From Proximity to Utility: A Voronoi Partition of Pareto Optima

Hsien-Chih Chang†  Sariel Har-Peled‡  Benjamin Raichel§

April 15, 2014

Abstract

We present an extension of Voronoi diagrams where not only the distance to the site is taken into account when considering which site the client is going to use, but additional attributes (i.e., prices) are also considered. A cell in this diagram is then the locus of all points (i.e., clients) that consider the same set of sites to be relevant. In particular, the precise site a client might use from this candidate set depends on parameters that might change between usages, and the candidate set lists all of the relevant sites. The resulting diagram is significantly more expressive than Voronoi diagrams, but naturally has the drawback that its complexity, even in the plane, might be quite high.

Nevertheless, we show that if the attributes of the sites are drawn from the same distribution (note that the locations are fixed), then the expected complexity of the candidate diagram is near linear. To derive this result, we derive several new technical results, which are of independent interest.

1. Introduction

Informal description of the candidate diagram. Suppose you open your refrigerator one day, to discover it is time to go grocery shopping\(^1\). Which store you go to will be determined by a number of different factors. For example, what items you are buying, and do you want the cheapest price or highest quality, and how much time you have for this chore. Naturally the distance to the store will also be a factor. On different days which store is the best to go to will differ based on that day’s preferences. However, there are certain stores you will never shop at. These are stores which are worse in every way than some other store (i.e. further, more expensive, lower quality, etc.). Therefore, the stores that are relevant (i.e. Pareto optima) are those that are not strictly worse in every way than some other

---

\(^1\)Work on this paper was partially supported by NSF AF award CCF-1217462. The paper is also available online at http://sarielhp.org/p/14/beer/, and would hopefully be posted to the arxiv soon.

†Department of Computer Science, University of Illinois; 201 N. Goodwin Avenue, Urbana, IL 61801, USA; hchang17@illinois.edu; http://web.engr.illinois.edu/~hchang17/.

‡Department of Computer Science, University of Illinois; 201 N. Goodwin Avenue, Urbana, IL 61801, USA; sariel@uiuc.edu; http://sarielhp.org.

§Department of Computer Science, University of Illinois; 201 N. Goodwin Avenue, Urbana, IL 61801, USA; raichel2@uiuc.edu; http://illinois.edu/~raichel2.

\(^1\)Unless you are feeling adventurous enough that day to eat the frozen mystery food stuck to the back of the freezer, which we strongly discourage you from doing.
store. Thus, every point in the plane is mapped to the set of stores that a client at that location might use. The candidate diagram is the partition of the plane into regions, where this candidate set is the same for all points in the same region. Naturally, if your only consideration is distance, then this is the (classical) Voronoi diagram of the sites. However, here, deciding which shop to use is an instance of multi-objective optimization — as there are multiple, potentially competing, objectives to be optimized, and the decision might change as the weighting and influence of these objectives mutate over time.

**Pareto optima in welfare economics.** Pareto efficiency, named after Vilfredo Pareto, is a core concept in economic theory and more specifically in welfare economics. Here each point in \( \mathbb{R}^d \) represents the corresponding utilities of a set of \( d \) players for a particular allocation of finite resources. A point is said to be Pareto optimal if there is no other allocation which increases the utility of any individual without decreasing the utility of another. The First Fundamental Theorem of Welfare Economics states that any competitive equilibrium (i.e. supply equals demand) is Pareto optimal. The origins of this theorem date back to 1776 with Adam Smith’s famous (and controversial) work, “The Wealth of Nations,” but was not formally “proven” until the 20th century by Lerner, Lange, and Arrow (see [Fel08]). Naturally such proofs rely on simplifying (i.e. potentially unrealistic) assumptions such as perfect knowledge, or absence of externalities. The Second Fundamental Theorem of Welfare Economics states that any Pareto optimum is achievable through lump-sum transfers (i.e. taxation and redistribution). In other words each Pareto optima is a “best solution” under some set of societal preferences, and is achievable through redistribution in one form or another (see [Fel08] for a more in depth discussion).

**Pareto optima in computer science.** In Computational Geometry such points relate to the orthogonal convex hull [OSW84], which is a subset of the well known convex hull (though the input points that lie on the rectilinear hull is a super set of those which lie on the convex hull). Pareto optima are also of importance to the database community [BKS01, HTC13], in which context such points are called maximal or skyline points. Such points are of interest as they can be seen as the relevant subset of the (potentially much larger) result of a relational database query. The standard example is that of querying a database of hotels for the cheapest and closest hotel, where naturally hotels which are farther and more expensive than an alternative hotel are not relevant results. There is a significant amount of work on computing these points, see Kung et al. [KLP75]. More recently, Godfrey et al. [GSG07] compared various approaches for the computation of these points (from a databases perspective), as well as introduce their own new external algorithm.\(^2\)

**Modeling uncertainty.** Recently, there is a growing interest in modeling uncertainty in data. As real data is acquired via physical measurements, noise and errors are introduced. This can be addressed by treating the data as coming from a distribution (e.g., a point location might be interpreted as a center of a Gaussian), and computing desired classical quantities adapted for such settings. Thus, a nearest-neighbor query becomes a probabilistic question — what is the expected distance to the nearest-neighbor? What is the most likely point to be the nearest-neighbor? Etc. See [AAH+13] and references therein for more information.

This in turn gives rise to the question of what is the expected complexity of geometric structures defined over such data. If the data is a set of points, and the locations of the points are chosen randomly then there is a lot research that was done on this problem, see [SW93, WW93, HR14] and references therein. The problem when the locations are fixed but weights associated with the points are chosen

\(^2\)There is of course a lot of other work on Pareto optimal points, from connections to Nash equilibrium to scheduling. We resisted the temptation of including many such references which are not directly related to our paper.
randomly is relatively new. Agarwal et al. [AHKS13] showed that for a set of disjoint segments in the plane, if they are being expanded randomly, then the expected complexity of the union is near linear. This can be interpreted as the level set of multiplicatively weighted Voronoi diagrams, where the weights are being chosen randomly. This result is somewhat surprising as in the worst case the complexity of such a union is quadratic. More recently, the authors [HR14] gave a simpler proof that works for the complexity of the whole diagram.

Specifically, Har-Peled and Raichel [HR14] showed that the complexity of the multiplicative Voronoi diagram of randomly weighted points in the plane is $O(n \text{polylog } n)$ (in expectation). The current write-up is a continuation of this work, to handle more general kinds of Voronoi diagrams.

1.1. Our contributions

Introducing the candidate diagram. We define formally the candidate diagram in Section 2.2. For every point $x$ in the plane, the diagram associates a candidate set $L(x)$ of sites that are relevant for $x$; that is, all the sites that are Pareto optima for $x$. Putting it differently, a site is irrelevant for $x$ (that is, not in $L(x)$) if it is further away and worse in all parameters than some other site. In particular, the client at $x$ may decide which site it prefers using any function that respects this domination relationship. This diagram is a significant extension of the notion of Voronoi diagrams, and includes other extensions of Voronoi diagrams as special subcases, like multiplicative weighted Voronoi diagrams.

Significantly, unlike the traditional Voronoi diagram, the candidate diagram allows the user to change their distance function, as long as the distance function respects the dominating condition mentioned above. Not surprisingly, the worst case complexity of this diagram can be quite high.

Sampled attributes/prices. We consider the case where each site chooses its $j$th attribute, from some distribution $D_j$ independently for each $j$. We show that the candidate diagram in expectation has near linear complexity, and that with high probability the candidate set has polylog size for any point in the plane. In the process we derive several results which are interesting in their own right.

(A) Low complexity of the minima for random points in the hypercube. We prove that if $n$ points are sampled from a fixed distribution over the $d$-dimensional hypercube then, with probability $\geq 1 - 1/n^{\Omega(1)}$, the number of staircase points $X$ (i.e. Pareto optima) is $O(\log^{d-1} n)$ (see Theorem 6.8). Previously, this result was only known in a weaker form (see [BDHT05, BR10a]). Specifically, it is known that after normalization the cumulative distribution function of $X$ is normal, up to an additive error $O(1/\text{polylog } n)$. In particular, these results (which are quite nice and mathematically involved), can imply only that this probability $\geq 1 - 1/\Omega(\text{polylog } n)$.

To the best of our knowledge this result is new — we emphasize, however, that for our purposes a weaker bound of $O(\log^d n)$ is sufficient, and such a weaker result follows readily from the $\varepsilon$-net theorem. Naturally, however, this weaker bound would add some polylog factors to later results in the paper.

