Université Nice Sophia Antipolis

École Doctorale STIC
sciences et technologies de l'information et de la communication

Thèse

pour l'obtention du grade de

Docteur en Sciences

de l'Université Nice Sophia Antipolis

Mention : Informatique

présentée et soutenue par

Rémy Thomasse

Analyse de complexité d'enveloppes convexes aléatoires

Thèse dirigée par Olivier Devillers

soutenue le 18 Décembre 2015

Jury :

Rapporteurs :	Imre Bárány	Hungarian Academy of Sciences
	Christian Sohler	Technische Universität Dortmund
Examinateurs :	Pierre Calka	Université de Rouen
	Philippe Duchon	Université de Bordeaux
	Olivier Devillers	Inria Nancy-Grand Est

Université Nice Sophia Antipolis

École Doctorale STIC

sciences et technologies de l'information et de la communication

PhD Thesis

to obtain the title of

Doctor of Science

of the University Nice Sophia Antipolis

Specialty: Computer Science

prepared and defended by

Rémy Thomasse

Complexity analysis of random convex hulls

supervised by Olivier Devillers

defended on December 18, 2015

Jury:

Reviewers:	Imre Bárány	Hungarian Academy of Sciences
	Christian Sohler	Technische Universität Dortmund
Examiners:	Pierre Calka	Université de Rouen
	Philippe Duchon	Université de Bordeaux
	Olivier Devillers	Inria Nancy-Grand Est

Résumé

Analyses de complexité d'enveloppes convexes aléatoires

Dans cette thèse, nous donnons de nouveaux résultats sur la taille moyenne (c'est à dire le nombre moyen de faces de toutes dimensions) d'enveloppes convexes de points choisis dans un convexe. La taille moyenne de l'enveloppe convexe est connue lorsque les points sont choisis uniformément (et indépendamment) dans un polytope convexe, ou un convexe suffisamment «lisse»; ou encore lorsque les points sont choisis indépendamment selon une loi normale centrée.

Dans la première partie de cette thèse, nous développons une technique nous permettant de donner de nouveaux résultats lorsque les points sont choisis arbitrairement dans un convexe, puis «bruités» par une perturbation aléatoire. Ce type d'analyse, appelée analyse *lissée*, a initialement été développée par Spielman et Teng dans leur étude de l'algorithme du simplexe. Pour un ensemble de points arbitraires dans une boule, nous obtenons une borne inférieure et une borne supérieure de cette complexité *lissée* dans le cas de perturbations uniformes dans une boule en dimension arbitraire , ainsi que dans le cas de perturbations gaussiennes en dimension 2.

Le comportement asymptotique de la taille moyenne d'une enveloppe convexe de points choisis uniformément dans un convexe est *polynomial* pour un convexe «lisse» et *polylogarithmique* pour un polytope. Dans la deuxième partie, nous montrons comment construire un convexe tel que la taille moyenne de l'enveloppe convexe de points choisis uniformément dans ce convexe oscille entre ces deux comportements lorsque le nombre de points augmente.

Dans la dernière partie, nous présentons un algorithme pour engendrer efficacement une enveloppe convexe aléatoire de points choisis uniformément et indépendamment dans un disque, ainsi que l'analyse de sa complexité moyenne en temps et en mémoire. Cet algorithme permet d'obtenir rapidement une enveloppe convexe aléatoire sans avoir à engendrer explicitement tous les points. Il a été implémenté en C++ et intégré dans la bibliothèque CGAL.

Abstract

Complexity analysis of random convex hulls

In this thesis, we give some new results about the average size (that is, the expected number of faces of all dimensions) of convex hulls made of points chosen in a convex body. The average size of the convex hull is known when the points are chosen uniformly (and independently) in a convex polytope or in a "smooth" enough convex body. This average size is also known if the points are independently chosen according to a centered Gaussian distribution.

In the first part of this thesis, we introduce a technique that will give new results when the points are chosen arbitrarily in a convex body, and then noised by some random perturbations. This kind of analysis, called *smoothed* analysis, has been initially developed by Spielman and Teng in their study of the simplex algorithm. For an arbitrary set of point in a ball, we obtain a lower and an upper bound for this *smoothed* complexity, in the case of uniform perturbation in a ball (in arbitrary dimension) and in the case of Gaussian perturbations in dimension 2.

The asymptotic behavior of the expected size of the convex hull of uniformly random points in a convex body is *polynomial* for a "smooth" body and *polylogarithmic* for a polytope. In the second part, we construct a convex body so that the expected size of the convex hull of points uniformly chosen in that body oscillates between these two behaviors when the number of points increases.

In the last part, we present an algorithm to generate efficiently a random convex hull made of points chosen uniformly and independently in a disk. We also compute its average time and space complexity. This algorithm can generate a random convex hull without explicitly generating all the points. It has been implemented in C++ and integrated in the CGAL library.

Remerciements

Je tiens tout d'abord à remercier Olivier Devillers de m'avoir encadré tout au long de cette thèse. Son experience, ses conseils, ses idées, son soutien et sa sympathie m'ont été indispensables.
Je tiens également à remercier mes autres co-auteurs Philippe Duchon, Xavier Goaoc et Marc Glisse, avec qui ce fut un plaisir de travailler.

Je remercie Imre Bàràny, Pierre Calka, Philippe Duchon et Christian Sohler d'avoir accepté de faire partie de mon jury de thèse.

J'aimerais également remercier les gens qui ont partagé mon bureau et qui ont dû supporter ma présence au quotidien : David, Sonali, Mael (ainsi Clément et Manish à mi-temps), mais également tous les membres de l'équipe GEOMETRICA.

Je remercie vivement Virginie pour son aide à la relecture.

Enfin, je remercie tout particulièrement Aurélie pour son soutien quotidien.

Contents

Introduction

In computational geometry, an important interest is to evaluate the size complexity of the geometric objects used by an algorithm. For a given structure, the worst-case can be very pessimistic, making the structure complicated to deal with. However, the worst-case, sometimes, does not reflect the behavior *in practice*. For example, a *realistic* terrain of triangles usually have a linear visibility map complexity, even if the worst-case is quadratic. The same problem appears when dealing with the running time of algorithm: the worst-case of the simplex algorithm is known to be exponential, even if it works well in practice. In these cases, the worst-case complexity does not seem to be a good quality measure of the general behavior.

A first attempt to explain a good behavior in practice is the *average* case. Suppose that the input data are randomly distributed, what is the expected complexity? If the worst cases appears with small probability, you can hope that the expected complexity becomes more optimistic. One of the most famous example for every computer science student is the complexity of the quicksort algorithm: the worst-case is quadratic (for example if the datas are already almost sorted), however if the pivot is chosen uniformly at random, the expected complexity becomes quasilinear.

Even if the average case is sometimes more optimistic, it can still be a bad quality measure, if the random hypothesis is not realistic. Especially, asking the input data to be randomly and independently distributed[1] can be too much to ask when the data are supposed to come from the "real world".

A new notion of complexity, the *smoothed complexity*, has been introduced as an intermediate notion between the worst-case complexity and the average complexity. The main insight is to suppose that the input data are arbitrary (like in the worst-case) but are slightly perturbed by a random noise. Now, computing the smoothed complexity resumes to bound the expected complexity of the noisy data (the expectation being taken according to the noise). It's a kind of interpolation between the worst-case and the average case: taking no perturbations corresponds to the study of the worst-case, and taking a huge perturbation roughly corresponds to the average case (the perturbation is so huge that the initial input value does not play a role anymore). This kind of complexity analysis works well when the

[1] For quicksort, the randomness comes only from the choice of the pivot.

worst-case is really unstable: perturbing the data a little bit decreases the order of magnitude of the complexity. As real data come from measures and are rounded to be stored in a computer, they will be sligthly perturbed so it's really unlikely that the user get a bad input. We describe some results of the smoothed complexity in Section 1.2.

In this thesis, we focus on the complexity analysis of convex hulls. This geometric structure is fairly used in computational geometry, computer graphics, so that's an interesting question to get informations about its number of faces. The worst-case is known and increases exponentially with the dimension, so this is a very pessimistic bound.

Stochastic geometry gives several results on the average complexity of convex hulls, under several random distributions in arbitrary dimension. We describe some of their result in Section 1.1. In the case where the points are initially chosen uniformly in an arbitrary convex set, the exact behavior is known only if the set is a polytope or a smooth convex set. If the convex set is not in these categories, then the expected size of the convex hull is, most of the time, unpredictable. In Chapter 3, we construct such a bad convex set, where the expected size of a convex hull of points uniformly chosen in oscillates asymptotically between two regimes.

We propose, in Chapter 4, an algorithm that generates a random convex hull, made of random points in a disk. A trivial way to do it would be to generate all the points in the disk, and then compute the convex hull. However, once you generate a small number of points, you already know that every point falling inside the convex hull of these first points will not be part of the final convex hull. Our algorithm reduces the number of points to generate, by estimating how many points will fall on a region where we already know that they will not be part of the output.

The main results of the thesis are describe in Chapter 2, where we focus on the smoothed complexity of convex hulls. These results, proved using the *Witness & Collector* technique, give an upper and lower bound on the smoothed complexity for two kind of random perturbations: the uniform distribution in a ball and the Gaussian distribution. The uniform case will be considered in arbitrary dimension, and the Gaussian case will be done in dimension 2. Since the Witness & Collector technique is defined with general settings, we can adapt it to obtain classic stochastic geometry results, but with a very simpler proof.

Publications

- O.Devillers, M.Glisse, X.Goaoc, R.Thomasse. *Smoothed Complexity of Convex Hulls by Witnesses And Collectors* [29]. Submitted, 2015.
- O.Devillers, M.Glisse, X.Goaoc, R.Thomasse. *On the Smoothed Complexity of Convex Hulls*, 31st International Symposium on Computational Geometry (SoCG 2015) [27].
- O.Devillers, M.Glisse, R.Thomasse. *A Chaotic Random Convex Hull*, 25th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2014) [26]. Poster paper.
- O.Devillers, P.Duchon, R.Thomasse. *A Generator of Random Convex Polygons in a Disc*, 25th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2014) [25]. Poster paper.

Chapter 1

State of the Art

In this chapter, we describe some results on stochastic geometry, computational geometry and algorithmics, that are related to the presented work.

Section 1.1 presents some results in stochastic geometry, where we consider the convex hull of points indentically distributed. In particular, we describe the tools and intermediate results used to find the expected number of faces of these convex hulls.

Section 1.2 and Section 1.3 describe the notion of smoothed analysis, and give applications to several problems. We focus on the analysis of algorithms with bad worst-case complexity, as well as the analysis of geometric structures. We also outline the proof for each example.

Outline

1.1 Stochastic Geometry

1.1.1 Sylvester's four-point problem

The Sylvester's four-point problem is, maybe after the Buffon's needle problem, the first and most famous problem in stochastic geometry. In 1865, Sylvester [50] asked the probability, for four points randomly chosen in the plane, to be in convex position. Several answers have been sent with different results. Actually, the result depends on the method chosen to "pick a random point in a plane", so the problem is not well-defined.

Since then, numerous results have been proved, if we restrict the random points to be chosen uniformly in a convex set, for example, so that the probability measure is well defined. Blaschke [11] showed that for any convex set (with non-empty interior) $K \subset \mathbb{R}^2$, the probability, for four points independently and uniformly chosen in K, to be in convex position, is lower bounded by the case where K is a triangle, and upper bounded by the case where K is a disk.

Valtr [52] found the explicit formula for the probability, for n points uniformly chosen in a triangle, to be in convex position. Recently, Marckert [39] found a recursive formula to compute the probability, for n points uniformly chosen in a disk, to be in convex position. The formula is explicit only for n very small. This result shows that even in very simple settings, the problem is still not completely understood.

More generally, the Sylvester problem started the fertile research on the properties of the convex hull of n random points.

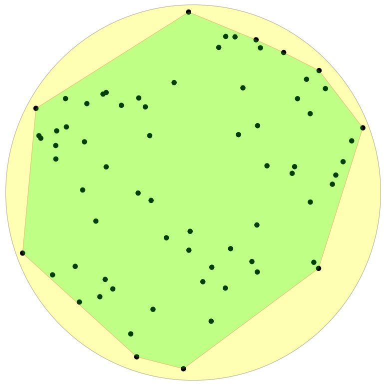
1.1.2 Random polytopes

Let $X_1, \dots, X_n$ be n points independently and identically chosen according to some distribution function.
Let

$$K_n = \mathrm{CH}\,(X_1, \dots, X_n) \tag{1.1}$$

be the convex hull of these random points.
We use the standard notation $f_\ell(K_n)$ to be the number of ℓ- dimensional faces of K_n and $\mathrm{vol}\,(K_n)$ its d-dimensional volume.

Properties of such a polytope (*e.g.* its number of faces or its volume) have been a well studied topic during the last decades and has several applications in algorithmics, optimization, biology...
References for such applications can be find in the introduction of the survey of Reitzner [43]. If several properties have been studied (*e.g.* central limit theorems), in this section we only describe some results about the expected number of faces of K_n.



A random polytope defined as the convex hull of points uniformly chosen in a disk

Convex Hulls of Uniform Points. Let K a compact convex set of $\mathbb{R}^d$ with non-empty interior (such a set is called a *convex body*). We choose the distribution of the points $X_1, \ldots, X_n$ to be uniform in K. The properties of K_n will (in general) depend on the shape of K. However, a very useful formula, proved by Efron [31] in 1965, transforms the computation of the expected number of vertices of K_n into the computation of an expected volume for any arbitrary convex body K:

$$\mathbb{E}\left[f_0(K_n)\right] = \frac{n}{\mathrm{vol}\,(K)}\,\mathbb{E}\left[\mathrm{vol}\,(K \setminus K_{n-1})\right]. \tag{1.2}$$

The first result on the expected volume of K_n (and so on its number of vertices) comes from Rényi and Sulanke in 1963 [44, 45]. They computed the asymptotic behavior of $\mathbb{E}\left[\mathrm{vol}\,(K_n)\right]$ when K is "smooth" and when K is a polygon, in dimension 2.

These results has been generalized for higher dimensions, the best estimation being the one given by Reitzner [42]:

- If K is a *smooth* convex body, i.e K has a twice differentiable boundary with positive Gaussian curvature, for any $\ell \in \{0, \ldots, d-1\}$,

$$\mathbb{E}\left[f_\ell(K_n)\right] = c_1(d, \ell, K) n^{\frac{d-1}{d+1}} + o(n^{\frac{d-1}{d+1}}) \tag{1.3}$$

- If K is a polytope, then

$$\mathbb{E}\left[f_\ell(K_n)\right] = c_2(d, \ell, K) \ln^{d-1} n + o(\ln^{d-1} n) \tag{1.4}$$

where $c_1(d, \ell, K)$ and $c_2(d, \ell, K)$ are positive constants depending only on d, ℓ and on K. The dependence in K can be given explicitly: c_1 depends on the curvature of ∂K, while c_2 depends on the combinatorial structure of K.

