Most Likely Voronoi Diagrams in Higher Dimensions

Nirman Kumar[1], Benjamin Raichel[2], Subhash Suri[3], and Kevin Verbeek[4]

1 **University of Memphis, Memphis**
nkumar8@memphis.edu
2 **University of Texas, Dallas**
bar150630@utdallas.edu
3 **University of California, Santa Barbara**
suri@cs.ucsb.edu
4 **TU Eindhoven**
k.a.b.verbeek@tue.nl

Abstract

The Most Likely Voronoi Diagram is a generalization of the well known Voronoi Diagrams to a stochastic setting, where a stochastic point is a point associated with a given probability of existence, and the cell for such a point is the set of points which would classify the given point as its most likely nearest neighbor. We investigate the complexity of this subdivision of space in d dimensions. We show that in the general case, the complexity of such a subdivision is $\Omega(n^{2d})$ where n is the number of points. This settles an open question raised in a recent (ISAAC 2014) paper of Suri and Verbeek [24], which first defined the Most Likely Voronoi Diagram. We also show that when the probabilities are assigned using a random permutation of a fixed set of values, in expectation the complexity is only $\tilde{O}\left(n^{\lceil d/2 \rceil}\right)$ where the $\tilde{O}(\cdot)$ means that logarithmic factors are suppressed. In the worst case, this bound is tight up to polylog factors.

1998 ACM Subject Classification I.3.5 Computational Geometry and Object Modeling

Keywords and phrases Uncertainty, Lower bounds, Voronoi Diagrams, Stochastic

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2016.

1 Introduction

Voronoi diagrams are a well known data structure in Computer Science. Given a finite set of points S, say in Euclidean d-space, the Voronoi diagram is a partition of space into cells, one for each point of S, such that the cell for each point $s \in S$ contains all points closer to s than to any other point of S. We study this classical data structure in the presence of uncertainty. For example, consider a situation in which a set of facilities (say car repair shops) are modeled as points, and the probability that a particular facility can provide a desired service (repairing your car) is known. A natural question is then, which facility is the most likely to be the nearest facility that can provide the desired service.

While uncertainty has been modeled in several ways in different contexts, we investigate the problem in the presence of *existential uncertainty*. The input in this model is a set $\mathcal{P} = \{p_1, \ldots, p_n\}$ of n stochastic points, where each stochastic point p_i is a tuple (s_i, π_i) where s_i is a regular point in $\mathbb{R}^d$ (we will call them sites, or simply points when it is clear they are not meant as probabilistic points) and π_i is its probability of existence. Consider the product distribution induced by these individual distributions. Under this distribution, we

can compute for any query point x, its most likely nearest neighbor (MLNN). The partition of $\mathbb{R}^d$ where for each p_i we have the associated region of points which would classify it as its most likely nearest neighbor, is called the Most Likely Voronoi Diagram (MLVD). This data structure was introduced in [24], where the authors investigated its properties in the simplest setting of $d = 1$ dimensions. Even in this simple setting, it was not obvious how to bound the worst case complexity of this data structure, but it was shown to be $\Theta(n^2)$. However, the lower bound construction relied on very carefully chosen probability values and location of the sites involved. Under certain conditions, such as the probabilities assigned to the sites coming from a random permutation of a fixed set of values, the authors showed an upper bound of $O(n \log n)$ on the complexity. We investigate the complexity of this data structure in higher dimensions.

Results. Our contributions can be summarized as follows.

- We show that in the worst case, the complexity of the MLVD is $\Omega(n^{2d})$. This settles an open question raised in [24].
- When the probabilities assigned to the stochastic points come from a random permutation on a fixed set of n values, we show that the expected complexity of the MLVD is $\tilde{O}(n^{\lceil d/2 \rceil})$ where the $\tilde{O}(\cdot)$ means that terms poly-logarithmic in n are suppressed. Note that this includes the case when all values are independently sampled from a single distribution (as one can first sample and then randomly permute the values). This generalizes a result of [24]. In the worst case, this bound is tight up to polylog factors.

Related work. The work most closely related to ours is of course the paper [24] which defined the MLVD and investigated upper and lower bounds for it in 1-d. Under the aegis of proximity searching, there has also been work under different uncertainty models [5, 3]. These papers investigate different definitions of closeness in the presence of uncertainty, and thereby the resulting Voronoi diagrams are different. For example the Expected Nearest Neighbor(ENN) Voronoi diagram is defined in [5], and the nonzero Nearest Neighbor Voronoi diagram is studied in [3]. These works focus on the 2-dimensional case, and even there tight bounds on the complexity of the Voronoi diagrams defined are not known. More generally, there has been a lot of work on uncertainty in several communities including databases, machine learning, optimization, and computational geometry [8, 9, 16]. Several fundamental problems have been studied in uncertain settings; see [5, 3] for work on nearest neighbors, [4] for work on range searching, [2] for skylines, and [1] for work on coresets. In the existential uncertainty model there has been a flurry of recent work on convex hulls [6], separability [14, 25] (see also [12] for work on separability in a different model of uncertainty), containment and evasion problems [21], arrangements [22], skylines [10] and optimization problems such as minimum spanning trees, and closest pair problems [17, 18].

For the upper bound, we use the candidate set technique, developed by Har-Peled and Raichel [15], and in particular the notation and background we present to use this technique closely follows that of [15]. Their aim was to bound the expected complexity of the multiplicative Voronoi diagram. While such diagrams differ from the MLVD, the candidate set technique is general enough, that with some modification and generalization, it can also be used to bound the complexity of the MLVD. We remark that while our upper bound proof is a significant contribution of this paper, in particular as it generalizes the bounds of [24] to higher dimensions, the main new technical contribution is our lower bound proof.

2 Preliminaries

Notation. Let S be a set of n points in general position in $\mathbb{R}^d$, which we call ***sites***. Throughout the paper we assume d is a constant. Let Π be a set of n values in the interval $(0, 1)$. The values in Π are indexed in decreasing order, $\pi_1 \geq \ldots \geq \pi_n$.

We use $T = \langle s_1, \ldots, s_n \rangle$ to denote a random permutation of the sites in S. Let $T_i = \langle s_1, \ldots, s_i \rangle$ denote the ***prefix*** of this permutation of length i.

When we care only about what elements appear in a permutation, T, but not their internal ordering, we use the notation $S = \text{set}(T)$ to denote the associated set. As such, $S_i = \text{set}(T_i)$ is the ***unordered prefix*** of length i of T.

We let $[n]$ denote the set $\{0, 1, \ldots, n-1\}$ for any natural number n.

Arrangements. As it will be used throughout the paper, we now define the standard terminology of arrangements (see [7]). A set H of n hyperplanes in $\mathbb{R}^d$, induces a partition of $\mathbb{R}^d$ into connected cells, called the arrangement of H, and denoted $\mathcal{A}(\mathsf{H})$. Specifically, each cell is a maximal connected region of the intersection of a (possibly empty) subset of H, which does not intersect any hyperplane not in this subset. In particular, the d-dimensional cells are the maximal connected subsets of $\mathbb{R}^d$ which do not intersect any hyperplanes in H. The combinatorial complexity of $\mathcal{A}(\mathsf{S})$ is the total number of cells of all dimensions.

