Average case complexity of Voronoi diagrams of n sites from the unit cube*

Marcin Bienkowski† (young@upb.de) Valentina Damerow† (vio@upb.de)
Friedhelm Meyer auf der Heide† (fmadh@upb.de) Christian Sohler† (csohler@upb.de)

Abstract

We consider the expected number of Voronoi vertices (or number of Delaunay cells for the dual structure) for a set of n i.i.d. random point sites chosen uniformly from the unit d-hypercube $[0,1]^d$. We show an upper bound for this number which is linear in n, the number of random point sites, where d is assumed to be a constant. This result matches the trivial lower bound of n.

This is an open problem since several years. In 1991, Dwyer [2] showed that for a uniform distribution from the unit d-ball the average number of Voronoi vertices is linear in n and it is commonly assumed that this holds for any reasonable probability distribution.

1 Introduction

Voronoi diagrams are a fundamental structure in several fields of science besides mathematics and computer science such as physics, geology, agriculture, geography, etc. Named after the Russian mathematician Voronoi [10] they have been 'reinvented' by other researchers, e.g., by the physicists Wigner and Seits [11] or the meteorologist Thiessen [9].

The Voronoi diagram of a set $\mathcal{S}$ of n points – called sites – partitions space into n regions, one per site. The region of a site s consists of all points that are closer to s than to any other site. The straight-line dual of the Voronoi diagram in the plane and its extension to higher dimensions is called the *Delaunay triangulation*. A triangulation of a set $\mathcal{S}$ of sites is a complete partition of the convex hull of $\mathcal{S}$ into fully dimensional simplices having the sites as vertices. The Delaunay triangulation is the unique triangulation of the set of sites such that the circumsphere of every simplex contains no other site in its interior. The Voronoi diagram can be computed in linear time from the Delaunay triangulation, using the one-to-one correspondence between their faces.

Voronoi diagrams have the great advantage to be a rather simple but quite elegant structure with many extensions obtained by varying metric, sites, environment, and constraints. In computer science they are widely used in clustering, mesh generation, graphics, curve and surface reconstruction, and other applications.

A vast variety of basic and (relatively) simple algorithms exists for their construction such as the plane sweep, the divide-and-conquer, the incremental, and the gift-wrapping algorithm, see also Chapter 20 in the Handbook of Discrete and Computational Geometry [5]. In fact most of these algorithms are actually specialized convex hull algorithms since there is a close connection with convexity. Any $(d+1)$-dimensional convex hull algorithm can be used to compute a d-dimensional Delaunay triangulation. All these algorithms depend in their run time on the number of faces of the Delaunay triangulation. Unfortunately, in d dimensions this number is $\Theta(n^{\lceil d/2 \rceil})$ in the worst case [7, 8] (for the usual diagram with the Euclidean metric).

Recent research attempts to quantify situations when the complexity of the Voronoi diagram is low or when it is high [4]. The average case complexity was considered by Dwyer [2] who showed that for n i.i.d. random point sites chosen uniformly from the unit d-ball the expected number of Delaunay simplices is $\Theta(n)$. It has been conjectured that this bound also holds for any uniform distribution in a convex domain but until now no explicit proofs were given [2, 6].

For further reading on Voronoi diagrams and Delaunay triangulations we refer to the survey by Franz Aurenhammer [1], the book by Herbert Edelsbrunner [3] and Chapter 20 in the Handbook of Discrete and Computational Geometry [5].

2 Average case

In this section we will present an average case analysis for the number of Delaunay cells. Let $\mathcal{P}$ be a set of n i.i.d. random points chosen uniformly from the unit d-hypercube $[0,1]^d$. Let $\mathbf{D}(\mathcal{P})$ be the Delaunay triangulation of $\mathcal{P}$. Generally, we will use that

$$\mathbf{E}\big[\text{number of Delaunay simplices of } \mathbf{D}(\mathcal{P})\big] = \binom{n}{d+1} \cdot \mathbf{Pr}\big[\text{ c-ball}(\Delta) \text{ is empty}\big]$$

*Research is partially supported by DFG grant 872/8-2 and by the DFG Research Training Group GK-693 of the Paderborn Institute for Scientific Computation (PaSCo) and the International Graduate School of Dynamic Intelligent Systems.