Floating Bodies. A way to compute an asymptotical estimation of $\mathbb{E}[f_0(K_n)]$ is the notion of *floating body.* Let the function $v : K \mapsto \mathbb{R}$:

$$v(x) = \min\{\text{vol}(K \cap H) : H \text{ is a half-space and } x \in H\}$$

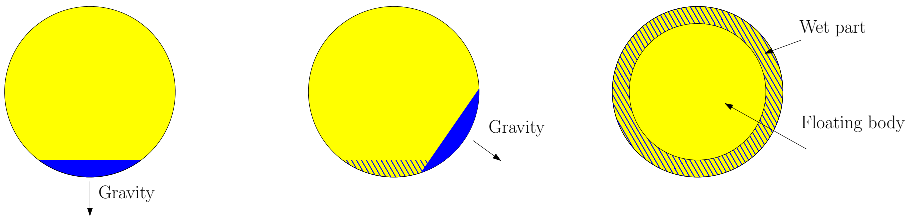
to be the function that, given a point of K, measures the minimum volume of a *cap* containing the point. The *floating body* is the level set

$$K(v \geq t) = \{x \in K \ : v(x) \geq t\}$$

and the *wet part* is

$$K(t) = K(v \leq t) = \{x \in K \ : v(x) \leq t\}.$$

Imagine that K is, say, an inflated balloon containing a small quantity t of water. Thanks to gravity, the part that contains the water will corresponds to some cap $K \cap H$ (where H is a half-space) with volume t. Then, if you turn the balloon on all directions, the water will stand on every possible cap $K \cap H$ with volume t. The union of these "wet" caps corresponds to $K(t)$. The floating body corresponds to the volume inside the balloon that is never wet, and so floats above the water.



Efron's formula (1.2) allows to consider only the expected volume $\mathbb{E}[\text{vol}(K \setminus K_n)]$ in order to compute $\mathbb{E}[f_0(K_{n+1})]$. Bárány and Larman [9], showed that $K\left(\frac{1}{n}\right)$ and $\mathbb{E}[\text{vol}(K \setminus K_n)]$ are related, so computing $\mathbb{E}[f_0(K_n)]$ becomes a pure geometrical problem:

Theorem 1 (Bárány, Larman [9]). *Let $K \subset \mathbb{R}^d$ be a convex body,* $\text{vol}(K) = 1$. *There exist some constants $c_1, c_2(d)$ and $n_1(d)$ such that for $n \geq n_1(d)$,*

$$c_1 \,\text{vol}\left(K\left(\tfrac{1}{n}\right)\right) \leq \mathbb{E}[\text{vol}(K \setminus K_n)] \leq c_2(d)\,\text{vol}\left(K\left(\tfrac{1}{n}\right)\right).$$

As a result, $\mathbb{E}[f_0(K_n)]$ is $\Theta\left(n \,\text{vol}\, K\left(\frac{1}{n}\right)\right)$, where the constants depend only on the dimension.

Economic cap covering An important tool used in the proof of Theorem 1 is the so-called *economic cap covering.* Bàràny and Larman [9] proved that we can cover the wet part of K, without "over covering too much".
More formally:

Theorem 2 (Bárány, Larman [9]). *Let K be a convex body in $\mathbb{R}^d$, $\mathrm{vol}\,(K) = 1$, and $0 < \varepsilon < (d^{2d})^{-1}$. There are caps $C_1, \ldots, C_m$ and pairwise disjoint convex sets $C'_1, \ldots, C'_m$ such that for each i, $C'_i \subset C_i$ and:*

- $\cup_{i=1}^m C'_i \subset K(\varepsilon) \subset \cup_{i=1}^m C_i$,
- $\mathrm{vol}\,(C'_i) = \Omega(\varepsilon)$ *and* $\mathrm{vol}\,(C_i) = O(\varepsilon)$,
- *Every cap C of volume ε is contained in some C_i.*

The caps C_i cover the wet part $K(\varepsilon)$ economically *in the sense that*

$$K(\varepsilon) = \Theta(m\varepsilon).$$

This technique appears to be, in some ways, similar to the *Witness & Collector* technique developped in Section 2.2.

Estimation of the wet part In this paragraph we suppose $\mathrm{vol}\,(K) = 1$ for simplicity. The computation of the volume of the wet part has been done by Schütt and Werner [47] if K is smooth, and by Schütt [48] if K is a polytope:

- If K is a smooth convex body, then for $t \geq 0$,

$$\mathrm{vol}\,(K(t)) = c_3(d, K) t^{\frac{2}{d+1}}(1 + o(1)),$$

- If K is a polytope, then for $t \geq 0$,

$$\mathrm{vol}\,(K(t)) = c_4(d, K) t \left(\ln \frac{1}{t}\right)^{d-1} (1 + o(1)).$$

These results, combined with Theorem 1, give the order of magnitude of $\mathbb{E}\,[f_0(K_n)]$ given in Equation (1.3) and (1.4). It appears that these two behaviors are extremal, since for any convex body K, and for t small enough,

$$K(t) = \Omega\left(t\left(\ln\frac{1}{t}\right)^{d-1}\right)$$

and

$$K(t) = O\left(t^{\frac{2}{d+1}}\right)$$

where the constants involved depends only on the dimension. This result has been proved by Bàràny and Larman [9]. Actually, if K is neither a polytope nor a smooth convex body, the behavior can be proved to be usually unpredictable. We construct such a convex body in Chapter 3.

Convex Hulls of Gaussian Points. Let's now define $X_1, \ldots, X_n$ as n random points in $\mathbb{R}^d$ chosen according to the centered Gaussian distribution with variance - covariance matrix I_d. Again we define K_n as the convex hull of these points. This kind of polytopes are called *Gaussian polytopes*. The expected number of ℓ-faces of K_n has been studied in arbitrary dimensions:

$$\mathbb{E}\left[f_\ell(K_n)\right] = c(d,\ell)\ln^{\frac{d-1}{2}} n\,(1+o(1)) \tag{1.5}$$

where $c(d,\ell)$ is a constant depending only on d and ℓ that can be explicited. This result comes from the work of Rényi and Sulanke [44, 45] and Raynaud [41].

1.2 Smoothed Analysis

1.2.1 Motivation

To understand and predict the practical behavior of an algorithm, a first step is to analyze how the amount of resources it requires grows with the size of the input. The basic building blocks of geometric algorithms are combinatorial structures induced by geometric data such as convex hulls or Voronoi diagrams of finite point sets, lattices of polytopes obtained as intersections of half-spaces, intersection graphs or nerves of families of balls... The size of these structures usually depends not only on the number of geometric primitives (points, half-spaces, balls...), but also on their relative position: for instance, the number of faces of the Voronoi diagram of n points in $\mathbb{R}^d$ is $\Theta(n)$ if these points form a regular grid but $\Theta\left(n^{\lceil d/2\rceil}\right)$ if they lie on the moment curve. (We assume here a *Real RAM* model of computation, so the points have arbitrary real coordinates and the input size is simply the number n of points.)

There are two traditional approaches to account for how the complexity of a structure depends on the position of the points that induce it: the *worst-case complexity*, which measures the maximum of the complexity function over the input space, and the *average-case analysis*, which averages the complexity function against a suitable probability distribution on the space of inputs. Unfortunately, both approaches have shortcomings: the worst-case may be exceedingly pessimistic when the maximum is achieved only by constructions that are so brittle that it is unlikely they arise in practice,[1] whereas the input distributions considered for the average complexity are often unconvincing for lack of relevant and tractable statistical models to work with.

The *smoothed complexity* model, proposed by Spielman and Teng [49] in the early 2000's, interpolates between the worst-case and the average case model.

[1]For instance, while Delaunay triangulations in $\mathbb{R}^3$ have quadratic worst-case complexity, they appear to have near-linear size for the point sets arising in practice in the context of reconstruction [14]; one should thus not consider Delaunay-based reconstruction methods inefficient on the sole ground of worst-case analysis. The worst-case analysis can sometimes be refined by introducing additional parameters such as fatness [19] or spread [32], but *realistic input models* remain elusive in many contexts (*eg.* computer graphics scenes).

Informally, it is defined as the maximum over the inputs of the expected complexity over small perturbations of that input. Intuitively, this "local averaging" mechanism disposes of configurations that vanish under small perturbation and models more accurately the behavior on "real data", which is usually given with bounded precision and subject to measurement noise. In other words, the smoothed complexity quantifies the stability of bad configurations.

If we are given a complexity measure $\mathcal{C}(x)$ for a given input x, and if X_n is the set of possible inputs with size n, then the smoothed complexity function $f(n, \sigma)$ is

$$f(n, \sigma) = \max_{x \in X_n} \mathbb{E}_g \left[\mathcal{C}(x + \sigma g)\right]$$

where σg is a vector of random variables of mean 0 and standard deviation $\sigma\|x\|$; $\|x\|$ being some measure of the magnitude of x. (This measure of magnitude depends on the problem, and can sometimes be simply 1.) Taking $\sigma = 0$ gives the worst-case complexity. On the other side, if we choose σ large enough (so that the contribution of x is negligeable compared to σg), one would expect to obtain the same behavior as the average-case. The smoothed complexity is said to be *polynomial* if $f(n, \sigma)$ is a polynomial in n and $1/\sigma$.

1.2.2 Smoothed Analysis of the Simplex Algorithm

We consider the problem of solving linear programming problems in the form

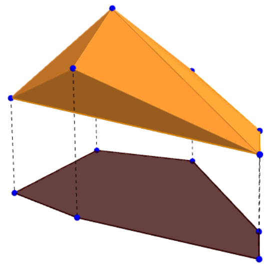
$$\begin{array}{cc} \max & z^t x \\ \text{subject to} & Ax \leq y \end{array} \tag{1.6}$$

using the simplex algorithm, introduced in 1947 by Dantzig [18]. This algorithm is known to work well in practice, even if its number of steps can be exponential in the dimensions of the matrix A [37]. Spielman and Teng [49] investigate the case of arbitrary data noised with small Gaussian noise.

The Simplex Algorithm. The basic idea of the simplex algorithm is to walk over the vertices of the polytope defined by the constrains, so that the objective function $z^t x$ is increasing at each step. If there is a solution to the linear problem, then the algorithm will stop at one vertex of the polytope, which will be the optimal solution for the objective function. At each step, the algorithm has to take a decision in order to choose the next vertex. There are several methods to do that, but the one considered in [49] is called the *Shadow Vertex Method*.

The Shadow Vertex Method. The Shadow Vertex Method consists on projecting the polytope defined by the inequalities of Problem (1.6) onto a plane spanned by the objective function z and another objective function t, such that we know a vertex x of the polytope that optimizes t. This projection (which is a polygon) is called the *shadow*.

We start from the image of x on the shadow, and we walk along the vertices of the shadow, until we reach the vertex whose pre-image vertex on the polytope optimizes the objective function z. As we are only walking on the vertices that are the vertices of the shadow, the running time of this method will depends on the size of the shadow.



Shadow of a polytope

Smoothed Complexity. Spielman and Teng computed the expected running time of the simplex algorithm (with a shadow vertex pivot) for linear programing problems in the form

$$\begin{array}{ll} \max & z^t x \\ \text{subject to} & \langle (A+G)_i, x\rangle \leq (y+h)_i \text{ for } 1 \leq i \leq n \end{array} \quad (1.7)$$

with G and h be a matrix and a vector of independant Gaussian random variables with mean 0 and standard deviation $\sigma \max_i \|(y_i, a_i)\|$.

Theorem 3 (Spielman, Teng [49])**.** *Let $n > d \geq 3$, $\sigma > 0$, $A \in \mathbb{R}^{n\times d}$ and $z \in \mathbb{R}^d$. The expected number of steps for the simplex algorithm to solve the problem* (1.7) *is at most polynomial in n, d and $\frac{1}{\sigma}$.*

Thus, the Simplex algorithm has a polynomial smoothed complexity for Gaussian perturbations.

Outline of the Proof. The first step of the proof is to show that if z and t are independent (non random) vectors, then we can bound the expected size of the shadow of the polytope defined by the noised inequalities of Problem (1.7), projected on the plane spanned by z and t.

Note that in this case, z and t are fixed and do not depend on the data. For the Shadow Vertex Method, this will not be the case, so the first step cannot be applied straightforwardly. The second part of the proof consists of dealing with this issue, using the first part as a black box.

1.2.3 Smoothed Analysis of the Gaussian Elimination

One of the first other result of smoothed analysis has been performed by Sankar, Spielman and Teng [46]. Let $\bar{A}$ an $n \times n$ matrix, let A be a random Gaussian perturbation of $\bar{A}$.

They consider the problem of solving the linear system $Ax = y$ by Gaussian elimination with a final accuracy of b bits, and compute the expected number of bits needed during the elimination. The worst-case is $O(bn)$ bits of precision [35], and we can construct examples that produce large entries [51]. However, in practice, the Gaussian elimination does not usually require more than a double precision. For example, the LAPACK library uses only 64 bits of precision [4]. The authors proved that the expected number of bits of precision to solve the perturbed linear system $Ax = y$ is only $O(b + \ln n)$:

Theorem 4 (Sankar, Spielman, Teng [46]). *For $n > e^4$, let $\bar{A}$ be an $n \times n$ matrix with $\|\bar{A}\|_2 \le 1$. Let $A = \bar{A} + G$ with $G_{i,j} \sim \mathcal{N}(0, \sigma^2)$, with $\sigma^2 \le \frac{1}{4}$.*

Then, the expected number of bits of precision necessary to solve $Ax = y$ to b bits of accuracy using Gaussian elimination without pivoting is at most

$$b + \frac{11}{2}\log_2 n + 3\log_2\left(\frac{1}{\sigma}\right) + \log_2(1 + 2\sqrt{n}\sigma) + \frac{1}{2}\log_2\log_2 n + 6.83$$

Outline of the Proof. The strategy is to use the result of Wilkinson [54], which states that the number of necessary bits is

$$b + \log_2(5n\kappa(A)\rho_U(A) + 3),$$

where:

- LU is the LU-decomposition of A obtained without pivoting,
- $\kappa(A)$ is the condition number of A, $\kappa(A) = \|A\|_2\|A^{-1}\|_2$,
- $\rho_U(A)$ is the growth factor of A in U:

$$\rho_U(A) = \frac{\|U\|_\infty}{\|A\|_\infty},$$

- $\rho_L(A)$ is the growth factor of A in L: $\rho(L) = \|L\|_\infty$.

It remains to show that with high probability, the condition number and the growth factors small when A is a perturbed matrix.