To get this result, we prove a lemma providing high probability bounds when applying backwards analysis [Sei93]. Such tail estimates are known in the context of randomized incremental algorithms [CMS93, BCKO08], but our proof is arguably more direct and cleaner, and should be applicable to more cases. See Section 6 for details.

(B) Overlay of the $k$th order Voronoi cells in randomized incremental construction. We prove that the overlay of cells during a randomized incremental construction of the $k$th order Voronoi diagram is of complexity $O(n \text{polylog } n)$ for any $k = O(\text{polylog } n)$ (see Theorem 5.1).

(C) Complexity of the candidate diagram. Combining the above two results carefully, and
inserting the sites in the right order, yields a bound on the complexity of the candidate diagram (see Theorem 5.2).

Outline of the paper. In Section 2 we formally introduce the candidate diagram. In Section 3.1 we state that for a set of \( n \) sites with randomly sampled attributes, with high probability, the candidate set for every point in the plane is of size \( O(\log^d n) \) (see Corollary 3.2). The more general proof which implies this bound is left to Section 6. To compliment this bound, the goal is then to bound the complexity of the candidate diagram (i.e. both the planar partition, and the set of all distinct candidate sets). As doing so directly is difficult, instead we introduce an enlarged candidate set, called the proxy set, in Section 3.2. Specifically, in opting to use the proxy set instead of the candidate set, we must pay an extra logarithmic factor. However, in Section 4 we show that the appropriate diagram for this enlarged set is the overlay of cells during the randomized incremental construction of the \( k \)th order Voronoi diagram, for which we can provide an expected bound of \( O(n \text{polylog} n) \). In Section 5 we then show that such a bound on the diagram for the proxy set implies a similar bound on the candidate diagram.

2. The candidate diagram

2.1. Preliminaries

Definition 2.1. Consider two points \( p = (p_1, \ldots, p_d) \) and \( q = (q_1, \ldots, q_d) \) in \( \mathbb{R}^d \). The point \( p \) dominates \( q \) (denoted by \( p \preceq q \)) if \( p_i \leq q_i \) for all \( i = 1, \ldots, d \).

Given a point set \( P \subseteq \mathbb{R}^d \), there are numerous terms for the subset of \( P \) that is not dominated as already discussed above, such as Pareto optima or minima. In this write-up, we prefer the term staircase point.

Definition 2.2. For a point set \( P \subseteq \mathbb{R}^d \), a point \( p \in P \) is a staircase point of \( P \) if no other point of \( P \) dominates it. We refer to the set of all staircase points as the staircase of \( P \), which we denote by \( S(P) \).

2.2. Formal definition of the candidate diagram

The input. Let \( S = \{s_1, \ldots, s_n\} \) be a set of \( n \) sites in the plane. For each site \( s_i \in S \), there is an associated list, \( a_i \), of \( d \) real valued attributes\(^3\), each in the interval \([0, 1]\). When viewed as a point in the unit hypercube, i.e. \( a_i \in [0, 1]^d \), we refer to this list of attributes as the parametric point of \( s_i \).

Preferences. Fix a client location \( x \in \mathbb{R}^2 \). When choosing a site, there are \( d+1 \) variables for the client to consider, namely, its distance to that site as well as the list of \( d \) attributes associated with the site. For each of these variables, the goal of the client is simple: pay as little as possible for that variable.

Definition 2.3. A client \( x \in \mathbb{R}^2 \) has a dominating preference if for any two sites \( s, s' \in \mathbb{R}^2 \), with attributes \( a, a' \in \mathbb{R}^d \) respectively, the client would prefer the site \( s \) over \( s' \) if \( \|x - s\| < \|x - s'\| \) and \( a \preceq a' \) (that is, \( a \) dominates \( a' \)).

\(^3\)We are assuming that \( d \) is a small constant. Throughout the paper, the \( O \) notation would hide constants that are potentially exponential (or worse) in \( d \).
Note that a client having a dominating preference does not identify a specific optimum site for the client, but rather a set of potential optimum sites. Specifically, given a client location \( x \in \mathbb{R}^2 \), let its distance to the \( i \)th site be \( \ell_i = \|x - s_i\| \), for \( i = 1, \ldots, n \). The set of sites the client might possibly use (assuming the client uses a dominating preference) are the staircase points of the set \( S'(x) = \{(a_1, \ell_1), \ldots, (a_n, \ell_n)\} \) (i.e., we are adding the distance to the \( i \)th site as an additional attribute of the \( i \)th site — this attribute depends on the location of \( x \)). We refer to the set of sites realizing the staircase of \( S'(x) \) (i.e., all the sites relevant for \( x \)) as the candidate set \( L(x) \) of \( x \):

\[
L(x) = \{s_i \in S \mid (a_i, \ell_i) \text{ is a staircase point of } S'(x)\}.
\]

The cell of \( x \) is the set of all the points in the plane that have the same candidate set associated with them; that is, cell(\( x \)) = \( \{p \in \mathbb{R}^2 \mid L(p) = L(x)\} \). The decomposition of the plane into these cells is the candidate diagram. We define any function that identifies a specific site in \( L(x) \) for each client point \( x \in \mathbb{R}^2 \) as a dominating distance function.

**Complexity.** We define the diagram complexity of the candidate diagram to be the complexity of the arrangement of all these cells, i.e. the total numbers of edges, faces, and vertices. The explicit description complexity (or complexity for short) of the diagram refers to the sum of the sizes of candidate sets over all cells (which is potentially much larger).

**Lower bound.** It is not hard to show that in the worst case, the (explicit description) complexity of the diagram is at least \( \Omega(n^2) \) if the attributes are not chosen randomly. The naive upper bound is significantly worse; that is, \( O(n^5) \). Specifically, the naive upper bound on the diagram complexity is achieved by considering the partition induced by the overlay of all bisectors of all pairs of sites, which has \( O(n^4) \) cells. The bound of \( O(n^5) \) on the complexity then follows as each candidate set in the worst case has linear size. We leave the problem of closing this gap as an open problem for further research.

### 3. Candidate diagrams for sampled parametric points

In this section we investigate the expected complexity of the candidate diagram when the associated parametric point (i.e. attributes) of each input site is randomly sampled.

**The input.** Let \( S \subseteq \mathbb{R}^2 \) be a set of \( n \) sites in the plane. For each point \( s_i \in S \), a point \( a_i \) is sampled uniformly at random from \([0, 1]^d\) as the parametric point of \( s_i \).

All we require for the sampling is that it is done independently for each point. In particular, we only assume that a \( d \)-dimensional parametric point \( a = (a_1, \ldots, a_d) \) is sampled from the product of distributions \( D_1 \times \cdots \times D_d \), such that when sorting the \( n \) parametric points in increasing order of a specific coordinate, each permutation is equally likely to occur. Formally, \( D_i \) is a continuous distribution over \([0, 1]\), such that the density function is bounded (this is to avoid repeated values) for each \( i \).

**Section outline.** After first defining some relevant terminology, in Section 3.1 we provide a high probability bound of \( O(\log^d n) \) on the size of every candidate sets. Thus if we could get a good bound on the diagram complexity of the candidate diagram, this would also yield a good bound on the explicit description complexity of the diagram. However, proving a reasonable bound directly is a rather difficult task. Therefore in Section 3.2 we instead define for each point \( x \) in the plane a slightly larger set, called
the proxy set (formal definition below), which with high probability contains the candidate set of \( x \). The advantage is that proving a bound on the complexity of the partition induced by this enlarged set is a more manageable task. Later we show that a bound on the complexity for this alternative partition implies a bound on the diagram complexity of the candidate diagram. The following definition will be needed.

**Definition 3.1.** Given a point \( p = (p_1, \ldots, p_d) \) in the \( d \)-dimensional unit hypercube, the **point volume** \( \text{pv}(p) \) of \( p \) is defined to be \( p_1 p_2 \cdots p_d \); that is, the volume of the hyperrectangle (or orthotope) whose opposite corners are \( p \) and the origin. When \( p \) is specifically the associated parametric point of an input site \( s \), we refer to the point volume of \( p \) as the **parametric volume** of \( s \).

Observe that if \( p \) dominates \( q \) then \( p \) must have smaller point volume (i.e. \( p \) lies in the hyperrectangle defined by \( q \)).