2.1 Voronoi diagrams

Let $S = \{s_1, \ldots, s_n\}$ be a set of n point sites in the $\mathbb{R}^d$. For a closed set $Y \subseteq \mathbb{R}^d$, and any point $x \in \mathbb{R}^d$, let $\mathsf{d}(x, Y) = \min_{y \in Y} \|x - y\|$ denote the ***distance*** of x to the set Y. For any two sites $s, r \in S$, we define their ***bisector*** $\beta(s, r)$ as the set of points $x \in \mathbb{R}^d$ such that $\mathsf{d}(x, s) = \mathsf{d}(x, r)$. Clearly, $\beta(s, r)$ is a hyperplane, passing through the midpoint of the segment $[s, r]$ and orthogonal to it.

Each $s \in S$, induces the function $f_s(x) = \mathsf{d}(x, s)$, where x is any point in $\mathbb{R}^d$. For any subset $H \subseteq S$ and any site $s \in H$, the ***Voronoi cell*** of s with respect to H, denoted $\mathcal{V}_{\text{cell}}(s, H)$, is the subset of $\mathbb{R}^d$ whose closest site in H is s, i.e. $\mathcal{V}_{\text{cell}}(s, H) = \left\{ x \in \mathbb{R}^d \;\middle|\; \forall r \in H \;\; f_s(x) \leq f_r(x) \right\}$. Finally, for any subset $H \subseteq S$, the Voronoi diagram of H, denoted $\mathcal{V}(H)$, is the partition of space into Voronoi cells induced by the minimization diagram of the set of functions $\left\{ f_s \;\middle|\; s \in H \right\}$.

2.2 Most Likely Voronoi diagrams

We now consider a set $\mathcal{P}$ of n stochastic points where the ith stochastic point $p_i = (s_i, \pi_i)$, and $\pi_i > 0$.

For a given query point x, let $B_i(x)$ denote the set of sites in the open ball with center x and radius $\|x - s_i\|$. The probability that a site s_i is the nearest neighbor to a query point x is given by the expression

$$\Pi_{nn}(s_i, x) = \pi_i \prod_{s_j \in B_i(x)} (1 - \pi_j) \tag{1}$$

The ***most likely nearest neighbor*** of the query point x is then $MLNN(x) = \arg\max_{s_i \in \mathcal{P}} \Pi_{nn}(p_i, x)$.

For any subset $H \subseteq S$ and any point $s \in H$, the ***most likely Voronoi cell*** of s with respect to H, denoted $\mathcal{M}_{\text{cell}}(s, H)$, is the subset of $\mathbb{R}^d$ whose most likely nearest neighbor in H is s, i.e. $\mathcal{M}_{\text{cell}}(s, H) = \left\{ x \in \mathbb{R}^d \;\middle|\; \forall r \in H \;\; \Pi_{nn}(s, x) \geq \Pi_{nn}(r, x) \right\}$. Finally, for any subset

$H \subseteq S$, the ***most likely Voronoi diagram*** (MLVD) of H, denoted $\mathcal{M}(H)$, is the partition of space into Voronoi cells induced by the maximization diagram of the set of functions $\{\Pi_{nn}(s,\cdot) \mid s \in H\}$. The MLVD is a polyhedral partition of space such that the cell for each site is the union of a set of polyhedral sets, where some of these sets may possibly be open. The cell for a given site is not necessarily a connected set. The ***complexity*** of the MLVD is the total number of faces, over all dimensions, of this polyhedral partition.

Let $g(n)$ denote the worst case complexity of the most likely Voronoi diagram. Consider the arrangement of all the bisectors of the points involved. This has complexity $O(n^{2d})$. Within each cell of the arrangement, the ordering of distances is fixed and as such the probability of each site to be a MLNN is also fixed, where we will resolve ties using a fixed (but arbitrary ordering of the points). Therefore, within a cell the MLNN is the same for all points. Thus, $O(n^{2d})$ is an upper bound on the total complexity of the MLVD. However, the total complexity of the MLVD can be smaller as adjacent cells can merge and the MLVD can be a true coarsening of the arrangement.

3 Bounding the expected complexity of the MLVD

For most of this section, we follow closely the presentation in [15] to develop the relevant machinery. Let S be a set of sites and $T = \langle s_1, \ldots, s_n \rangle$ be a random permutation of the sites in S. Let $\pi_1 \geq \pi_2 \geq \ldots \geq \pi_n$ be a fixed set of n probability values in $(0,1)$. Consider the (random) set of stochastic points $\mathcal{P}$ where the ith stochastic point $p_i = (s_i, \pi_i)$. In this section we show that the expected value of the complexity of the MLVD of $\mathcal{P}$ where the expectation is over the random permutation T, is given by $\tilde{O}(n^{\lceil d/2 \rceil})$ where the $\tilde{O}(\cdot)$ suppresses factors logarithmic in n.

3.1 Candidate sets

▶ Definition 1. Let $T = \langle s_1, \ldots, s_n \rangle$ be an ordered set of n sites in $\mathbb{R}^d$. For any point x in the $\mathbb{R}^d$, the ***candidate set*** of x, denoted by $L(x,T)$, is the set of all sites $s_i \in T$, such that $\|x - s_i\| = \mathsf{d}(x, T_i)$, for $i = 1, \ldots, n$. In words, s_i is in $L(x,T)$ if it is the closest site to x in its prefix T_i.

Suppose we assign probabilities to the sites s_i such that π_i is assigned to s_i, where recall that $\pi_1 \geq \pi_2 \geq \ldots \geq \pi_n$. A prerequisite for a site s_j of S to be the most likely nearest neighbor to x, is that s_j is in the candidate set $L(x,T)$.

▶ **Lemma 2.** *For a point x in $\mathbb{R}^d$, if $MLNN(x) = s_j$, then s_j is in $L(x,T)$, where T is the ordering of S by decreasing probabilities.*

Proof. Let s_j be the most likely nearest neighbor of x, and suppose that $\|x - s_j\| \neq \mathsf{d}(x, T_j)$. This implies there exists some $i < j$ such that $\|x - s_i\| < \|x - s_j\|$. However, by the definition of T, for $i < j$, we have $\pi_i \geq \pi_j$, and so

$$\begin{aligned}\Pi_{nn}(s_j, x) &= \pi_j \prod_{s_k \in B_j(x)}(1 - \pi_k) \leq \pi_i \prod_{s_k \in B_j(x)}(1 - \pi_k) \\ &< \pi_i \prod_{s_k \in B_i(x)}(1 - \pi_k) = \Pi_{nn}(s_i, x),\end{aligned}$$

which is a contradiction. Therefore, s_j must be the closest point to x in its prefix T_j. ◀

Consider a random permutation T of S, and the candidate set as a random variable. Next, we investigate the size of this set for all points of space. We prove that with high probability,

the candidate set is logarithmic in size for all points in space. To this end, we need the following well known fact. The expectation bound is well known from [20, 11], and the high probability result was probably well known before as well, but see [15] for a recent written proof.