Application: A Robust Algorithm for Linear System. The above result can lead to an algorithm that solves efficiently a linear system [1]. Given a matrix $\bar{A}$, add a Gaussian perturbation with a well chosen standard deviation. Then, solve the perturbed linear system with a well chosen precision. Then, with probability as close to 1 as you want, the perturbed solution will be a solution of the original linear system with an accuracy of b bits:

Theorem 5 (Sankar, Spielman, Teng [1]). *Let $\bar{A}$ be an $n \times n$ matrix and $y \in \mathbb{R}^n$. Let A be the perturbed version of $\bar{A}$ with Gaussian noise of standard deviation $\delta\left(2^{b+3}n^{\frac{1}{2}}\kappa(A)\right)^{-1}$. Let's solve the perturbed system $Ax = y$ using Gaussian elim-*

ination without pivoting, with

$$4b + 10 \ln n + 3 \ln \kappa(A) + 5 \ln \left(\frac{1}{\delta}\right) + 7$$

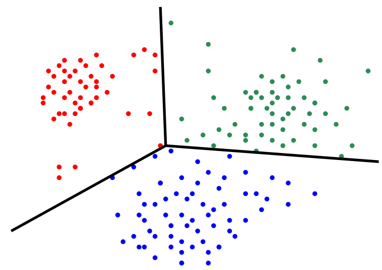
bits of precision. Then, with probability at least $1 - \delta$*, the obtained solution is a solution of* $\bar{A}x = y$ *accurate to* b *bits.*

1.2.4 Smoothed Analysis of the k-means Algorithm

Let $\mathcal{X}$ be a set of n points in $\mathbb{R}^d$. Given a parameter k, the k*-clustering problem* consists of finding a partition of $\mathcal{X}$ into k clusters $C_1, \dots, C_k$ such that every point is "close" to the centroid of its cluster. More formally, an optimal solution would be to find the set of cluster $\mathcal{C}_{opt}$ such that

$$C_{opt} = \underset{\substack{C=\{C_1,\dots,C_k\} \\ \cup_{i=1}^k C_i = \mathcal{X}}}{\arg\min} \phi(C, \mathcal{X}), \qquad \phi(C, \mathcal{X}) = \sum_{i=1}^{k} \sum_{x \in \mathcal{X} \cap C_i} \|x - \mu_i\|^2 \tag{1.8}$$

where μ_i is the centroid of the cluster C_i.
This problem is known to be NP-hard [2, 38], even if k or d is fixed (but if both k and d are fixed, the problem can be solved in polynomial time [36]). The k-means algorithm is a well used iterative heuristic algorithm to solve this problem, even if the optimal solution is not guaranteed. This algorithm is said to be efficient in practice, but the number of iterations, in the worst-case, is exponential in n [53]. Thus, this algorithm is a good candidate for the smoothed analysis.



An example of clustering with $k = 3$

Algorithm. The basic idea of the algorithm is to start from an arbitrary clustering, and then try to modify it so that it decreases the potential function ϕ of Equation (1.8). Once the potential cannot be decreased anymore, the algorithm stops.

1. Select k points $c_1, \dots, c_k$ arbitrary in $\mathbb{R}^d$.
2. $x \in \mathcal{X}$ is in C_i if for all $j \neq i$, $\|x - c_i\| < \|x - c_j\|$
3. $c_i := \frac{1}{\operatorname{card} C_i} \sum_{x \in C_i} x$
4. If the c_i's or the clusters have changed, goto 2, otherwise return $C_1, \dots, C_k$.

Smoothed Analysis. Arthur, Manthey and Röglin [5], performed a smoothed analysis on the k-means algorithm:

Theorem 6 (Arthur, Manthey, Röglin [5]). *If the set of point $\mathcal{X}$ is noised by a Gaussian perturbation of mean 0 and standard deviation σ, then the expected number of iterations of the k-means algorithm is bounded by a polynomial in n, d and $\frac{1}{\sigma}$.*

In other words, the smoothed complexity is polynomial.

Outline of the Proof. The idea of the proof is to have a look on how much the potential function of Equation (1.8) decreases in every sequence of a few consecutive iterations. The authors proved that the minimal improvement appears only with a very small probability.

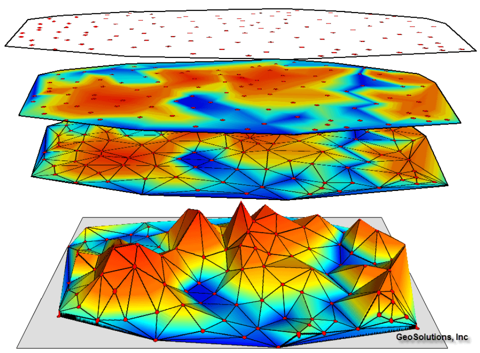
The main insight is to encode each possible iteration in a *transition blueprint*. A transition blueprint is basically a graph[2] where each vertex corresponds to a cluster, and each oriented edge between two clusters corresponds to a point that moves from a cluster to another one. Then, the transition blueprints are classified with only six possible cases (a general case and five special cases that do not fit in the general one), according to some of their combinatorial properties. The main challenge is then, under these combinatorial assumptions, to bound the probability of the smallest improvement to be small.

1.3 Smoothed Analysis of Geometric Structures

1.3.1 Smoothed Analysis of Visibility Map of Realistic Terrains

A *terrain* is a surface obtained by, given a triangulation in the plane, assigning to each vertices an elevation. This geometric structure can be used to model moutainous regions.

Let $\mathcal{T}$ be a terrain of n triangles, and p be a viewing point. The visibility map of $\mathcal{T}$ from p is the projection of the triangles of $\mathcal{T}$, visible from p, onto a plane. We define the complexity of the visibility map as the number of vertices of the visibility map from p (such vertices can be vertices or edge intersection in $\mathcal{T}$). In the worst case, this complexity is known to be quadratic. However, in practice, the complexity seems to be close to linear.



Credits: GeoSolutions, Inc.

de Berg, Haverkort and Tsirogiannis [20] showed that the smoothed complexity of visibility maps, under some random uniform perturbation of the vertices elevation:

[2] A blueprint keeps also informations about the cluster centers.

Theorem 7 (de Berg, Haverkort, Tsirogiannis [20]). *Let $\mathcal{T}$ be a terrain of n triangles. Then, a visibility map of $\mathcal{T}$ under perspective projection has smoothed complexity*

$$O\left(\left(\frac{\tan\theta}{c} + \frac{1}{\sin\phi}\right)\sigma n\right)$$

under the addition of noise to each vertex's elevation that is uniformly distributed in an interval $[-c, c]$, where:

- *The fatness $\phi > 0$ is the smallest angle in the triangles of the underlying triangulation;*
- *The steepness $\theta < \frac{\pi}{2}$ is the largest dihedral angle between any triangle and the horizontal plane,*
- *The scale factor $\sigma \geq 1$ is the length of the longest edge divided by the length of the shortest edge of the triangulation;*
- *The constant c is a fixed constant fraction of the minimum edge length of the triangulation.*

If we assume the fatness, the steepness and the scale factor to be constant in n, the smoothed complexity of visibility maps is linear. These assumptions on the terrain are used in several contexts and seem to be experimentally realistic. However, they are not sufficient, alone, to explain the linear behavior in practice, since the worst-case of visibility maps of realistic terrains is still $\Theta(n\sqrt{n})$.

Outline of the Proof. The number of terrain vertices is already $O(n)$, so it remains to compute the expected number of visible edge-intersections. Take two edges e and f of $\mathcal{T}$ that do not share a vertex (otherwise these edges cannot intersect). A first step is to bound the probability, by perturbing e, that e creates a visible intersection with f. This probability can be bounded by some local geometric constants. Then, the authors deduced that the expected number of visible intersections with e can be bounded as well. Summing up and bounding the local constants by the global ones θ, σ and ϕ give the result.

1.3.2 Smoothed Analysis of the Number of Extremal Points

Let $p_1^*, \ldots, p_n^*$ be n points chosen in a d-dimensional hypercube. Let $p_1, \ldots, p_n$ be the noisy version of the point set under some random distribution and compute the expected size of the convex hull of the noisy points. The problem seems to be similar to Section 1.1.2, however in this case the points are not identically distributed: the mean value of p_i is p_i^*. In other words, this corresponds to the smoothed complexity of the convex hulls.

In the worst case, the number of vertices of a convex hull of n points in $\mathbb{R}^d$ is $\Theta\left(n^{\lceil \frac{d}{2} \rceil}\right)$, and the average-case corresponds to Section 1.1.2. Damerow and Sohler

[17] computed an upper-bound of the expected number of vertices of the convex hull (called *extremal points*) for Gaussian perturbations and uniform perturbations in a cube. They proceed by bounding the expected number of *maximal points* (a point is maximal if each of its orthants contains no points). Since every extremal point is maximal, their bound is also valid for the expected number of extremal points.

Theorem 8 (Damerow, Sohler [17]). *Let $P^* = \{p_1^*, \dots, p_n^*\}$ be a set of n points of $\mathbb{R}^d$ for fixed dimension d, in the d-dimensional unit hypercube. Let $P = \{p_1, \dots, p_n\}$ its noisy version under some noise distribution Δ.*

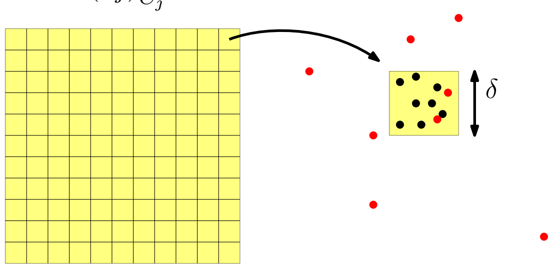
- *if Δ is the d-dimensional Gaussian distribution with standard deviation σ, then the expected number maximal points is*

$$O\left(\frac{1}{\sigma^d}\ln^{3/2\cdot d-1} n\right),$$

- *if Δ is the uniform distribution in a hypercube of side length ε, then the expected number of maximal points is*

$$O\left(\left(\frac{n\ln n}{\varepsilon}\right)^{\frac{d}{d+1}}\right).$$

Outline of the Proof. First, the authors compute the *average* number of maximal points if the points are chosen according to the distribution Δ. Under some assumption on Δ (which are valid for the one considered in the final theorem), they proved that the average complexity is $O\left(\ln^{d-1} n\right)$. Now, fix a parameter δ, and split the unit hypercube into $\frac{1}{\delta^d}$ smaller hypercubes of size δ. For each small hypercube C_j, we consider the points $(p_i^*)_{C_j}$ that are in C_j, and compute the expected number of maximal points of $(p_j)_{C_j}$



For each cell C_j, we consider the initial points in C_j (in black) and their noisy version (in red).

The main idea is that if δ is small enough, then the points $(p_i)_{C_j}$ will behave almost like in the average-case. Summing up the maximal points of all the C_j gives an upper bound of the total number of maximal points. This idea, summerized in a lemma, can be applied for the two considered distributions.

Chapter 2

Smoothed Complexity of Convex Hulls by Witnesses and Collectors

Outline

2.1 Introduction

Let P^* be a finite set of points in $\mathbb{R}^d$ and consider a random perturbation $P = \{p^* + \eta(p^*) : p^* \in P^*\}$ where each point p^* is moved by some random vector $\eta(p^*)$, typically chosen independently. We are interested in the asymptotic behavior of the expected number of faces (of all dimensions) of the convex hull of P, as a function of the number n of points and some parameter that describes the amplitude of the perturbations.

Formally, the *smoothed complexity of convex hulls* relative to a probability distribution μ on $\mathbb{R}^d$ is defined as

$$\mathcal{S}(n,\mu) = \max_{\substack{p_1^*, p_2^*, \ldots, p_n^* \in \mathbb{R}^d \\ \operatorname{diam}\{p_1^*, p_2^*, \ldots, p_n^*\} \leq 1}} \mathbb{E}\left[\operatorname{card} \operatorname{CH}\left(\{p_1^* + \eta_1, p_2^* + \eta_2, \ldots, p_n^* + \eta_n\}\right)\right]$$

where diam denotes the diameter, card S denotes the cardinality of a set S, $\operatorname{CH}(X)$ denotes the set of faces, of all dimensions, of the convex hull of X, and $\eta_1, \eta_2, \ldots, \eta_n$

are random variables chosen independently from the distribution μ.We present upper and lower bounds on $\mathcal{S}(n, \mathcal{U}_{\delta\mathbb{B}})$, where $\mathcal{U}_{\delta\mathbb{B}}$ is the uniform distribution on the ball of radius δ centered in the origin in $\mathbb{R}^d$, and $\mathcal{S}(n, \mathcal{N}(0, \sigma^2 I_2))$, where $\mathcal{N}(0, \sigma^2 I_2)$ is the Gaussian distribution centered in the origin and with covariance matrix $\sigma^2 I_2$.

2.1.1 New Results

Our main results are a technique to analyze random geometric hypergraphs, which we call the *witness-collector technique*, as well as its application to the analysis of the smoothed complexity of convex hulls. Before we spell them out we need to clarify some terminology.

Random Geometric Hypergraphs. Let $\mathcal{X}$ be a set, $(\mathcal{X}, \mathcal{R})$ a range space (i.e. $\mathcal{R}$ is a family of subsets (ranges) of $\mathcal{X}$) and P a finite set of random elements of $\mathcal{X}$. The *random geometric hypergraph* induced by $(\mathcal{X}, \mathcal{R})$ on P is the set $\mathcal{H} = \{P \cap r : r \in \mathcal{R}\}$; that is, a subset $Q \subset P$ is a hyperedge of $\mathcal{H}$ if and only if there exists $r \in \mathcal{R}$ such that $r \cap P = Q$. Our analyses of random convex hulls proceed by analyzing random geometric hypergraphs where $\mathcal{X} = \mathbb{R}^d$, $\mathcal{R}$ is the set of all half-spaces of $\mathbb{R}^d$, and the elements of P are chosen independently (but not identically distributed!). Any face of the convex hull of P is a hyperedge of $\mathcal{H}$, but the converse is not true. It turns out, however, that the average size of $\mathcal{H}$ is close enough to that of CH(P) that our technique yields meaningful upper and lower bounds on the smoothed complexity of convex hulls (*cf.* Section 2.2.3).

Notations for Orders of Magnitude. Our goal is to understand how the order of magnitude of the smoothed complexity depends on the number n of points and the amplitude δ or σ of the perturbation. For the sake of the presentation, we do not keep track in our analyses of additive or multiplicative constants depending on fixed quantities such as the dimension of the space. Throughout the chapter, we therefore write $a = O(b)$, $a = \Omega(b)$ and $a = \Theta(b)$ to mean that there exist positive reals c and c' such that, respectively, $a \le cb$, $a \ge cb$ and $cb \le a \le c'b$; we also use $\Theta(b)$ (and similarly for $O()$ and $\Omega()$) as a shorthand for a quantity x for which $x = \Theta(b)$ holds. These notations do *not* carry any asymptotic meaning (since several variables may assume large and unrelated values); when used without stating any condition on n, σ or δ, these notations mean inequalities that hold for any $n \ge d+1$, $\delta > 0$ and $\sigma > 0$.