### 3.1. On the size of the staircase and correspondingly the candidate set

We need the following result that bounds with high probability the number of staircase points in a random point set sampled from the hypercube \([0, 1]^d\). While, surprisingly, this result can not be found in the literature in exactly this form, significantly deeper results are known which “almost” imply it. Specifically, the distribution of minimal points is almost a normal distribution [BDHT05] up to an additive \( 1/\text{polylog } n \) error (thus, implying the theorem for a weaker high-probability bound). Also, high probability bounds on the size of the tail of the distribution are known [Vu05] in terms of the variance (on the number of vertices of the convex-hull), but these bounds seem to be too weak in this case.

The following corollary follows immediately from Theorem 6.7, whose proof is the main goal of Section 6. The proof presented is simpler and more direct and might be of independent interest.

**Corollary 3.2.** Let \( c > 1 \) be a constant, and let \( S \) be a set of \( n \) sites in the plane, where for each \( s \in S \) we independently sample an associated parametric point from a distribution over \([0, 1]^d\). Then, with probability \( \geq 1 - 1/n^c \), simultaneously for all points in the plane, the candidate set has size \( O(\log^d n) \).

**Proof:** Consider the arrangement of bisectors of all pairs of points of \( S \). This arrangement has complexity \( O(n^d) \); inside each cell the candidate set is the same. Now for any point in a cell of this arrangement Theorem 6.7 (and Observation 6.9) immediately gives us the stated bound with polynomially-high probability. Therefore picking a representative point from each cell in this arrangement and applying the union bound imply the claim.

### 3.2. The Proxy Set

#### 3.2.1. Definitions

As before, the input is a set of sites \( S \). For each site \( s \in S \), we randomly pick uniformly and independently an attribute \( a(s) \in [0, 1]^d \).

**Definition 3.3.** The **volume ordering** of the sites of \( S \) is a permutation \( T = \langle s_1, \ldots, s_n \rangle \) ordered by increasing point volume of the attributes; that is, \( \text{pv}(a_1) \leq \text{pv}(a_2) \leq \ldots \leq \text{pv}(a_n) \), where \( a_i = a(s_i) \) (see Definition 3.1). For any ordering \( T \) of a site set \( S \), we use the notation \( T_i \) (resp. \( S_i \)) to denote the ordered (resp. unordered) \( i \)th prefix of \( T \).
Corollary 3.7. Let \( \mathbf{a}_i \) dominates \( \mathbf{a}_j \) then \( \mathbf{s}_i \) precedes \( \mathbf{s}_j \) in the volume ordering (see Definition 3.1). So if we add the sites in volume order, then when we add the \( i \)th site we can ignore all later sites when determining its region of influence (i.e. the region of points whose candidate set it belongs to), as no later point can dominate it.

Definition 3.4. For a set of points \( \mathcal{P} \subseteq \mathbb{R}^d \) and a point \( \mathbf{x} \in \mathbb{R}^d \), let \( d_k(\mathbf{x}, \mathcal{P}) \) denote the \( k \)th nearest neighbor distance to \( \mathbf{x} \) of \( \mathcal{P} \). Formally, \( d_k(\mathbf{x}, \mathcal{P}) \) is the \( k \)th smallest value in the multiset \( D = \{ \|\mathbf{x} - \mathbf{p}\| \mid \mathbf{p} \in \mathcal{P} \} \). The \( k \) nearest neighbors to \( \mathbf{x} \) in \( \mathcal{P} \) is the set \( \mathcal{P}_{\leq k}(\mathbf{x}) = \{ \mathbf{p} \in \mathcal{P} \mid \|\mathbf{x} - \mathbf{p}\| \leq d_k(\mathbf{x}, \mathcal{P}) \} \).

\[ \text{\[3.2.2\] The proxy set and its relation to the candidate set} \]

For each point in the plane we now define a subset of sites such that with high probability this set contains the point’s weighted nearest site (under a dominating distance function).

Definition 3.5. Let \( \mathcal{S} \) be a set of sites in the plane, and let \( \mathcal{T} = \langle \mathbf{s}_1, \ldots, \mathbf{s}_n \rangle \) be the volume ordering of \( \mathcal{S} \). For a parameter \( k \) and a point \( \mathbf{x} \in \mathbb{R}^d \), the \((k)\)th proxy set of \( \mathbf{x} \) is the set \( \mathcal{C}_k(\mathbf{x}, \mathcal{T}) = \bigcup_{i=1}^{n} (\mathcal{T}_{i,k}(\mathbf{x}) \subseteq \mathcal{S}) \), where \( \mathcal{T}_i = \langle \mathbf{s}_1, \ldots, \mathbf{s}_i \rangle \). In words, \( \mathbf{s}_i \) is in \( \mathcal{C}_k(\mathbf{x}, \mathcal{T}) \) if it is one of the \( k \) nearest neighbors to \( \mathbf{x} \) in the prefix \( \mathcal{T}_i \).

Observe that all the points in \( \mathcal{T}_k \) are also in \( \mathcal{C}_k(\mathbf{x}, \mathcal{T}) \) for all \( \mathbf{x} \in \mathbb{R}^2 \).

Lemma 3.6. Let \( c > 1 \) be a constant, and let \( \mathcal{S} \) be a set of sites in the plane. Let \( k \geq 1 \) be a fixed parameter. With probability \( \geq 1 - 1/n^c \), we have \( |\mathcal{C}_k(\mathbf{x}, \mathcal{T})| = O(k \log n) \), simultaneously for all points \( \mathbf{x} \) in the plane.

Proof: Fix a point \( \mathbf{x} \) in the plane. Ordering the sites by increasing parametric volume creates a random permutation \( \mathcal{T} \) on the distances of the sites from \( \mathbf{x} \). A site \( \mathbf{s}_i \) gets added to the \((k)\)th proxy set \( \mathcal{C}_k(\mathbf{x}, \mathcal{T}) \) if the distance \( \|\mathbf{x} - \mathbf{s}_i\| \) is one of the \( k \) smallest values in prefix \( \mathcal{T}_i \) of this permutation. Therefore a direct application of Lemma 6.1,p14 implies, with probability \( \geq 1 - 1/n^c \), that \( |\mathcal{C}_k(\mathbf{x}, \mathcal{T})| = O(k \log n) \).

Now we must argue that this holds for all points in the plane simultaneously. Consider the arrangement determined by all bisectors of all pairs of sites in \( \mathcal{S} \). This arrangement is a simple planar map with \( O(n^4) \) vertices and \( O(n^4) \) faces. Observe that within each face the proxy set cannot change since all points in this face have the same ordering of their distances to the sites in \( \mathcal{S} \). Therefore, picking a representative point from each of these \( O(n^4) \) faces, applying the high probability bound to each one of them, and then applying the union bound implies the claim.\(^4\) \( \blacksquare \)

The corollary below follows from a careful (but straightforward) integration argument, and the proof is delegated to the appendix — see Lemma A.1,p18 for the proof.

Corollary 3.7. Let \( F_d(\Delta) \) be the total measure of the points \( \mathbf{p} = (\mathbf{p}_1, \ldots, \mathbf{p}_d) \) in the hypercube \([0,1]^d\), such that \( \mathcal{P}(\mathbf{p}) = \mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_d \leq \Delta \). Then for \( \Delta \geq (\log n)/n \) we have \( F_d(\Delta) = O(\Delta \log^{d-1} n) \) and \( F_d(\log n/n) = O((\log^d n)/n) \).

Lemma 3.8. Let \( c > 1 \) be a prespecified constant, \( k = \Theta(\log^d n) \) be a fixed parameter, and \( \mathcal{S} \) be a set of \( n \) sites. For any point \( \mathbf{x} \in \mathbb{R}^2 \), let \( \mathbf{s} \) be the site that \( \mathbf{x} \) decides to use (assuming that \( \mathbf{x} \) uses a dominating distance function). Then, we have \( \mathbf{s} \in \mathcal{C}_k(\mathbf{x}, \mathcal{T}) \) with polynomially high probability. Formally, we have \( \mathcal{L}(\mathbf{x}) \subseteq \mathcal{C}_k(\mathbf{x}, \mathcal{T}) \) (see Eq. (2.1),p5) with probability \( \geq 1 - 1/n^c \) for all the points in the plane simultaneously.

\(^4\)Note that picking representatives based on the faces of the arrangement is valid as the arrangement is fixed and only the weights are being sampled.
Proof: Let \( s_j \) be any site not in \( C_k(x, T) \). We now show that with high probability \( s_j \) is (parametrically) dominated by some other site which is closer to \( x \), and hence by the definition of dominating preference (Definition 2.3.4), \( s_j \) cannot be a site used by \( x \). Taking the union bound over all sites not in \( C_k(x, T) \) then implies the claim.