▶ **Lemma 3.** *Let $\Pi = \langle \pi_1, \ldots, \pi_n \rangle$ be a random permutation of $\{1, \ldots, n\}$, and let X_i be an indicator variable which is 1 if π_i is the smallest number among $\pi_1, \ldots, \pi_i$, for $i = 1, \ldots, n$. Let $Z = \sum_{i=1}^{n} X_i$, then $Z = O(\log n)$, with high probability (i.e., for any constant $c > 0$, one can choose the constant in the $O(\cdot)$ above such that the probability is at least $\geq 1 - 1/n^c$).*

▶ **Corollary 4.** *Let $\pi_1 \geq \ldots \geq \pi_n$ be a set of n probabilities in $(0, 1)$. Let S be a set of n points in $\mathbb{R}^d$, and let $T = \langle s_1, \ldots, s_n \rangle$ be a random permutation of S. Assign the probability π_i to s_i, for all i. Then simultaneously for all points in $\mathbb{R}^d$, their candidate set for T is of size $O(\log n)$, with high probability.*

Proof. Fix a point $x \in \mathbb{R}^d$. Since $T = \langle s_1, \ldots, s_n \rangle$ is a random permutation of S, the sequence $\|x - s_1\|, \ldots, \|x - s_n\|$ is a random permutation of the distance values from x to the sites in S. Therefore, by the definition of the candidate set and Lemma 3, we have $|L(x, T)| = O(\log n)$ with high probability.

Consider the arrangement of all the bisectors of all the pairs of sites in S of complexity $O(n^{2d})$. Within each face of this arrangement, the candidate set cannot change, for any permutation, since all points in this face have the same ordering of their distances to the sites in S. So pick a representative point for each of the $O(n^{2d})$ faces. For any such representative, with probability $\leq 1/n^c$, the candidate set has $> \alpha \log(n)$ sites, for any constant c of our choosing (where α is a constant that depends only on c). Therefore, by choosing c to be sufficiently large, taking the union bound on the bad events (where a bad event is that the size of the candidate set for some face exceeds $\alpha \log(n)$), and then taking the complement, the claim follows. ◀

3.2 Getting a compatible partition

The goal now is to find a low complexity subdivision of space, such that within each cell of the subdivision the candidate set is fixed. As we know, the arrangement of the bisectors already provides such a subdivision. However, the complexity of this subdivision is high. The main insight is that by using the standard Voronoi diagram one can get such a subdivision, which (A) is sensitive to an ordering of the sites (thus, it can intuitively save on certain permutations over the worst case), and, (B) its complexity in expectation can be bounded by $\tilde{O}(n^{\lceil d/2 \rceil})$. Let K_i denote the Voronoi cell of s_i in the usual Voronoi diagram of the ith prefix $S_i = \{s_1, \ldots, s_i\}$. Let $\mathcal{A}$ denote the arrangement formed by the overlay of the regions $K_1, \ldots, K_n$. The complexity of $\mathcal{A}$, denoted by $|\mathcal{A}|$, is the total number of these faces of all dimensions in the arrangement.

▶ **Lemma 5.** *For any face F of $\mathcal{A} = \mathcal{A}(K_1, \ldots, K_n)$, the candidate set is the same, for all points in F.*

Proof. Consider adding points in the order $s_1, \ldots, s_n$. Initially, before any points are added, all points in $\mathbb{R}^d$ have the same candidate set, namely the empty set. When the site s_i is added, the only points in $\mathbb{R}^d$ whose candidate set changes are those for which s_i is their nearest neighbor in S_i. However, these are precisely the points in the Voronoi cell of s_i in the usual Voronoi diagram of S_i. That is, the candidate set changes only for the points covered by K_i.

The claim now easily follows, as $\mathcal{A}$ is the overlay arrangement of these regions. ◀

▶ **Theorem 6.** *Let $\pi_1 \geq \ldots \geq \pi_n$ be a set of n probabilities in $(0,1)$. Let S be a set of n points in $\mathbb{R}^d$, and let $T = \langle s_1, \ldots, s_n \rangle$ be a random permutation of S, where probability π_i is assigned to s_i, for all i. Let $K_i = \mathcal{V}_{\text{cell}}(s_i, T_i)$, for $i = 1, \ldots, n$. Finally, let $\mathcal{A} = \mathcal{A}(K_1, \ldots, K_n)$ be the arrangement formed by the overlay of all these cells.*

Then, the expected complexity of the most likely Voronoi diagram is $O\Big(\mathbf{E}\big[|\mathcal{A}|\big]\, g(\log n)\Big)$, where $|\mathcal{A}|$ is the total complexity of $\mathcal{A}$, and $g(m)$ denotes the worst case complexity of the most likely Voronoi diagram of m sites.

Proof. Let Z be the set of all permutations of S. For any $z \in Z$, let $\mathsf{C}(z)$ denote the size of the largest candidate set of any point in $\mathbb{R}^d$ determined by z. We first argue that for z sampled uniformly at random from Z, $\mathbf{E}[|\mathcal{A}|\, g(\mathsf{C}(z))] = O(\mathbf{E}[|\mathcal{A}|]\, g(\log n))$.

Partition Z into two sets, good and bad, such that for any $z \in Z$, $z \in \mathsf{good}$ if $\mathsf{C}(z) \leq \alpha \log n$, and $z \in \mathsf{bad}$ otherwise, for some constant α. Using Corollary 4, we choose α large enough such that for z sampled uniformly at random from Z, $\mathbf{P}[z \in \mathsf{bad}] \leq 1/n^\beta$, where $\beta = \beta(\alpha)$ is some sufficiently large constant to be determined shortly. We then have:

$$\begin{aligned}
&\mathbf{E}[|\mathcal{A}|\, g(\mathsf{C}(z))] \\
&= \mathbf{E}[|\mathcal{A}|\, g(\mathsf{C}(z)) \mid z \in \mathsf{good}]\, \mathbf{P}[z \in \mathsf{good}] + \mathbf{E}[|\mathcal{A}|\, g(\mathsf{C}(z)) \mid z \in \mathsf{bad}]\, \mathbf{P}[z \in \mathsf{bad}] \\
&\leq \mathbf{E}[|\mathcal{A}|\, g(\alpha \log n) \mid z \in \mathsf{good}]\, \mathbf{P}[z \in \mathsf{good}] + \mathbf{E}[|\mathcal{A}|\, g(\mathsf{C}(z)) \mid z \in \mathsf{bad}]\, / n^\beta \\
&= g(\alpha \log n)\, \mathbf{E}[|\mathcal{A}| \mid z \in \mathsf{good}]\, \mathbf{P}[z \in \mathsf{good}] + \mathbf{E}[|\mathcal{A}|\, g(\mathsf{C}(z)) \mid z \in \mathsf{bad}]\, / n^\beta.
\end{aligned}$$

