

A near-linear time approximation scheme for geometric transportation with arbitrary supplies and spread

Kyle Fox*

Jiashuai Lu[†]

November 24, 2020

Abstract

The geometric transportation problem takes as input a set of points P in d -dimensional Euclidean space and a supply function $\mu : P \rightarrow \mathbb{R}$. The goal is to find a transportation map, a non-negative assignment $\tau : P \times P \rightarrow \mathbb{R}_{\geq 0}$ to pairs of points, so the total assignment leaving each point is equal to its supply, i.e., $\sum_{r \in P} \tau(q, r) - \sum_{p \in P} \tau(p, q) = \mu(q)$ for all points $q \in P$. The goal is to minimize the weighted sum of Euclidean distances for the pairs, $\sum_{(p,q) \in P \times P} \tau(p, q) \cdot \|q - p\|_2$.

We describe the first algorithm for this problem that returns, with high probability, a $(1 + \varepsilon)$ -approximation to the optimal transportation map in $O(n \text{ poly}(1/\varepsilon) \text{ polylog } n)$ time. In contrast to the previous best algorithms for this problem, our near-linear running time bound is independent of the spread of P and the magnitude of its real-valued supplies.

*Department of Computer Science, The University of Texas at Dallas; kyle.fox@utdallas.edu.

[†]Department of Computer Science, The University of Texas at Dallas; jiashuai.lu@utdallas.edu.

1 Introduction

We consider the *geometric transportation problem* in d -dimensional Euclidean space for any constant d . In this problem, we are given a set $P \subset \mathbb{R}^d$ of n points. Each point is assigned a real *supply* $\mu : P \rightarrow \mathbb{R}$ where $\sum_{p \in P} \mu(p) = 0$. A *transportation map* is a non-negative assignment $\tau : P \times P \rightarrow \mathbb{R}_{\geq 0}$ to pairs of points such that for all $q \in P$ we have $\sum_{r \in P} \tau(q, r) - \sum_{p \in P} \tau(p, q) = \mu(q)$. The *cost* of the transportation map is the weighted sum of Euclidean distances across all pairs, i.e. $\sum_{(p,q) \in P \times P} \tau(p, q) \cdot \|q - p\|_2$. Our goal is to find a transportation map of minimum cost, and we denote this minimum cost as $\text{COST}(P, \mu)$.

One may imagine the points with positive supply as piles of earth and those with negative supplies as holes in the ground. A transportation map describes how to transfer the earth to the holes without overfilling any hole, and its cost is the total number of “earth-miles” used to do the transfer. Consequently, $\text{COST}(P, \mu)$ is often referred to as the earth mover’s distance, although it can also be called the 1-Wasserstein distance between measures over the positively and negatively supplied points. The continuous version of the problem is sometimes called the *optimal transport* or *Monge-Kantorovich* problem, and it has been studied extensively by various mathematics communities [23]. The discrete version we study here has applications in shape matching, image retrieval, and graphics [5, 7, 9, 10, 17, 22].

Computing an optimal transportation map is easily done in polynomial time by reduction to the uncapacitated minimum cost flow problem in a complete bipartite graph between points with positive supply and those with negative supply. The graph has as many as $\Omega(n^2)$ edges, so this approach takes $O(n^3 \text{polylog } n)$ time using a combinatorial minimum cost flow algorithm of Orlin [16]. Assuming integral supplies with absolute values summing to U , we can use an algorithm of Lee and Sidford [15] instead to reduce the running time to $O(n^{2.5} \text{polylog}(n, U))$. Taking advantage of the geometry inherent in the problem, Agarwal *et al.* [1] describe how to implement Orlin’s algorithm for arbitrary supplies to find the optimal transportation map in $O(n^2 \text{polylog } n)$ time, but only for $d = 2$.

We can significantly reduce these running times by accepting a small loss in optimality. However, many results along this line focus on merely estimating the earth mover’s distance instead of computing the associated transportation map. Indyk [12] describes an $O(n \text{polylog } n)$ time algorithm that estimates the earth mover’s distance within a constant factor assuming unit supplies. Andoni *et al.* [3] describe an $O(n^{1+o(1)})$ time algorithm for arbitrary supplies that estimates the cost within a $1 + \varepsilon$ factor (the dependency on ε is hiding in the $o(1)$). As pointed out by Khesin, Nikolov, and Paramonov [13], a $1 + \varepsilon$ factor estimation of the distance is possible in $O(n^{1+o(1)} \varepsilon^{-O(d)})$ time (without the $o(1)$ hiding dependencies on ε) by running an approximation algorithm for minimum cost flow by Sherman [20] on a sparse Euclidean spanner over the input points. However, it is not clear how to extract a nearly optimal transportation map using the spanner’s flow.

Finding an actual transportation *map* may be more difficult. Sharathkumar and Agarwal [18] describe a $(1 + \varepsilon)$ -approximation algorithm for the integral supply case (i.e., an algorithm returning a map of cost at most $(1 + \varepsilon) \cdot \text{COST}(P, \mu)$) in $O(n\sqrt{U} \text{polylog}(U, \varepsilon, n))$ time. Agarwal *et al.* [1] describe a randomized algorithm with expected $O(\log^2(1/\varepsilon))$ -approximation ratio running in $O(n^{1+\varepsilon})$ expected time for the arbitrary supply case and a deterministic $O(n^{3/2} \varepsilon^{-d} \text{polylog}(U, n))$ time $(1 + \varepsilon)$ -approximation algorithm for the bounded integral supply case. Lahn *et al.* [14] describe an $O(n(C\delta)^2 \text{polylog}(U, n))$ ($C = \max_{p \in P} |\mu(p)|$) time algorithm computing a map of cost at most $\text{COST}(P, \mu) + \delta U$. Recently, Khesin *et al.* [13] described a randomized $(1 + \varepsilon)$ -approximation algorithm for the arbitrary supply case running in $O(n\varepsilon^{-O(d)} \log^{O(d)}(\text{SP}(P)) \log(n))$ time, where $\text{SP}(P)$

is the spread of the point set P .¹ The *spread* (also called aspect ratio) of P is the ratio of the diameter of P to the smallest pairwise distance between points in P . As Khesin *et al.* point out, one can reduce an instance with unbounded spread but bounded integral supplies to the case of bounded spread to get a $(1 + \varepsilon)$ -approximation running in $O(n\varepsilon^{-O(d)} \log^{O(d)}(U) \log^2(n))$ time, generalizing a near-linear time $(1 + \varepsilon)$ -approximation algorithm by Sharathkumar and Agarwal [19] for the unit supply case. The unit supply case is sometimes referred to as the *geometric bipartite matching problem*. Agarwal and Sharathkumar [2] also describe a *deterministic* $(1/\varepsilon)$ -approximation algorithm for geometric bipartite matching that runs in $O(n^{1+\varepsilon} \log n)$ time.

Despite these successes, prior work still does not include a near-linear time $(1 + \varepsilon)$ -approximation algorithm for the general case of arbitrary spread and real valued supplies. Often, an algorithm designed for bounded spread cases can be extended to work with cases of arbitrary spread. For example, one might substitute in *compressed quadtrees* [11, Chapter 2] in places where the bounded spread algorithm uses standard (uncompressed) quadtrees. This straightforward approach does not appear to work for the geometric transportation problem, however. As detailed below, Khesin *et al.* [13] use a quadtree to build a sparse graph as part of a reduction to the minimum cost flow problem. Both their running time and approximation analysis rely heavily on the tree having low depth when the spread is bounded. Unfortunately, a compressed quadtree is only guaranteed to have small *size*; the depth can still be linear in the number of leaves. One may also try the strategy of separating out groups of points P' that are much closer to each other than to the rest of the point set P , routing as much supply as possible within P' , and then satisfying what remains of the supplies in P' by treating P' as a single point. In fact, the result described below does employ a variant of this strategy (see Section 2.3). However, the simplified instances of the problem one gets using this strategy still yield compressed quadtrees of very high depth.

1.1 Our results and approach

We describe a randomized $(1 + \varepsilon)$ -approximation algorithm for the geometric transportation problem that runs in near-linear time irrespective of the spread of P or the supplies of its points. Our specific result is spelled out in the following theorem. We say an event occurs with *high probability* if it occurs with probability at least $1 - 1/n^c$ for some constant c .

Theorem 1.1. *There exists a randomized algorithm that, given a set of n points $P \in \mathbb{R}^d$ and a supply function $\mu : P \rightarrow \mathbb{R}$, runs in time $O(n\varepsilon^{-O(d)} \log^{O(d)} n)$ and with high probability returns a transportation map with cost at most $(1 + \varepsilon) \cdot \text{COST}(P, \mu)$.*

At a high level, our algorithm follows the approach laid out by Khesin *et al.* [13] for the bounded spread case. However, removing the running time’s dependency on the spread introduces fundamental and technical issues to nearly every step in their approach.

Let ε_0 be a function of ε and n to be specified later. Taking a cue from prior work on geometric transportation and its specializations [3, 19], Khesin *et al.*’s algorithm begins by building a random sparse graph over $O(n\varepsilon_0^{-O(d)} \log \text{SP}(P))$ vertices including the points in P . In expectation, the shortest path distance between any pair of points in P is maintained up to an $O(\varepsilon_0 \log \text{SP}(P))$ factor, so computing a transportation map is done by setting ε_0 to $O(\varepsilon/(\log \text{SP}(P)))$ and running a minimum cost flow algorithm on the sparse graph.