The Witness-Collector Technique. Let $(\mathcal{X}, \mathcal{R})$ denote a range space. Our analyses are based on the following notion:

Definition 9. *A* system of witnesses and collectors *for a covering* $R_1 \cup R_2 \cup \ldots \cup R_m$ *of* $\mathcal{R}$ *is a family* $\{(W_i^j, C_i^j)\}_{\substack{1 \le i \le m \\ 1 \le j \le \ell}}$ *of pairs of subsets of* $\mathcal{X}$ *such that*

(a) *for all* i, j, *any* $r \in R_i$ *contains* W_i^j *or is contained in* C_i^j,
(b) *for all* i, $W_i^1 \subseteq W_i^2 \subseteq \ldots \subseteq W_i^\ell$,
(c) *for all* i, j, $W_i^j \subseteq C_i^j$.

We denote by $\mathcal{H}^{(k)}$ the set of hyperedges of cardinality k of a hypergraph $\mathcal{H}$. Our analyses are based on the following theorem, which we prove in Section 2.2:

Theorem 10. *Let* $(\mathcal{X}, \mathcal{R})$ *be a range space, let* P *be a set of* n *random elements of* $\mathcal{X}$ *chosen independently and let* $\mathcal{H}$ *denote the hypergraph induced by* $\mathcal{R}$ *on* P.

(i) If $\{(W_i^j, C_i^j)\}_{\substack{1 \le i \le m \\ 1 \le j \le \ln^2 n}}$ *is a system of witnesses and collectors for a covering* $R_1 \cup R_2 \cup \ldots \cup R_m$ *of* $\mathcal{R}$ *such that* $W_i^j \cap P$ *and* $C_i^j \cap P$ *have average size* $\Omega(j)$ *and* $O(j)$ *respectively then* $\mathbb{E}\left[\operatorname{card} \mathcal{H}^{(k)}\right] = O(m)$.

(ii) If every element of $\mathcal{H}^{(1)}$ *is in at least one element of* $\mathcal{H}^{(k)}$, *and* $\{W_i^1\}_{1 \le i \le m}$ *is a family of disjoint subsets of* $\mathcal{X}$ *such that* $\mathbb{E}\left[\operatorname{card}\left(W_i^1 \cap \mathcal{H}^{(1)}\right)\right] = \Omega(1)$ *then* $\mathbb{E}\left[\operatorname{card} \mathcal{H}^{(k)}\right] = \Omega(m)$.

In several of our applications we first construct a system $\{(W_i^j, C_i^j)\}$ of witnesses and collectors satisfying the assumptions of Theorem 10 (i), then use a subfamily of the W_i^1's that are disjoint to apply Theorem 10 (ii).

Applications. We present, in Sections 2.3 and 2.4, two designs of systems of witnesses and collectors suited to study the smoothed complexity of convex hulls relative to Euclidean and Gaussian perturbations with the following results (*cf.* Figures 2.1 and 2.2):

Smoothed Complexity. We obtain upper bounds on the smoothed complexity of convex hulls relative to Euclidean and Gaussian perturbations; in the Euclidean case we obtain sharper bounds for the smoothed number of vertices. We also analyze the convex hull of perturbations of points in convex position and delineate the main regimes in terms of the number of points and the amplitude of the perturbation; this provides lower-bounds on the Euclidean and Gaussian smoothed complexities of convex hulls.

Large Perturbations. We show that for $\delta = \Omega\left(n^{\frac{2}{d+1}}\right)$ the smoothed complexity of convex hulls relative to $\mathcal{U}_{\delta\mathbb{B}}$ is of the same order of magnitude as the expected complexity of the convex hull of random points chosen i.i.d. from $\mathcal{U}_{\delta\mathbb{B}}$, the classical model of random polytope. Our smoothed complexity upper bound also implies a similar result for Gaussian perturbation with $\sigma = \Omega(1)$.

Simple Analysis of Classical Random Polytopes. The classical model of random polytopes corresponds to the case where all points of p_i^* coincide. There, our systems of witnesses and collectors yield the order of magnitude of the expected number of faces with considerably less effort than earlier analyses.

A Surprising Phenomenon. We observed experimentally (Figure 2.2c) that the expected size of the convex hull of perturbations of points in convex position consistently decreases with the amplitude of the noise in the Gaussian model, whereas some non-monotonicity appears in the Euclidean model. Our analyses of perturbations of points in convex position provide a theoretical confirmation of this difference in behaviours (see Figures 2.1b and 2.2a).

As evidence that the witness-collector technique is relevant for the study of other geometric hypergraphs, we outline a design of witnesses and collectors that yields the order of magnitude of the number of faces in the Delaunay triangulations of a set of random points chosen uniformly and independently from the unit ball (Theorem 22); again, this is a well-known result but the proof (only sketched here) is considerably shorter than the original one.

2.1.2 Related Works

The results presented here appeared in preliminary form in research reports [6, 22] and proceedings of conferences [23, 28]. Note that the shift from *static* to *adaptative* witness-collectors in Section 2.2.2 is based on an idea which we learned from [33] and systematize here. We briefly position our results with respect to prominent related previous work.

Smoothed Number of Maximal Points. The only previous bound on the smoothed complexity of convex hulls is due to Damerow and Sohler [17], see Section 1.3.2. They study the number of *maximal* points under Gaussian and ℓ^∞ perturbations (we included the results for the Gaussian case in Figures 2.1d and 2.2b). Their technique requires that the perturbation acts independently on each coordinate (thus restricting possible perturbations) so that the analysis of point dominance reduces to considerations on independent random permutations. The number of maximal points bounds from above the number of extreme points, but in probabilistic setting these two quantities typically have different orders of magnitude. As a consequence, the upper bounds are not sharp and there is no lower bound.

One may expect that when the magnitude of the perturbations is sufficiently large compared to the scale of the initial input, the initial position of the points does not matter and smoothed complexity is subsumed by some average-case analysis (up to constant multiplicative factors). The main insight of Damerow and Sohler [17] is a quantitative version of this claim. Specifically, they show that if n points from a region of diameter r are perturbed by a Gaussian noise of standard deviation $\Omega(r\sqrt{\ln n})$ or a ℓ^∞ noise of amplitude $\Omega(r\sqrt[3]{n/\ln n})$ then the expected number of maximal points is the same as in the average-case analysis. A smoothed

any d	Range of δ	$\left[0, n^{\frac{2}{d+1}-\frac{1}{d-1}\lfloor\frac{d}{2}\rfloor}\right]$	$\left[n^{\frac{2}{d+1}-\frac{1}{d-1}\lfloor\frac{d}{2}\rfloor}, 1\right]$	$\left[1, 3n^{\frac{2}{d+1}}\right]$	$\left[3n^{\frac{2}{d+1}}, +\infty\right)$
	$\mathcal{S}(n, \mathcal{U}_{\delta\mathbb{B}})$	$O\left(n^{\lfloor\frac{d}{2}\rfloor}\right)$	$O\left(n^{2\frac{d-1}{d+1}}\delta^{-(d-1)}\right)$	$O\left(n^{2\frac{d-1}{d+1}}\right)$	$\Theta\left(n^{\frac{d-1}{d+1}}\right)$
any d	$\mathbb{E}\left[\text{card}\,\mathcal{H}^{(1)}\right] = O\left(n^{\frac{d-1}{d+1}} + \delta^{-\frac{2d}{d+1}} n^{1+2\frac{d-1}{(d+1)^2}}\right)$				

$d = 2$	Range of δ	$[0, \frac{1}{\sqrt{n}}]$	$[\frac{1}{\sqrt{n}}, 1]$	$[1, n^{5/12}]$	$[n^{5/12}, n^{2/3}]$	$[n^{2/3}, +\infty]$
	$\mathcal{S}(n, \mathcal{U}_{\delta\mathbb{B}})$	$O(n)$	$O\left(\delta^{-\frac{2}{3}} n^{\frac{2}{3}}\right)$	$O\left(n^{2/3}\right)$	$O\left(\delta^{-\frac{4}{3}} n^{\frac{11}{9}}\right)$	$O\left(n^{1/3}\right)$

(a) Upper bounds on the smoothed complexity relative to Euclidean perturbations (Theorem 13 and Corollary 14).

Range of δ	$0 \leq \delta \leq n^{\frac{2}{1-d}}$	$n^{\frac{2}{1-d}} \leq \delta \leq 1$	$1 \leq \delta \leq n^{\frac{2}{d+1}}$	$n^{\frac{2}{d+1}} \leq \delta$
$\mathbb{E}\left[\text{card}\,\text{CH}(P)\right]$	$\Theta(n)$	$\Theta\left(n^{\frac{d-1}{2d}} \delta^{\frac{1-d^2}{4d}}\right)$	$\Theta\left(n^{\frac{d-1}{2d}} \delta^{\frac{(1-d)^2}{4d}}\right)$	$\Theta\left(n^{\frac{d-1}{d+1}}\right)$

(b) Expected complexity of a Euclidean perturbation P of a regular sample of the unit sphere in $\mathbb{R}^d$ (Theorem 15). This gives a lower bound on the smoothed complexity for Euclidean perturbation.

	any d	$d = 2$
$\delta \geq \delta_0 \Rightarrow$ average-case behavior	$\delta_0 = O\left(n^{\frac{2}{d+1}}\right)$	$\delta_0 = O\left(n^{2/3}\right)$

(c) Amplitude of a Euclidean perturbation for which the smoothed complexity behaves as the average-case complexity (Lemma 2.3.8).

	Our bounds ($d = 2$)	Previous bound [17][a]
$\sigma \geq \sigma_0 \Rightarrow$ average-case behavior	$\sigma_0 = O(1)$	$\sigma_0 = O\left(\sqrt{\ln n}\right)$
$\mathcal{S}(n, \mathcal{N}(0, \sigma^2 I_2))$	$O\left(\sqrt{\ln n} + \sigma^{-1}\sqrt{\ln n}\right)$	$O\left(\ln n + \sigma^{-2}\ln^2 n\right)$

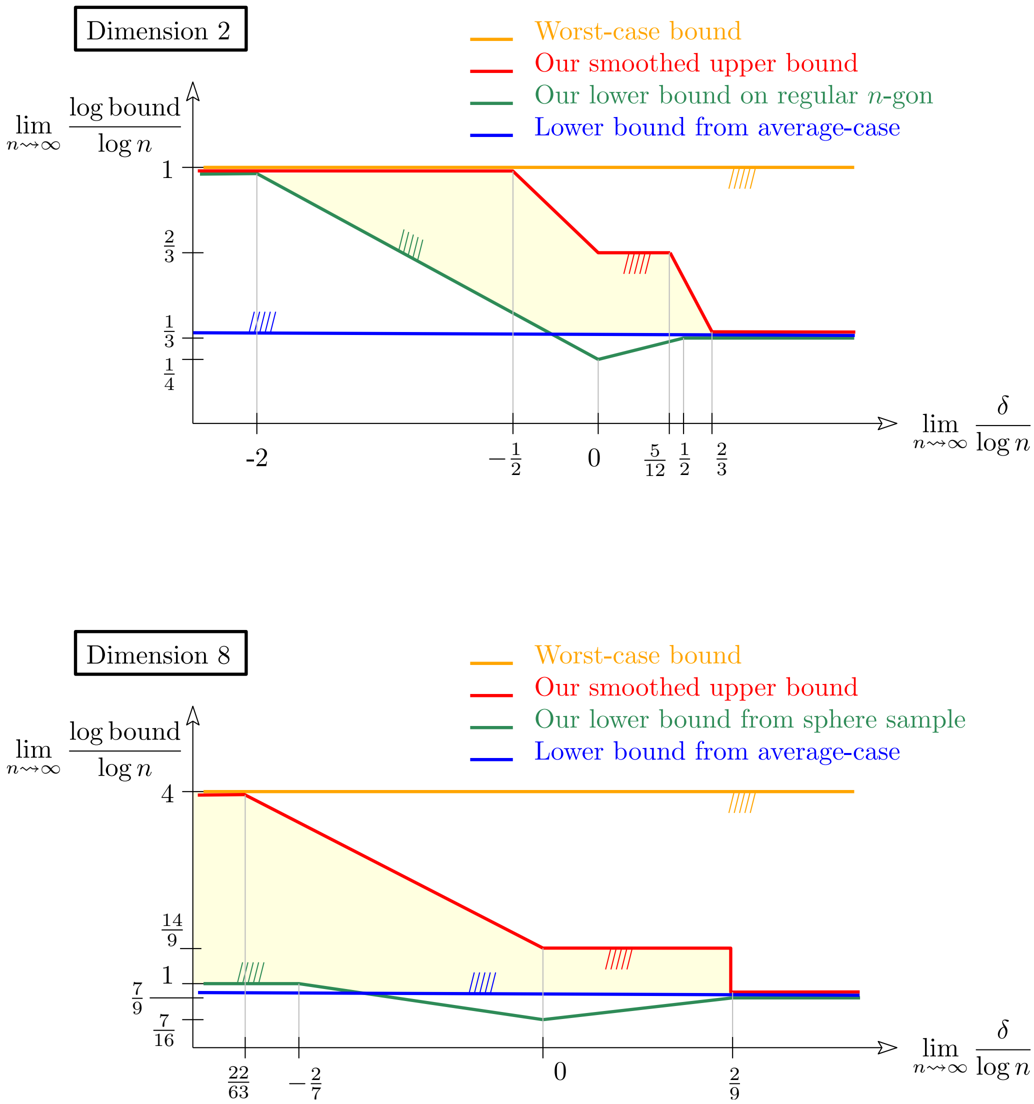
(d) Upper bounds for Gaussian perturbations (Theorem 18).

[a] This bound applies to *maximal point*, *cf.* the comparison to earlier work.