Now the first term in the above sum is bounded by $g(\alpha \log n)\, \mathbf{E}[|\mathcal{A}|]$ because of the following equality: $\mathbf{E}[|\mathcal{A}| \mid z \in \mathsf{good}]\, \mathbf{P}[z \in \mathsf{good}] = \mathbf{E}[|\mathcal{A}|] - \mathbf{E}[|\mathcal{A}| \mid z \in \mathsf{bad}]\, \mathbf{P}[z \in \mathsf{bad}] \leq \mathbf{E}[|\mathcal{A}|]$. To bound the second term in the sum, observe that both $|\mathcal{A}|$ and $g(\mathsf{C}(z))$ are in the worst case bounded by $O(n^{2d})$, as both the MLVD and $|\mathcal{A}|$ are coarsenings of the arrangement of all bisectors of the sites, which itself has complexity $O(n^{2d})$. Hence by choosing $\beta > 4d$ we can bound the second term of the above expectation as: $\mathbf{E}[|\mathcal{A}|\, g(\mathsf{C}(z)) \mid z \in \mathsf{bad}]\, / n^\beta = O(n^{4d}/n^\beta) = O(1)$. Combining the bounds on each term of the sum we get:

$$\mathbf{E}[|\mathcal{A}|\, g(\mathsf{C}(z))] \leq g(\alpha \log n)\, \mathbf{E}[|\mathcal{A}|] + O(1) = O(\mathbf{E}[|\mathcal{A}|]\, g(\log n)).$$

We now argue that for any fixed $z \in Z$, the corresponding complexity of $\mathcal{M}(S)$ is $O(|\mathcal{A}|\, g(\mathsf{C}(z)))$, which combined with the above bound on the expectation will complete the proof. First decompose all faces (of all dimensions) of $\mathcal{A}$ into constant complexity simplices. (Note that the simplices are constant complexity since d is constant). This can be done by computing a bottom vertex triangulation (see for example [23]). Again since d is assumed to be constant, this triangulation has the same asymptotic complexity as $|\mathcal{A}|$.

Now we have a partition of space into $O(|\mathcal{A}|)$ constant complexity simplices, and by Lemma 5 within each such simplex the candidate set is fixed. So consider such a simplex Δ, and let L be its candidate set. Lemma 2 implies that the only sites whose most likely Voronoi cells can have non-zero area in Δ are the sites in L. That is, the most likely Voronoi diagram restricted to Δ is the intersection of Δ with the most likely Voronoi diagram of some subset of L. Now the most likely Voronoi diagram of $\leq |L|$ points has worst case complexity $g(|L|)$. Since Δ is a constant complexity region this implies that the complexity of the most likely Voronoi diagram in Δ is $O(g(|L|))$. By definition, no candidate set has size more than $\mathsf{C}(z)$, and hence for any simplex Δ, $O(g(|L|)) = O(g(\mathsf{C}(z)))$. Hence the total complexity over all simplices is $O(|\mathcal{A}|\, g(\mathsf{C}(z)))$. ◀

A naive bound on the worst case complexity is $g(m) = O(m^{2d})$. Kaplan *et al.* [19] showed that for a random permutation of n points (as is the case here) the expected total

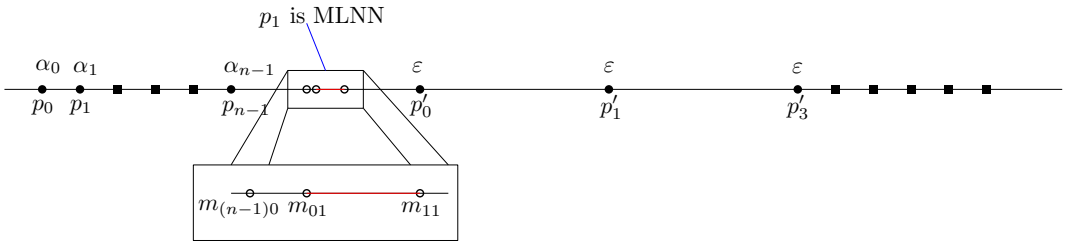

Figure 1 The lower bound example in 1-d with $\Omega(n^2)$ complexity.

complexity of $\mathcal{A}$ is $O(n^{\lceil d/2 \rceil} \log n)$ when d is even, and $O(n^{\lceil d/2 \rceil})$ when d is odd. We therefore readily have the following result.

▶ **Theorem 7.** *Let $\pi_1 \geq \ldots \geq \pi_n$ be a set of n probabilities in $(0, 1)$. Let S be a set of n points in $\mathbb{R}^d$, and let $T = \langle s_1, \ldots, s_n \rangle$ be a random permutation of S, where probability π_i is assigned to s_i, for all i.*

Then the expected complexity of the most likely Voronoi diagram is $O(n^{\lceil d/2 \rceil} \log^{2d+1} n)$ when d is even, and $O(n^{\lceil d/2 \rceil} \log^{2d} n)$ when d is odd.

▶ **Corollary 8.** *Let S be a set of n points in $\mathbb{R}^d$, where independently for each site s we sample a probability value from a single fixed distribution over $(0, 1)$. Then the expected complexity of the most likely Voronoi diagram is $O(n^{\lceil d/2 \rceil} \log^{2d+1} n)$ when d is even, and $O(n^{\lceil d/2 \rceil} \log^{2d} n)$ when d is odd.*

Proof. The distribution induced by choosing the π_i from a fixed distribution over $(0, 1)$ is the same as would be induced by first choosing π_1 then, choosing $\pi_2 \leq \pi_1$ and so on, and then permuting them randomly. Under every random permutation for any fixed choice of the π_i the complexity of the most likely Voronoi diagram is bounded by Theorem 7, and this expression is independent of the choice of the π_i. As such in expectation over the π_i we have the same bound. ◀

Notice that the regular Voronoi diagram of any set of sites is a special case of the MLVD since if all the probabilities are equal, the MLNN is always the nearest neighbor. Our bound above holds for *any* set of sites and a random assignment of probability values from an arbitrary set, while it is known that the worst case complexity of the regular Voronoi diagram is $\Omega(n^{\lceil d/2 \rceil})$ [13], hence it follows that the bound we establish is tight up to polylog factors in the worst case.

4 Worst-case lower bound

In this section we show that $g(n) = \Omega(n^{2d})$. Since the construction is a generalization and uses part of the construction for the lower bound in $d = 1$ dimensions from [24], we briefly recall it here.

The construction of [24] uses two groups of stochastic points each with n points. In the first group S the stochastic points are $p_i = ((i + 1)/n, \alpha_i)$ for $i = 0, 1, \ldots, n - 1$, where in [24] they choose $\alpha_i = 1/(i + 1)$, but as we show different sets of values also work. In the second group T the stochastic points are $p'_i = (2 + j, \varepsilon)$ for ε sufficiently small, for $j = 0, 1, \ldots, n - 1$. They also follow the convention that if two points have the same minimum probability of being the MLNN at a point the one with the smaller index is designated as the MLNN. With this convention they basically show that the

MLNN changes across each of the bisector points of S and T. The $\Omega(n^2)$ bisection points m_{ij} are $m_{ij} = i/(2n) + j/2 + (1 + 1/(2n))$ and these are ordered as per the sequence $m_{00}, m_{10}, \ldots, m_{(n-1)0}, m_{01}, m_{11}, \ldots, m_{(n-1)1}, \ldots, m_{0(n-1)}, m_{1(n-1)}, \ldots, m_{(n-1)(n-1)}$, i.e. in lexicographical order first by j and then by i, see Figure 1 for an illustration. They choose ε so small that a probabilistic point in T can never be the MLNN. For their choice of the $\alpha_i = 1/(i+1)$, it turns out it is sufficient to have ε be so small as to satisfy $(1-\varepsilon)^n/n > \varepsilon$ which is achieved for $\varepsilon \leq 1/n^2$. Next, observe that, for the choice $\alpha_i = 1/(i+1)$, at a point infinitesimally to the left of m_{ij} the probability of p_k being the MLNN is precisely $(1-\varepsilon)^j/n$ for $k \geq i$ and it is $(1-\varepsilon)^{j+1}/n$ for $0 \leq k < i$. Thus, to the left of midpoint m_{ij} the MLNN is p_i and therefore it assumes n values before j increments by 1 at which point it cycles through all of p_i again for $i = 0, \ldots, n-1$.