†Heinz Nixdorf Institute, Computer Science Department, University of Paderborn, 33102 Paderborn, Germany

where Δ is a random d-simplex, i.e., it is the convex hull of $d+1$ random point sites chosen uniformly from $[0,1]^d$ and c-ball(Δ) is the smallest d-ball enclosing Δ.

Unfortunately, in general it is

$$\mathbf{Pr}\big[\text{ c-ball}(\Delta) \text{ is empty}\big] \neq \big(1 - \text{vol}(\text{c-ball}(\Delta))\big)^{n-(d+1)}$$

for the following reason. All random point sites are from inside $[0,1]^d$ while some part of c-ball(Δ) might lie outside of $[0,1]^d$. Of course, the probability for a random point site to be in the outer part of c-ball(Δ) is equal to 0 and therefore we must not consider the outer part. This causes the main difficulty in our analysis namely to bound the volume of c-ball$(\Delta) \cap [0,1]^d$.

Fortunately, we can show the following crucial lemma though we postpone the proof to Section 3.

Lemma 1 *Let Δ be a random d-simplex, i.e., its $d+1$ vertices are i.i.d. random points chosen uniformly from $[0,1]^d$. Then for any constant $a \in [0,1]$ it is*

$$\mathbf{Pr}\left[\text{vol}\left(\text{ c-ball}(\Delta) \cap [0,1]^d\right) \le a\right] \;\le\; \text{const}_d \cdot a^d$$

where const_d *is a constant depending only on d.*

Based on this lemma we will now establish the main theorem of this section.

Theorem 2 *For n i.i.d. random points sites chosen uniformly from $[0,1]^d$ it holds that*

$$\mathbf{E}\big[\textit{number of Delaunay simplices}\big] \;=\; \mathcal{O}(n) \;.$$

Proof. The main idea of the proof is to consider (classes of) simplices with a 'large' circumball that are very likely to have another point site in their circumball, i.e., these simplices are not Delaunay simplices. Then we show that the remaining simplices with a 'small' circumball are very few.

Let us assume w.l.o.g. that n is a power of 2. Let us consider the $\binom{n}{d+1}$ possible simplices that have $d+1$ of the given n random point sites as vertices. For the simplices with 'large' circumball we define classes $\mathcal{S}_0, \ldots, \mathcal{S}_{\log n-1}$ s.t. for a simplex Δ we have that

$$\Delta \in \mathcal{S}_i \;\Leftrightarrow\; \frac{1}{2^{i+1}} \;<\; \text{vol}\left(\text{ c-ball}(\Delta) \cap [0,1]^d\right) \;\le\; \frac{1}{2^i} \;.$$

From Lemma 1 it follows immediately that

$$\begin{aligned}\mathbf{Pr}\big[\Delta \in \mathcal{S}_i\big] \;&\le\; \mathbf{Pr}\left[\text{vol}\left(\text{ c-ball}(\Delta) \cap [0,1]^d\right) \le \frac{1}{2^i}\right] \\ &\le\; \text{const}_d \cdot \left(\frac{1}{2^i}\right)^d \;.\end{aligned}$$

The probability for a simplex $\Delta \in \mathcal{S}_i$ to be a Delaunay simplex is

$$\begin{aligned}\mathbf{Pr}\big[\text{ c-ball}(\Delta) \text{ is empty} \mid \Delta \in \mathcal{S}_i\big] \;&\le\; \left(1 - \frac{1}{2^{i+1}}\right)^{n-(d+1)} \\ \le\; \left(\frac{1}{e}\right)^{\frac{n-(d+1)}{2^{i+1}}} \;&\le\; \left(\frac{1}{2}\right)^{\frac{n-(d+1)}{2^{i+1}}} \;.\end{aligned}$$

Now we can bound the expected number of Delaunay simplices for each class $\mathcal{S}_i$.