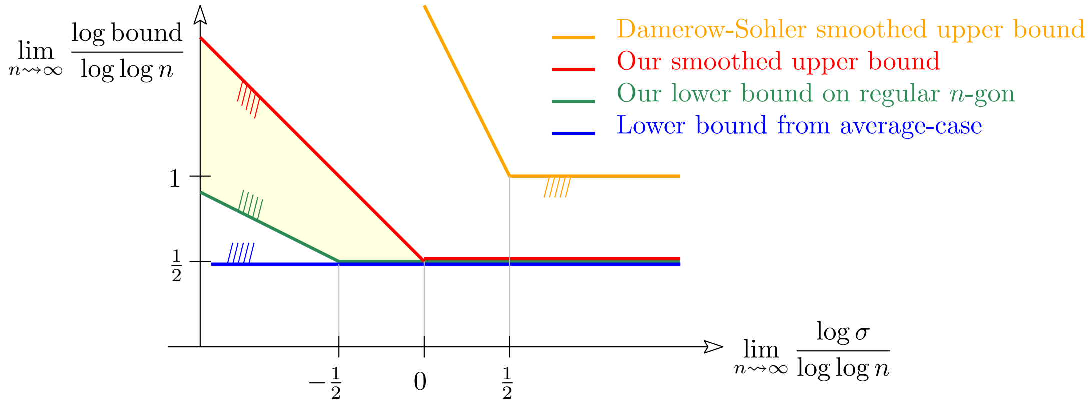
Range of σ	$0 \leq \sigma \leq \frac{1}{n^2}$	$\frac{1}{n^2} \leq \sigma \leq \frac{2\ln^4 n}{n^2}$	$\frac{2\ln^4 n}{n^2} \leq \sigma \leq \frac{1}{\sqrt{\ln n}}$	$\frac{1}{\sqrt{\ln n}} \leq \sigma$
$\mathbb{E}\left[\text{card}\,\text{CH}(P)\right]$	$\Theta(n)$	$O(n)$, $\Omega\left(\frac{\sqrt[4]{\ln(n\sqrt{\sigma})}}{\sqrt{\sigma}}\right)$	$\Theta\left(\frac{\sqrt[4]{\ln(n\sqrt{\sigma})}}{\sqrt{\sigma}}\right)$	$\Theta\left(\sqrt{\ln n}\right)$

(e) Expected complexity of a Gaussian perturbation P of a regular n-gon in $\mathbb{R}^2$ (Theorem 19 and Theorem 20). The lower bound gives a lower bound on the smoothed complexity for Gaussian perturbation.

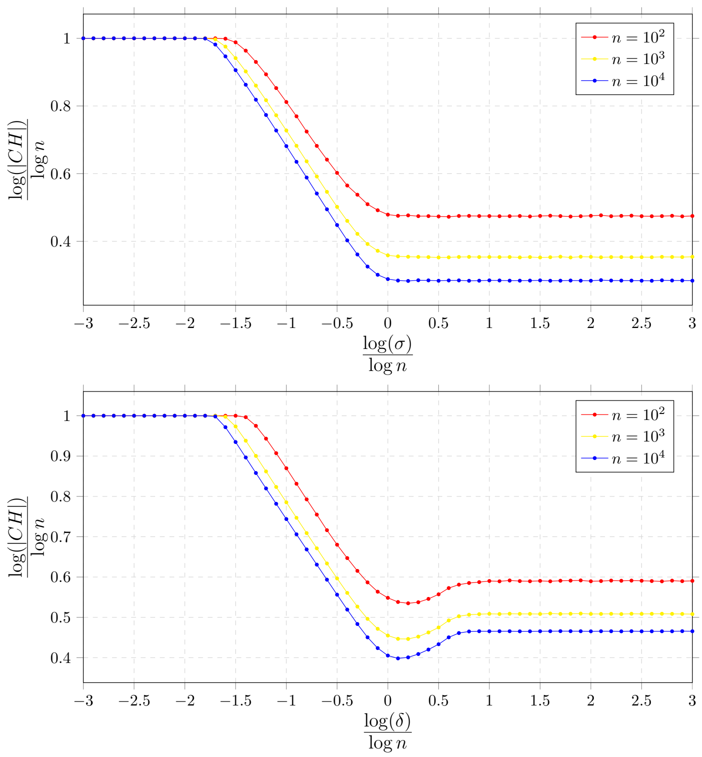
Figure 2.1: Summary of our bounds.



(a) A comparison of our smoothed complexity bound for Euclidean perturbation (Theorem 13 and Corollary 14) and two lower bounds, where the initial points are placed respectively at the vertices of a unit-size n-gon (Theorem 15) and in the origin. A data point with coordinates (x, y) means that for a perturbation with δ of magnitude n^x the expected size of the convex hull grows as n^y, subpolynomial terms being ignored. The worst-case bound is given as a reference. The constants in the $O()$ and $\Omega()$ have been ignored as their influence vanishes as $n \to \infty$ in this coordinate system.



(b) Comparison of the smoothed bounds for Gaussian perturbation in dimension 2 (Theorem 18 and [17]) and the lower bound perturbing the regular n-gon (Theorem 19). A data point with coordinates (x, y) means that for a perturbation of magnitude $\sigma = \ln^x n$ the expected size of the convex hull grows as $\ln^y n$.



(c) Experimental results for the complexity of the convex hull of a perturbation of the regular n-gon inscribed in the unit circle. Left: Gaussian perturbation of variance σ^2. Right: Euclidean perturbation of amplitude δ. Each data point corresponds to an average over 1000 experiments.

Figure 2.2: Plots summarizing the main results.

complexity bound then follows by a simple rescaling argument: split the input domain into cells of size $r = O(\sigma/\sqrt{\ln n})$, assume that *each* cell contains all of the initial point set, and charge each of them with the average-case bound.

Our technique yields a similar subsuming of the smoothed complexity analysis by the average-case analysis for the number of faces (Lemma 2.4.5) with the same threshold, thus extending Damerow and Sohler's main insight; we also obtain a similar statement in the Euclidean model (Lemma 2.3.8). The smoothed complexity bound we obtain by the rescaling argument (Corollary 17) is better than the one of Damerow and Sohler because the average number of extreme points is asymptotically smaller than the number of maximal points. It should be noted that the rescaling argument only applies to bound the number of vertices of the convex hull since faces of higher dimension may come from more than one cell; in any case, we further improve the bound obtained from the rescaling argument by a more direct analysis that accounts for faces of arbitrary dimension (Theorem 18).

Smoothed Complexity of a Simplex Algorithm. A substantial literature in the analysis of algorithms was devoted to explain the very good practical performance of the simplex algorithm, given that most of the pivoting rules had exponential worst-case complexity. This motivated the study of various models of random polytopes, and eventually the introduction of the smoothed complexity analysis model by Spielman and Teng [49]. We encourage the interested reader to consult their discussion of earlier literature, and simply compare our work to the smoothed complexity bound for convex hulls that is at the core of their analysis of the shadow-vertex pivot rule. As seen on Section 1.2.2, they estimate the expected number of vertices of an arbitrary two-dimensional projection of a polytope given as an intersection of n halfspaces in d dimensions and perturbed by a Gaussian noise of standard deviation σ using techniques quite different from ours, see [49, Th 4.1]. Neither n nor d are fixed, so the number of vertices may be exponential in the input; their analysis shows that it is polynomial in n, d and $\frac{1}{\sigma}$. The question we consider is therefore, from the point of view of the model, of a rather different nature: we consider the dimension to be fixed rather than variable, specify the polytope as a convex hull of vertices rather than intersection of half-spaces, and estimate the number of faces rather than the two-dimensional silhouettes. More importantly, our intent is to understand a transition within the polynomial domain rather than identify a polynomial behavior in place of an exponential worst-case bound.

Floating Bodies and Economic Cap Coverings. As seen on Section 1.1.2, Bárány and Larman [9] established that the expected number of faces of the convex hull of n random points chosen uniformly from a convex body K is $\Theta\left(nK\left(\frac{1}{n}\right)\right)$, where $K(t)$ denotes the volume of the *wet part* of K with parameter t: the union of the intersections of K with a half-space that intersects it with volume at most t. This connection allowed them to transfer to the study of random polytopes various results from convex geometry, for which wet parts, or their complements the floating bodies, are classical objects.

When the ranges are half-spaces in $\mathbb{R}^d$, our systems of witnesses and collectors are essentially equivalent to the *economic cap covers* on which Bárány and Larman's proof is based (Bárány and Vu [10, § 5] also use the same idea in the proving of a central limit theorems for Gaussian polytopes). A first difference is that the analogue of our Condition (a) for economic cap covers is formulated in terms of wet parts, so the role of the range space is implicit. This has little effect as far as the ranges are half-spaces, but we note that the analogue of wet parts for other range spaces is not straightforward to define and study, whereas our presentation naturally extends to other range spaces (as the case of Delaunay triangulation sketched in Section 2.5 demonstrates). We also note that the constructions of systems of witnesses and collectors differ from the constructions of economic cap covers, but believe that this is a less essential distinction.

2.2 Witnesses and Collectors

In this section we first explain the idea behind Theorem 10 in a simpler setting in Section 2.2.1, then prove Theorem 10 in Section 2.2.2, then clarify its use for the analysis of convex hulls of random point sets in Section 2.2.3.

2.2.1 Principle: Static Witnesses and Collectors

Let $(\mathcal{X}, \mathcal{R})$ be a range space, P a random set of n elements of $\mathcal{X}$ chosen independently, $\mathcal{H}$ the hypergraph induced by $\mathcal{R}$ on P, and $k \in \mathbb{N}$. Let $R_1 \cup R_2 \cup \ldots \cup R_m$ be a covering of $\mathcal{R}$ and $\{(W_i^1, C_i^1)\}_{1 \le i \le m}$ a system of witnesses and collectors for that covering. Since $\ell = 1$, we shorten W_i^1 into W_i and C_i^1 into C_i and note that Condition (b) is trivial.

Conditioning on Loaded Witnesses. If $\operatorname{card}(W_i \cap P)$ is at least k then Condition (a) ensures that every hyperedge of size k in $\{r \cap P : r \in R_i\}$ is contained in C_i, so there are at most $\mathbb{E}\left[\operatorname{card}(C_i \cap P)^k\right]$ such hyperedges; otherwise we can use the trivial upper bound $\binom{n}{k}$. Conditioning on the event that $\operatorname{card}(W_i \cap P)$ is at least k for all i we therefore get

$$\begin{aligned}\mathbb{E}\left[\operatorname{card}\mathcal{H}^{(k)}\right] \;\le\; & \mathbb{P}\left[\exists i, \operatorname{card}(W_i \cap P) < k\right]\binom{n}{k} \\ & + \mathbb{P}\left[\forall i, \operatorname{card}(W_i \cap P) \ge k\right] \cdot \sum_{i=1}^{m} \mathbb{E}\left[\operatorname{card}(C_i \cap P)^k\right]\end{aligned}$$

so if the witnesses are chosen so that $\operatorname{card}(W_i \cap P) \ge k$ with probability $1 - O(n^{-k})$,

$$\begin{aligned}\mathbb{E}\left[\operatorname{card}\mathcal{H}^{(k)}\right] \;=\; & O\left(\left(\sum_{i=1}^{m} \mathbb{P}\left[\operatorname{card}(W_i \cap P) < k\right]\right) \cdot \binom{n}{k}\right. \\ & \left. + \mathbb{P}\left[\forall i, \operatorname{card}(W_i \cap P) \ge k\right] \cdot \sum_{i=1}^{m} \mathbb{E}\left[\operatorname{card}(C_i \cap P)^k\right]\right) \\ =\; & O\left(m + \sum_{i=1}^{m} \mathbb{E}\left[\operatorname{card}(C_i \cap P)^k\right]\right). \qquad (2.1)\end{aligned}$$

Role of $\mathbf{W_i \cap P}$ and $\mathbf{C_i \cap P}$. Chernoff's multiplicative bound implies that if $W_i \cap P$ has average size $\Omega(k \ln n)$ then indeed $\operatorname{card}(W_i \cap P) \geq k$ with probability $1 - O(n^{-k})$. More generally:

Lemma 2.2.1. *Let P be a set of random elements of $\mathcal{X}$ chosen independently and W a subset of $\mathcal{X}$.*
(a) $\mathbb{P}[W \cap P = \emptyset] \leq e^{-\mathbb{E}[\operatorname{card}(W \cap P)]}$.
(b) If $\mathbb{E}[\operatorname{card}(W \cap P)] \geq k+1$ *then* $\mathbb{P}[\operatorname{card}(W \cap P) < k] \leq e^{-\Omega(\mathbb{E}[\operatorname{card}(W \cap P)])}$.

(We defer the proof to Section 2.2.4.) The bound in Equation (2.1) is expressed in terms of the $\mathbb{E}\left[\operatorname{card}(C_i \cap P)^k\right]$ but can be controlled by $\mathbb{E}[\operatorname{card}(C_i \cap P)]$ since the elements of P are chosen independently:

Lemma 2.2.2. *If $V = \sum_{i=1}^{n} V_i$, where the V_i are independently distributed random variables with value in $\{0,1\}$ and $\mathbb{E}[V] \geq 1$ then $\mathbb{E}\left[V^k\right] = O\left(\mathbb{E}[V]^k\right)$.*

(Again, the proof is postponed to Section 2.2.4.) In the situations we consider, one can construct witnesses and collectors such that $W_i \cap P$ and $C_i \cap P$ both have expected size $\Theta(k \ln n)$; see [23] for several examples. Equation (2.1) and Lemma 2.2.2 then yield that $\mathbb{E}\left[\operatorname{card} \mathcal{H}^{(k)}\right]$ is of order m up to some logarithmic factors.

Shaving Log Factors. The use of a Chernoff bound to control the probability that witnesses contain fewer than k elements increases the expected size of the $W_i \cap P$ so that *all of them* are large for *most* realizations of P. By Condition (c), $W_i \subseteq C_i$, so this also overloads the collectors, resulting in the extra log factors. The idea that leads to the sharper bounds of Theorem 10, which we learned from [33], is to make W_i and C_i random variables depending on P. By adapting the witness-collector pairs used in the analysis to each realization of P, very few collectors will need to be large, and their contribution to the total will remain negligible.

It is perhaps worth pointing out that the above analysis holds for several of our constructions when only the first layer ($j = 1$) of witnesses and collectors is considered. Our proofs can therefore be further simplified should one not care about some extra logarithmic factors.

2.2.2 Proof of Theorem 10: Adaptative Witnesses and Collectors

We first prove the upper bound, in a format that will allow slightly more flexibility.

Lemma 2.2.3. *Let $(\mathcal{X}, \mathcal{R})$ be a range space, let P be a set of n random elements of $\mathcal{X}$ chosen independently and let $\mathcal{H}$ denote the hypergraph induced by $\mathcal{R}$ on P. If $R_1 \cup R_2 \cup \ldots \cup R_m$ is a covering of $\mathcal{R}$ and $\{(W_i^j, C_i^j)\}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq \ln^2 n}}$ is a system of* *witnesses and collectors for that covering with*

$$\mathbb{P}\left[\operatorname{card}\left(W_i^j \cap P\right) < k\right] = O\left(e^{-\Omega(j)}\right) \quad \textit{and} \quad \mathbb{E}\left[\operatorname{card}\left(C_i^j \cap P\right)\right] = O(j)$$

then $\mathbb{E}\left[\operatorname{card} \mathcal{H}^{(k)}\right]$ *is* $O(m)$.