We need a small modification of the above construction for our proof. Notice that the probabilities for being the MLNN are always of the form $\frac{(1-\varepsilon)^j}{n}$. However, as we show for any p sufficiently small (depending on a function of n) we can choose probabilities for the points in S and T so that the probabilities of the MLNN are always of the form $(1-\varepsilon)^j p$. This can be seen from the following lemma, and by choosing ε small enough, so that a probabilistic point in T is never the MLNN. Intuitively, we can use the α_i for the probabilities of the sites in S in the construction outlined above.

▶ **Lemma 9.** *For any $n \geq 1$ and $0 < \alpha_0 \leq 1/n$, there exist numbers $\alpha_1, \ldots, \alpha_{n-1} \in (0, 1]$ such that*

$$\alpha_0 = \alpha_1(1-\alpha_0) = \alpha_2(1-\alpha_1)(1-\alpha_0) = \ldots = \alpha_{n-1}(1-\alpha_{n-2})\ldots(1-\alpha_0).$$

Proof. Given $\alpha_0 \in (0, 1/n]$ we can define $\alpha_1 = \alpha_0/(1-\alpha_0)$ and then continue inductively by $\alpha_i = \alpha_{i-1}/(1-\alpha_{i-1})$. This ensures that,

$$\alpha_0 = \alpha_1(1-\alpha_0) = \alpha_2(1-\alpha_1)(1-\alpha_0) = \ldots = \alpha_{n-1}(1-\alpha_{n-2})\ldots(1-\alpha_0).$$

It can be verified that, $\alpha_i = \alpha_0/(1-i\alpha_0)$ and clearly $\alpha_0 \leq \alpha_1 \leq \ldots \leq \alpha_{n-1}$. The condition $\alpha_{n-1} \leq 1$ implies $\alpha_0 \leq 1/n$, and moreover, any such α_0 leads to a valid sequence. ◀

We need another lemma for the proof below.

▶ **Lemma 10.** *For any δ small enough, there exist numbers $\pi_0, \ldots, \pi_{n-1}$ and $\alpha_0, \ldots, \alpha_{n-1}$ all in $(0, 1)$, such that for any ε small enough the following are satisfied:*

(A) $\alpha_0 = \alpha_1(1-\alpha_0) = \alpha_2(1-\alpha_1)(1-\alpha_0) = \ldots = \alpha_{n-1}(1-\alpha_{n-2})\ldots(1-\alpha_0)$.
(B) $\pi_0 = \pi_1(1-\pi_0)(1-\delta) = \pi_2(1-\pi_1)(1-\pi_0)(1-\delta)^2 = \ldots = \pi_{n-1}(1-\pi_{n-2})\ldots(1-\pi_0)(1-\delta)^{n-1}$.
(C) $\varepsilon < (1-\varepsilon)^n\alpha_0$, $\delta < \pi_0(1-\delta)$, $\pi_0 > \alpha_0$, *and* $\pi_0(1-\delta) < (1-\varepsilon)^n\alpha_0$, *i.e.,* $[(1-\varepsilon)^n\alpha_0, \alpha_0] \subseteq ((1-\delta)\pi_0, \pi_0)$.

Proof. Start with a symbolic δ and π_0 and compute $\pi_1, \ldots, \pi_{n-1}$ iteratively by $\pi_i = \frac{\pi_{i-1}}{(1-\pi_{i-1})(1-\delta)}$ for $i = 1, \ldots, n-1$. This recursive definition guarantees that $\pi_0 = \pi_1(1-\pi_0)(1-\delta) = \pi_2(1-\pi_1)(1-\pi_0)(1-\delta)^2 = \ldots = \pi_{n-1}(1-\pi_{n-2})\ldots(1-\pi_0)(1-\delta)^{n-1}$. It can be verified that,

$$\pi_i = \frac{\pi_0}{(1-\pi_0)(1-\delta)^i - \left(\pi_0 \sum_{j=1}^{i-1}(1-\delta)^j\right)}$$

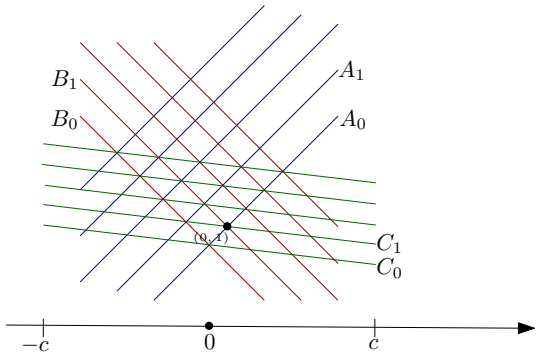

Figure 2 The grid of lines A_i, B_j, C_k.

It can be seen that $\pi_0 < \pi_1 < \ldots < \pi_{n-1}$. The constraint $\pi_{n-1} \leq 1$ gives us,

$$\pi_0 \leq \frac{(1-\delta)^{n-1}}{1 + (1-\delta) + \ldots + (1-\delta)^{n-1}}$$

When $\delta \to 0$ the upper bound for π_0 goes to $1/n$ and when δ increases the upper bound decreases. By choosing a very small δ and π_0 to be the upper bound we get for it, we can ensure $1/n > \pi_0(1-\delta) > \delta$. The first condition can be satisfied for any small enough $\alpha_0 \leq 1/n$ by Lemma 9. In particular we choose α_0 such that α_0 lies in $((1-\delta)\pi_0, \pi_0)$ and then choose ε small enough so that $\pi_0(1-\delta) < \alpha_0(1-\varepsilon)^n$ as well as $\varepsilon < (1-\varepsilon)^n\alpha_0$. Thus, all the desired conditions can be satisfied, and moreover they are satisfied for every small enough ε. Also, any starting choice of δ which is small enough will work. ◀

▶ **Theorem 11.** *There is a set of $(3d-1)n$ probabilistic points in $\mathbb{R}^d$ whose MLVD has complexity at least $\Omega(n^{2d})$.*

Proof. The proof is an inductive argument. The base case, the result for $d = 1$ was already shown in [24], and is sketched above. We show how the induction step works for $d = 2$ dimensions; the general case is similar and we sketch the details later.