By Corollary 3.7 the total measure of the points in \([0,1]^d\) with point volume \( \leq \Delta = \log n/n \) is \( O\left(\frac{(\log^d n)}{n}\right) \). As such, by Chernoff’s inequality, with probability \( \geq 1 - 1/n^\epsilon \), there are \( N = O\left(\log^d n\right) \) sites in \( S \) that their attribute points have point volume smaller than \( \Delta \). In particular, by choosing \( k \) to be sufficiently large (i.e. \( k > N \)), \( T_k \) will contain all these small point volume sites, and since \( T_k \subseteq C_k(x, T) \), so will \( C_k(x, T) \). Therefore, from this point on, we will assume that \( s_j \notin C_k(x, T) \) and \( \Delta = \text{pv}(a_j) = \Omega(\log n/n) \).

Now for any site \( s_i \) with smaller parametric volume than \( s_j \), we have that \( i < j \). In particular, the \( k \) nearest neighbors of \( x \) in \( T_j \) all have smaller parametric volume than \( s_j \). Hence \( C_k(x, T) \) contains \( k \) points all of which have smaller parametric volume than \( s_j \), and which are closer to \( x \). Therefore, the claim will be implied if one of these \( k \) points dominates \( s_j \).

Now, the probability of a site \( s_i \) (that is closer to \( x \) than \( s_j \)) to dominate \( s_j \) is the probability that \( a_i \preceq a_j \), given that \( \text{pv}(a_i) \leq \text{pv}(a_j) \). The probability that a random attribute point picked in \([0,1]^d\) to be dominated by \( a_j \) is exactly \( \text{pv}(a_j) \), and as such the desired probability is \( \Delta/F_d(\Delta) = O\left(1/\log^d n \right) \), by Corollary 3.7. As each one of the \( \geq k = \Theta(\log^d n) \) points has equal probability to be anywhere in the region of the hypercube with smaller volume than \( s_j \), this implies the expected number of points which dominate \( s_j \) is \( \Theta(\log n) \). Therefore by choosing the constant appropriately for \( k \), Chernoff’s inequality implies the desired result.

One can now extend this result to hold for all points in the plane simultaneously, by making a similar argument to that shown in the proof of Lemma 3.6. \( \blacksquare \)

4. Overlaying the \( k \)th order Voronoi diagram cells

We now show that the appropriate partition of the plane for the proxy set is determined by the overlay of cells during the randomized incremental construction of the \( k \)th order Voronoi diagram.

4.1. Preliminaries — the \( k \)th order Voronoi diagrams

4.1.1. From \( k \)th order Voronoi diagram to arrangement of planes

Let \( P \) be a set of \( n \) points in the plane. The \( k \)th order Voronoi diagram of \( P \), denoted \( V_k(S) \), is the partition of the plane into cells such that each cell is the locus of points which have the same set of \( k \) nearest points of \( P \) (the internal ordering of these \( k \) sites, by distance to the query point, may vary within the cell however). It is well known that the worst case complexity of this diagram is \( \Theta(k(n-k)) \).

One can also interpret the \( k \)th order Voronoi diagram in terms of an arrangement of planes in \( \mathbb{R}^3 \). Specifically, lift each site to the paraboloid defined by the set of points \((x, y, x^2 + y^2)\). Consider the arrangement of planes tangent to the paraboloid at the lifted locations of the sites. We say a point on the union of these planes is on the exact \( k \)th-level if there are exactly \( k \) planes strictly below it (i.e., the positive \( z \)-axis is pointing down under this interpretation). Let \( E_k(S) \) denote the set of edges from the exact \( k \)th level, where an edge is a maximal length portion of the \( k \)th level that lies on the intersection of the planes of two fixed sites. One can show the increasing \( z \)-ordering of the planes above any point in the \( xy \)-plane is the same as the increasing distance ordering of that point to the sites, and hence the
orthogonal projection onto the \( xy \)-plane of the edges in \( E_{k-1}(S) \) gives the edges of the \( k \)th order Voronoi diagram. The set of edges from all levels of height at most \( k \), is denoted by \( E_{\leq k}(S) = \cup_{i=0}^{k} E_i(S) \).

4.1.2. Arrangement of Planes

Let \( H \) be a set of \( n \) planes in general position in \( \mathbb{R}^3 \). For any subset \( R \subseteq H \), let \( \mathcal{A}(R) \) denote the arrangement of \( R \), and let \( V(R) \) denote the set of vertices in \( \mathcal{A}(R) \). Moreover, let \( V_{\leq k}(R) \) denote the subset of vertices on any \( \leq k \) level of the arrangement, i.e., vertices with at most \( k \) planes of \( R \) below them (similarly \( V_k(R) \) will denote vertices on the exact \( k \) level). For any vertex \( v \in V(R) \), let its **below conflict list**, be \( B(v) = \{ h \in H \mid h \text{ is below } v \} \), and \( b_v = |B(v)| \).

4.1.3. The Clarkson-Shor technique

We will now use a well known result of Clarkson-Shor [CS89], which we state here without proof (see [Har11] for details), in order to bound the sum of the below conflict lists over all vertices in the \( \leq k \) levels. Specifically, let \( S \) be a set of elements such that any subset \( R \subseteq S \) defines a corresponding set of objects \( T(R) \) (e.g., \( S \) is a set of planes and any subset \( R \subseteq S \) induces the set vertices in their arrangement). Each potential object, \( \tau \), has a defining set and a stopping set. The **defining set**, \( D(\tau) \), is a subset of \( S \) that must appear in \( R \) in order for the object to be present in \( T(R) \). We require that this set has at most a constant size for all objects. The **stopping set**, \( \kappa(\tau) \), is a subset of \( S \) such that if any of its members appear in \( R \) then \( \tau \) is not present in \( T(R) \) (we also naturally require that \( \kappa(\tau) \cap D(\tau) = \emptyset \), for all \( \tau \)). Surprisingly, this already implies the following.

**Theorem 4.1 (Bounded Moments, [CS89]).** Using the above notation, let \( S \) be a set of \( n \) objects, and let \( R \) be a random sample of size \( r \) from \( S \). Let \( f(\cdot) \) be a polynomially growing function\(^5\). We have that \( \mathbb{E}\left[ \sum_{\tau \in T(R)} f(|\kappa(\tau)|) \right] = O\left( \mathbb{E}\left[ |T(R)| \right] f(n/r) \right) \), where the expectation is over the sample \( R \).

4.2. Bounding the size of the below conflict-lists

We next bound the sum of the sizes of the below conflict lists. The proof of Lemma 4.2 is technically interesting as it does not follow in a straightforward fashion from the Clarkson-Shor technique. Indeed, the below conflict list is not the standard conflict list. To circumvent this issue, we use a second random sample and then deploy the Clarkson-Shor technique – this is reminiscent of the proof of the at most \( k \)-level bound of Clarkson-Shor [CS89], and the proof of the exponential decay lemma of Chazelle and Friedman [CF90].

**Lemma 4.2.** Let \( R \) be a random sample (without replacement) of size \( r \) from a set of \( H \) of \( n \) planes in \( \mathbb{R}^3 \), we have \( \mathbb{E}\left[ \sum_{v \in V_{\leq k}(R)} b_v \right] = O(nk^3) \).

**Proof:** From the sake of simplicity of exposition, let us assume that the sampling here is done by picking every element into the random sample \( R \) with probability \( r/n \). Doing the computations below using sampling without replacement (so we get the exact size) requires modifying the calculations so that the probabilities are stated using binomial coefficients — this makes the calculation messier, but the results remain the same. See [Sha03] for further discussion of this minor issue.

\(^5\)A function \( f(n) \) is a **polynomially growing** function, if (i) \( f(\cdot) \) is monotonically increasing, (ii) for any integers \( i, n \geq 1 \), \( f(i \cdot n) = i^{O(1)} f(n) \). This holds for example if \( f(n) \) is a constant degree polynomial of \( n \), with all its coefficients being positive. Of course, it holds for a much larger family of functions, e.g. \( f(i) = i \log i \).
So, consider taking a second random sample \( R' \) from \( R \). Specifically, sample each plane in \( R \) with probability \( 1/k \). Let us consider the probability that a vertex \( v \in V_{\leq k}(R) \) ends up on the lower envelope of \( R' \). A lower bound can be achieved by the standard argument of Clarkson-Shor. Specifically, if a vertex \( v \) is on the lower envelope then its three defining planes must be in \( R' \) and moreover as \( v \in V_{\leq k}(R) \) by definition there are at most \( k \) planes below it that must not be in \( R' \). So let \( X_v \) be an indicator variable for whether \( v \) appears on the lower envelope of \( R' \), we then have \( \mathbb{E}[X_v] \geq \frac{1}{k^3}(1 - 1/k)^k \geq \frac{1}{e^2 k^3} \).