For $0 \le i \le \log n - 1$ it is

$$\begin{aligned}&\mathbf{E}\big[\text{number of Delaunay simplices} \in \mathcal{S}_i\big] \\ &\le\; \binom{n}{d+1} \cdot \mathbf{Pr}\big[\Delta \in \mathcal{S}_i\big] \\ &\qquad\qquad \cdot \mathbf{Pr}\big[\text{ c-ball}(\Delta) \text{ is empty} \mid \Delta \in \mathcal{S}_i\big] \\ &\le\; \binom{n}{d+1} \cdot \text{const}_d \cdot \left(\frac{1}{2^i}\right)^d \cdot \left(\frac{1}{2}\right)^{\frac{n-(d+1)}{2^{i+1}}} \;.\end{aligned}$$

The expected number of Delaunay simplices for all classes $\mathcal{S}_0, \ldots, \mathcal{S}_{\log n-1}$ is

$$\begin{aligned}&\sum_{i=0}^{\log n-1} \mathbf{E}\big[\text{number of Delaunay simplices} \in \mathcal{S}_i\big] \\ &\le\; \binom{n}{d+1} \cdot \text{const}_d \sum_{i=0}^{\log n-1} \left(\frac{1}{2}\right)^{i\cdot d + \frac{n-(d+1)}{2^{i+1}}} \\ &=\; \binom{n}{d+1} \cdot \text{const}_d \sum_{i=0}^{\log n-1} \left(\frac{1}{2}\right)^{(\log n-(i+1))\cdot d + \frac{n-(d+1)}{2^{\log n-(i+1)+1}}} \\ &=\; \binom{n}{d+1} \cdot \text{const}_d \cdot \frac{1}{n^d} \sum_{i=0}^{\log n-1} \left(\frac{1}{2}\right)^{2^i\cdot\left(1-\frac{d+1}{n}\right)-(i+1)\cdot d} \\ &\le\; n \cdot \text{const}_d \cdot \left((d+2)\cdot 2^{(d+3)\cdot d} + 1\right) \;=\; \mathcal{O}(n) \;.\end{aligned}$$

The last step follows immediately if $d+2 \ge \log n - 1$ since

$$\begin{aligned}\sum_{i=0}^{d+2} \left(\frac{1}{2}\right)^{2^i\cdot\left(1-\frac{d+1}{n}\right)-(i+1)\cdot d} \;&\le\; \sum_{i=0}^{d+2} 2^{(i+1)\cdot d} \\ &\le\; (d+2)\cdot 2^{(d+3)\cdot d} \;.\end{aligned}$$

In the other case it is

$$\sum_{i=d+3}^{\log n-1} \left(\frac{1}{2}\right)^{2^i\cdot\left(1-\frac{d+1}{n}\right)-(i+1)\cdot d} \;\le\; \sum_{i=d+3}^{\log n-1} \left(\frac{1}{2}\right)^i \;\le\; 1 \;,$$

where we assume that $n \ge 2\cdot(d+1)$.

The expected number of remaining simplices with 'small' circumball can be bounded using lemma 1, too. Let $\mathcal{S}_{\text{re}}$ denote the set of simplices s.t. for a simplex Δ we have

$$\Delta \in \mathcal{S}_{\text{re}} \;\Leftrightarrow\; \text{vol}\left(\text{ c-ball}(\Delta) \cap [0,1]^d\right) \;\le\; \frac{1}{n} \;.$$

Then it is

$$\begin{aligned}&\mathbf{E}\big[\text{number of simplices } \in \mathcal{S}_{\text{re}}\big] \\ &\le\; \binom{n}{d+1} \cdot \mathbf{Pr}\left[\text{vol}\left(\text{ c-ball}(\Delta) \cap [0,1]^d\right) \le \frac{1}{n}\right] \\ &\le\; n^{d+1} \cdot \text{const}_d \cdot \frac{1}{n^d} \;\le\; n \cdot \text{const}_d \;.\end{aligned}$$