We consider four sets of sites P, Q, Q_l, and Q_r, where $|P| = 2n$ and $|Q| = |Q_l| = |Q_r| = n$. We first explain how we place the sites in $Q_l \cup Q_r \cup Q$. To understand how the sites are placed we need to understand how the distance to them varies from a point $(x, 0)$ on the x-axis. Since it is equivalent to consider squared distance functions we will work with them instead. Consider the function $f_{(a,b)}(x)$ which is the square of the distance function of point (a, b) to point $(x, 0)$ on the x-axis. We have $f_{(a,b)}(x) = (a-x)^2 + b^2$. The graph of each such function is a parabola but when x is small the graph is approximately a straight line. In order to define the placement of sites in $Q_l \cup Q_r \cup Q$ we will choose sites such that the graphs of the corresponding distance functions are approximately the lines of the grid we define below. We let $Q_l = \{(a_0, b_0), \ldots, (a_{n-1}, b_{n-1})\}, Q_r = \{(c_0, d_0), \ldots, (c_{n-1}, d_{n-1})\}, Q = \{(e_0, f_0), \ldots, (e_{n-1}, f_{n-1})\}$ and let the corresponding distance functions be F_i, G_j, H_k i.e., $F_i(x) = f_{(a_i,b_i)}(x), G_j(x) = f_{(c_j,d_j)}(x), H_k(x) = f_{(e_k,f_k)}(x)$ for $i, j, k \in [n]$. It will be clear towards the end of the proof what the values of the coordinates are.

To define the grid, consider the lines A_i, B_j, C_k for $i, j, k \in \{0, \ldots, n-1\}$, where A_i has the equation: $y = x + ci/n + ci/n^2 + M$ where $c > 0$ is a number depending on n we define later, and $M > 0$ is some constant we fix later. The line B_j has the equation $y = -x + cj/n + M$. Notice that the lines $A_0, \ldots, A_{n-1}$ are parallel and equally spaced; so

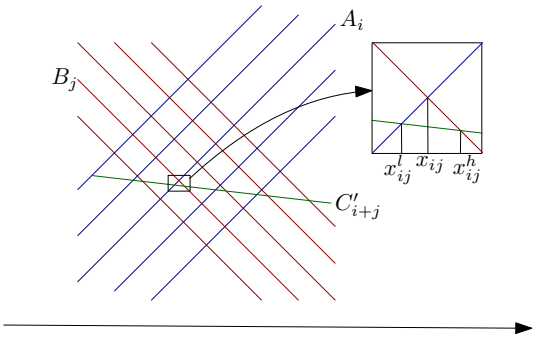

Figure 3 The grid with C_k pushed down to C'_k. In $[x^l_{ij}, x^h_{ij}]$ the line C'_k where $k = i + j$ lies above k of the A_i, B_j otherwise, not near an intersection point it lies strictly above $k + 1$ of them.

are the lines B_j, though they have a different spacing. The intersection point of A_i, B_j is the point (x_{ij}, y_{ij}) where,

$$x_{ij} = \frac{c(j-i)}{2n} - \frac{ci}{2n^2}, \quad y_{ij} = \frac{c(j+i)}{2n} + \frac{ci}{2n^2} + M.$$

All of the x_{ij} lie within $[-c, c]$. Moreover, all the x_{ij} are distinct. We can number the point (x_{ij}, y_{ij}) by (i, j) (see Figure 2). The points (x_{ij}, y_{ij}) for a fixed value of $j - i$ lie on a line with large negative slope.

It turns out that the points for the same value of $(i + j)$ for $0 \leq i + j \leq n - 1$ also lie on lines with small negative slope. These are defined by lines C_k for $k = 0, 1, \ldots, n - 1$ where C_k is defined by the equation: $y = \frac{-1}{2n+1}x + kc\frac{n+1}{n(2n+1)} + M$. It can be verified easily that C_k passes through all the intersection points (x_{ij}, y_{ij}) with $i + j = k$. See Figure 2.

We now slightly push down the lines C_k to C'_k where C'_k is defined by the equation $y = \frac{-1}{2n+1}x + kc\frac{n+1}{n(2n+1)} + M - c'$ where $c' = O(c/n^3)$. Notice that for each intersection point (i, j) there is an interval on the line C'_k cut off by the lines A_i, B_j - this interval is also of length $O(c/n^3)$. Consider the projection of this interval onto the x-axis and denote it by $[x^l_{ij}, x^h_{ij}]$. This interval contains x_{ij}, see Figure 3. It can be verified that the x_{ij} are all separated by at least $c/2n^2$, as such the intervals corresponding to all the x_{ij} are disjoint if c' is chosen $O(c/n^3)$. This property is crucial to us. This grid structure defined by the lines A_i, B_j, C'_k is what we need for the remainder of the proof, and the crucial properties are the following: (i) the intervals for each x_{ij} where $0 \leq i + j \leq n - 1$, i.e., $[x^l_{ij}, x^h_{ij}]$ are all disjoint, and, (ii) for $x \in [x^l_{ij}, x^h_{ij}]$ there are k lines strictly below C'_k where $k = i + j$, while if x is not in such an interval then there are at least $k + 1$ lines among the A_i, B_j strictly below C'_k.

We want the following correspondence between the distance functions F_i, G_j, H_k and the set of lines A_i, B_j, C'_k: $F_i \leftrightarrow A_i, G_j \leftrightarrow B_j, H_k \leftrightarrow C'_k$. Intuitively we assume that the distance functions look like the lines as per the correspondence. The rest of the proof uses precisely the above two properties of the lines, and so we continue the proof assuming these two properties of the distance functions. Unfortunately, distance functions are parabolas and we cannot assume that they will behave like the lines. Fortunately, we can show that if c is small enough, within the interval $[-c, c]$ the distance functions can be made to behave precisely like the lines as desired. Since this issue is somewhat of a technicality we defer the formal demonstration of it to the end of the proof. Thus, in what follows, we assume that the distance functions F_i, G_j, H_k for the points of $Q_l \cup Q_r \cup Q$ behave like the grid of lines A_i, B_j, C'_k and have the properties as desired. We continue calling the intersection points of

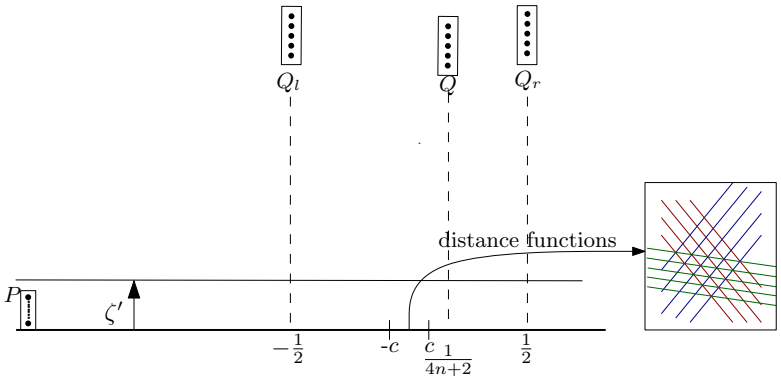

Figure 4 The main construction of the 2d worst case example.

F_i, G_j as x_{ij} and the intervals around them as $[x^l_{ij}, x^h_{ij}]$, where $0 \leq i + j \leq n - 1$.