This implies,

\[
\mathbb{E}_R \left[ \sum_{v \in V_0(R')} b_v \right] = \mathbb{E}_R \left[ \mathbb{E}_R \left[ \sum_{v \in V_0(R')} b_v \mid R \right] \right] = \mathbb{E}_R \left[ \mathbb{E}_R \left[ \sum_{v \in V_{\leq k}(R)} X_v b_v \mid R \right] \right].
\]

(4.1)

Now fix the value of \( R \), and observe that

\[
\mathbb{E}_R \left[ \sum_{v \in V_{\leq k}(R)} X_v b_v \right] = \sum_{v \in V_{\leq k}(R)} \mathbb{E}_R[Y_v b_v] = \sum_{v \in V_{\leq k}(R)} b_v \mathbb{E}_R[X_v] \geq \sum_{v \in V_{\leq k}(R)} \frac{b_v}{e^2 k^3}.
\]

by linearity of expectation, and as \( b_v \) is a fixed quantity for \( v \). Plugging this bound into Eq. (4.1), we have

\[
\mathbb{E}_R \left[ \sum_{v \in V_0(R')} b_v \right] \geq \mathbb{E}_R \left[ \sum_{v \in V_{\leq k}(R)} \frac{b_v}{e^2 k^3} \right] = \frac{1}{e^2 k^3} \mathbb{E}_R \left[ \sum_{v \in V_{\leq k}(R)} b_v \right].
\]

(4.2)

We next bound \( \mathbb{E}_R \left[ \sum_{v \in V_0(R')} b_v \right] \). To this end, observe that \( R' \) is a random sample of \( R \) which itself is a random sample of \( H \), and so we can interpret \( R' \) as a random sample of \( H \) directly\(^6\), and hence we can apply the moments technique. Specifically, each vertex in \( \mathcal{A}(R') \) is defined by the intersection of 3 planes, and let the stopping set of a vertex be any plane in \( H \) which lies strictly below it. As the lower envelope of a set of planes has linear complexity, the moments technique (i.e. Theorem 4.1) then implies

\[
\mathbb{E}_R \left[ \sum_{v \in V_0(R')} b_v \right] = O \left( |R'| \frac{n}{|R'|} \right) = O(n). \text{ And plugging this into Eq. (4.2) implies the claim.} \]

For a subset \( R \subseteq H \), let \( E_{\leq k}(R) \) denote the set of edges of level at most \( k \) in \( A(R) \). Analogous to the vertex case, for an edge \( e \in E_{\leq k}(R) \), \( B(e) \) is set of planes of \( H \) which lie below \( e \) (i.e., there is at least one point on \( e \) that lies above the plane), and \( b_e = |B(e)| \).

**Corollary 4.3.** Let \( R \) be a random sample (without replacement) of size \( r \) from a set of \( H \) of \( n \) planes in \( \mathbb{R}^3 \). We have that \( \mathbb{E}_R \left[ \sum_{e \in E_{\leq k}(R)} b_e \right] = O(nk^3) \).

**Proof:** Under general position assumption every vertex of the arrangement \( \mathcal{A}(H) \) is adjacent to 8 edges. For an edge \( e = vu \), it is easy to verify that \( B(e) \subseteq B(v) \cup B(u) \), and as such we charge the conflict list of \( e \) to its two vertices, and every vertex get charged \( O(1) \) times. Now, the claim follows by Lemma 4.2.

This argument fails to capture edges that are rays in the arrangement, but this is easy to overcome by clipping the arrangement to a bounding box that contains all the vertices of the arrangement. We omit the easy but tedious details. \( \blacksquare \)

---

\(^6\)There are subtle issues here about whether \( R' \) is a sample from \( R \) or \( H \). To avoid confusion, first create a subset \( S \) by sampling each plane in \( H \) with probability \( 1/k \). Then once \( R \) has been determined, let \( R' = R \cap S \).
4.3. Environments and overlays

For a site \( s \in S \), the \( k \) environment of \( s \), denoted by \( \text{env}_k(s, S) \), is the set of all the points in the plane, such that \( s \) is one of their \( k \) nearest neighbors in \( S \), where \( k \) is some fixed value; that is,

\[
\text{env}_k(s, S) = \left\{ p \in \mathbb{R}^2 \mid \| p - s \| \leq d_k(p, S) \right\}.
\]

The following was previously observed by Aurenhammer and Schwarzkoff [AS92].

Claim 4.4. The set \( \text{env}_k(s, S) \) is a star shaped polygon, with respect to the point \( s \).

Proof: Consider the set of all \( n-1 \) bisectors determined by \( s \) and any other site in \( S \). Now consider any point \( p \) in the plane. Clearly \( p \in \text{env}_k(s, S) \) if when walking from \( s \) to \( p \) along their common line one crosses at most \( k-1 \) of these bisectors. Now the star shaped property follows as when walking along any ray emanating from \( s \), the number of bisectors crossed is a monotonically increasing function of distance from \( s \). Moreover, \( \text{env}_k(s, S) \) is a polygon as its boundary is composed of subsets of straight line bisectors.

Observation 4.5. Let \( S = \{s_1, \ldots, s_n\} \). The overlay of the polygons \( \text{env}_k(s_1, S), \ldots, \text{env}_k(s_n, S) \) produces the \( k \)th order Voronoi diagram of \( S \). Indeed, for any point \( p \) in the plane, if \( s \) is one of \( p \)'s \( k \) nearest sites, then by definition it is covered by \( \text{env}_k(s, S) \), and conversely if it is covered by \( \text{env}_k(s, S) \) then \( s \) is one of \( p \)'s \( k \) nearest neighbors.

Going back to our original problem, let \( k = O\left( \log^d n \right) \), and let \( T = \langle s_1, \ldots, s_n \rangle \) be the permutation of \( S \) by increasing parametric volume. We use \( T_i \) (resp. \( S_i \)) to denote the ordered (resp. unordered) \( i \)th prefix of \( T \). Let \( \text{env}_i = \text{env}_k(s_i, S_i) \), that is \( \text{env}_i \) is the union of all Voronoi cells in \( \mathcal{V}_k(S_i) \) that \( s_i \) participates in. Let \( \mathcal{A} = \mathcal{A}(\text{env}_1, \ldots, \text{env}_n) \) denote the arrangement determined by the overlay of the polygons \( \text{env}_1, \ldots, \text{env}_n \), and let \( |\mathcal{A}| \) denote its combinatorial complexity.

Observation 4.6. For any two points in the same cell of \( \mathcal{A} \), their proxy set is the same. This follows immediately from the definition of proxy set.

4.4. Putting it together

Our goal here is to bound \( |\mathcal{A}| \), and our proof is similar in spirit to the argument of [HR14].

First we need the following helper lemmas.

Lemma 4.7. Let \( L \) be a set of lines in general position in the plane. Let \( \ell \) be any line in \( L \), then at most \( k+2 \) edges from \( E_k(L) \) can lie on \( \ell \), where \( E_k(L) \) denotes the exact \( k \) level of the arrangement of \( L \).

Proof: This claim is well known, and it is included here for the sake of completeness.

One can perform a linear transformation such that \( \ell \) is horizontal and the \( k \)th level is preserved. As we go from left to right along the now horizontal line \( \ell \) (starting at infinity), we may leave and enter the exact \( k \) level multiple times. However, every time we leave and then return to the exact \( k \) level we must intersect a negative slope line. Specifically, both when we leave and return to the \( k \)th level, there must be an intersection with another line. If when leaving, this intersection is with a negative slope line then we are done, so assume it has positive slope. In this case the level on \( \ell \)
decreases as we leave the $k$th level, therefore when we return to the $k$th level, the point of return must be at an intersection with a negative slope line (since only negative slope intersections can increase the level), see figure on the right.

So after leaving and returning to the $k$th level $k + 1$ times, there must be at least $k + 1$ negative slope lines below, which implies that the remaining part of $\ell$ must always be on at least level $k + 1$. □

Lemma 4.8. Let $L$ be a set of $n$ lines in general position in the plane. Fix any arbitrary insertion ordering of the lines in $L$. Let $m$ be the total number of distinct vertices on the exact $k$ level seen over all iterations of this insertion process, then $m = O(nk)$.