The graph is constructed by first building a randomly shifted quadtree over P . The quadtree is constructed by surrounding P with an axis-aligned box called a cell, partitioning it into 2^d equal

¹Khesin *et al.* [13] and Agarwal *et al.* [1] present geometric transportation with integer supplies, but their unbounded supply algorithms work without modification when presented with real valued supplies.

sized child cells, and recursively building a quadtree in each child cell; the whole tree has depth $\log \text{SP}(P)$. After building the quadtree, they add ε_0^d Steiner vertices within each cell along with a carefully selected set of edges. While other methods are known for constructing such a sparse graph even without Steiner vertices [6], the hierarchical structure of Khesin *et al.*'s construction is necessary for extracting the transportation map after a minimum cost flow is computed. Observe that not only is the quadtree's size dependent on $\text{SP}(P)$, but so is the number of Steiner vertices added to each cell.

As suggested earlier, the natural approach for reducing the quadtree's size is to remove subtrees containing no members of P and to *compress* the tree by replacing each maximal path of cells with exactly one non-empty child each with a single link to the lowest cell in the path. This approach does result in a quadtree of size $O(n)$, but its depth could also be as large as $\Omega(n)$. This large depth introduces many issues, the worst of which is that standard analyses only guarantee point-to-point distances are maintained up to an $O(\varepsilon_0 n)$ factor. We cannot afford to set ε_0 to ε/n , because the sparse graph would have $\Omega(n^d)$ vertices!

The solution to avoiding such a large increase in expected distances is to use the idea of *moats* around the points as done in the almost-linear time constant factor approximation algorithm of Agarwal *et al.* [1]. In short, we modify the quadtree construction so that, with high probability, all points are sufficiently far away from the boundary of every quadtree cell they appear in. Assuming this condition holds, there are only a limited number of quadtree "levels" at which a pair of points can be separated, and we use this fact to show distances increase by only an $O(\varepsilon_0 \log n)$ factor in expectation. It turns out modifying the quadtree construction correctly is a surprisingly subtle task. Guaranteeing the moats are avoided potentially requires us to perform independent random shifts at several places throughout the quadtree. However, we need to be selective with where the independent shifts occur so that we can successfully analyze the expected distances between points in the sparse graph.

The second stage of Khesin *et al.*'s [13] algorithm solves the minimum cost flow problem in the sparse graph using a framework of Sherman [20]. First, they encode the minimum cost flow problem as finding a flow vector f of minimum cost subject to linear constraints $Af = b$ where A is the vertex-edge incidence matrix and b is a supply vector (not necessarily equal to μ). Sherman's framework involves repeatedly finding flows f of approximately optimal cost that approximately satisfy such constraints. Each iteration of this algorithm requires an application of A and A^T to a pair of vectors, and the number of iterations needed in this approach is polynomial in the *condition number* of A . Unfortunately, A may not be well-conditioned, so Khesin *et al.* describe a *preconditioner* matrix B such that BA has low condition number and is still sparse. They proceed to use Sherman's framework under the equivalent constraints $BAf = Bb$.

One interpretation of Khesin *et al.*'s [13] preconditioner is that it describes a way to charge each Steiner vertex an amount based on the supply of "descendent" vertices below it so that the sum of charges bound the cost of an optimal flow from below. Consequently, both the number of non-zero entries in each column of B and the condition number of B are proportional to the quadtree's depth.

The high depth of our quadtree again appears to cause issues. However, our use of moats implies additional structure to the sparse graph that we can take advantage of. Our preconditioner B is based on essentially the same charging scheme as Khesin *et al.*, but thanks to the moats, we prove the condition number remains proportional to $O(\varepsilon_0^{-1} \log(n/\varepsilon_0))$ instead of the quadtree depth. This charging scheme still results in a precondition B that *is not* sparse, so a naive implementation of Sherman's [20] framework may take quadratic time per iteration. To address this issue, we describe a pair of algorithms based on the hierarchical structure of the graph that let us apply both BA and its transpose in only linear time.

The final stage of the algorithm is the extraction of an approximately minimum cost transportation map from an approximately minimum cost flow in the sparse graph. Khesin *et al.*'s [13]'s original procedure modifies the graph's flow by iteratively reassigning flow to travel directly from input points to each of their many ancestor Steiner vertices or vice versa. We use binary search tree based data structures in a novel way to do flow reassignments in bulk, allowing us to extract the transportation map in time near-linear in the graph size.

Our result relies on a computation model supporting basic combinatoric operations over real numbers including addition, subtraction, multiplication, division, and comparisons. In particular, we do not rely on the floor function.² Our results (and those of Khesin *et al.* [13]) can be extended to work with any L_p metric instead of just Euclidean distance. The rest of the paper proceeds as follows. We describe our sparse graph construction and describe the reduction to minimum cost flow in Section 2. We describe our preconditioner and its use Section 3. Finally, we describe how to extract the approximately optimal transportation map from a flow on the sparse graph in Section 4.

2 Reduction to minimum cost flow in a sparse graph

In this section, we present a way to build a sparse graph $G^* = (V^*, E^*)$ based on P and reduce the transportation problem to finding a minimum cost flow in this sparse graph. Similar to the one presented by Khesin *et al.* [13], our sparse graph G^* is based on a randomly shifted quadtree whose cells have been subdivided into smaller *subcells*. However, the quadtree we use is compressed under certain conditions to guarantee the number of nodes in it is nearly linear in n . We also independently shift subtrees under these compression points so we may guarantee a low expected distortion for point-to-point distance.

2.1 Construction of the sparse graph

Our sparse graph construction begins by building a variant of the compressed quadtree on P we call a **conditionally-compressed quadtree**. Let T^* denote this tree. Let \square_P be the minimum bounding square of P . We fix an $\varepsilon_0 = O(\varepsilon/\log n)$ such that $1/\varepsilon_0$ is a power of 2^{-1} . Suppose the side length of \square_P is Δ^* . Let \square be a square of side length $3\Delta^*$ such that \square_P and \square are concentric. We shift \square by a vector chosen uniformly at random from $[0, \Delta^*)^d$. See Figure 1, left.

Each node of T^* is a square cell in \mathbb{R}^d . Set \square to be the root of T^* . We recursively process each cell C as follows. Suppose C has side length Δ and the subset of P in C is P' . Let $\Delta_{P'}$ be the side length of the minimum bounding square $\square_{P'}$ of P' .

- 1) If $|P'| = 1$, then C is a leaf node.
- 2) If $|P'| > 1$ and $\Delta_{P'} < \frac{\varepsilon_0 \Delta}{3n^8}$, we find the minimum bounding square $\square_{P'}$ of P' . Let $\Delta_{P'}$ be the side length of $\square_{P'}$. We recursively build a conditionally-compressed quadtree over P' with an independently shifted root square \square' with side length $3\Delta_{P'}$ that is concentric to $\square_{P'}$ before the shift. We connect the root of this sub-quadtree to T^* as a child of C .
- 3) If $|P'| > 1$ and $\Delta_{P'} \geq \frac{\varepsilon_0 \Delta}{3n^8}$, we evenly divide C into 2^d squares in \mathbb{R}^d each of side length $\frac{\Delta}{2}$, and make each square that contains at least one point of P' a child cell of C .

²A previous version [8] of this paper relied on a slightly stronger model of computation that allowed one to quickly compute the location of points within arbitrary grids (see Bern *et al.* [4] and Har-Peled [11, Chapter 2]). The current version uses a slightly modified algorithm and more careful analysis to avoid needing this operation.

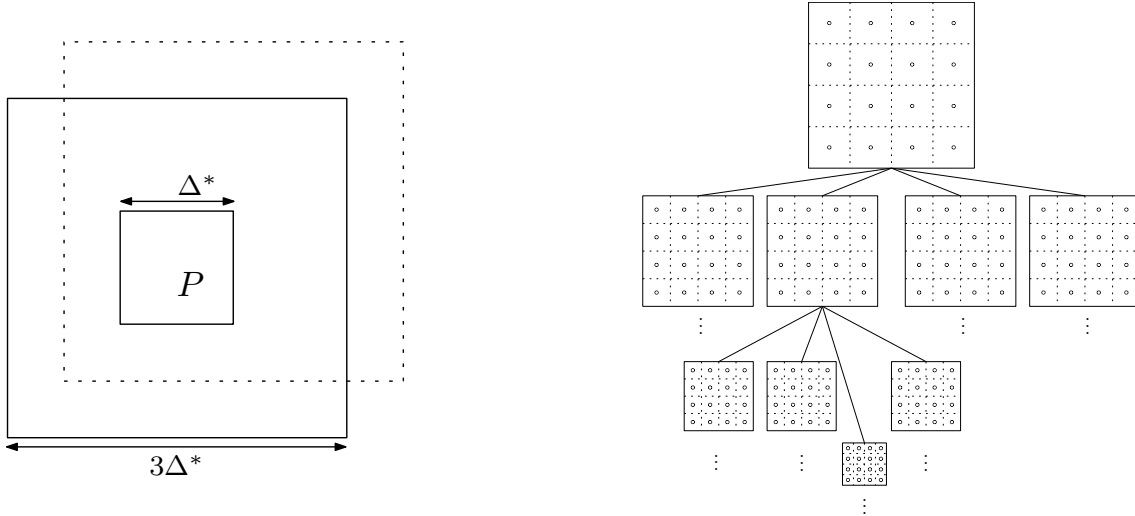


Figure 1. Left: Randomly shifting a box around P . Right: The quadtree cells form a hierarchy. Each cell is partitioned into ε_0^{-d} sub cells, and each subcell has a single net point at its center.

Lemma 2.1. *Let m be an upper bound on the number of nodes in T^* . Conditionally-compressed quadtree T^* can be constructed in $O(m + n \log n)$ time.*

Proof: Suppose we are processing a cell C containing point subset P' . Following standard practice [6], we assume access to d doubly-linked lists containing the points of P' . The points in each list are sorted by distinct choices of one of their d coordinates.