Now we can combine everything and by linearity of expectation we get that

$$\begin{aligned} &\mathbf{E}\big[\text{number of Delaunay cells}\big] \\ \leq\; &\sum_{i=0}^{\log n-1} \mathbf{E}\big[\text{number of Delaunay simplices} \in \mathcal{S}_i\big] \\ &\qquad + \mathbf{E}\big[\text{number of simplices } \in \mathcal{S}_{\text{re}}\big] \\ \leq\; & n \cdot \text{const}_d \cdot \Big((d+2)^{(d+3)\cdot d} + 2\Big) \;=\; \mathcal{O}(n) \ , \end{aligned}$$

which concludes the proof of Theorem 2. □

3 Proof of Lemma 1

Let $p_1, \ldots, p_{d+1} \in [0,1]^d$ be the vertices of simplex $\Delta = \Delta(p_1, \ldots, p_{d+1})$, i.e., Δ is the convex hull of $p_1, \ldots, p_{d+1}$. The volume of c-ball(Δ) is given by $\mathcal{V}_d \cdot r^d$ where $r = r(\Delta)$ is the radius of the circumball of Δ and $\mathcal{V}_d = \frac{\pi^{d/2}}{\Gamma(1+d/2)}$ is the volume of the unit d-ball. We can approximate the radius $r(\Delta)$ and the volume of c-ball(Δ) by the following observation:

Observation 1 *It holds that*

$$\begin{aligned} 2 \cdot r(\Delta) \;&\geq\; \max_{1\leq i<j\leq d+1} \|p_i - p_j\|_2 \\ &\geq\; \max_{1\leq i<j\leq d+1} \|p_i - p_j\|_\infty \\ &=:\; \text{maxwidth}(\Delta) \end{aligned}$$

and therefore it is

$$\text{vol}\left(\text{c-ball}(\Delta)\right) \;\geq\; \mathcal{V}_d \cdot \frac{1}{2^d} \cdot \text{maxwidth}(\Delta)^d \ .$$

In other words we approximate the volume of c-ball(Δ) by a fraction of the volume of a smallest hypercube containing all the point sites $p_1, \ldots, p_{d+1}$, cf. Figure 1.

In a next step we will reformulate our random process. Instead of considering $d+1$ many d-dimensional random variables (= point sites) we will combine the random variables coordinate-wise leading to d sets of $d+1$ random numbers each. In more detail, let us again consider the point sites $p_1, \ldots, p_{d+1} \in [0,1]^d$ where $p_i = (p_i^{(1)}, \ldots, p_i^{(d)})$ for $1 \leq i \leq d+1$. Let $\mathcal{P}_1, \ldots, \mathcal{P}_d$ be the sets s.t. $\mathcal{P}_j = \{p_1^{(j)}, \ldots, p_{d+1}^{(j)}\}$ for $1 \leq j \leq d$ and let

$$\text{width}(\mathcal{P}_j) \;:=\; \max \mathcal{P}_j - \min \mathcal{P}_j$$

denote the maximal distance between two elements in $\mathcal{P}_j$. We can now define the variable maxwidth in another way as

$$\text{maxwidth}(\mathcal{P}_1, \ldots, \mathcal{P}_d) \;:=\; \max_{1\leq j\leq d} \text{width}(\mathcal{P}_j) \ ,$$

which is consistent with the earlier definition, i.e., it is $\text{maxwidth}(\Delta) = \text{maxwidth}(\mathcal{P}_1, \ldots, \mathcal{P}_d)$. (Therefore we sometimes write only maxwidth.)

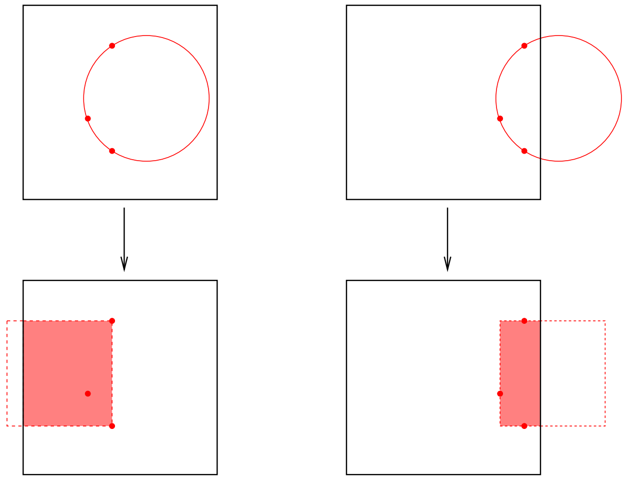

Figure 1: The 2 dimensional case: 3 point sites and their circumcircles in the unit square. Generally, the position of the smallest hypercube containing all point sites is not uniquely defined. When intersected by $[0,1]^d$ consider the hypercube that has smallest intersection volume.

