mass can be transported at a distance at least n^3.

Note that each vertical strip $\mathcal{V}_i$ has a length of n^4. Hence only a very small $\frac{1}{n}$ fraction of the mass of each box $R_{i,j}$ from $\mathcal{V}_i$ in Ξ' can be transported to a box in any other vertical strip $\mathcal{V}_{i'}$ at a cost less than n^3. Combining these two results, we then conclude that a $\left(1 - \frac{2}{n}\right)$-fraction of the mass from each rectangle $R_{i,j}$ in Ξ' must be transported to a box $R_{i,j'}$ in the same vertical strip $\mathcal{V}_i$.

We note that Ξ' has quadratic complexity. Therefore, if σ were to have subquadratic complexity, then some of the segments of Ξ' that subdivide Ξ do not actually increase the complexity of the transport plan σ. That is, if mass from R_{i,j_1} is transported to R_{i,j_2}, then we must also have that mass from R_{i+1,j_1} is transported to R_{i+1,j_2}. We bound the complexity of the transport plan by proving a lower bound on the number of such events in σ.

Consider the transition from vertical strip $\mathcal{V}_{i+1}$ to vertical strip $\mathcal{V}_i$. In this sweep, we notice that two boxes were added: one at the top of the vertical strip and one at the bottom. At least a $(1 - \frac{2}{n})$ fraction of the mass of every box of $\mathcal{V}_i$ (resp. $\mathcal{V}_{i+1}$) must be routed to another box of $\mathcal{V}_i$ (resp. $\mathcal{V}_{i+1}$). Let σ_i (resp. σ_{i+1}) denote the transport plan restricted to pairs u, v such that $u \in R_{i,j}$ and $v \in R_{i,j'}$ for two values j, j'. Let $V_i = \bigcup_{R_{i,j} \in \mathcal{V}_i} R_{i,j}$ be the set of points covered by rectangles in $\mathcal{V}_i$ for convenience. Define π_i to be the measure of changes in the transport plans σ_i and σ_{i+1}. That is,

$$
\begin{aligned}
&\pi_i(A \times B) \\
&\quad = \max_{A_1, A_2} \left\{ \left| \sigma_i(A_1 \times B) - \sigma_{i+1}((A_1 + \vec{e}_1) \times B) \right| + \left| \sigma_i(A_2 \times B) - \sigma_{i+1}((A_2 + \vec{e}_1) \times B) \right| \right\}
\end{aligned}
$$

for all measurable $A, B \subseteq V_i$ where the maximum is taken over all pairs of disjoint measurable sets A_1, A_2 that partition A and $\vec{e}_1$ is the unit vector in the first coordinate direction and

$A + \vec{v} = \{a + \vec{v} : a \in A\}$ denotes the Minkowski sum. It is possible to show that π_i is indeed a measure. We can view π_i as the analogue of an "augmenting path", though this analogy is not quite precise. Our goal is to lower bound $\sum_i \sum_{j_1,j_2} \pi_i(R_{i,j_1} \times R_{i,j_2})$.

For any choice of i, we observe that the rectangles $R_\uparrow = R_{i,n-i+1}$ and $R_\downarrow = R_{i,-n+i-1}$ (i.e. the top- and bottom-most rectangles in $\mathcal{V}_i$ that have positive mass) both necessarily satisfy

$$\sum_j \pi_i(R_\uparrow \times R_{i,j}) + \pi_i(R_{i,j} \times R_\uparrow) \geq \left(1 - \frac{2}{n}\right) \cdot \frac{1}{n(n+1)}$$

and

$$\sum_j \pi_i(R_\downarrow \times R_{i,j}) + \pi_i(R_{i,j} \times R_\downarrow) \geq \left(1 - \frac{2}{n}\right) \cdot \frac{1}{n(n+1)}$$

since either $\mu_1(R_\uparrow) = \mu_2(R_\downarrow) = 0$ or $\mu_2(R_\uparrow) = \mu_1(R_\downarrow) = 0$ by construction of μ_1 and μ_2. This observation alone is sufficient to deduce

$$\begin{aligned}
&\sum_{j_1,j_2} \pi_i(R_{i,j_1} \times R_{i,j_2}) \cdot \inf_{x \in R_{i,j_1}, y \in R_{i,j_2}} \{\|x-y\|\} \\
&\geq \left(1 - \frac{2}{n}\right) \cdot \frac{1}{n(n+1)} \cdot \left(1 - \frac{1}{n^6}\right)(2n - 2i + 1) \\
&= \left(1 - \frac{2}{n}\right) \cdot \frac{2n - 2i + 1}{n(n+1)} \cdot \left(1 - \frac{1}{n^6}\right)
\end{aligned}$$

since the masses of $R_\uparrow$ and $R_\downarrow$ must be transported somewhere by σ_i.

We additionally observe that for any j_1, j_2, the quantity

$$\sigma_i(R_{i,j_1} \times R_{i,j_2}) + \sigma_{i+1}(R_{i+1,j_1} \times R_{i+1,j_2}) - \pi_i(R_{i,j_1} \times R_{i,j_2})$$

denotes the measure of mass that that both σ_i sends from R_{i,j_1} to R_{i,j_2} and σ_{i+1} sends from R_{i+1,j_1} to R_{i+1,j_2}. Note that $\inf_{x \in R_{i,j_1}, y \in R_{i,j_2}} \{\|x-y\|\} \geq \left(1 - \frac{1}{n^6}\right)$. The cost of this portion of the transport plan is therefore necessarily at least