We now describe how to process C . We determine if $|P'| = 1$ in constant time. If so, we stop processing C and discard its data structures. Otherwise, we use the lists to determine $\Delta_{P'}$ and $\square_{P'}$ in $O(1)$ time. If $\Delta_{P'} < \frac{\varepsilon_0 \Delta}{3n^8}$, we pass along the lists for C to the recursive quadtree constructions as Rule 2 describes.

Suppose $\Delta_{P'} \geq \frac{\varepsilon_0 \Delta}{3n^8}$ and Rule 3 applies. We compute the point subsets and their lists going into each child cell by splitting P' one dimension at a time. Specifically, for each dimension, we search the relevant linked list from both ends simultaneously for the position of the split. In doing so, we find the right position to split in time proportional to the less populated side of the split. In the same amount of time, we also perform individual deletions and insertions to make the other linked lists for the points on the less populated side; what remains of the original linked lists are the points going to the more populated side. Eventually, we pass along the lists we construct when computing subtrees for children of C .

We spend $O(\log n)$ time per node in addition to the time spent searching, inserting, and deleting points from lists when applying Rule 3. However, every time a point moves to a new data structure, the number of points in its cell drops by a factor of at least 2. We spend $O(m + n \log n)$ time total implementing Rule 3. \square

We define a type of sub-quadtree of T^* . A *simple sub-quadtree* is a sub-quadtree consisting of a cell C that either is the root of T^* or is randomly shifted independently of its parent along with a maximal set of descendent cells of C that were *not* shifted independently of C (i.e., Rule 2 was never applied to create descendent cells of C in the sub-quadtree).

For every cell C in T^* , we perform a secondary subdivision on C . Let Δ_C denote the side length of C . We divide C into ε_0^{-d} square regions with equal side length $\varepsilon_0 \Delta_C$. If a sub-region of C contains a point $p \in P$, we say it is a subcell \tilde{C} of C and we use C^+ to denote the set of subcells of C . Again, see Figure 1.

Utilizing an idea of Agarwal *et al.* [1], we define the *moat of size h* around a point p as an axis-parallel square of side length h around p . Consider a randomly shifted grid with cells of side length Δ . The probability of any of the grid lines hitting a moat of size $\frac{2\Delta}{n^4}$ around any points $p \in P$ is at most $\frac{2\Delta}{n^4} \cdot n \cdot \frac{d}{\Delta} = O(\frac{1}{n^3})$.

Lemma 2.2. *With probability at least $1 - O((1/n) \log(n/\varepsilon_0))$, the conditionally-compressed quadtree T^* has the following properties:*

1. *The total number of cells is $O(n \log(n/\varepsilon_0))$.*
2. *Suppose cell C with side length Δ_C contains $p \in P$ and let \tilde{C} be the subcell of C that contains p . Then p is at least $\frac{\Delta_C}{n^4}$ distance away from any side of C and is at least $\frac{\Delta_C \varepsilon_0}{n^4}$ distance away from any side of \tilde{C} . In other words, the moats of p with respect to the uniform grids containing C and \tilde{C} as cells do not touch the grid lines.*
3. *Let T' be any simple sub-quadtree of T^* , and let C' be a child cell of some leaf C of T' . Cell C' lies entirely within a subcell of C .*

Proof: The condition to trigger Rule 2 guarantees every path of descendent cells with one child each has length $O(\log(n/\varepsilon_0))$. We get Property 1.

Let T_0 be the simple sub-quadtree containing the root cell of T^* . With Property 1, the number of cells in T_0 is $O(n \log(n/\varepsilon_0))$ and hence the smallest cell in T_0 lies $O(n \log(n/\varepsilon_0))$ quadtree levels down. Therefore, at most $O(n \log(n/\varepsilon_0))$ shifted grids in \mathbb{R}^d determine the boundaries of T_0 's (sub)cells. We see Property 2 is violated for at least one cell in T_0 with probability at most $c \cdot \frac{n}{n^3} \cdot \log \frac{n}{\varepsilon_0}$ for some constant c .

Assume from here on that Property 2 holds for all cells in T_0 . Let C be any leaf cell with side length Δ_C of T_0 such that C has some descendent cell in T^* . Let C' be a child cell with side length $\Delta_{C'}$ of C and T' be the simple sub-quadtree rooted at C' . Then we have $\Delta_{C'} < \frac{\varepsilon_0 \Delta_C}{n^8}$ by Rule 2. Moreover, all points in T' are at distance at least $\frac{\varepsilon_0 \Delta_C}{n^4}$ away from subcell boundaries of C . Any random shift of T' using a shift vector in $[0, \frac{\Delta_{C'}}{3}]^d$ will keep all points in T' inside the same subcell of C .

Let $\{T_1, T_2, \dots\}$ denote the distinct sub-quadtrees, each consisting of a child of a leaf in T_0 and all the child's descendents in T^* . For each T_i , let n_i be the number of points over which T_i is built. We have $n_i \leq n - 1$ for each i and $\sum_i n_i \leq n$. Inductively assume Properties 2 through 3 fail to hold for T_i with probability at most $c \cdot \frac{n_i^2}{n^3} \cdot \log \frac{n}{\varepsilon_0} \leq c \cdot \frac{(n-1)n_i}{n^3} \cdot \log \frac{n}{\varepsilon_0}$. (For each T_i , on the path from its root to each of its leaves in T^* , at most n_i distinct simple sub-quadtrees exists. Each of those sub-quadtrees has at most $n_i \log(n/\varepsilon_0)$ levels. Summing them all, we may assume Properties 2 through 3 fail to hold for T_i with probability at most $c \cdot \frac{n_i^2}{n^3} \cdot \log \frac{n}{\varepsilon_0}$.) Taking a union bound, the probability of Properties 2 through 3 failing to hold for either T_0 or any T_i is at most $c \cdot \frac{n^2}{n^3} \cdot \log \frac{n}{\varepsilon_0} = c \cdot \frac{1}{n} \cdot \log \frac{n}{\varepsilon_0}$. \square

Lemma 2.3. *Conditioned on Property 2 of Lemma 2.2, the expected distance between any pair $p, q \in P$ in G^* is at most $(1 + O(\varepsilon_0 \log n)) \|p - q\|_2$.*

Proof: Let $\lambda = \|p - q\|_2$, and let $\text{dist}_{G^*}(p, q)$ be the distance between p and q in G^* . Points p and q must be connected through the net points of some cell containing both of them. Let $C(p, q)$ be the lowest common ancestor cell of p and q . Let $N_{C(p, q)}(p)$ and $N_{C(p, q)}(q)$ be the net points of the subcell of $C(p, q)$ that contains p and q , respectively. Then $\text{dist}_{G^*}(p, q) = \text{dist}_{G^*}(p, N_{C(p, q)}(p)) + \text{dist}_{G^*}(N_{C(p, q)}(p), N_{C(p, q)}(q)) + \text{dist}_{G^*}(q, N_{C(p, q)}(q))$. Value $\text{dist}_{G^*}(p, N_{C(p, q)}(p))$ is the distance from

$N_{C(p,q)}(p)$ to p through its descendent net points. We have $\sum_{i \geq 1} 2^{-i} \sqrt{d} \varepsilon_0 \Delta_{C(p,q)} \leq \sqrt{d} \varepsilon_0 \Delta_{C(p,q)}$, because subcell side lengths at least halve every level down in T^* . Similarly, $\text{dist}_{G^*}(q, N_{C(p,q)}(q)) \leq \sqrt{d} \varepsilon_0 \Delta_{C(p,q)}$. By the triangle inequality, $\text{dist}_{G^*}(N_{C(p,q)}(p), N_{C(p,q)}(q)) \leq \|p - q\|_2 + \|p - N_{C(p,q)}(p)\|_2 + \|q - N_{C(p,q)}(q)\|_2 \leq \|p - q\|_2 + \sqrt{d} \varepsilon_0 \Delta_{C(p,q)}$. Then we have $\text{dist}_{G^*}(p, q) \leq \|p - q\|_2 + 3\sqrt{d} \varepsilon_0 \Delta_{C(p,q)}$.

We define the *extra cost* to be $\Phi_{p,q} = \text{dist}_{G^*}(p, q) - \|p - q\|_2$. We have $\Phi_{p,q} \leq 3\sqrt{d} \varepsilon_0 \Delta_{C(p,q)}$. Let e be the event that Property 2 of Lemma 2.2 holds. We have the conditional expectation of the extra cost $\mathbb{E}(\Phi_{p,q} | e) \leq \mathbb{E}(3\sqrt{d} \varepsilon_0 \Delta_{C(p,q)} | e) \leq 3\sqrt{d} \varepsilon_0 \mathbb{E}(\Delta_{C(p,q)} | e)$.

Let Δ be any value in $[\frac{\lambda}{\sqrt{d}}, n^4 \lambda]$. Our current goal is to bound the probability that $\Delta_{C(p,q)} \in [\Delta, 2\Delta)$ conditioned on Property 2 of Lemma 2.2. As we'll see later, we need not consider other possible values of Δ given we condition on that property. To find our probability bound, we first bound the probability conditioned on a weaker event. Let C' denote the *highest* common ancestor cell of p, q in T^* with side length at most $n^8 \lambda$. Let Δ' be its side length and let P' be the point set in C' . Let T be the simple sub-quadtrees in which C' lies. In each dimension, the shift of T in that dimension that determines C' constitutes a continuous unique interval. Let ℓ_j denote the length of this interval for the j th dimension. For our proof, we must consider different shifts of T that lead to our current choice of C' , and argue $\mathbb{E}(\Delta_{C(p,q)})$ is not too large. To do so, however, we must condition on the following event, denoted e' , that there exists a shift of T such that no point of P' lies at distance less than $\frac{\Delta'}{n^4}$ from the sides of C' . Note that e' is necessary *but not sufficient* for Property 2 of Lemma 2.2 to hold. We now consider the probability of $\Delta_{C(p,q)} \in [\Delta, 2\Delta)$ conditioned on e' , denoted by $\mathbb{P}[\Delta_{C(p,q)} \in [\Delta, 2\Delta) | e']$.