Proof. Let $i \in \{1, 2, \ldots, m\}$. We let d_i denote the smallest j such that W_i^j contains at least k points and $C_i = C_i^{d_i}$, or, if no such W_i^j exists, $d_i = \infty$ and $C_i = \mathcal{X}$. (So d_i and C_i are random variables depending on P.) By Condition (a) and the definition of d_i, every hyperedge of $\mathcal{H}$ of size k induced by R_i is contained in C_i so:

$$\mathbb{E}\left[\operatorname{card} \mathcal{H}^{(k)}\right] \leq \sum_{i=1}^{m} \mathbb{E}\left[\operatorname{card}\left(C_i \cap P\right)^k\right]. \tag{2.2}$$

Moreover, by Condition (b) we have

$$\mathbb{P}\left[d_i \geq j\right] = \mathbb{P}\left[\operatorname{card}\left(W_i^j \cap P\right) < k\right] = O\left(e^{-\Omega(j)}\right).$$

We claim that

$$\mathbb{E}\left[\operatorname{card}\left(C_i^j \cap P\right) \mid d_i \geq j\right] \leq \mathbb{E}\left[\operatorname{card}\left(C_i^j \cap P\right)\right] + k = O(j). \tag{2.3}$$

Indeed, working with the complement $\bar{C}_i^j$ of C_i^j,

$$\mathbb{E}\left[\operatorname{card}\left(\bar{C}_i^j \cap P\right)\right] = \sum_{p \in P} \mathbb{P}\left[p \notin C_i^j\right].$$

For any $T \subset P$ we have

$$\mathbb{E}\left[\operatorname{card}\left(\bar{C}_i^j \cap P\right) \mid W_i^j \cap P = T\right] = \sum_{p \in P \setminus T} \mathbb{P}\left[p \notin C_i^j \mid p \notin W_i^j\right] \geq \sum_{p \in P \setminus T} \mathbb{P}\left[p \notin C_i^j\right],$$

the last inequality following from Condition (c). Thus,

$$\mathbb{E}\left[\operatorname{card}\left(\bar{C}_i^j \cap P\right)\right] \leq \mathbb{E}\left[\operatorname{card}\left(\bar{C}_i^j \cap P\right) \mid W_i^j \cap P = T\right] + \operatorname{card} T.$$

By Condition (b), $d_i \geq j$ if and only if $\operatorname{card}\left(W_i^{j-1} \cap P\right) < k$. Total probabilities let us decompose this event:

$$\begin{aligned}
&\mathbb{E}[\operatorname{card}\left(\bar{C}_i^j \cap P\right) \mid d_i \geq j] \\
&= \sum_{T : \operatorname{card} T < k} \mathbb{E}\left[\operatorname{card}\left(\bar{C}_i^j \cap P\right) \mid W_i^{j-1} \cap P = T\right] \mathbb{P}\left[W_i^{j-1} \cap P = T \mid \operatorname{card}\left(W_i^{j-1} \cap P\right) < k\right] \\
&\geq \left(\mathbb{E}\left[\operatorname{card}\left(\bar{C}_i^j \cap P\right)\right] - k\right) \sum_{T : \operatorname{card} T < k} \mathbb{P}\left[W_i^{j-1} \cap P = T \mid \operatorname{card}\left(W_i^{j-1} \cap P\right) < k\right] \\
&= \mathbb{E}\left[\operatorname{card}\left(\bar{C}_i^j \cap P\right)\right] - k
\end{aligned}$$

Moving back to the complement yields Inequation (2.3). Now, since P has n points in total, conditioning on the value of d_i we obtain

$$\begin{aligned}
\mathbb{E}\left[\operatorname{card}\left(C_i \cap P\right)\right] &= \sum_{j=1}^{\ln^2 n} \mathbb{E}\left[\operatorname{card}\left(C_i^j \cap P\right) \cdot \mathbb{1}_{d_i=j}\right] + \mathbb{E}\left[n \cdot \mathbb{1}_{d_i=\infty}\right] \\
&\leq \sum_{j=1}^{\ln^2 n} \mathbb{E}\left[\operatorname{card}\left(C_i^j \cap P\right) \cdot \mathbb{1}_{d_i \geq j}\right] + \mathbb{E}\left[n \cdot \mathbb{1}_{d_i=\infty}\right] \\
&= \sum_{j=1}^{\ln^2 n} \mathbb{E}\left[\operatorname{card}\left(C_i^j \cap P\right) \;\mid\; d_i \geq j\right] \mathbb{P}\left[d_i \geq j\right] + n \cdot \mathbb{P}\left[d_i = \infty\right] \\
&= \sum_{j=1}^{\ln^2 n} O\left(j e^{-\Omega(j)}\right) + O\left(n e^{-\Omega(\ln^2 n)}\right)
\end{aligned}$$

so each collector C_i contains on average a constant number of elements of P. Lemma 2.2.2 and Equation (2.2) imply that $\mathbb{E}\left[\operatorname{card} \mathcal{H}^{(k)}\right] = O(m)$. □

We now wrap-up the proof of our witness-collector theorem.

Proof of Theorem 10. Since $\mathbb{E}\left[\operatorname{card}\left(W_i^j \cap P\right)\right] = \Omega(j)$, there exists some constant $c > 0$ such that $\mathbb{E}\left[\operatorname{card}\left(W_i^j \cap P\right)\right] \geq cj$. For $j \geq \frac{k+1}{c}$, the Chernoff bound of Lemma 2.2.1 (b) thus ensures that $\mathbb{P}\left[\operatorname{card}\left(W_i^j \cap P\right) < k\right]$ is at most $e^{-\Omega(j)}$. Bounding that probability from above by 1 in the cases $j < \frac{k+1}{c}$ we get that $\mathbb{P}\left[\operatorname{card}\left(W_i^j \cap P\right) < k\right]$ is $O\left(e^{-\Omega(j)}\right)$. Statement (i) then follows readily from Lemma 2.2.3.

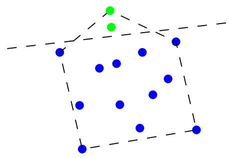
Now consider Statement (ii). We can charge each element of $\mathcal{H}^{(1)}$ to an element of $\mathcal{H}^{(k)}$ that contains it. Since each element of $\mathcal{H}^{(k)}$ is charged at most k times, we have $\operatorname{card} \mathcal{H}^{(k)} \geq \frac{1}{k} \operatorname{card} \mathcal{H}^{(1)}$. The assumptions ensure that each W_i^1 contains on average $\Omega(1)$ elements of $\mathcal{H}^{(1)}$ and that these elements are distinct. It follows that $\mathbb{E}\left[\operatorname{card} \mathcal{H}^{(1)}\right]$ and $\mathbb{E}\left[\operatorname{card} \mathcal{H}^{(k)}\right]$ are $\Omega(m)$. □

2.2.3 The Special Case of Convex Hulls

Unless indicated otherwise, in the remainder of this chapter the range space $(\mathcal{X}, \mathcal{R})$ considered is that of half-spaces in $\mathbb{R}^d$, where d is a constant. Every element of $\mathcal{H}^{(1)}$ belongs to some element of $\mathcal{H}^{(k)}$, so the first condition of Theorem 10 (ii) holds for this range space.

In this setting, the elements of $\mathcal{H}^{(k)}$ are also called the *k-sets* of the point set P. The bounds that we establish are expressed with $O()$, $\Omega()$ and $\Theta()$ in which the multiplicative constants depend on k; they are therefore valid for any *fixed* k.

For $k \leq d$, any $(k-1)$-dimensional face of $\mathrm{CH}(P)$ is a k-set, so the upper bound of Theorem 10 (i) applies to the size of the convex hull. The reverse is not true (*cf.* the figure on the right) but we remark that $\mathcal{H}^{(1)}$ is exactly the set of vertices of $\mathrm{CH}(P)$ and that every element of $\mathcal{H}^{(1)}$ belongs to an actual $(k-1)$-dimensional face of $\mathrm{CH}(P)$; the proof of Statement (ii) of Theorem 10 therefore provides, *mutatis mutandis*, a lower bound on the number of $(k-1)$-dimensional faces of $\mathrm{CH}(P)$. In the rest of the chapter, we will navigate without further justification between the convex hull of a random point set P and the associated random geometric hypergraph.



2.2.4 Proofs of Lemmas 2.2.1 and 2.2.2

Proof of Lemma 2.2.1. Let V_i be the indicator function of the event that the i^{th} point from P belongs to W. We write $V = V_1 + \ldots + V_n$ and let $t = \mathbb{E}[V]$. Chernoff's bound for lower tails yields that for any $\delta \in (0,1)$

$$\mathbb{P}[V < (1-\delta)t] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^t = e^{-t(1-(1-\delta)(1-\ln(1-\delta)))}. \tag{2.4}$$

In particular,

$$\mathbb{P}[V = 0] \leq \lim_{\delta \to 1} \mathbb{P}[V < (1-\delta)t] = \lim_{\delta \to 1} e^{-t(1-(1-\delta)(1-\ln(1-\delta)))} = e^{-t}$$

which proves Statement (a). Moreover, for $1 - \delta = \frac{k}{t}$, Equation (2.4) specializes into

$$\mathbb{P}[V < k] < e^{-t\left(1-\frac{k}{t}\left(1-\ln\frac{k}{t}\right)\right)}$$

Since $x \mapsto x(1-\ln x)$ is increasing on $(0,1)$, for $t \geq k+1$ we have

$$1 - \frac{k}{t}\left(1 - \ln\frac{k}{t}\right) \geq 1 - \frac{k}{k+1}\left(1 - \ln\frac{k}{k+1}\right) > 0$$

and Statement (b) follows. $\square$

Proof of Lemma 2.2.2. The statement is a special case of a classical inequality for sums of random variables [40, Th 2.12]; we give a simple, elementary, proof.

Expanding $V^k = \left(\sum_{i=1}^n V_i\right)^k$ we obtain

$$\begin{aligned} \mathbb{E}\left[V^k\right] &= \sum_{1\leq i_1, i_2 \ldots i_k \leq n} \mathbb{E}\left[V_{i_1} \cdot V_{i_2} \ldots V_{i_k}\right] \\ &= \sum_{\ell=1}^{k} \sum_{\substack{1\leq i_1, i_2 \ldots i_k \leq n \\ |\{i_1, i_2 \ldots i_k\}| = \ell}} \mathbb{E}\left[V_{i_1} \cdot V_{i_2} \ldots V_{i_k}\right]. \end{aligned}$$

Since the V_i's have values in $\{0,1\}$, for any positive integers $a_1, a_2, \dots a_t$ and $i_1, i_2, \dots i_t$

$$\mathbb{E}\left[V_{i_1}^{a_1} \cdot V_{i_2}^{a_2} \dots V_{i_t}^{a_t}\right] = \mathbb{E}\left[V_{i_1} \cdot V_{i_2} \dots V_{i_t}\right].$$

Letting $p(\ell, k)$ denote the number of partition of $\{1, 2, \dots, k\}$ in ℓ subsets, we can thus write

$$\mathbb{E}\left[V^k\right] = \sum_{\ell=1}^{k} \sum_{\substack{1 \le i_1, i_2 \dots i_\ell \le n \\ i_a \neq i_b \text{ if } a \neq b}} p(\ell, k)\, \mathbb{E}\left[V_{i_1} \cdot V_{i_2} \dots V_{i_\ell}\right].$$

Since V_i and V_j are independent if $i \neq j$ the previous identity rewrites as

$$\mathbb{E}\left[V^k\right] = \sum_{\ell=1}^{k} \left(p(\ell, k) \sum_{\substack{1 \le i_1, i_2 \dots i_\ell \le n \\ i_a \neq i_b \text{ if } a \neq b}} \mathbb{E}\left[V_{i_1}\right] \cdot \mathbb{E}\left[V_{i_2}\right] \dots \mathbb{E}\left[V_{i_\ell}\right] \right).$$

Thus,

$$\mathbb{E}\left[V^k\right] \le \sum_{\ell=1}^{k} \left(p(\ell, k) \sum_{1 \le i_1, i_2 \dots i_\ell \le n} \mathbb{E}\left[V_{i_1}\right] \cdot \mathbb{E}\left[V_{i_2}\right] \dots \mathbb{E}\left[V_{i_\ell}\right] \right)$$

and since

$$\sum_{1 \le j_1, j_2, \dots, j_\ell \le n} \mathbb{E}\left[V_{j_1}\right] \cdot \mathbb{E}\left[V_{j_2}\right] \dots \mathbb{E}\left[V_{j_\ell}\right] = \left(\sum_{i=1}^{n} \mathbb{E}\left[V_i\right] \right)^{\ell} = \mathbb{E}\left[V\right]^{\ell}$$

we finally obtain that

$$\mathbb{E}\left[V^k\right] \le \sum_{\ell=1}^{k} p(\ell, k)\, \mathbb{E}\left[V\right]^{\ell} \le \left(\sum_{\ell=1}^{k} p(\ell, k) \right) \mathbb{E}\left[V\right]^{k}$$

the last inequality following from the fact that $\mathbb{E}\left[V\right] \ge 1$. □

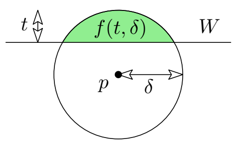
2.3 Euclidean Perturbation

We first consider the complexity of convex hulls of points perturbed under Euclidean perturbations.

Terminology and Notations. Let $d \geq 2$ be a constant. We denote by $\rho\mathbb{B}$ the ball of radius ρ centered at the origin of $\mathbb{R}^d$. Given $X \subset \mathbb{R}^d$ we denote by $\mathrm{vol}_k(X)$ its k-dimensional volume and by ∂X its boundary. We say that two half-spaces are *parallel* if they have the same inner normal. The *intersection depth* of a half-space W and a ball $p + \delta\mathbb{B}$ is $\delta - \bar{d}(p, W)$, where $\bar{d}(p, W)$ is the signed distance of p to ∂W (positive if and only if $p \notin W$).

2.3.1 Preliminaries: Ball/Half-space Intersection

We denote by $f(t, \delta)$ the volume of the intersection of $p+\delta\mathbb{B}$ with a half-space that intersects it with depth t. Note that $t \mapsto f(t, \delta)$ is increasing on $[0, 2\delta]$ for any fixed δ.