Proof: Let $\ell_i$ be the $i$th line inserted, and let $L_i$ be the set of the first $i$ inserted lines. Any new vertex on the exact $k$ level created by this insertion, must lie on $\ell_i$. However, by Lemma 4.7 at most $k + 2$ edges from $E_k(L_i)$ can lie on $\ell_i$. As each such edge has at most two endpoints, the insertion of $\ell_i$ contributes $O(k)$ vertices to the exact $k$ level. The bound now follows by summing over all $n$ lines. □

We are now ready for the main proof.

Lemma 4.9. Let $T = \langle s_1, \ldots, s_n \rangle$ be a random permutation of a set $S$ of points in the plane. Let $k$ be a fixed number. The expected total complexity of the overlay arrangement $A = A(\text{env}_1, \ldots, \text{env}_n)$ is $O(k^4 n \log n)$, where $\text{env}_i = \text{env}_{\ell}(s_i, S_i)$.

Proof: As $A$ is a planar map it suffices to bound the number of edges in the arrangement.

Let arcs(\text{env}_i) be the edges in $E_{\leq k}(T_i)$ that appear on the boundary of $\text{env}_i$ (for simplicity we do not distinguish between edges in $E_{\leq k}(T_i)$ in $\mathbb{R}^3$ and their projection in $\mathbb{R}^2$). Such an edge $e \in$ arcs(\text{env}_i), created in the $i$th iteration, is going to be broken into several pieces in the final overlay arrangement $A$, and let $Z_e$ be the number of such pieces that arise from $e$.

Fix the prefix $S_i$: that is, fix the sites that are the first $i$ sites in the permutation $T$ (but not their internal ordering in the permutation). First observe that since $S_i$ is fixed, $B(e)$ is determined for any $e \in E_{\leq k}(S_i)$. Moreover, $Z_e = O(k b_e)$. To see this, observe that $Z_e$ counts the number of future intersections of $e$ with edges in arcs(\text{env}_j) for any $j > i$. As the edge $e$ is on the exact $k$ level at the time of creation, and edges in arcs(\text{env}_j) are on the exact $k$ level when created in the future, edges in arcs(\text{env}_j) must lie below $e$, i.e. they arise from intersections of planes in $B(e)$. So consider the intersection of all planes in $B(e)$, on the vertical plane containing $e$. On this vertical plane, we now have a set of $b_e$ lines, whose insertion ordering is defined by $T$. Now any edge of arcs(\text{env}_j), for some $j > i$, that intersects $e$ must appear as a vertex on the exact $k$ level at some point during the insertion of these lines. However, by Lemma 4.8 for $b_e$ lines, under any insertion ordering there are at most $O(k b_e)$ vertices that ever appear on the exact $k$ level.

For an edge $e \in E_{\leq k}(S_i)$, let $X_e$ be an indicator variable that is one if $e$ was created in the $i$th iteration, and furthermore, it lies on the boundary of $\text{env}_i$. Observe that $\mathbb{E}[X_e] \leq 4/i$, as an edge appears for the first time in round $i$ only if one of its (at most) four defining sites was the $i$th site inserted.

Let $Y_i = \sum_{e \in \text{env}_i} Z_e X_e$ be the total (forward looking) complexity contribution to the final arrangement $A$ of edges added in round $i$. We thus have

$$
\mathbb{E}[Y_i \mid S_i] = \mathbb{E}\left[\sum_{e \in E_{\leq k}(S_i)} Z_e X_e \mid S_i\right] = \mathbb{E}\left[\sum_{e \in E_{\leq k}(S_i)} O(k b_e X_e) \mid S_i\right] = \sum_{e \in E_{\leq k}(S_i)} O(k b_e) \mathbb{E}[X_e \mid S_i] = O\left(\frac{k}{i} \sum_{e \in E_{\leq k}(S_i)} b_e\right).
$$

12
The total complexity of $A$ is asymptotically bounded by $\sum_i Y_i$, and so by Corollary 4.3 we have

$$E\left[\sum_i Y_i \right] = \sum_i E\left[Y_i \right] = \sum_i E\left[E\left[Y_i \mid S_i \right] \right] = \sum_i O\left(\frac{k}{i} E\left[\sum_{c \in E_{\leq A}(S_i)} b_c \right] \right) = O\left(\sum_i \frac{n k^4}{i} \right) = O\left(k^4 n \log n \right).$$

5. On the complexity of the candidate diagram

We now use the above bound on the complexity of the diagram for the proxy set, as well as our knowledge of the relationship between the candidate set and the proxy set to bound the complexity of the candidate diagram.

Theorem 5.1. Let $S$ be a set of $n$ sites in the plane, where for each $s \in S$ we sample an associated parametric point $p \in [0, 1]^d$. Then the expected (diagram) complexity of the candidate diagram is $O(n \log^{8d+5} n)$

Proof: Let $T = (s_1, \ldots, s_n)$ be the permutation of $S$ by increasing parametric volume. Let $A = A(\text{env}_1, \ldots, \text{env}_n)$ denote the arrangement determined by the overlay of the polygons $\text{env}_1, \ldots, \text{env}_n$, where $\text{env}_i = \text{env}_k(s_i, S_i)$, and $k = \Theta(\log^d n)$.

By Observation 4.6, we know that the proxy set is the same for all points in a fixed face of this arrangement. Moreover, by Lemma 4.9 the expected complexity of this arrangement is $O(k^4 n \log n)$. Triangulating each polygonal cell in this arrangement does not increase this asymptotic bound.

Now Lemma 3.6 implies, with polynomially high probability, that simultaneously for all points in the plane, the proxy set has size $O(k \log n)$, so assume this is the case. Now, for $k = \Omega(\log^d n)$, Lemma 3.8 implies that, with high probability for all points in the plane, the proxy set contains the candidate set, and so assume that this also holds.

Therefore, we have a set of $O(k^4 n \log n)$ triangular faces, such that within each face the exact points in the candidate set may vary as we move around the face, however, the possible different points that can appear in the candidate set come from a fixed set of size $O(k \log n)$ (i.e. the proxy set). From Section 2.2 we have the (trivial) bound that the worst case complexity of the candidate diagram for $m$ sites is $O(m^4)$ (i.e. the number of faces in the arrangement of all bisectors). Therefore the diagram complexity of the candidate diagram in a cell is $O((k \log n)^4)$ (note that as each face has constant complexity, i.e. it is a triangle, clipping the diagram of these sites to the cell does not increase the complexity). Multiplying the number of cells by this diagram complexity bound within each cell yields the desired result.

Theorem 5.2. Let $S$ be a set of $n$ point sites in the plane, where for each $s \in S$ we sample an associated parametric point $p \in [0, 1]^d$. Then the expected explicit description complexity of the candidate diagram is $O(n \log^{9d+5} n)$.

Proof: Theorem 5.1 bounded the diagram complexity, and so all we need is a bound on the size of the candidate set, which readily follows by Corollary 3.2.
6. Backward analysis with high probability

6.1. Framework

Let $P$ be a set of $n$ elements (for example, a set of points in the plane). Let $X \subseteq P$. Let $\mathcal{P}(X)$ be the subset (we call it a property) of $P$ of all points that have some desired property (for example, all the vertices of the convex-hull of $P$).

Lemma 6.1. Let $P$ be a set of $n$ elements, and let $\mathcal{P}(X)$ be a property defined over any subset $X \subseteq P$. Furthermore, assume that there is a fixed $k$, such that if $|X| \geq k$ then $|\mathcal{P}(X)| = k$. Now, consider a random permutation $Q = \langle p_1, \ldots, p_n \rangle$ of $P$, and let $X_i$ be an indicator variable of $p_i \in \mathcal{P}(Q_i)$ for each $i = 1, \ldots, n$. Let $Y = \sum_{i=1}^n X_i$. We have the following:

(A) $\Pr[Y > 12ck \ln n] \leq n^{-6ck}$ for any $c \geq 1$.

(B) The bound in (A) holds under the weaker condition: for all $X \subseteq P$ we have that $|\mathcal{P}(X)| \leq k$.

(C) The bound in (A) holds under the even weaker condition: for each $i = 1, \ldots, n$, we have $|\mathcal{P}(Q_i)| \leq k$ (i.e., the bound on the size of $\mathcal{P}$ is only for these sets, and no other sets).

Proof: (A) Let $\mathcal{E}_i$ be the event that $X_i = 1$, that is, $p_i \in \mathcal{P}(Q_i)$ for each $i$. We now show the events $\mathcal{E}_1, \ldots, \mathcal{E}_n$ are independent.