Since we actually want to bound the volume of $\text{c-ball}(\Delta) \cap [0,1]^d$, we consider the (smallest) hypercube containing all point sites that has minimal volume when intersected by $[0,1]^d$. Therefore, we introduce the variable value that indicates how much each dimension contributes to the volume of the minimal intersection between a smallest hypercube containing all the point sites and $[0,1]^d$. If for a fixed dimension the coordinates of all point sites lie close to 0 (or 1) then the dimension contributes less than maxwidth to the volume, namely only the distance of the maximal coordinate to 0 (or the minimal coordinate to 1), cf. also figure 1. In other words, the hypercube then sticks out of $[0,1]^d$ in this dimension.

We define now the value of set $\mathcal{P}_j$ to be

$$\text{value}(\mathcal{P}_j) \;:=\; \begin{cases} \text{maxwidth}(\mathcal{P}_1, \ldots, \mathcal{P}_d) & \textbf{if } \max \mathcal{P}_j - \text{maxwidth} \geq 0 \\ & \text{and } \min \mathcal{P}_j + \text{maxwidth} \leq 1 \\ \min\{\max \mathcal{P}_j, 1 - \min \mathcal{P}_j\} & \textbf{else} \ . \end{cases}$$

With these definitions we can formulate the following lemma.

Lemma 3 *It holds that*

$$\begin{aligned} &\text{vol}\left(\text{c-ball}(\Delta) \cap [0,1]^d\right) \\ &\quad \geq\; \min\left\{\mathcal{V}_d \cdot \left(\frac{1}{2}\right)^{2d}, \frac{1}{d!}\right\} \cdot \prod_{j=1}^{d} \text{value}(\mathcal{P}_j) \ . \end{aligned}$$

Due to space limitations we defer the proof of Lemma 3 to a later full version of this paper.

From now on we will consider the following random process: we have d sets $\mathcal{P}_1, \ldots, \mathcal{P}_d$ of $d+1$ i.i.d. random numbers chosen uniformly from the interval $[0,1]$. From Lemma 3 it follows that if we show that

$$\mathbf{Pr}\left[\prod_{j=1}^{d} \text{value}(\mathcal{P}_j) \leq a\right] = \mathcal{O}\left(a^d\right) , \quad (1)$$

Lemma 1 is also shown.

In order to show (1) we will now establish two lemmas. The first one covers the case that maxwidth is smaller than $\sqrt[d]{a}$, then it follows immediately that $\prod_{j=1}^{d} \text{value}(\mathcal{P}_j) \leq a$. The second lemma covers the case that maxwidth is larger than $\sqrt[d]{a}$ and we have to spend some more effort to show (1).

Lemma 4 *For any value $a \in [0,1]$ it holds that*

$$\mathbf{Pr}\left[\text{maxwidth}(\mathcal{P}_1, \ldots, \mathcal{P}_d) \leq \sqrt[d]{a}\right] \leq \mathcal{O}\left(a^d\right) .$$

Proof. It suffices to bound the probability that $\text{width}(\mathcal{P}_j) \leq \sqrt[d]{a}$ for $1 \leq j \leq d$. For set $\mathcal{P}_j$ we fix the two elements with the maximal distance, i.e., we fix $\max \mathcal{P}_j$ and $\min \mathcal{P}_j$ where the distance between both mustn't exceed $\sqrt[d]{a}$. The remaining $d-1$ elements in $\mathcal{P}_j$ must have values between $\max \mathcal{P}_j$ and $\min \mathcal{P}_j$. Now we can write

$$\mathbf{Pr}\left[\text{width}(\mathcal{P}_j) \leq \sqrt[d]{a}\right] \leq$$
$$(d+1) \cdot d \cdot \int_0^1 \int_{\max\{0, Y - \sqrt[d]{a}\}}^{Y} (Y-X)^{d-1} \, \mathrm{d}X \, \mathrm{d}Y \quad (2)$$

where the outer intergral denotes the range of element $\max \mathcal{P}_j (= Y)$ and the inner integral the range of element $\min \mathcal{P}_j (= X)$. The integration boundaries assure that their distance is at most $\sqrt[d]{a}$. The integrand $(Y-X)^{d-1}$ denotes exactly the probability that all remaining $d-1$ elements of $\mathcal{P}_j$ are between Y and X. The factor before the integral is due to fixing the maximal and minimal element in $\mathcal{P}_j$.