Let $\Delta'' = 2^{-i} \cdot \Delta'$ for some $i \in \mathbb{N}$ such that $\Delta'' \in [\Delta, 2\Delta)$. Conditioned on a particular choice of C' , each $(d-1)$ -dimensional hyperplane perpendicular to the j -th axis in the grid of size $\frac{\Delta''}{2}$ aligned with the shift of T has a j -coordinate somewhere on an interval of length ℓ_j . Therefore, there are at most $\lceil \frac{2\ell_j}{\Delta''} \rceil$ $(d-1)$ -dimensional hyperplanes possibly shifted to lie between p and q . Summing over all d choices of j , the probability that p and q are separated by some $(d-1)$ -dimensional hyperplane in the grid of size $\frac{\Delta''}{2}$ is at most $\sum_{j=1}^d \lceil \frac{2\ell_j}{\Delta''} \rceil \cdot \lambda \cdot \frac{1}{\ell_j}$ conditioned on e' . Therefore, $\mathbb{P}[\Delta_{C(p,q)} \in [\Delta, 2\Delta) | e'] \leq \sum_{j=1}^d \lceil \frac{2\ell_j}{\Delta''} \rceil \cdot \lambda \cdot \frac{1}{\ell_j}$.

Suppose $\Delta' \in (\frac{n^8 \lambda}{2}, n^8 \lambda]$. Recall e' requires some shift of T such that each point of P' lies distance $\frac{\Delta'}{n^4}$ or greater from each side of C' . Therefore, $\ell_j \geq \frac{2\Delta'}{n^4} > n^4 \lambda$ for each j . Further,

$$\begin{aligned} \mathbb{P}[\Delta_{C(p,q)} \in [\Delta, 2\Delta) | e'] &\leq \sum_{j=1}^d \lceil \frac{2\ell_j}{\Delta''} \rceil \cdot \lambda \cdot \frac{1}{\ell_j} \\ &\leq \sum_{j=1}^d \frac{2\ell_j + \Delta''}{\Delta''} \cdot \lambda \cdot \frac{1}{\ell_j} \\ &< O\left(\frac{\lambda}{\Delta''} + \frac{1}{n^4}\right). \end{aligned}$$

Because $\Delta \in [\frac{\lambda}{\sqrt{d}}, n^4 \lambda]$, we see $\Delta'' < 2n^4 \lambda$, implying $\frac{\lambda}{\Delta''} > \frac{2}{n^4}$. We have $\mathbb{P}[\Delta_{C(p,q)} \in [\Delta, 2\Delta) | e'] < O(\frac{\lambda}{\Delta})$.

If $\Delta' \leq \frac{n^8 \lambda}{2}$, C' must be the root cell of the sub-quadtrees T . In this case, all shifts of T will

satisfy e' assuming n is sufficiently large. We have $\ell_j = \frac{\Delta'}{3}$ and $\lceil \frac{2\ell_j}{\Delta''} \rceil \leq \frac{3\ell_j}{2\Delta''}$ for all j .

$$\begin{aligned} \mathbb{P}[\Delta_{C(p,q)} \in [\Delta, 2\Delta] \mid e'] &\leq \sum_{j=1}^d \lceil \frac{2\ell_j}{\Delta''} \rceil \cdot \lambda \cdot \frac{1}{\ell_j} \\ &= \sum_{j=1}^d \frac{3\lambda}{2\Delta''} \\ &= O\left(\frac{\lambda}{\Delta}\right). \end{aligned}$$

Either way, $\mathbb{P}[\Delta_{C(p,q)} \in [\Delta, 2\Delta] \mid e'] \leq O\left(\frac{\lambda}{\Delta}\right)$.

Again, e immediately implies e' , but the converse is not necessarily true. Therefore,

$$\begin{aligned} \mathbb{P}[\Delta_{C(p,q)} \in [\Delta, 2\Delta] \mid e] &= \frac{\mathbb{P}[\Delta_{C(p,q)} \in [\Delta, 2\Delta] \wedge e]}{\mathbb{P}[e]} \\ &\leq \frac{\mathbb{P}[\Delta_{C(p,q)} \in [\Delta, 2\Delta] \wedge e']}{\mathbb{P}[e]} \\ &= \frac{\mathbb{P}[\Delta_{C(p,q)} \in [\Delta, 2\Delta] \mid e'] \cdot \mathbb{P}[e']}{\mathbb{P}[e]} \\ &\leq \frac{\mathbb{P}[\Delta_{C(p,q)} \in [\Delta, 2\Delta] \mid e']}{\mathbb{P}[e]} \\ &\leq O\left(\frac{\lambda}{\Delta}\right) \cdot \frac{1}{1 - O(n/\log n)} \\ &= O\left(\frac{\lambda}{\Delta}\right). \end{aligned}$$

Assuming Property 2 of Lemma 2.2, the value of $\Delta_{C(p,q)}$ is in $[\frac{\lambda}{\sqrt{d}}, n^4\lambda]$; smaller cells cannot fit both points, and larger cells cannot separate them without one or both points lying too close to a cell side. Let \lg denote the logarithm of base 2. We have

$$\begin{aligned} \mathbb{E}(\Delta_{C(p,q)} \mid e) &\leq \sum_{\Delta=2^i \cdot \frac{\lambda}{\sqrt{d}}, 0 \leq i \leq \lg(n^4\sqrt{d}), i \in \mathbb{N}} \mathbb{P}[\Delta_{C(p,q)} \in [\Delta, 2\Delta] \mid e] \cdot 2\Delta \\ &< \sum_{\Delta=2^i \cdot \frac{\lambda}{\sqrt{d}}, 0 \leq i \leq \lg(n^4\sqrt{d}), i \in \mathbb{N}} O\left(\frac{\lambda}{\Delta}\right) \cdot 2\Delta \\ &\leq O(\log n) \cdot \lambda. \end{aligned}$$

We conclude

$$\begin{aligned} \mathbb{E}(\text{dist}_{G^*}(p, q) \mid e) &= \|p - q\|_2 + \mathbb{E}(\Phi_{p,q} \mid e) \\ &\leq \|p - q\|_2 + 3\sqrt{d}\varepsilon_0 \mathbb{E}(\Delta_{C(p,q)} \mid e) \\ &\leq (1 + O(\varepsilon_0 \log n)) \cdot \|p - q\|_2. \end{aligned}$$

□

2.2 Reduction to minimum cost flow

Having built our sparse graph, we now reduce to a minimum cost flow problem in G^* . We model the minimum cost flow problem as follows to simplify later discussions.

Let $G = (V, E)$ be an arbitrary undirected graph with $V \in \mathbb{R}^d$. Let \vec{E} be the set of edges in E oriented arbitrarily. We call $f \in \mathbb{R}^{\vec{E}}$ a **flow vector** or more simply, a **flow**. Let A be a $|V| \times |\vec{E}|$ **vertex-edge incidence matrix** where $\forall (u, (v, w)) \in V \times \vec{E}$, $A_{u, (v, w)} = 1$ if $u = v$, $A_{u, (v, w)} = -1$ if $u = w$, and $A_{u, (v, w)} = 0$ otherwise. Given f , we define the **divergence** of a vertex v as $(Af)_v = \sum_{(v, w)} f_{(v, w)} - \sum_{(u, v)} f_{(u, v)}$. For simplicity of exposition, we may sometimes refer to $f_{(v, u)}$ even though $(u, v) \in \vec{E}$. In such cases, it is assumed $f_{(v, u)} = -f_{(u, v)}$.

Let $\|\cdot\|_{\vec{E}}$ be a norm on $\mathbb{R}^{\vec{E}}$ such that $\|f\|_{\vec{E}} = \sum_{(u, v) \in \vec{E}} |f_{(u, v)}| \cdot \|v - u\|_2$. Let $b \in \mathbb{R}^V$ denote a set of divergences for all $v \in V$. We define an instance of **uncapacitated minimum cost flow** as the pair (G, b) . We seek a flow vector f minimizing $\|f\|_{\vec{E}}$ subject to $Af = b$.

In particular, set $b^* \in \mathbb{R}^V$ such that $b_p^* = \mu(p)$, $\forall p \in P$ and $b_v^* = 0$, $\forall v \in V \setminus P$. Ultimately, we will find an approximate solution to the instance (G^*, b^*) . Let $\text{COST}(G^*, b^*) := \|f^*\|_{\vec{E}}$ for some optimal solution f^* of this instance. From construction of G^* , $\text{COST}(P, \mu) \leq \text{COST}(G^*, b^*)$. With high probability, the conditions of Lemma 2.2 hold true, and by Lemma 2.3, $\mathbb{E}(\text{COST}(G^*, b^*)) \leq (1 + O(\varepsilon_0 \log n)) \text{COST}(P, \mu)$. In particular, $\mathbb{E}(\text{COST}(G^*, b^*) - \text{COST}(P, \mu)) \leq O(\varepsilon_0 \log n) \text{COST}(P, \mu)$. We can guarantee that expected bound holds with high probability as well by doubling the constant in the big-Oh and taking the best result from $O(\log n)$ runs of our algorithm. From here on, we assume both that the conditions Lemma 2.2 hold and $\text{COST}(G^*, b^*) \leq (1 + O(\varepsilon_0 \log n)) \text{COST}(P, \mu)$.