Recall that as we move from $-c$ to c on the x-axis the graphs of F_i, G_j, H_k indicate the square of distances from sites of $Q_l \cup Q_r \cup Q$ to $(x, 0)$. We will now describe how to place the sites of P – this depends on having the sites of $Q_l \cup Q_r \cup Q$ placed as described. Consider the x-coordinates of all intersections among F_i, G_j, H_k for $x \in [-c, c]$ – this includes all the $x_{ij}, x^l_{ij}, x^h_{ij}$, for $0 \leq i + j \leq n - 1$. Call this set $\mathcal{I}$. By the properties above of the distance functions in $[-c, c]$ all of these points are distinct. We now want that from a point (x, y) where $x \in [-c, c]$ the ordering of distances to sites among $Q_l \cup Q_r \cup Q$ is the same as that for $(x, 0)$. Intuitively, this should be true if y is small enough, but at intersection points of the distance functions it may not be true. So, consider a number ζ much smaller than (say it is 1/10 of) the minimum distance between any two of these points in $\mathcal{I}$. Consider the points of the x-axis between $[-c, c]$ but at least ζ away from all these intersection points [1]. We call this set of points X. Notice that each of $[x^l_{ij}, x^h_{ij}] \cap X$ is still non-empty. For any such point $(x, 0) \in X$ we have the following important property: if $(x, 0) \notin X \cap [x^l_{ij} \cap x^h_{ij}]$ where $i + j = k$ then at least $k + 1$ among the $Q_l \cup Q_r$ lie strictly closer to it than (e_k, f_k), otherwise, if $(x, 0) \in X \cap [x^l_{ij}, x^h_{ij}]$ then only k of them are strictly closer. Moreover, there is a number ζ' depending on ζ such that if we move to a point (x, y) where $x \in X$, and, $-\zeta' \leq y \leq \zeta'$ this property is still true. This number $\zeta' > 0$ will be used below for placement of the sites in P.

The sites of P are arranged according to the lower bound example for $d = 1$ on a vertical line parallel to the y-axis to the left of the origin, such that the sites in $Q_l \cup Q_r \cup Q$ are closer for $x \in [-c, c]$, see Figure 4 for what the placement of sites looks like. (The assigned x-coordinates in Figure 4 are derived later on, when we discuss how the grid lines are approximated by distance functions.) However, they have been "squished" to all lie very close to the x-axis; the 1-d lower bound construction works for any squishing. In particular, the squishing ensures all points lie within the strip of width ζ' around the x-axis. All bisectors of the points among P, which are now lines parallel to the x-axis, lie within the strip as well (these bisector lines are important to the 1-d construction, being the places where the MLNN among the sites of P changes).

We now indicate the probabilities assigned to the sites. Consider $\varepsilon, \delta, \pi_0, \ldots, \pi_{n-1}$, and,

[1] The number ζ is not really significant for this proof, rather a technicality. If we are ζ away from the intersection points, the distance functions have some gap amongst themselves. Then, as we move away from the x-axis to (x, y) for small y, the relative ordering of distances to sites in $Q_l \cup Q_r \cup Q$ from (x, y) remains same as that for $(x, 0)$.

$\alpha_0, \ldots, \alpha_{n-1}$ from Lemma 10. We let all sites in $Q_l \cup Q_r$ have associated probability δ. With the site $(e_k, f_k) \in Q$ we associate π_k, for $k \in [n]$, and to the sites in P we assign the probabilities α_i and ε, for $i \in [n]$, see the 1-d construction above. We fix our attention on $x \in X$. We have the property that if x is within $X \cap [x_{ij}^l, x_{ij}^h]$, there are k sites among $Q_l \cup Q_r$ closer to $(x, 0)$ than (e_k, f_k) for $k = i + j$, otherwise there are at least $k + 1$ of them closer than (e_k, f_k). By the properties of the numbers from Lemma 10, and the properties of distances to sites in $Q_l \cup Q_r \cup Q$ we have the following: If $x \in X \cap [x_{ij}^l, x_{ij}^h]$, then the probability of (e_k, f_k) being a MLNN where $k = i + j$ is precisely $\pi_0 > \alpha_0$. Otherwise, the probability of (e_k, f_k) being a MLNN is at most $\pi_0(1 - \delta)$. Moreover, for a site in $Q_l \cup Q_r$ the probability is $\delta < \pi_0(1 - \delta)$. Recall from the 1-d construction that the probabilities of the MLNN are of the form $\alpha_0(1 - \varepsilon)^j$ for $0 \leq j < n$. Therefore, when $x \in X \cap [x_{ij}^l, x_{ij}^h]$ then (e_k, f_k) is the MLNN where $k = i + j$ since $\pi_0 > \alpha_0$. If $x \in X$ but not within any of the intervals $[x_{ij}^l, x_{ij}^h]$ then the MLNN will be a site in P, since $\pi_0(1 - \delta) < \alpha_0(1 - \varepsilon)^n$ and $\alpha_0(1 - \varepsilon)^n$ is a lower bound on the minimum possible probability for a site in P to be the MLNN.

We now analyze the complexity of the resulting MLVD. Note that P has n^2 bisectors at which the MLNN in P changes. Furthermore, a site $(e_k, f_k) \in Q$ becomes the MLNN briefly at $k + 1$ different "intervals" $X \cap [x_{ij}^l, x_{ij}^h]$ (we will call this a pseudo-interval, since this set is not actually an interval but lies within $[x_{ij}^l, x_{ij}^h]$ and all such pseudo-intervals are disjoint from each other by construction), where $i + j = k$ as argued above. Therefore the most likely nearest neighbor is in Q for $n(n + 1)/2$ different pseudo-intervals, corresponding to the intersection points from the lower half of the grid, see Figure 2. Moreover, between any two of the pseudo-intervals the MLNN lies in P. Since there are n^2 bisectors of P, and each of these bisectors causes a vertex of the MLVD for each of the $n(n + 1)/2$ pseudo-intervals, the total complexity of the MLVD is $\Omega(n^4)$.

For $d > 2$, consider the same construction, but now with P replaced by the lower bound construction for $d - 1$. The sites in $Q_l \cup Q_r \cup Q$ can be placed in the x_1x_2 plane and we reason with the $[-c, c]$ interval on the x_1-axis now. In this case, the strips corresponding to the MLVD for the sites in P, will be replaced by tubes (or higher dimensional versions of it), which may lie all "around" the x_1-axis. However, moving to a point close to x_1-axis in any direction orthogonal to it, will still preserve the ordering of distances as before to sites in Q (away from intersection points as before), i.e., if we are within a ζ' tube of the x_1 axis within the set X then we are in a similar situation as in the 2-d case. We will need a more general version of Lemma 10, but the crucial point to observe is that the possible probability values for a site in P to be MLNN will lie in a very small interval depending on δ, and then one can choose a π_0 and δ such that $(\pi_0(1 - \delta), \pi_0)$ entirely contains this interval. The rest of the construction details are tedious but work similar to the 2-d case. This leads to a total complexity of $\Omega(n^{2d})$ for the MLVD, as by assumption the complexity of the MLVD of P was $\Omega(n^{2d-2})$. The number of points involved in the construction increase by $3n$ for each dimension and start with $2n$, thus only $(3d - 1)n$ points are involved.