In order to solve this integral we will split it up in the following way to remove the maximum expression from the integration boundary of the inner integral.

$$\begin{aligned}
& \int_0^1 \int_{\max\{0, Y - \sqrt[d]{a}\}}^{Y} (Y-X)^{d-1} \, \mathrm{d}X \, \mathrm{d}Y \\
= \; & \int_0^{\sqrt[d]{a}} \int_0^{Y} (Y-X)^{d-1} \, \mathrm{d}X \, \mathrm{d}Y \\
& + \int_{\sqrt[d]{a}}^1 \int_{Y - \sqrt[d]{a}}^{Y} (Y-X)^{d-1} \, \mathrm{d}X \, \mathrm{d}Y \\
= \; & \frac{1}{d} \cdot \left(\int_0^{\sqrt[d]{a}} Y^d \, \mathrm{d}Y + \int_{\sqrt[d]{a}}^1 a \, \mathrm{d}Y \right) \\
= \; & \frac{1}{d} \cdot \left(\frac{1}{d+1} \cdot a^{\frac{d+1}{d}} + a - a^{\frac{d+1}{d}} \right) \leq \frac{1}{d} \cdot a
\end{aligned}$$

It follows that

$$\mathbf{Pr}\left[\text{width}(\mathcal{P}_j) \leq \sqrt[d]{a}\right] \leq (d+1) \cdot a \quad \Rightarrow$$
$$\mathbf{Pr}\left[\text{maxwidth} \leq \sqrt[d]{a}\right] \leq \mathcal{O}\left(a^d\right) .$$

□

Lemma 5 *For any value of $a \in [0,1]$ it holds that*

$$\mathbf{Pr}\left[\text{maxwidth}(\mathcal{P}_1, \ldots, \mathcal{P}_d) > \sqrt[d]{a} \quad \textit{and} \quad \prod_{j=1}^{d} \text{value}(\mathcal{P}_j) \leq a\right] \leq \mathcal{O}\left(a^d\right) .$$

The proof of Lemma 5 is very involved and rather lengthy. We defer it also to a full version of this paper.

From Lemma 4 and Lemma 5 it follows that Equation (1) holds and thus Lemma 1 is shown.

References

[1] Aurenhammer, F. Voronoi diagrams – a survey of a fundamental geometric data structure. *ACM Comput. Surv. 23* (1991), pp. 345–405.

[2] Dwyer, R. A. Higher-dimensional Voronoi diagrams in linear expected time. In *Proceedings of the 5th Annual ACM Symposium on Computational Geometry* (1989), pp. 326–333.

[3] Edelsbrunner, H. *Algorithms in Combinatorial geometry.* Springer Verlag, Berlin, 1987.

[4] Erickson, J. Nice point sets can have nasty Delaunay triangulations. In *Proceedings of the 17th Annual ACM Symposium on Computational Geometry* (2001), pp. 96–105.

[5] Goodmann J. E. and O'Rourke, J. *Handbook of Discrete and Computational Geometry.* CRC Press LLC, 1997.

[6] Golin, M. J. and Na, H.-S. On the average complexity of 3D-Voronoi diagrams of random points on convex polytopes. *Computational Geometry 25* (2003), pp. 197 231.

[7] Klee, V. On the complexity of d-dimensional Voronoi diagrams. *Archiv. Math. 34* (1980), pp. 75–80.

[8] Seidel, R. On the number of faces in higher-dimensional Voronoi diagrams. In *Proceedings of the 3rd Annual ACM Symposium on Computational Geometry* (1987), pp. 181–185.

[9] Thiessen, A. H. Precipitation averages for large areas. *Monthly Weather Review 39* (1911), pp. 1082–1084.

[10] Voronoi, G. Nouvelles applications des parametres continus à la theorie des formes quadratique, deuxième mémoire: recherches sur les parallélloèdres primitifs. *J. Reine und Angewandte Mathematik 134* (1908), pp. 198–287.

[11] Wigner, E. and Seits, F. On the constitution of metallic sodium. *Physical Reviews 43* (1933), pp. 804–810.