2.3 Decomposition into simpler subproblems

In the sequel, we apply Sherman's generalized preconditioning framework [13, 20] to find an approximate solution to the minimum cost flow instance (G^*, b^*) . For technical reasons, however, we cannot afford to run the framework on the entire sparse graph G^* at once.

Here, we reduce finding an approximately optimal flow for minimum cost flow instance (G^*, b^*) to finding $O(n)$ approximately optimal flows, each within an induced subgraph defined by the net points within a simple sub-quadtrees.

Recall, for each point $p \in P$, $\tilde{C}(p)$ denotes the smallest subcell containing p , and $N_{\tilde{C}}$ denotes the net point of subcell \tilde{C} . Let f be the flow such that $f_{(p, N_{\tilde{C}(p)})} = b_p^*$ for all $p \in P$. Let $G' = (V', E')$ and A' be the restriction of G^* and its vertex-edge incidence matrix A after removing all vertices $p \in P$. Let b' be the restriction of $b^* - Af$ to vertices of G' . Every vertex $p \in P$ of G^* has exactly one incident edge, so an optimal solution to our original minimum cost flow instance consists of f along with an optimal solution to the instance defined on A' and b' . From here on, we focus on finding an approximately minimum cost flow in G' .

Suppose there are multiple simple sub-quadtrees. Let $G_0 = (V_0, E_0)$ be the subgraph induced by the m net point vertices of a simple sub-quadtrees with no descendent sub-quadtrees. Let C be the root cell of the simple sub-quadtrees for G_0 , let u be a net point for an arbitrary subcell of C , and let v be the parent net point of u in G' where C' is the subcell with v as its net point. In $O(m)$ time, we compute $B = \sum_{w \in V_0} b_w'$, the total divergence of vertices within G_0 . We then let f' be the flow in G' that is 0 everywhere except for $f'_{(u, v)} := B$. Finally, let $b'' = b' - A'f'$.

Notice that at least B units of flow in G_0 need to leave or enter C by edges of side length at least $\Delta_{\tilde{C}}$. Given $\Delta_C \leq O(1/n^8) \Delta_{\tilde{C}}$, we can lazily assign the flow between net points of C and v , increasing the cost by at most $2\sqrt{d} \Delta_C B \leq O(1/n^8) \Delta_{\tilde{C}} B$. We have the following lemma.

Lemma 2.4. *There exists a flow f'' in G' such that $f''_{(w,x)} = 0$ for all $w \in V_0, x \notin V_0$; $Af'' = b''$; and $\|f'' + f'\|_{\vec{E}'} \leq (1 + O(1/n^8)) \cdot \text{COST}(G', b')$.*

Proof: Let \tilde{C} be the subcell for which v is a net point. Let $\Delta_{\tilde{C}}$ be the side length of \tilde{C} . By construction of G' , at least B units of flow must travel to or from vertex v from G_0 at a cost of $\Delta_{\tilde{C}}$. Specifically, G_0 is totally inside \tilde{C} , v is the only vertex in \tilde{C} incident to some edge crossing the side of \tilde{C} , and the nearest vertex $x \notin V_0$ is at least $\Delta_{\tilde{C}}$ far from v . So $\text{COST}(G', b') \geq \Delta_{\tilde{C}}B$.

Suppose f'^* is a flow in G' with cost $\text{COST}(G', b')$. Let N_C be the set of net points of subcells of C . We may assume there is no pair $y, z \in N_C$ such that $f'_{(y,v)} > 0$ and $f'_{(v,z)} > 0$, because we could send the flow directly between y and z more cheaply. We create flow f''' as follows starting with $f'' = f'^*$. While there exists some vertex $u' \in N_C \setminus \{u\}$ with $f'''_{(u',v)} \neq 0$, let $\delta = f'''_{(u',v)}$. We divert flow by setting $f'''_{(u,v)} \leftarrow f'''_{(u,v)} + \delta$, $f'''_{(u',u)} \leftarrow f'''_{(u',u)} + \delta$, and $f'''_{(u',v)} \leftarrow 0$. This increases the cost by at most twice of the length of the diagonal of C per diverted unit of flow. Overall, we divert at most B units. The total cost increase is at most $2\sqrt{d}\Delta_C B \leq O(1/n^8)\text{COST}(G', b')$ where Δ_C is side length of C , because $\Delta_C \leq O(1/n^8)\Delta_{\tilde{C}}$. We have $\|f'''\|_{\vec{E}} \leq (1 + O(1/n^8)) \cdot \text{COST}(G', b')$. Finally, let $f'' = f''' - f'$. \square

The above lemma implies we can use the following strategy for approximating a minimum cost flow in G' : Let b_0 be the restriction of b'' to V_0 . We find a flow in G_0 with divergences b_0 of cost at most $(1 + O(\varepsilon)) \cdot \text{COST}(G_0, b_0)$ using the algorithm described in the next section. Then, we recursively apply our algorithm on $G'' = (V'', E'')$, the induced subgraph over $V'' = V' \setminus V_0$. The depth of recursion is $O(n)$, so the total cost from combining our separately computed flows is $(1 + O(\varepsilon))(1 + O(1/n^7)) \cdot \text{COST}(G', b') = (1 + O(\varepsilon))\text{COST}(G', b')$. We must emphasize that simple sub-quadtrees may still have linear depth, so we still need to apply our own techniques to make Sherman's framework run within the desired time bounds.

3 Approximating the minimum cost flow

Let $G = (V, E)$ be an induced subgraph of sparse graph G^* where V is the subset of net points for one simple sub-quadtrees T as defined above. Let $m = |E|$, and let A be the vertex-edge incidence matrix for G . We now describe the ingredients we need to provide to efficiently approximate the minimum cost flow problem in G using Sherman's generalized preconditioning framework [13, 20]. We then provide those ingredients one-by-one to achieve a near-linear time $(1 + O(\varepsilon))$ -approximate solution for the minimum cost flow instance.

3.1 The preconditioning framework

Consider an instance of the minimum cost flow problem in G with an arbitrary divergence vector $\tilde{b} \in \mathbb{R}^V$, and let $f_{\tilde{b}}^* := \operatorname{argmin}_{f \in \mathbb{R}^{\vec{E}}, Af = \tilde{b}} \|f\|_{\vec{E}}$. A flow vector $f \in \mathbb{R}^{\vec{E}}$ is an (α, β) **solution** to the problem if

$$\begin{aligned} \|f\|_{\vec{E}} &\leq \alpha \|f_{\tilde{b}}^*\|_{\vec{E}} \\ \|Af - \tilde{b}\|_1 &\leq \beta \|A\| \|f_{\tilde{b}}^*\|_{\vec{E}} \end{aligned}$$

where $\|A\|$ is the norm of the linear map represented by A . An algorithm yielding an (α, β) -solution is called an (α, β) -**solver**.

By arguments in [13], we seek a preconditioner $B \in \mathbb{R}^{V \times V}$ of full column rank such that, for any $\tilde{b} \in \mathbb{R}^V$ with $\sum_{v \in V} \tilde{b}_v = 0$, it satisfies

$$\|B\tilde{b}\|_1 \leq \min\{\|f\|_{\vec{E}} : f \in \mathbb{R}^{\vec{E}}, Af = \tilde{b}\} \leq \kappa \|B\tilde{b}\|_1 \quad (1)$$

for some sufficiently small function κ of n , ε , and d .

Let M be the time it takes to multiply BA and $(BA)^T$ by a vector. Then there exists a $(1 + \varepsilon, \beta)$ -solver for any $\varepsilon, \beta > 0$ for this problem with running time bounded by $O(\kappa^2(|V| + |\vec{E}| + M) \log |\vec{E}|(\varepsilon^{-2} + \log \beta^{-1}))$ [20]. Moreover, if a feasible flow $f \in \mathbb{R}^{\vec{E}}$ with cost $\|f\|_{\vec{E}} \leq \kappa B\tilde{b}$ can be found in time K , there is a $(\kappa, 0)$ -solver with running time K . By setting $\beta = \varepsilon \kappa^{-2}$ [13], the composition of these two solvers is a $(1 + 2\varepsilon, 0)$ -solver with running time bounded by

$$O(\kappa^2(|V| + |\vec{E}| + M) \log |\vec{E}|(\varepsilon^{-2} + \log \kappa) + K).$$

3.2 Preconditioning the minimum cost flow

We present a way to construct such a preconditioner B similar to the one of Khesin *et al.* [13] that guarantees κ in (1) is sufficiently small for our performance objective. Our algorithm *does not* compute B directly, because B is not sparse. However, the time for individual applications of BA or $(BA)^T$ is $O(|V| + |\vec{E}|)$.

Let $\tilde{\mathcal{C}}$ denote the set of all subcells defining the net points of G . For any subcell $\tilde{C} \in \tilde{\mathcal{C}}$, let $N_{\tilde{C}}$ denote its net point and let $\Delta_{\tilde{C}}$ denote its side length.

Let B be a matrix indexed by $(u, v) \in V \times V$ such that, for every net point ν in V where ν is the net point of some subcell \tilde{C} , we set $B_{\nu, v} = \frac{\Delta_{\tilde{C}}}{\Lambda}$ for all descendent net points v of ν , where $\Lambda = 22 \lg(\frac{n}{\varepsilon_0})$. $B_{\nu, v} = 0$ for all other v . Matrix B has full column rank, because each column specifies exactly which ancestor net points each vertex has in G .