The insight is to think about the sampling process in a different way, and then the result readily follows. Indeed, we randomly pick a permutation of the given elements, and set the first element to be $p_n$. We then, again, pick a random permutation of the remaining elements and set the first element as the penultimate element (i.e., $p_{n-1}$) in the output permutation. We repeat this process till we generate the whole permutation.

Now, consider $1 \leq i_1 < i_2 < \ldots < i_t \leq n$, and observe that

$$\Pr[\mathcal{E}_{i_1} \mid \mathcal{E}_{i_1} \cap \ldots \cap \mathcal{E}_{i_{t-1}}] = \Pr[\mathcal{E}_{i_t}] = \frac{|\mathcal{P}(Q_{i_t})|}{i_t} = \min\left(\frac{k}{i_t}, 1\right)$$

since by our thought experiment, $\mathcal{E}_{i_t}$ is determined before all the other variables $\mathcal{E}_{i_{t-1}}, \ldots, \mathcal{E}_{i_1}$, and these variables are inherently not affected by this event happening or not. As such, we have by induction that

$$\Pr[\mathcal{E}_{i_1} \cap \mathcal{E}_{i_2} \cap \ldots \cap \mathcal{E}_{i_t}] = \Pr[\mathcal{E}_{i_t} \mid \mathcal{E}_{i_1} \cap \ldots \cap \mathcal{E}_{i_{t-1}}] \Pr[\mathcal{E}_{i_1} \cap \ldots \cap \mathcal{E}_{i_{t-1}}]$$

which implies that these variables are mutually independent. As such, we have, for $Y = \sum_i X_i$, that

$$\mu = \mathbb{E}[Y] \leq k + \sum_{i=k+1}^n \frac{k}{i} \leq k \left(1 + \ln n + 1 - \ln(k + 1)\right) \leq k \left(2 + \ln \frac{n}{k}\right) \leq 2k \ln n.$$

For $t \geq 2e$, we have by Chernoff’s inequality that $\Pr[Y > t\mu] \leq 2^{-t\mu}$. As such, we have that

$$\Pr[Y > 12ck \ln n] = \Pr\left[Y > \frac{12ck \ln n}{\mu} \mu\right] \leq n^{-6ck}.$$

(B) The idea is to extend the given property $\mathcal{P}$ so that it holds for exactly $k$ elements. So, fix an arbitrary ordering $\prec$ on the elements of $\mathcal{P}$. Now given any set $X$ (with $|X| \geq k$), if $|\mathcal{P}(X)| = k$ then
set $\mathcal{P}'(X) = \mathcal{P}(X)$. Otherwise, let $L$ be the $k - |\mathcal{P}(X)|$ smallest elements in $X \setminus \mathcal{P}(X)$ according to the ordering of $\prec$, and set $\mathcal{P}'(P) = \mathcal{P}(X) \cup L$. Clearly, the new property $\mathcal{P}'$ complies with (A), and as such the bound of (A) holds, and clearly it also provides an upper bound to the desired quantity.

(C) Follows readily by observing that the required conditions on $\mathcal{P}$ applies only to the prefix sets $Q_1, \ldots, Q_n$. ■

The result of Lemma 6.1 is known in the context of randomized incremental construction algorithms (see [BCKO08, Section 6.4]). However, the known proof is more convoluted – indeed, if the property $\mathcal{P}(X)$ has different sizes for different sets $X$, then it is no longer true that the $X_i$ variables in the proof of Lemma 6.1 are independent. Thus the padding idea in proving Lemma 6.1 (B) is crucial in making the result more widely applicable.

Example 6.2. To see the power of Lemma 6.1 we provide two easy applications — both results are of course known, and are included here to make it clearer in what settings Lemma 6.1 can be applied. The impatient reader is encouraged to skip this example.

(A) QuickSort: Consider the execution of QuickSort when sorting a set $P$ of $n$ numbers. Specifically, in the $i$th iteration, it randomly picks a number that was not handled yet, pivots based on this number, and then recursively handles the left and then the right subproblems. We conceptually can think about QuickSort as being a randomized incremental algorithm, building up a list of numbers in the order they are used as pivots. Now, let $Q_n = \langle p_1, \ldots, p_n \rangle$ be the random permutation of the numbers used by the above version of QuickSort, and consider a specific element $x \in P$. For a subset $X \subseteq P$, let $\mathcal{P}(X)$ be the set of two numbers in $X$ having $x$ in between them and are closest to each other. Let $X_i$ be the indicator variable of $p_i \in \mathcal{P}(Q_i)$ — that is, $x$ got compared to the $i$th pivot when it was inserted. Clearly, the total number of comparisons $x$ participates in is $\sum X_i$, and by Lemma 6.1 the number of such comparisons is $O(\log n)$ with high probability. Implying that QuickSort takes $O(n \log n)$ with high probability.

(B) Point-location queries in a history DAG: Consider a set of lines in the plane, and build their vertical decomposition using randomized incremental construction. Let $L_n = \langle \ell_1, \ldots, \ell_n \rangle$ be the permutation used by the randomized incremental construction. Given a query point $p$, the point-location time is the number of times the vertical trapezoid containing $p$ changes in the vertical decomposition of $L_i = \langle \ell_1, \ldots, \ell_i \rangle$, as $i$ increases. Thus, let $X_i$ be one if $\ell_i$ is one of the (at most) four lines defining the vertical trapezoid containing $p$ the vertical decomposition of $L_i$. Again, Lemma 6.1 applies and implies that the query time is $O(\log n)$, with high probability. This result is well-known, see [CMS93] and [BCKO08, Section 6.4], but our proof is arguably more direct and cleaner.

6.2. Bounding the size of the staircase

6.2.1. The two dimensional case

Lemma 6.3. Let $\Pi = \{\pi_1, \ldots, \pi_n\}$ be a random permutation of $\{1, \ldots, n\}$ (or more generally any set of $n$ distinct real values), and let $X_i$ be an indicator variable which is 1 if $\pi_i$ is the smallest number in $\Pi_i = \{\pi_1, \ldots, \pi_i\}$, for $i = 1, \ldots, n$. Let $Z = \sum_{i=1}^{n} X_i$, then with probability $\geq 1 - 1/n^{\Omega(1)}$, we have that $Z = O(\log n)$.

Proof: Follows immediately from Lemma 6.1, by settings $\mathcal{P}(\Pi_i)$ to be the minimum number of this prefix. ■
If \( P \) is a set of \( n \) points sampled uniformly at random from the unit square, then if we order the points in \( P \) by increasing x-coordinate, then the staircase points are exactly the points which have the smallest \( y \)-values out of all points in their prefix in this ordering. As the \( x \)-coordinates are sampled uniformly at random, this ordering is a random permutation of the \( y \)-values, and we thus have the following.

**Corollary 6.4.** Let \( c > 1 \) be an arbitrary constant, and let \( P \) be a set of \( n \) points sampled uniformly at random from the unit square, \([0,1]^2\). Then with probability \( \geq 1 - 1/n^c \), the number of staircase points in \( P \) is \( |S(P)| = O(\log n) \).

### 6.2.2. Higher dimensions

**Lemma 6.5.** Let \( t, n \) be parameters, such that \( t \leq n \). Let \( Q = \{q_1, \ldots, q_t\} \) be an ordered set of \( t \) points in \([0,1]^d\), for \( d \geq 2 \), where \( q_i \) is randomly and uniformly picked from \([0,1]^d\). Furthermore, assume that with probability \( \geq 1 - 1/n^c \), for some constant \( c \), we have that \( |S(Q_i)| \leq k = c' \ln^{d-1} n \), for \( i = 1, \ldots, t \), where \( Q_i = \{q_1, \ldots, q_i\} \) and \( c' \) is a constant. Then, the sets \( S_j = \cup_{i=1}^j S(Q_i) \) are of size \( O(c' \ln^d n) \), for \( j = 1, \ldots, t \), and this bound holds with probability \( \geq 1 - 2/n^c \).

**Proof:** By Lemma 6.1, by setting \( P(Q_i) = S(Q_i) \), we have that \( \Pr[|S_i| > 12\Delta k \ln t] \leq t^{-6\Delta k} \), for any \( \Delta \geq 1 \). Setting \( \Delta = \Omega((\ln n)/(\ln t)) \) then implies the claim. \( \blacksquare \)

**Lemma 6.6.** Let \( t, n \) be parameters, such that \( t \leq n \). Let \( P \) be a set of \( t \) points picked randomly, uniformly and independently in \([0,1]^d\), for \( d \geq 2 \) and let \( c > 1 \) be an arbitrary constant. Then, \( |S(P)| \leq c_d \log^{d-1} n \), and this holds with probability larger than \( 1 - 1/n^c \). Here, \( c_d \) is a constant that depends only on \( d \) and \( c \).