$$\begin{aligned}
&\sum_{j_1,j_2} \left(\sigma_i(R_{i,j_1} \times R_{i,j_2}) + \sigma_{i+1}(R_{i+1,j_1} \times R_{i+1,j_2}) - \pi_i(R_{i,j_1} \times R_{i,j_2})\right) \cdot \left(1 - \frac{1}{n^6}\right) \\
&\geq \left(1 - \frac{2}{n}\right) \cdot \left(\frac{1}{n(n+1)} \cdot ((2n - 2i + 1) + (2n - 2(i+1) + 1))\right) \cdot \left(1 - \frac{1}{n^6}\right) \\
&\quad - \left(\sum_{j_1,j_2} \pi_i(R_{i,j_1} \times R_{i,j_2})\right) \cdot \left(1 - \frac{1}{n^6}\right) \\
&= \left(\left(1 - \frac{2}{n}\right) \frac{4n - 4i}{n(n+1)} - \sum_{j_1,j_2} \pi_i(R_{i,j_1} \times R_{i,j_2})\right) \cdot \left(1 - \frac{1}{n^6}\right)
\end{aligned}$$

Combining these two inequalities gives

$$\begin{aligned}
&w(\sigma_i) + w(\sigma_{i+1}) \\
&= \sum_{j_1,j_2} \left[\int_{R_{i,j_1} \times R_{i,j_2}} \|x-y\| d\sigma_i(x,y) + \int_{R_{i+1,j_1} \times R_{i+1,j_2}} \|x-y\| d\sigma_{i+1}(x,y)\right]
\end{aligned}$$

$$
\begin{aligned}
&\geq \sum_{j_1,j_2} \inf_{x\in R_{i,j_1}, y\in R_{i,j_2}} \{\|x-y\|\} \cdot (\sigma_i(R_{i,j_1}\times R_{i,j_2}) + \sigma_{i+1}(R_{i+1,j_1}\times R_{i+1,j_2}))\\
&\geq \sum_{j_1,j_2} \inf_{x\in R_{i,j_1}, y\in R_{i,j_2}} \{\|x-y\|\}\cdot\\
&\quad \Big(\sigma_i(R_{i,j_1}\times R_{i,j_2}) + \sigma_{i+1}(R_{i+1,j_1}\times R_{i+1,j_2}) - \pi_i(R_{i,j_1}\times R_{i,j_2}) + \pi_i(R_{i,j_1}\times R_{i,j_2})\Big)\\
&\geq \left(\left(1-\frac{2}{n}\right)\left(\frac{2n-2i+1}{n(n+1)} + \frac{4n-4i}{n(n+1)}\right) - \sum_{j_1,j_2}\pi_i(R_{i,j_1}\times R_{i,j_2})\right)\cdot\left(1-\frac{1}{n^6}\right)\\
&= \left(\left(1-\frac{2}{n}\right)\frac{6n-6i+1}{n(n+1)} - \sum_{j_1,j_2}\pi_i(R_{i,j_1}\times R_{i,j_2})\right)\cdot\left(1-\frac{1}{n^6}\right)
\end{aligned}
$$

Now after summation over the vertical strips $i \leq n$, we conclude

$$
\begin{aligned}
2w(\sigma) &\geq \sum_{i=1}^{n-1} w(\sigma_i) + w(\sigma_{i+1})\\
&\geq \sum_{i=1}^{n-1}\left(\left(1-\frac{2}{n}\right)\frac{6n-6i+1}{n(n+1)} - \sum_{j_1,j_2}\pi_i(R_{i,j_1}\times R_{i,j_2})\right)\cdot\left(1-\frac{1}{n^6}\right)\\
&= \left(\left(1-\frac{2}{n}\right)\frac{6n^2-3n(n+1)+n}{n(n+1)} - \sum_{i=1}^{n-1}\sum_{j_1,j_2}\pi_i(R_{i,j_1}\times R_{i,j_2})\right)\cdot\left(1-\frac{1}{n^6}\right)\\
&= \left(\frac{3n^2-8n+4}{n(n+1)} - \sum_{i=1}^{n}\sum_{j_1,j_2}\pi_i(R_{i,j_1}\times R_{i,j_2})\right)\cdot\left(1-\frac{1}{n^6}\right)\\
&\geq \left(3 - \frac{11}{n+1} - \sum_{i=1}^{n}\sum_{j_1,j_2}\pi_i(R_{i,j_1}\times R_{i,j_2})\right)\cdot\left(1-\frac{1}{n^6}\right).
\end{aligned}
$$

Since σ is assumed to be a $(1+\varepsilon)$-approximate transport plan and there exists a box-to-box transport plan with cost 1, we again note $2w(\sigma) \leq 2(1+\varepsilon) = 2+2\varepsilon$. Therefore after some algebra, we obtain the inequality

$$
\sum_{i=1}^{n}\sum_{j_1,j_2}\pi_i(R_{i,j_1}\times R_{i,j_2}) \geq 3 - \frac{11}{n+1} - (2+2\varepsilon)\left(1-\frac{1}{n^6}\right)^{-1} = 1 - \frac{12}{n+1} - 2\varepsilon \geq 1-3\varepsilon
$$

assuming $n \geq 2+2\varepsilon$ and $\varepsilon \geq \frac{12}{n+1}$. Therefore by definition of π_i, the difference in the amount of mass $R_{i,j}$ transports mass to each $R_{i,j'}$ and the amount of mass $R_{i+1,j}$ transports mass to each $R_{i+1,j'}$ totaled across all rectangles $R_{i,j}$ is equal to $1-3\varepsilon$. There are $O(n^2)$ many such rectangles $R_{i,j}$. We conclude that the complexity of σ must necessarily be $\Omega((1-3\varepsilon)n^2)$. ◀