Now, fix any $\tilde{b} \in \mathbb{R}^V$ such that $\sum_{v \in V} \tilde{b}_v = 0$. Observe,

$$\|B\tilde{b}\|_1 = \sum_{\tilde{C} \in \tilde{\mathcal{C}}} \frac{\Delta_{\tilde{C}}}{\Lambda} \left| \sum_{v \in \tilde{C}} \tilde{b}_v \right|. \quad (2)$$

Lemma 3.1. *We have $\|B\tilde{b}\|_1 \leq \min\{\|f\|_{\vec{E}} : f \in \mathbb{R}^{\vec{E}}, Af = \tilde{b}\}$.*

Proof: Let $f_b^* := \operatorname{argmin}_{f \in \mathbb{R}^{\vec{E}}, Af = \tilde{b}} \|f\|_{\vec{E}}$. We arbitrarily decompose f_b^* into a set of flows $F = \{f^1, f^2, \dots\}$ with the following properties: 1) each flow follows a simple path between two vertices u and v ; 2) for each flow $f^i \in F$ and edge $(u, v) \in \vec{E}$ either $f^i(u, v) = 0$ or its sign is equal to the sign of $f_b^*(u, v)$; 3) for each flow $f^i \in F$ and vertex v , either $(Af^i)_v = 0$ or its sign is equal to \tilde{b}_v ; and 4) for each edge $(u, v) \in \vec{E}$, we have $f_b^*(u, v) = \sum_{f^i \in F} f^i(u, v)$. The existence of such a decomposition is a standard part of network flow theory and one can be computed in a simple greedy manner (however, our algorithm does not actually need to compute one). From construction, we have $\sum_{f^i \in F} \|f^i\|_{\vec{E}} = \|f_b^*\|_{\vec{E}}$. We describe a way to charge summands of $\sum_{\tilde{C} \in \tilde{\mathcal{C}}} \Delta_{\tilde{C}} \left| \sum_{v \in \tilde{C}} \tilde{b}_v \right|$ to the summands of $\sum_{f^i \in F} \|f^i\|_{\vec{E}}$. Our charges will cover each of the former and exceed each of the latter by at most a Λ factor. Consider a subcell \tilde{C} . For each vertex $u \in \tilde{C}$, for each flow f^i sending flow to or from u , we charge $\Delta_{\tilde{C}} |(Af^i)_u|$. Clearly, we charge at least $\Delta_{\tilde{C}} \left| \sum_{v \in \tilde{C}} \tilde{b}_v \right|$ for each subcell \tilde{C} .

It remains to prove we did not overcharge by too large a factor. Consider an arbitrary flow $f^i \in F$ sending flow from some vertex u to some vertex v . Let $C(u, v)$ be the lowest common ancestor cell containing u and v . Let $\Delta_{C(u, v)}$ be its side length, and let $C(\hat{u}, v)$ be the child cell of $C(u, v)$ that includes u . Let Δ be the side length of $C(\hat{u}, v)$.

Suppose there exists a descendant cell C' of $C(\hat{u}, v)$ containing u that is at least $5 \lg n$ levels down from $C(\hat{u}, v)$. Its side length $\Delta_{C'}$ is at most $\frac{\Delta}{n^5}$. By Property 2 of Lemma 2.2, u is at least $\frac{\Delta}{n^4}$ distance away from any side of $C(\hat{u}, v)$ and therefore v as well. Therefore, we charge at most an $\frac{\varepsilon_0}{n}$ fraction of $\|f^i\|_{\bar{E}}$ to cover u 's subcell in C' . The amounts charged by similar subcells of smaller side length containing u form a decreasing geometric series evaluating to at most that value, so all these small subcells charge at most a $\frac{2\varepsilon_0}{n}$ fraction total.

Now, consider the cells with larger side length. Suppose there exists an ancestor cell C'' of $C(\hat{u}, v)$ at least $\lg \varepsilon_0^{-1} + 1$ levels up from $C(\hat{u}, v)$, and let \tilde{C}'' be the subcell of C'' containing u . Then the side length of \tilde{C}'' is at least $\Delta_{C(u,v)}$ and all points in $C(u, v)$ will be included in \tilde{C}'' also. Therefore, we do not charge to $\|f^i\|_{\bar{E}}$ for subcell \tilde{C}'' , and there are at most $5 \lg n + \lg \varepsilon_0^{-1} \leq 5 \lg \frac{n}{\varepsilon_0}$ subcells in addition to those handled above for which we do charge to $\|f^i\|_{\bar{E}}$. Consider any such subcell \tilde{C} . The path carrying f^i leaves \tilde{C} through an edge of length at least $\Delta_{\tilde{C}}/2$, so we charge at most $2 \cdot \|f^i\|_{\bar{E}}$ to cover \tilde{C} . Summing over all $5 \lg \frac{n}{\varepsilon_0}$ choices of \tilde{C} and accounting for the tiny cells as discussed above, we charge at most $(10 \lg \frac{n}{\varepsilon_0} + 2\varepsilon_0/n) \|f^i\|_{\bar{E}} \leq 11 \lg(\frac{n}{\varepsilon_0}) \cdot \|f^i\|_{\bar{E}}$ to cover subcells containing u . We also charge to $\|f^i\|_{\bar{E}}$ to cover subcells containing v , so we overcharge by a factor of at most $22 \lg(\frac{n}{\varepsilon_0}) = \Lambda$. The lemma follows. \square

Lemma 3.2. *We have $\min\{\|f\|_{\bar{E}} : f \in \mathbb{R}^{\bar{E}}, Af = \tilde{b}\} \leq \kappa \|B\tilde{b}\|_1$ for some $\kappa = O(\varepsilon_0^{-1} \log(n/\varepsilon_0))$. Moreover, a flow vector f satisfying $Af = \tilde{b}$ of cost at most $\kappa \|B\tilde{b}\|_1$ can be computed in $O(m)$ time.*

Proof: We describe a greedy algorithm based on one by Khesin *et al.* [13] to iteratively construct a feasible flow f satisfying $Af = \tilde{b}$ with a cost $\|f\|_{\bar{E}} \leq \kappa B\tilde{b}$ in $O(m)$ time. At any point during f 's construction, we say the **surplus** of vertex $u \in V$ is $\pi(u, f) = (Af)_u - \tilde{b}_u$, the difference between the current and desired divergences of u .

1. For every cell C in a postorder traversal of G 's simple sub-quadtrees, for every subcell \tilde{C} of C , we do the following. Let $\nu = N_{\tilde{C}}$. We choose any two child net points v, w of ν such that $\pi(v, f) > 0 > \pi(w, f)$. We then add $\min\{|\pi(v, f)|, |\pi(w, f)|\}$ to $f_{(v,w)}$. In doing so, we make the surplus of at least one child net point of ν equal to 0, and we decrease the absolute values of surpluses of both v and w . Therefore, after at most a number of steps equal to the number of child net points of ν , either all child net points have non-negative surplus or all child net points have non-positive surplus. Finally, for each vertex v among child net points with non-zero surplus, we set $f_{(\nu,v)} = \pi(v, f)$. Afterward, every child net point of ν has surplus 0. In other words, the unbalance among those child net points is collected into ν . Each net point ν has at most 2^d child net points. Therefore, the total running time for this step is $O(m)$.
2. After performing step 1), all net points with parents have a surplus of 0. We pick up any two net points u, v of subcells of T 's root cell with two different surplus signs as described in step 2 and add $\min\{|\pi(u, f)|, |\pi(v, f)|\}$ to $f_{(v,u)}$. After $O(\varepsilon_0^{-d}) = O(m)$ steps, all points $v \in V$ will have surplus 0, and f is a feasible flow satisfying $Af = \tilde{b}$.

We now analyze $\|f\|_{\bar{E}}$. Consider a subcell \tilde{C} of some cell C with net point ν . Flow does not leave or enter \tilde{C} until we move flow between ν and either another net point in C or ν 's parent net point. Therefore, $\pi(\nu, f) = -\sum_{v \in \tilde{C}} \tilde{b}_v$ immediately after moving flow from ν 's children to ν in step 1) above. All subsequent steps moving flow to or from ν involve an edge of length at most $\varepsilon_0^{-1} \sqrt{d} \Delta_{\tilde{C}}$ and only serve to reduce $|\pi(\nu, f)|$.

Summing over all subcells, we get

$$\|f\|_{\vec{E}} \leq \sum_{\tilde{C} \in \tilde{\mathcal{C}}} \varepsilon_0^{-1} \sqrt{d} \Delta_{\tilde{C}} \left| \sum_{v \in \tilde{C}} \tilde{b}_v \right| \leq \varepsilon_0^{-1} \sqrt{d} \Lambda \|B\tilde{b}\|_1.$$

Therefore, $\|f_b^*\|_{\vec{E}} \leq \|f\|_{\vec{E}} \leq \kappa \|B\tilde{b}\|_1$, where $\kappa = O(\varepsilon_0^{-1} \log(n/\varepsilon_0))$. \square

Lemma 3.3. *Applications of BA and $(BA)^T$ to arbitrary vectors $f \in \mathbb{R}^{\vec{E}}$ and $\tilde{b} \in \mathbb{R}^V$, respectively, can be done in $O(m)$ time.*

Proof: Both applications can be performed using dynamic programming algorithms.

Computing BAf Let $A' = Af$. Recall, $\forall v \in V$, A'_v is the divergence of v given flow f . Matrix A has m non-zero entries, so A' can be computed in $O(m)$ time.