Claim 2.3.1. *For any $\lambda \geq 1$ and any $t \geq 0$, $f(\lambda t, \delta) \leq \lambda^{\frac{d+1}{2}} f(t, \delta)$.*

Proof. First assume that $\lambda t \leq 2\delta$. Let ν_{d-1} denote the volume of a $(d-1)$-dimensional ball of radius 1. By integrating along the direction of the inner normal to the half-space, we find

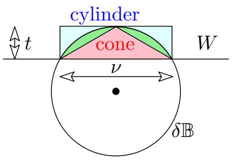
$$f(\lambda t, \delta) = \nu_{d-1} \int_0^{\lambda t} \left(2x\delta - x^2\right)^{\frac{d-1}{2}} dx = \nu_{d-1} \int_0^{t} \lambda^{\frac{d-1}{2}} \left(2x\delta - \lambda x^2\right)^{\frac{d-1}{2}} \lambda dx$$

$$\leq \nu_{d-1} \int_0^{t} \lambda^{\frac{d+1}{2}} \left(2x\delta - x^2\right)^{\frac{d-1}{2}} dx = \lambda^{\frac{d+1}{2}} f(t, \delta)$$

which proves the claim. The case $\lambda t > 2\delta$ then follows easily:

$$f(\lambda t, \delta) = \mathrm{vol}_d\left(\delta\mathbb{B}\right) = f\left(\frac{2\delta}{t} t, \delta\right) \leq \left(\frac{2\delta}{t}\right)^{\frac{d+1}{2}} f(t, \delta) \leq \lambda^{\frac{d+1}{2}} f(t, \delta).$$ □

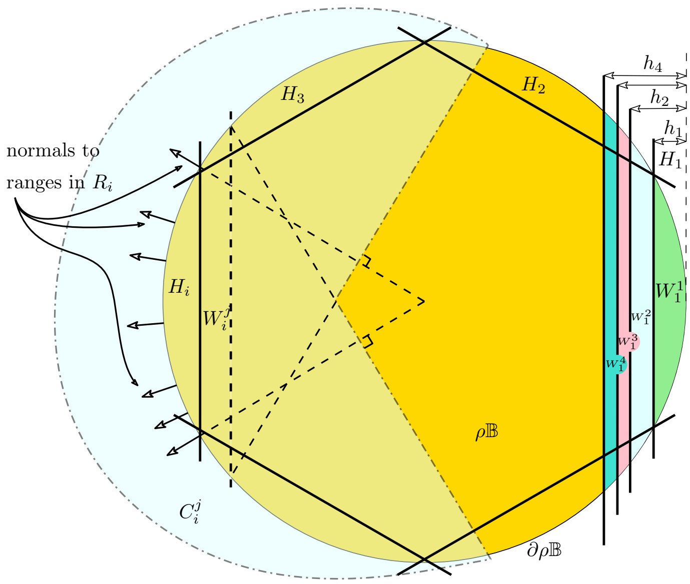
Claim 2.3.2. *For* $t \in [0, \delta]$, $f(t, \delta) = \Theta\left(t^{\frac{d+1}{2}} \delta^{\frac{d-1}{2}}\right)$.



Proof. Let W be a half-space that intersects $\delta\mathbb{B}$ with depth t and let $\nu = (\partial W) \cap \delta\mathbb{B}$. The region $W \cap \delta\mathbb{B}$ is sandwiched between a cone and a right cylinder with heights t and bases ν, with respective volumes $t\,\mathrm{vol}_{d-1}(\nu)/d$ and $t\,\mathrm{vol}_{d-1}(\nu)$. Since ν is a ball of radius r, with $r^2 = \delta^2 - (\delta - t)^2 = 2t\delta - t^2$, it has $(d-1)$-dimensional volume $\Theta\left((t\delta)^{\frac{d-1}{2}}\right)$ and the claim follows. □

2.3.2 Witness-Collector Construction

Our systems of witnesses and collectors for Euclidean perturbations are based on the construction summarized in the following figure.



Definition. Our construction is parameterized by a radius ρ, usually chosen so that the *perturbed* point set remains inside $\rho\mathbb{B}$, and a sequence of positive reals $h_1 < h_2 < \ldots < h_\ell$. We let $H_1, H_2, \ldots H_m$ be an inclusion-minimal cover of $\partial(\rho\mathbb{B})$ by half-spaces of intersection depth h_1 with $\rho\mathbb{B}$.

Claim 2.3.3. $m = \Theta\left((\rho/h_1)^{\frac{d-1}{2}}\right)$.

Proof. If $h_1 \geq \rho$ then $m \leq 2$ so assume $h_1 < \rho$. The intersection $H_i \cap \partial(\rho\mathbb{B})$ is a spherical cap with radius $r = \Theta\left(\sqrt{h_1\rho}\right)$. Since the H_i form a minimal cover of $\partial(\rho\mathbb{B})$,

$$m = \Theta\left(\frac{\mathrm{vol}_{d-1}(\partial(\rho\mathbb{B}))}{\mathrm{vol}_{d-1}(\partial(\rho\mathbb{B}) \cap H_i)}\right)$$

The statement then follows from the fact that $\mathrm{vol}_{d-1}(\partial(\rho\mathbb{B}))$ and $\mathrm{vol}_{d-1}(\partial(\rho\mathbb{B}) \cap W_i^1)$ are, respectively, proportional to ρ^{d-1} and r^{d-1}. □

We then define the range R_i as the set of half-spaces whose inner normal is parallel to a vector from the origin to a point of $H_i \cap \partial(\rho\mathbb{B})$. We define W_i^j as the intersection of $\rho\mathbb{B}$ with the half-space parallel to H_i and with intersection depth h_j with $\rho\mathbb{B}$. We define C_i^j as the union of the half-spaces of R_i that do not contain W_i^j.

Lemma 2.3.4. *$R_1 \cup R_2 \cup \ldots \cup R_m$ covers the set of half-spaces in $\mathbb{R}^d$ and $\{(W_i^j, C_i^j)\}_{\substack{1\le i\le m\\ 1\le j\le \ell}}$ is a system of witnesses and collectors for that covering. Moreover, a constant fraction of the W_i^1 are pairwise disjoint.*

Proof. The definition readily ensures that the union of the R_i is the set of all half-spaces and that Condition (a) holds. The monotonicity of the h_i implies that Condition (b) is also satisfied. Let $x \in W_i^j$.If $x \notin \partial H_i$, then let H denote the half-space parallel to H_i with x on its boundary. If $x \in \partial H_i$, we have to tilt the plane slightly: let H be a half space in R_i with x on its boundary but not parallel to H. In both cases H is in R_i and does not contain W_i^j and thus $x \in H \subset W_i^j$ and Condition (c) holds. As the cover of $\partial\rho\mathbb{B}$ by $\{H_i\}_{i=1,2,\ldots,m}$ is inclusion-minimal, we can extract a family $I \subseteq \{1, 2, \ldots, m\}$ of size $\Omega(m)$ such that the $\{W_i^1\}_{i\in I}$ are pairwise disjoint. □

In our analysis we will need some control over the intersection of C_i^j with $\rho\mathbb{B}$:

Claim 2.3.5. *$C_i^j \cap \rho\mathbb{B}$ is contained in a half-space parallel to H_i with intersection depth at most $9h_j$ with $\rho\mathbb{B}$.*

Proof. For any half-space H, the region $H \cap \rho\mathbb{B}$ is the convex hull of $H \cap \partial\rho\mathbb{B}$. It follows that $H \in R_i$ does not contain W_i^j if and only if $H \cap \partial\rho\mathbb{B}$ does not contain $W_i^j \cap \partial\rho\mathbb{B}$. This implies that for any $H \in R_i$ the spherical cap $H \cap \partial\rho\mathbb{B}$ is contained in a cap with same center as $W_i^j \cap \partial\rho\mathbb{B}$ and three times its radius. A half-space cutting out a cap of radius r_x in $\partial\rho\mathbb{B}$ intersects $\rho\mathbb{B}$ with depth $h_x = \Theta\left(\frac{r_x^2}{\rho}\right)$. Tripling the radius of a cap thus multiplies the depth of intersection by 9, and the statement follows. □

Claim 2.3.6. *If $\mathbb{E}\left[\mathrm{card}\left(W_i^1 \cap P\right)\right] = \Omega(1)$ then $\mathbb{E}\left[\mathrm{card}\left(W_i^1 \cap \mathcal{H}^{(1)}\right)\right] = \Omega(1)$*

Proof. If $W_i^1 \cap P$ is non-empty then W_i^1 contains the point of P extreme in direction $\vec{u}_i$ and $W_i^1 \cap \mathcal{H}^{(1)}$ is therefore non-empty. We thus have

$$\begin{aligned}\mathbb{E}\left[\mathrm{card}\left(W_i^1 \cap \mathcal{H}^{(1)}\right)\right] &\ge \mathbb{P}\left[W_i^j \cap \mathcal{H}^{(1)} \neq \emptyset\right] \\ &\ge \mathbb{P}\left[W_i^1 \cap P \neq \emptyset\right] \ge 1 - e^{-\mathbb{E}\left[\mathrm{card}\left(W_i^1 \cap P\right)\right]} = \Omega(1),\end{aligned}$$

the last inequality following from the Chernoff bound of Lemma 2.2.1 (a). □

2.3.3 Warm-up: Average-Case Analysis Made Easy

As a first example, let us use a system of witnesses and collectors to give a short[1] proof of a classical result of Raynaud.

Theorem 11 (Raynaud [41]). *Let $d \geq 2$ be a constant and $P = \{p_1, p_2, \ldots, p_n\}$ be a set of random points uniformly and independently distributed in a ball of $\mathbb{R}^d$. For any fixed k, the expected number of $k-$dimensional faces of the convex hull of P is $\Theta\left(n^{\frac{d-1}{d+1}}\right)$.*

Proof. The problem is invariant under scaling, so we can choose the ball to be $\mathbb{B}$. We use our construction of Section 2.3.2 with $\rho = 1$. Using Claim 2.3.2, we find that setting $h_j = (j/n)^{\frac{2}{d+1}}$ yields

$$\mathbb{E}\left[\text{card}\left(W_i^j \cap P\right)\right] = n\frac{f(h_j, 1)}{\text{vol}\,(\mathbb{B})} = \Theta(j).$$

Claim 2.3.3 gives $m = \Theta\left((\rho/h_1)^{\frac{d-1}{2}}\right) = \Theta\left(n^{\frac{d-1}{d+1}}\right)$. With Claims 2.3.1 and 2.3.5 this implies

$$\mathbb{E}\left[\text{card}\left(C_i^j \cap P\right)\right] \leq n\ O\left(\frac{f(h_j, 1)}{\text{vol}\,(\mathbb{B})}\right) = O(j)$$

so $\mathbb{E}\left[\text{card}\,\text{CH}(P)\right] = O\left(n^{\frac{d-1}{d+1}}\right)$ by Theorem 10 (i). Moreover, a constant fraction of the W_i^1 are pairwise disjoint, and Claim 2.3.6 ensures that $\mathbb{E}\left[\text{card}\left(W_i^1 \cap \mathcal{H}^{(1)}\right)\right] = \Omega(1)$; Theorem 10 (ii) thus implies that $\mathbb{E}\left[\text{card}\,\text{CH}(P)\right] = \Omega\left(n^{\frac{d-1}{d+1}}\right)$. □

2.3.4 Upper Bounds on the Smoothed Complexity

We now bound from above $\mathcal{S}\left(n, \mathcal{U}_{\delta\mathbb{B}}\right)$, using various arguments whose effectiveness vary with the value of δ.

Charging Argument. Our first smoothed complexity bound relies on a charging argument between the witness and the collector that form a pair. Let P^* be some point set of diameter at most 1 in $\mathbb{R}^d$. Without loss of generality we assume that P^* is contained in $\mathbb{B}$, and use a system of witnesses and collectors similar to the one presented in Section 2.3.2 with $\rho = 1 + \delta$.

We make an important change, though: the depth of intersection of each witness W_i^j depends on i, and is adapted to P^*. We start with an inclusion-minimal covering $H_1, H_2, \ldots, H_m$ of $\partial(\rho\mathbb{B})$ by half-spaces whose intersection depth with $\rho\mathbb{B}$ is $\Theta\left(\left(\frac{r}{1+\delta}\right)^2\right)$. Each cuts out a spherical caps of radius $r = \delta n^{-\frac{2}{d+1}}$ on $\partial(\rho\mathbb{B})$,

[1]Raynaud's original argument was more than 7 pages long, still leaving substantial computations to the reader.

and $m = O\left(n^{2-\frac{4}{d+1}}\left(1+\frac{1}{\delta}\right)^{d-1}\right)$ by Claim 2.3.3. For $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, \lceil \ln^2 n \rceil$ we define:

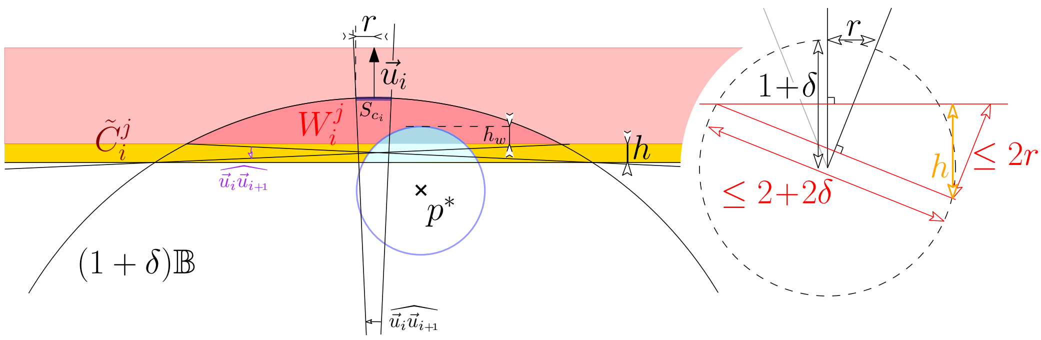
- R_i as the set of half-spaces whose inner normal is parallel to a vector from the origin to a point of $H_i \cap \partial(\rho\mathbb{B})$,
- W_i^j as the intersection of $\rho\mathbb{B}$ with a half-space parallel to H_i positioned so that $\mathbb{E}\left[W_i^j \cap P\right] = j$,
- C_i^j as the union of the half-spaces of R_i that do not contain W_i^j.

The proof of Lemma 2.3.4 readily implies that $\{(W_i^j, C_i^j)\}_{\substack{1\le i\le m\\ 1\le j\le \ell}}$ is a system of witnesses and collectors for the covering of the set of half-spaces in $\mathbb{R}^d$ by $R_1 \cup R_2 \cup \ldots \cup R_m$. To apply Theorem 10 (i) it remains to control the expected number of points of P in the collectors.

Claim 2.3.7. *For any perturbed point $p \in P$, with* card $P \geq 2^{\frac{d+1}{2}}$,

$$\mathbb{P}\left[p \in C_i^j\right] = O\left(\frac{1}{n} + \mathbb{P}\left[p \in W_i^j\right]\right).$$

Proof. Let $p^* \in P^*$ and p its perturbed copy. We fix some indices $1 \leq i \leq m$ and $1 \leq j \leq \lceil \ln^2 n \rceil$ and write $w = \mathbb{P}\left[p \in W_i^j\right]$ and $c = \mathbb{P}\left[p \in C_i^j\right]$.