In order to finish the proof we still need to show how to approximate the lines by actual distance functions. Consider a line $y = Px + Q$ where $Q \geq P^2/4$. Consider the point $(\beta, \gamma) = (-P/2, \sqrt{Q - P^2/4})$. It can be verified that the distance function $f_{(\beta,\gamma)}$ is $x^2 + Px + Q$. As such if $x \in [-c, c]$ we have that $0 \leq f_{(\beta,\gamma)}(x) - (Px + Q) \leq c^2$ and if c is "small" the parabolas are approximately lines. In particular, consider the grid of Figure 2. As we noticed the important intersection points are (x_{ij}, y_{ij}) for $0 \leq i+j \leq n-1$. Moreover, the x-coordinates of any two intersection points are separated by $\Omega(c/n^2)$. The intervals $[x_{ij}^l, x_{ij}^h]$ are of length $O(c/n^3)$. Consider replacing the lines of the grid by thin strips of

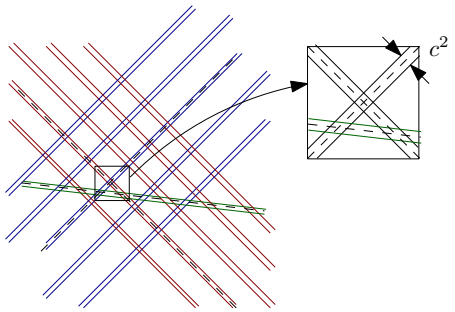

Figure 5 Grid of strips each of width c^2. The dotted lines show the actual distance functions trapped in the strip within $[-c, c]$. The lower strip line in each case is the earlier A_i, B_j or C'_k.

width c^2. If $c = O(1/n^4)$, then $c^2 \ll c/n^2, c/n^3$ and the picture looks like the modified picture of the grid, see Figure 5. As the distance functions lie inside the corresponding strips, the intersection points lie inside the "diamonds" that occur at the intersection of the strips. Interestingly, the x coordinates of the intersection points do not change, as can be verified. Moreover because the width of these strips is $O(c^2)$ the intersection points will be still distinct and the corresponding intervals will be disjoint as well. The distance functions otherwise behave like lines (i.e., they are continuous and intersect at most once), moreover all the F_i are "parallel" and equally spaced, as are the G_j and the H_k, as such the grid induced by them looks like and has the essential properties of the grid of lines A_i, B_j, C'_k of Figure 2 and will suffice for the proof. To fix the coordinates, notice that we can choose the constant M large enough so that $Q > P^2/4$ for each of the lines involved; in particular $M = 1$ works. The corresponding points can be computed by the formulas for β, γ above and it can be easily seen that the x coordinates of all points in Q_l is $-1/2$, in Q_r it is $1/2$ and it is $1/(4n+2)$ for each point of Q, as shown in Figure 4. This completes the proof. ◀

Acknowledgments

Work by N. Kumar and S. Suri was partially supported by NSF grants CCF-1161495 and CCF- 1525817. Work by B. Raichel was partially supported by NSF CRII award CCF-1566137. Work by K. Verbeek was partially supported by the Netherlands Organization for Scientific Research (NWO) under grant number 639.021.541.

References

1 A. Abdullah, S. Daruki, and J. M. Phillips. Range counting coresets for uncertain data. In *Proc. 29th Annu. Sympos. Comput. Geom.*, pages 223–232. ACM, 2013.

2 P. Afshani, P. K. Agarwal, L. Arge, K. G. Larsen, and J. M. Phillips. (Approximate) uncertain skylines. *Theory of Comput. Syst.*, 52:342–366, 2013.

3 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 PODS*, pages 115–126, 2013.

4 P. K. Agarwal, S. W. Cheng, and K. Yi. Range searching on uncertain data. *ACM Trans. on Algorithms*, 8(4):43:1–43:17, 2012.

5 P. K. Agarwal, A. Efrat, S. Sankararaman, and W. Zhang. Nearest-neighbor searching under uncertainty. In *Proc. 31st ACM PODS*, pages 225–236. ACM, 2012.

6 P. K. Agarwal, S. Har-Peled, S. Suri, H. Yıldız, and W. Zhang. Convex hulls under uncertainty. In *Proc. 22nd ESA*, pages 37–48, 2014.

7 P. K. Agarwal and M. Sharir. Arrangements and their applications. In J.-R. Sack and J. Urrutia, editors, *Handbook of Computational Geometry*, pages 49–119. North-Holland, 2000. URL: `http://www.cs.duke.edu/~pankaj/papers/arrangement-survey.ps.gz`.

8 C. C. Aggarwal. *Managing and Mining Uncertain Data*. Springer, 2009.

9 C. C. Aggarwal and P. S. Yu. A survey of uncertain data algorithms and applications. *IEEE TKDE.*, 21(5):609–623, 2009.

10 A. Agrawal, S. Rahul, Y. Li, J. Xue, and R. Janardan. Skyline problems for stochastic points: Algorithms and hardness results. Unpublished manuscript (personal communication), 2015.

11 Z. Bai, L. Devroye, H. Hwang, and T. Tsai. Maxima in hypercubes. *Random Structures & Algorithms*, 27(3):290–309, 2005.

12 M. de Berg, A. D. Mehrabi, and F. Sheikhi. Separability of imprecise points. In *Proc. 14th SWAT*, pages 146–157. Springer, 2014.

13 H. Edelsbrunner. *Algorithms in Combinatorial Geometry*. Springer-Verlag New York, Inc., 1987.

14 M. Fink, J. Hershberger, N. Kumar, and S. Suri. Hyperplane separability and convexity of probabilistic point sets. In *Proc. 32nd Annu. Sympos. Comput. Geom.*, pages 38:1–38:16, 2016.

15 S. Har-Peled and B. Raichel. On the complexity of randomly weighted Voronoi diagrams. In *Proc. 30th Annu. Sympos. Comput. Geom.*, pages 232–241, 2014.

16 A. Jørgensen, M. Löffler, and J. M. Phillips. Geometric computations on indecisive and uncertain points. *CoRR*, abs/1205.0273, 2012.

17 P. Kamousi, T. M. Chan, and S. Suri. Stochastic minimum spanning trees in Euclidean spaces. In *Proc. 27th Annu. Sympos. Comput. Geom.*, pages 65–74, 2011.

18 P. Kamousi, T. M. Chan, and S. Suri. Closest pair and the post office problem for stochastic points. *Comput. Geom. Theory Appl.*, 47(2):214–223, 2014.

19 H. Kaplan, E. Ramos, and M. Sharir. The overlay of minimization diagrams in a randomized incremental construction. *Discrete Comput. Geom.*, 45(3):371–382, 2011.

20 A. I. Kuksa and N. Z. Šor. The method of estimating the number of conditionally optimal trajectories of discrete separable dynamic programming (in russian). *Kibernetika (Kiev)*, 6:37–44, 1972.

21 N. Kumar and S. Suri. Containment and evasion in stochastic point data. In *Proc. 12th Latin Am. Th. Infor. Sympos.*, pages 576–589, 2016.

22 Y. Li, J. Xue, A. Agrawal, and R. Janardan. On the arrangement of stochastic lines in $\mathbb{R}^2$. Unpublished manuscript (personal communication), 2015.

23 J. Matoušek. *Lectures on Discrete Geometry*, volume 212 of *Grad. Text in Math.* Springer, 2002. URL: `http://kam.mff.cuni.cz/~matousek/dg.html`.

24 S. Suri and K. Verbeek. On the most likely voronoi diagram and nearest neighbor searching. In *Proceedings of the 25th Int. Sympos. Alg. and Comp. (ISAAC 2014)*, pages 338–350, 2014.

25 J. Xue, Y. Li, and R. Janardan. On the separability of stochastic geometric objects, with applications. In *Proc. 32nd Annu. Sympos. Comput. Geom.*, pages 62:1–62:16, 2016.