We compute BAf by computing BA' . Let ν be any net point of G , and let \tilde{C} be its subcell. From the definition of B , we have $(BA')_\nu = \frac{\Delta_{\tilde{C}}}{\Lambda} \sum_{v \in \tilde{C}} A'_v$. Now, let \tilde{C}^+ be the (possibly empty) set of all child subcells of \tilde{C} with net points in G . We have $\sum_{v \in \tilde{C}} A'_v = A'_\nu + \sum_{\tilde{C}' \in \tilde{C}^+} \sum_{v \in \tilde{C}'} A'_v$. Thus, we can use dynamic programming to compute BA' in $O(m)$ time. Each entry is filled in during a postorder traversal of the quadtree cells.

Computing $(BA)^T \tilde{b}$ Recall, $(BA)^T = A^T B^T$. Let $b' = B^T \tilde{b}$. We begin by computing b' . Let \tilde{C} be any subcell with a net point in G , and let $\nu = N_{\tilde{C}}$. Let \tilde{C}^- be the set of all ancestor subcells of \tilde{C} with net points in G including \tilde{C} . We have $b'_\nu = \sum_{\tilde{C}' \in \tilde{C}^-} \frac{\Delta_{\tilde{C}'}}{\Lambda} \tilde{b}_{N_{\tilde{C}'}} = \frac{\Delta_{\tilde{C}}}{\Lambda} \tilde{b}_\nu + b'_{N_{\tilde{C}'p}}$. Therefore, we can use dynamic programming to compute b' in $O(m)$ time. Each entry is filled in during a preorder traversal of the quadtree cells. Finally, A^T has m non-zero entries, so $A^T B^T \tilde{b} = A^T b'$ can be computed in $O(m)$ time as well. \square

We have shown there exists a $(1 + 2\varepsilon, 0)$ -solver for the minimum cost flow problem on G . Plugging in all the pieces, we get a running time bounded by

$$O(m\varepsilon_0^{-2} \log^3(n/\varepsilon_0)(\varepsilon^{-2} + \log(n/\varepsilon_0))).$$

Recall, $\varepsilon_0 = O(\varepsilon/\log n)$. We run the preconditioning framework algorithm in each graph G induced by a simple sub-quadtree's net points as described in Section 2.3. The final running time to compute a flow in G^* of cost at most $(1 + \varepsilon)\text{COST}(P, \mu)$ is

$$O(n\varepsilon^{-O(d)} \log^{O(d)} n).$$

4 Recovering a transportation map from the minimum cost flow

We now describe how to recover a transportation map of P using the approximately minimum cost flow $\hat{f} \in \vec{E}$ we computed for G^* . Unlike \hat{f} , the transportation map τ contains only weighted pairs of points in P . We will *implicitly* maintain a flow f of cost at most $\|\hat{f}\|_{\vec{E}^*}$ that will eventually describe our transportation map. Abusing notation, we extend the definition of $f_{(u,v)}$ to include any pair of vertices in G^* . Value $f_{(u,v)}$ is initially 0 for all $uv \notin E^*$. We follow the strategy of Khesin *et al.* [13] of iteratively rerouting flow going through each net point ν to instead go directly between vertices receiving from or sending flow to ν , eventually resulting in no flow going through any net point. Nearly every pair containing a point $p \in P$ and an ancestor net point may at some

moment carry flow during this procedure. Because quadtree T^* has such high depth, we must take additional care.

To quickly maintain these flow assignments with points in P , we store two data structures $pt(\nu)$ and $nt(\nu)$ for each net point $\nu \in V^* \setminus P$. We call these data structures the *prefix split trees* of ν . The prefix split tree is stored as an ordered binary tree data structure where each node has a weight. We let $w(x)$ denote the weight of node x in a tree S and $w(S)$ denote the total weight of all nodes in S . These trees support the standard operations of insertion and deletion. They support changing the weight of a single node. They support the $\text{MERGE}(S, S')$ operation which takes two trees S and S' and combines them into one tree with all members of S appearing in order before S' . Finally, they support the $\text{PREFIXSPLIT}(S, t)$ operation defined as follows. Given a target value t and a prefix split tree S , PREFIXSPLIT finds a maximal prefix of S 's nodes in order where the sum of node weights in the subset is less than or equal to t . If the sum is less than t , it splits the next node x into two nodes x_1 and x_2 where $w(x_1) + w(x_2) = w(x)$. The split guarantees adding x_1 to the maximal prefix subset makes the sum weight of the subset exactly equal to t . The operation then splits off all members of this subset, including x_1 if a node x was split, into their own tree S' and returns it, leaving S with only the remaining nodes. We emphasize that the order of nodes within the data structure is important for defining PREFIXSPLIT , but the nodes are not “sorted” in any meaningful sense; in particular, any two trees can be merged as defined above. All those operations can be done in amortized $O(\log m)$ time, where m is the number of nodes in the tree, by applying simple modifications to the splay tree data structure of Sleator and Tarjan [21]. We provide details on how to implement a prefix split tree in Appendix A.

In our setting, every node in $pt(\nu)$ and $nt(\nu)$ represents a point $p \in P$. Thanks to our use of the PREFIXSPLIT procedure, some points may be represented multiple times in a single tree. We use $pt(\nu)[p]$ to denote the set nodes representing p in $pt(\nu)$, and define $nt(\nu)[p]$ similarly. Our algorithm implicitly maintains the invariant that for all net points ν and points $p \in P$, $\sum_{x \in pt(\nu)[p]} w(x) - \sum_{x \in nt(\nu)[p]} w(x) = f_{(\nu,p)}$. We proceed with the algorithm given in Figure 2.

Lemma 4.1. *Our algorithm results in a transportation map of cost at most $\|\hat{f}\|_{\bar{E}^*}$, and it can be implemented to run in $O(n\epsilon_0^{-2d} \log^2(n/\epsilon_0))$ time.*

Proof: As stated, our algorithm implicitly maintains a flow f such that for all net points ν and points $p \in P$, $\sum_{x \in pt(\nu)[p]} w(x) - \sum_{x \in nt(\nu)[p]} w(x) = f_{(\nu,p)}$. After every iteration of any of the while loops, the divergences among all vertices in G^* remain the same. Further, after processing any net point v in the inner for loop, there are no other vertices u in V^* such that $f_{(u,v)} \neq 0$. Observe the algorithm never changes the flow coming into or out of a net point v unless $f_{(u,v)} \neq 0$ for some vertex u . Therefore, after v is processed, it *never* has flow going into or out of it again (Khesin *et al.* [13] refer to this property as v having *uniform flow parity*). Because we eventually process every net point in G^* , we eventually end up with a flow f such that $f_{(p,q)} \neq 0$ only if $p, q \in P$. We immediately see τ is a transportation map.

To analyze the cost of τ , observe that after every iteration of a while loop, we replace some δ units of flow passing through v , possibly between multiple sources and one destination or vice versa, with δ units going directly from the source(s) to the destination(s). By the triangle inequality, this new way to route flow is cheaper, so the final flow f , and subsequently τ has smaller cost than \hat{f} .

To implement our algorithm quickly, we only explicitly store new flow values whenever we have a line “ $f_{(u,w)} \leftarrow -$ ” for some pair of vertices (u, w) . Observe that every time we finish processing a cell, every one of its net points is also processed. By the above discussion, flow no longer passes through those net points. Therefore, as we process the net points for a cell C , we never send flow from a net point $v \in N_C$ to a net point outside $N_C \cup N'_C$. Every time we change flow going through

```

«Initialize data structures.»
For all net points  $v \in V^* \setminus P$  and  $p \in P$  where  $f_{(p,v)} > 0$ 
  Insert a node of weight  $f_{(p,v)}$  into  $nt(v)$  representing  $p$ 
For all net points  $v \in V^* \setminus P$  and  $p \in P$  where  $f_{(v,p)} > 0$ 
  Insert a node of weight  $f_{(v,p)}$  into  $pt(v)$  representing  $p$ 
Let  $\mathbb{C}$  be the set of all cells
For  $C \in \mathbb{C}$  in postorder
  Let  $N_C = \{N_{\tilde{C}} : \tilde{C} \in C\}$ ,  $N'_C = \{\text{parent of } v : v \in N_C\}$ 
  For each  $v \in N_C$ 
    «Cancel flow to/from other net points.»
    While  $\exists u, w \in N_C \cup N'_C : f_{(v,w)} > 0 > f_{(v,u)}$ 
       $\delta \leftarrow \min\{f_{(u,v)}, f_{(v,w)}\}$ 
       $f_{(u,w)} \leftarrow f_{(u,w)} + \delta$ 
       $f_{(u,v)} \leftarrow f_{(u,v)} - \delta$ 
       $f_{(v,w)} \leftarrow f_{(v,w)} - \delta$ 
    «Now, either all other net points send flow to  $v$  or all get flow from  $v$ .»
    While  $\exists u \in N_C \cup N'_C : f_{(u,v)} > 0$ 
      «Implicitly reduce  $f_{(v,p)}$  and increase  $f_{(u,p)}$  for several  $p \in P$ »
       $pt' \leftarrow \text{PREFIXSPLIT}(pt(v), f_{(u,v)})$ 
       $\text{MERGE}(pt', pt(u))$ 
    While  $\exists w \in N_C \cup N'_C : f_{(v,w)} > 0$ 
      «Implicitly reduce  $f_{(p,v)}$  and increase  $f_{(p,w)}$  for several  $p \in P$ »
       $nt' \leftarrow \text{PREFIXSPLIT}(nt(v), f_{(v,w)})$ 
       $\text{MERGE}(nt', nt(w))$ 
    «Now, all flow to/from  $v$  involves points  $p \in P$ .»
    While  $pt(v)$  and  $nt(v)$  are not empty
      Let  $x \in nt(v)[p], y \in pt(v)[q]$  for some  $p, q \in P$ 
       $\delta \leftarrow \min\{w(x), w(y)\}$ 
       $f_{(p,q)} \leftarrow f_{(p,q)} + \delta$ 
       $w(x) \leftarrow w(x) - \delta$ ; if  $w(x) = 0$ , delete  $x$  from  $nt(v)$ 
       $w(y) \leftarrow w(y) - \delta$ ; if  $w(y) = 0$ , delete  $y$  from  $pt(v)$ 
  For all  $(p, q) \in P \times P$  where  $f_{(p,q)} > 0$ 
     $\tau(p, q) \leftarrow f_{(p,q)}$ 