**Proof:** For \( d = 2 \) this follows by Corollary 6.4. The argument now follows by induction on dimension, so assume we proved the claim for dimensions \( 2, \ldots, d-1 \).

Now, sort \( P \) by increasing value of the \( d \)-th coordinate, and let \( p_i = (q_i, \ell_i) \) be the \( i \)-th point in this order, for \( i = 1, \ldots, t \). Observe that the points \( q_1, \ldots, q_t \) are uniformly picked in the hypercube \([0,1]^{d-1}\). Now, if \( p_i \) is a minima point of \( P \), then it is a minima point of \( \{p_1, \ldots, p_i\} \). But this implies that \( q_i \) is a minima of \( Q_i = \{q_1, \ldots, q_i\} \). Namely, \( q_i \in S_i = \cup_{i=1}^j S(Q_i) \). Now, by applying Lemma 6.5 on \( S_i \), we have that \( |S(P)| \leq |S_i| = O(\log^d n) \), as claimed. \( \blacksquare \)

The above two lemmas, now imply the following two theorems:

**Theorem 6.7.** Let \( c > 1 \) be an arbitrary constant, and let \( Q = \{q_1, \ldots, q_n\} \) be an ordered set of \( n \) points in \([0,1]^d\), for \( d \geq 2 \), where \( q_i \) is randomly and uniformly picked from \([0,1]^d\). Then, the set \( S_n = \cup_{i=1}^n S(Q_i) \) is of size \( O(\ln^d n) \), and this bound holds with probability \( \geq 1 - 1/n^c \).

**Theorem 6.8.** Let \( P \) be a set of \( n \) points picked randomly, uniformly and independently in \([0,1]^d\), for \( d \geq 2 \) and let \( c > 1 \) be an arbitrary constant. Then, \( |S(P)| \leq c_d \log^{d-1} n \), and this holds with probability larger than \( 1 - 1/n^c \). Here, \( c_d \) is a constant that depends only on \( d \) and \( c \).

**Observation 6.9.** The above theorems hold for any distribution, i.e. these results for the uniform case imply the result for any distribution. Specifically, whether a point is in \( S(P) \) or not only depends on the coordinate orderings of the points and not their actual values. In particular, as the uniform distribution has the property that in each coordinate each permutation of the points (by increasing coordinate value) is equally likely, proving the result for the uniform case implies the result for any other distribution in which each permutation is equally likely in each coordinate.
Remark 6.10. Note that the definition of the staircase can be made with respect to any corner of the hypercube (i.e. this corner would replace the origin in the definitions of dominance, point volume, etc.). Taking the union over all $2^d$ such staircases gives us subset of $P$ on the orthogonal convex hull of $P$. Therefore the above theorem also bounds the number of input points on the orthogonal convex hull. As the vertices on the convex hull of $P$ are a subset of the points in $P$ on the orthogonal convex hull, the above also implies the same bound on the number of vertices on the convex hull.

Acknowledgments. The authors would like to thank Pankaj Agarwal, Ken Clarkson, Nirman Kumar, and Raimund Seidel for useful discussions related to this work.

References

[AAH+13] P. K. Agarwal, B. Aronov, S. Har-Peled, J. M. Phillips, K. Yi, and W. Zhang. Nearest neighbor searching under uncertainty II. In Proc. 32nd ACM Sympos. Principles Database Syst. (PODS), pages 115–126, 2013.

[AHKS13] P. K. Agarwal, S. Har-Peled, H. Kaplan, and M. Sharir. Union of random Minkowski sums and network vulnerability analysis. Invited to the special issue of DCG for SoCG 2013, 2013.

[AS92] F. Aurenhammer and O. Schwarzkopf. A simple on-line randomized incremental algorithm for computing higher order Voronoi diagrams. Internat. J. Comput. Geom. Appl., pages 363–381, 1992.

[BCKO08] M. de Berg, O. Cheong, M. van Kreveld, and M. H. Overmars. Computational Geometry: Algorithms and Applications. Springer-Verlag, 3rd edition, 2008.

[BDHT05] Z.-D. Bai, L. Devroye, H.-K. Hwang, and T.-H. Tsai. Maxima in hypercubes. Random Struct. Alg., 27(3):290–309, 2005.

[BKS01] S. Börzsönyi, D. Kossmann, and K. Stocker. The skyline operator. In ICDE, pages 421–430, 2001.

[BR10a] I. Bárány and M. Reitzner. On the variance of random polytopes. Adv. Math., 225(4):1986–2001, 2010.

[BR10b] I. Bárány and M. Reitzner. Poisson polytopes. Annals. Prob., 38(4):1507–1531, 2010.

[CF90] B. Chazelle and J. Friedman. A deterministic view of random sampling and its use in geometry. Combinatorica, 10(3):229–249, 1990.

[CMS93] K. L. Clarkson, K. Mehlhorn, and R. Seidel. Four results on randomized incremental constructions. Comput. Geom. Theory Appl., 3(4):185–212, 1993.

[CS89] K. L. Clarkson and P. W. Shor. Applications of random sampling in computational geometry, II. Discrete Comput. Geom., 4:387–421, 1989.

[Fel08] A. Feldman. Welfare economics. In S. Durlauf and L. Blume, editors, The New Palgrave Dictionary of Economics. Palgrave Macmillan, 2008.
Lemma A.1. Let $F_d(\Delta)$ be the total measure of the points $p = (p_1, \ldots, p_d)$ in the hypercube $[0, 1]^d$, such that $pv(p) = p_1p_2\cdots p_d \leq \Delta$. That is, $F_d(\Delta)$ is the measure of all points in hypercube with point volume $\leq \Delta$. Then $F_d(\Delta) = \sum_{i=0}^{d-1} \frac{\Delta^i}{i!} \ln \frac{1}{\Delta}$.

Proof: The claim follows by tedious but relatively standard calculations. As such, the proof is included for the sake of completeness. First consider the simpler $d = 2$ case. Here the points whose point volume equals $\Delta$ are defined by the curve $xy = \Delta$. This curve intersects the unit square at the point $(\Delta, 1)$ (see Figure A.1). As $F_d(\Delta)$ is the total volume under this curve in the unit square we have that

$$F_2(\Delta) = \Delta + \int_{x=\Delta}^1 \frac{\Delta}{x} dx = \Delta + [\Delta \ln x]_{x=\Delta}^1 = \Delta + \Delta \ln \frac{1}{\Delta}. $$
Figure A.1: Graph of the function $xy = c$

Extending the argument one dimension higher we have,

$$F_3(\Delta) = \Delta + \int_{x_3=\Delta}^{1} F_2\left(\frac{\Delta}{x_3}\right) dx_3 = \Delta + \int_{x_3=\Delta}^{1} \left(\frac{\Delta}{x_3} + \frac{\Delta}{x_3} \ln \frac{x_3}{\Delta}\right) dx_3$$

$$= \Delta + \Delta \ln \frac{1}{\Delta} + \int_{x_3=\Delta}^{1} \left(\frac{\Delta}{x_3} \ln \frac{x_3}{\Delta}\right) dx_3 = \Delta + \Delta \ln \frac{1}{\Delta} + \left[ \frac{\Delta}{2} \ln^2 \frac{x_3}{\Delta}\right]_{x_3=\Delta} = \Delta + \Delta \ln \frac{1}{\Delta} + \frac{\Delta}{2} \ln^2 \frac{1}{\Delta}$$

More generally, we have

$$F_d(\Delta) = \Delta + \int_{x_d=\Delta}^{1} F_{d-1}\left(\frac{\Delta}{x_d}\right) dx_d = \Delta + \int_{x_d=\Delta}^{1} \left(\sum_{i=0}^{d-2} \frac{\Delta}{i! x_d} \ln^i \frac{x_d}{\Delta}\right) dx_d$$

$$= \Delta + \sum_{i=0}^{d-2} \frac{1}{i!} \left(\int_{x_d=\Delta}^{1} \frac{\Delta}{x_d} \ln^i \frac{x_d}{\Delta} dx_d\right) = \Delta + \sum_{i=1}^{d-1} \frac{\Delta}{i!} \ln^i \frac{1}{\Delta} = \sum_{i=0}^{d-1} \frac{\Delta}{i!} \ln^i \frac{1}{\Delta}. \quad \blacksquare$$