```

Figure 2. Recovering a transportation map from an approximately minimum cost flow in G^* .

another net point while processing a net point v , we decrease the number net points u such that $f_{(u,v)} \neq 0$ by one. There are $O(n\varepsilon_0^{-d} \log(n/\varepsilon_0))$ net points, and $O(\varepsilon_0^{-d})$ other net points in each $N_C \cup N'_C$, so the number of iterations total in the first three while loops is $O(n\varepsilon_0^{-2d} \log(n/\varepsilon_0))$. Finally, observe that we only do PREFIXSPLIT operations during these while loops, implying we create a total of $O(n\varepsilon_0^{-2d} \log(n/\varepsilon_0))$ nodes throughout all prefix split trees. Every iteration of the fourth while loop results in deleting a node from at least one of $nt(v)$ or $pt(v)$, so the number of iterations of this while loop is $O(n\varepsilon_0^{-2d} \log(n/\varepsilon_0))$ as well. Finally, every while loop iteration consists of a constant number of constant time operations in addition to a constant number of prefix split tree operations, each of which can be done in $O(\log(n/\varepsilon_0))$ amortized time. \square

Acknowledgement The authors would like to thank Hsien-Chih Chang for some helpful discussions that took place with the first author at Dagstuhl seminar 19181 “Computational Geometry”. We would also like to thank the anonymous reviewers of the conference version of this paper [8] for many helpful comments and suggestions.

References

- [1] P. K. Agarwal, K. Fox, D. Panigrahi, K. R. Varadarajan, and A. Xiao. Faster algorithms for the geometric transportation problem. In *Proc. 33rd Intern. Symp. Comput. Geom.*, pages 7:1–7:16, 2017.
- [2] P. K. Agarwal and R. Sharathkumar. Approximation algorithms for bipartite matching with metric and geometric costs. In *Proc. 46th Symp. Theory Comput.*, pages 555–564. ACM, 2014.
- [3] A. Andoni, A. Nikolov, K. Onak, and G. Yaroslavtsev. Parallel algorithms for geometric graph problems. In *Proc. 46th Symp. Theory Comput.*, pages 574–583, 2014.
- [4] M. W. Bern, D. Eppstein, and S. Teng. Parallel construction of quadtrees and quality triangulations. *Int. J. Comput. Geometry Appl.*, 9(6):517–532, 1999.
- [5] N. Bonneel, M. van de Panne, S. Paris, and W. Heidrich. Displacement interpolation using Lagrangian mass transport. *ACM Trans. Graph.*, 30(6):158, 2011.
- [6] P. B. Callahan and S. R. Kosaraju. Faster algorithms for some geometric graph problems in higher dimensions. In *Proc. 4th Ann. ACM/SIGACT-SIAM Symp. Discrete Algorithms*, pages 291–300, 1993.
- [7] M. Cuturi and A. Doucet. Fast computation of Wasserstein barycenters. In *Proc. 31st Intern. Conf. Machine Learning*, pages 685–693, 2014.
- [8] K. Fox and J. Lu. A near-linear time approximation scheme for geometric transportation with arbitrary supplies and spread. In *Proc. 36th Intern. Symp. Comput. Geom.*, pages 45:1–45:19, 2020.
- [9] P. Giannopoulos and R. C. Veltkamp. A pseudo-metric for weighted point sets. In *Proc. 7th Europ. Conf. Comput. Vision*, pages 715–730, 2002.
- [10] K. Grauman and T. Darrell. Fast contour matching using approximate earth mover’s distance. In *Proc. 24th IEEE Conf. Comput. Vision and Pattern Recog.*, pages I:220–I:227, 2004.
- [11] S. Har-Peled. *Geometric approximation algorithms*, volume 173. American Mathematical Soc., 2011.
- [12] P. Indyk. A near linear time constant factor approximation for Euclidean bichromatic matching (cost). In *Proc. 18th Ann. ACM-SIAM Symp. Discrete Algorithms*, pages 39–42, 2007.
- [13] A. B. Khesin, A. Nikolov, and D. Paramonov. Preconditioning for the geometric transportation problem. In *Proc. 35th Intern. Symp. Comput. Geom.*, pages 15:1–15:14, 2019.
- [14] N. Lahn, D. Mulchandani, and S. Raghvendra. A graph theoretic additive approximation of optimal transport. In *Proc 32nd Adv. Neur. Info. Proces. Sys.*, pages 13813–13823, 2019.
- [15] Y. T. Lee and A. Sidford. Path finding methods for linear programming: Solving linear programs in $\tilde{O}(\text{vrnk})$ iterations and faster algorithms for maximum flow. In *Proc. 55th IEEE Ann. Symp. Found. Comput. Sci.*, pages 424–433, 2014.
- [16] J. B. Orlin. A faster strongly polynomial minimum cost flow algorithm. *Operations Research*, 41(2):338–350, 1993.

- [17] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover’s distance as a metric for image retrieval. *Intern. J. Comput. Vision*, 40(2):99–121, 2000.
- [18] R. Sharathkumar and P. K. Agarwal. Algorithms for the transportation problem in geometric settings. In *Proc. 23rd Ann. ACM-SIAM Symp. Discrete Algorithms*, pages 306–317, 2012.
- [19] R. Sharathkumar and P. K. Agarwal. A near-linear time ϵ -approximation algorithm for geometric bipartite matching. In *Proc. 44th Symp. Theory Comput.*, pages 385–394, 2012.
- [20] J. Sherman. Generalized preconditioning and undirected minimum-cost flow. In *Proc. 28th Ann. ACM-SIAM Symp. Discrete Algorithms*, pages 772–780, 2017.
- [21] D. D. Sleator and R. E. Tarjan. Self-adjusting binary search trees. *J. ACM*, 32(3):652–686, 1985.
- [22] J. Solomon, R. M. Rustamov, L. J. Guibas, and A. Butscher. Earth mover’s distances on discrete surfaces. *ACM Trans. Graph.*, 33(4):67:1–67:12, 2014.
- [23] C. Villani. *Optimal Transport: Old and New*. Springer Science & Business Media, 2008.

A Prefix split trees

We implement our prefix split trees by modifying the splay tree data structure of Sleator and Tarjan [21]. Let S be a prefix split tree. We store the weight $w(x)$ of each node x directly with the node itself. Moreover, every node x keeps another value $W(x)$ equal to the sum weight of all the descendants of x including x itself.

A **splay** of a node x in S is a sequence of double rotations (possibly followed by a standard single rotation) that move x to the root of S . Only those nodes on the path from the root to x have their children pointers updated by a splay. We can update $W(y)$ for every such node y with only a constant factor overhead in the time to perform a splay. Let $s(x)$ denote the number of descendants of x in its prefix split tree, and let $r(x) = \lceil \lg s(x) \rceil$. Let $\Phi(S) = \sum_{x \in S} r(x)$. The **amortized time** for an operation on S can be defined as the real time spent on the operation plus the net change to $\Phi(S)$ after the operation. The amortized time for a splay in an m -node tree is $O(\log m)$ [21].

Recall, the order of nodes within a tree is largely irrelevant outside the definition of the PREFIXSPLIT operation. To insert a node x in S , we add x as the child of an arbitrary leaf of S and splay x to the root. The number of operations in the splay dominates, so the amortized cost of insertion is $O(\log m)$. To delete a node x , we splay x to the root and delete it, resulting in two disconnected subtrees S_1 and S_2 . We then perform a MERGE(S_1, S_2) in $O(\log m)$ amortized time as described below, so the whole deletion has amortized cost $O(\log m)$. To update the weight of a node x , we splay x to the root and update $w(x)$ and $W(x)$ in constant time each. The splay once again dominates, so the total amortized cost is $O(\log m)$.

The operation MERGE(S_1, S_2) is implemented as follows. Let x be the rightmost leaf of S_1 . We splay x to the root so it has exactly one child. We then make the root of S_2 the other child of x . Let m be the total number of nodes in S_1 and S_2 . Adding the root of S_2 as a child increases $\Phi(S_1) + \Phi(S_2)$ by $O(\log m)$, so the amortized time for the MERGE is $O(\log m)$.

Finally, we discuss the implementation of PREFIXSPLIT(S, t). We assume $t > 0$. We use the values $W(\cdot)$ to find the prefix of nodes desired. Let y be the next node in order after the prefix. We splay y to the root of S . Let x be the left child of y after the splay (if it exists). Suppose $W(x) < t$. We delete y , creating two trees S_1 and S_2 where S_1 contains the nodes in the prefix. We create a

new node y_1 of weight $t - W(y)$ and make the root of S_1 its child so that y_1 is the new root. We create a node y_2 of weight $w(y) - w(y_1)$ and make the root of S_2 its child. Now, suppose instead $W(x) = t$. In this case, we simply remove the edge between x and y to create a subtree S_1 with x as its root. Let S_2 be the remainder of S . Whether or not $W(x) = t$, we return S_1 and set $S = S_2$. The amortized time for the PREFIXSPLIT is the amortized time for a single splay and a constant number of edge changes, implying the PREFIXSPLIT takes $O(\log m)$ amortized time total.