Refer to the figure above and let $\tilde{C}_i^j$ be the halfspace with normal $\vec{u}_i$ containing $C_i^j \cap (1+\delta)\mathbb{B}$ and with minimal intersection depth with $(1+\delta)\mathbb{B}$. Let h denote the difference of the intersection depth of the half space cutting out W_i^j and $\tilde{C}_i^j$ with $(1+\delta)\mathbb{B}$ and h_w denote the intersection depth at which W_i^j intersects $B(p^*, \delta)$. Observe that $\tilde{C}_i^j$ intersects $B(p^*, \delta)$ with depth at most $h_w + h$. Since the diameter of $\tilde{C}_i^j \cap P$ is at most $2 + 2\delta$, considerations on similar triangles show that $h \leq 2r$.

If $h_w \leq 2r$ then we obtain the first part of the announced bound on c:

$$c \leq \frac{f(2r+h,\delta)}{f(2\delta,\delta)} \leq \frac{f(4\delta n^{-\frac{2}{d+1}},\delta)}{f(2\delta,\delta)} = \frac{f(4n^{-\frac{2}{d+1}},1)}{f(2,1)} = \frac{1}{f(2,1)} \int_0^{4n^{-\frac{2}{d+1}}} \left(2x - x^2\right)^{\frac{d-1}{2}} dx$$
$$\leq \frac{1}{f(2,1)} \int_0^{4n^{-\frac{2}{d+1}}} (2x)^{\frac{d-1}{2}} dx$$
$$= O\left(\frac{1}{n}\right).$$

If $h_w > 2r$ then we can assume that $c > 2w$, as otherwise the claim holds trivially. In particular $h_w \leq \delta$. Since $h \leq 2r = 2n^{-\frac{2}{d+1}}$, the hypothesis $n \geq 2^{\frac{d+1}{2}}$ ensures $h < \delta$ and the depths of intersection of both W_i^j and $\tilde{C}_i^j$ are in the interval $[0, 2\delta]$. We then have

$$c \leq \frac{f(h_w+h,\delta)}{f(2\delta,\delta)} = \frac{f\left(\left(1+\frac{h}{h_w}\right)h_w,\delta\right)}{f(2\delta,\delta)} \leq \left(1+\frac{h}{h_w}\right)^{\frac{d+1}{2}} w \leq 2^{\frac{d+1}{2}} w,$$

the last inequality coming from $h_w > 2r \geq h$. □

Claim 2.3.7 implies that, for n bigger than the constant $2^{\frac{d+1}{2}}$,

$$\mathbb{E}\left[\mathrm{card}\left(C_i^j \cap P\right)\right] = O\left(1 + \mathbb{E}\left[\mathrm{card}\left(W_i^j \cap P\right)\right]\right) = O(j)$$

and Theorem 10 (i) provides the following bound:

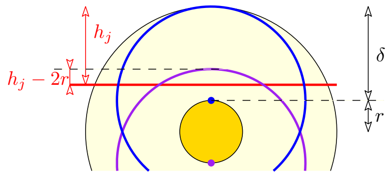
Proposition 12. $\mathcal{S}\left(n, \mathcal{U}_{\delta\mathbb{B}}\right) = O\left(n^{2\frac{d-1}{d+1}} + n^{2\frac{d-1}{d+1}}\delta^{-(d-1)}\right)$.

Large Perturbations. As $\delta \to \infty$ the bound of Proposition 12 does not tend to $\Theta\left(n^{\frac{d-1}{d+1}}\right)$, the average-case complexity bound. We thus complement it by a variation on the same system of witnesses and collectors better suited for the analysis of large perturbations.

Lemma 2.3.8. *For $\delta \geq 3n^{\frac{2}{d+1}}$ we have $\mathcal{S}(n, \mathcal{U}_{\delta\mathbb{B}}) = \Theta\left(n^{\frac{d-1}{d+1}}\right)$.*

Proof. We again assume, without loss of generality, that P^* is contained in $\mathbb{B}$ and use the construction of Section 2.3.2 with $\rho = 1 + \delta$ and $h_j = (1+\delta)\left(\frac{j}{n}\right)^{\frac{2}{d+1}}$. By Claim 2.3.3 we have

$$m = \Theta\left((1+\delta)/h_1\right)^{\frac{d-1}{2}} = \Theta\left(n^{\frac{d-1}{d+1}}\right).$$



For any point p^* in $\mathbb{B}$, we have

$$\frac{f(h_j-2,\delta)}{\mathrm{vol}\,(\delta\mathbb{B})} \leq \mathbb{P}\left[p \in W_i^j\right] \leq \frac{f(h_j,\delta)}{\mathrm{vol}\,(\delta\mathbb{B})}$$

Since $h_j \geq 3$, Claims 2.3.1 and 2.3.2 imply that $\mathbb{P}\left[p \in W_i^j\right] = \Theta(\frac{j}{n})$. By Claims 2.3.1 and 2.3.5 we get $\mathbb{P}\left[p \in C_i^j\right] = \Theta(\frac{j}{n})$ as well, so Theorem 10 (i) applies. A constant fraction of the W_i^1 are pairwise disjoint, by Lemma 2.3.4, and $\mathbb{E}\left[\mathrm{card}\,(W_i^1 \cap P)\right] = \Omega(1)$. Using Claim 2.3.6, it follows that Theorem 10 (ii) also applies, and

$$\mathbb{E}\left[\mathrm{card}\,\mathcal{H}^{(k)}\right] = \Theta(m) = \Theta\left(n^{\frac{d-1}{d+1}}\right).$$

□

Smoothed Number of Faces. Combining Proposition 12 and Lemma 2.3.8 we obtain the following upper bound on the smoothed number of faces of any dimension:

Theorem 13. *Let $d \geq 2$ be a constant. Then, we have:*

Range of δ	$\left[0, n^{\frac{2}{d+1}-\frac{1}{d-1}\lfloor\frac{d}{2}\rfloor}\right]$	$\left[n^{\frac{2}{d+1}-\frac{1}{d-1}\lfloor\frac{d}{2}\rfloor}, 1\right]$	$\left[1, 3n^{\frac{2}{d+1}}\right]$	$\left[3n^{\frac{2}{d+1}}, +\infty\right)$
$\mathcal{S}(n, \mathcal{U}_{\delta\mathbb{B}})$	$O\left(n^{\lfloor\frac{d}{2}\rfloor}\right)$	$O\left(n^{2\frac{d-1}{d+1}}\delta^{-(d-1)}\right)$	$O\left(n^{2\frac{d-1}{d+1}}\right)$	$\Theta\left(n^{\frac{d-1}{d+1}}\right)$

In dimension 2, a Euclidean noise of amplitude above $n^{-1/3}$ suffices to guarantee an expected sub-linear complexity. In dimension 3, the second bound is uninteresting as it exceeds the worst-case bound. In dimension d, a Euclidean noise of amplitude above $n^{-4/(d^2-1)}$ suffices to guarantee an expected sub-quadratic complexity.

Smoothed Number of Vertices. The bounds of Theorem 13 may be improved by a rescaling argument like the one used by Damerow and Sohler [17]: splitting the input into small cells and accounting separately for the contribution of each cell using a scaled version of Lemma 2.3.8. This only applies to the number of vertices, as a face of dimension 1 or more may involve perturbation of points coming from more than one cell.

Corollary 14. *For any constant $d \geq 2$, $\mathbb{E}\left[\mathrm{card}\,\mathcal{H}^{(1)}\right] = O\left(n^{\frac{d-1}{d+1}} + \delta^{-\frac{2d}{d+1}} n^{1+2\frac{d-1}{(d+1)^2}}\right)$, and for $d = 2$ we have:*

Range of δ	$[0, \frac{1}{\sqrt{n}}]$	$[\frac{1}{\sqrt{n}}, 1]$	$[1, n^{5/12}]$	$[n^{5/12}, n^{2/3}]$	$[n^{2/3}, +\infty]$
$\mathcal{S}(n, \mathcal{U}_{\delta\mathbb{B}})$	$O(n)$	$O\left(\delta^{-\frac{2}{3}} n^{\frac{2}{3}}\right)$	$O\left(n^{2/3}\right)$	$O\left(\delta^{-\frac{4}{3}} n^{\frac{11}{9}}\right)$	$O\left(n^{1/3}\right)$

Proof. We continue to assume that $P^* \subset \mathbb{B}$ and we cover $\mathbb{B}$ with $m' = \Theta(1 + r^{-d})$ disjoint cells of size $r = \frac{1}{3}\delta n^{-\frac{2}{d+1}}$. We partition P^* into $P_1^* \cup P_2^* \cup \ldots \cup P_{m'}^*$ by taking its intersection with each of the covering cells; we let P_i denote the perturbation of P_i^* and $n_i = \operatorname{card} P_i$. Every vertex of $\mathrm{CH}(P)$ is a vertex of some $\mathrm{CH}(P_i)$, and we can apply Lemma 2.3.8 to bound the number of vertices of $\mathrm{CH}(P_i)$ from above by ${n_i}^{\frac{d-1}{d+1}}$. If $m' > 1$, the sum is maximized with $\forall i,\ n_i = \frac{n}{m'}$ which bounds from above the number of vertices of $\mathrm{CH}(P)$ by

$$m'O\left(\left(\tfrac{n}{m'}\right)^{\frac{d-1}{d+1}}\right) = O\left(\left(\left(\delta n^{-\frac{2}{d+1}}\right)^{-d}\right)^{\frac{2}{d+1}} n^{\frac{d-1}{d+1}}\right) = O\left(\delta^{-\frac{2d}{d+1}} n^{\frac{4d}{(d+1)^2}+\frac{d-1}{d+1}}\right)$$
$$= O\left(\delta^{-\frac{2d}{d+1}} n^{1+2\frac{d-1}{(d+1)^2}}\right).$$

This proves the first statement. For the second statement, in two dimensions, we proceed differently in each regime:

$\delta \le \frac{1}{\sqrt{n}}$. In this case, the worst-case bound is used.

$1 \le \delta \le n^{5/12}$. This case is solved using Proposition 12.

$n^{2/3} \le \delta$. Here, Lemma 2.3.8 yields the result.

$n^{5/12} \le \delta \le n^{2/3}$. This case is handled through the first statement of the present corollary.

$\frac{1}{\sqrt{n}} \le \delta \le 1$. For the remaining case, we apply the same partitioning idea, but using Proposition 12 instead of Lemma 2.3.8 as an upper bound for one cell. Namely, considering a partitioning induced by covering cells of size δ, we get sets P_i^* whose convex hull has size $n_i^{\frac{2}{3}}$. Summing on the $\frac{1}{\delta^2}$ cells and using the concavity of $x \mapsto x^{\frac{2}{3}}$, we have

$$\sum_{i=1}^{O(\delta^{-2})} n_i^{\frac{2}{3}} = O\left(\delta^{-2}(\delta^2 n)^{\frac{2}{3}}\right) = O\left(\left(\tfrac{n}{\delta}\right)^{\frac{2}{3}}\right) \qquad \square$$

2.3.5 Lower Bound: Points in Convex Position

We finally analyze the expected complexity of Euclidean perturbations of some particular point configuration: points in convex position that are "nicely spread out"; more precisely, we take P^* to be an (ε, κ)-*sample* of a sphere with fixed radius, *ie.* a sample such that any ball of radius ε centered on the sphere contains between 1 and κ points of the sample.

Our motivation for studying this class of configurations is that they are natural candidates to realize the smoothed complexity of convex hulls in 2 and 3 dimensions and therefore provide an interesting lower bound. In light of Theorem 10 (ii), setting up the witnesses W_i^1 is enough to obtain a lower bound on the expected size of the convex hull; we give a complete analysis since at this stage it comes easily and makes it clear that the lower bound obtained by our choice of W_i^1 is sharp for these configurations.

Theorem 15. *Let $d \geq 2$ be a constant and $P^* = \{p_i^* : 1 \leq i \leq n\}$ be an $\left(\Theta\left(n^{\frac{1}{1-d}}\right), \Theta(1)\right)$-sample of the unit sphere in $\mathbb{R}^d$ and let $P = \{p_i = p_i^* + \eta_i\}$ where $\eta_1, \eta_2, \ldots, \eta_n$ are random variables chosen independently from $\mathcal{U}_{\delta\mathbb{B}}$. For any fixed k, $\mathbb{E}\left[\operatorname{card}\mathcal{H}^{(k)}\right]$ is*

Range of δ	$[0, n^{\frac{2}{1-d}}]$	$[n^{\frac{2}{1-d}}, 1]$	$[1, n^{\frac{2}{d+1}}]$	$[n^{\frac{2}{d+1}}, +\infty)$
$\mathbb{E}\left[\operatorname{card}\mathcal{H}^{(k)}\right]$	$\Theta(n)$	$\Theta\left(n^{\frac{d-1}{2d}}\delta^{\frac{1-d^2}{4d}}\right)$	$\Theta\left(n^{\frac{d-1}{2d}}\delta^{\frac{(1-d)^2}{4d}}\right)$	$\Theta\left(n^{\frac{d-1}{d+1}}\right)$

The last bound corresponds to the average-case behavior which applies for δ sufficiently large, as follows from Lemma 2.3.8. We thus only have to analyze the range $\delta \leq n^{\frac{2}{d+1}}$. Note that the first bound merely reflects that a point remains extreme when the noise is small compared to the distance to the nearest hyperplane spanned by points in its vicinity, and that the bounds reveal that as the amplitude of the perturbation increases, the expected size of the convex hull does not vary monotonically (see Figures 2.2a and 2.2c): the lowest expected complexity is achieved by applying a noise of amplitude roughly the diameter of the initial configuration.

The following claim will be useful to position the witnesses and control the collectors.

Claim 2.3.9. *Under the assumptions of Theorem 15, let $j \leq \ln^2 n$, let H be a half-space such that $\mathbb{E}[\operatorname{card}(H \cap P)] = \Theta(j)$ and let h denote its depth of intersection with $(1+\delta)\mathbb{B}$.*

(i) If $\delta = O\left(\frac{j}{n}\right)^{\frac{2}{d-1}}$ then $h = \Theta\left(\left(\frac{j}{n}\right)^{\frac{2}{d-1}}\right)$
and if $\Omega\left(\frac{j}{n}\right)^{\frac{2}{d-1}} \leq \delta \leq O\left(n^{\frac{2}{d+1}}\right)$ then $h = \Theta\left(\left(\frac{j}{n}\right)^{\frac{1}{d}}\delta^{\frac{d+1}{2d}}\right)$.

(ii) If H' is a half-space that intersects $(1+\delta)\mathbb{B}$ with depth $9h$ then

$$\mathbb{E}[\operatorname{card}(H' \cap P)] = O(\mathbb{E}[\operatorname{card}(H \cap P)]).$$

Proof. The region $S \subseteq \partial\mathbb{B}$ in which we can center a ball of radius δ that intersects H is the intersection of $\partial\mathbb{B}$ with a half-space parallel to H and that intersects it with depth h; S is thus a spherical cap of $\mathbb{B}$ of radius $\sqrt{2h-h^2} = \Theta\left(\sqrt{h}\right)$ and $(d-1)$-dimensional area $\Theta\left(h^{\frac{d-1}{2}}\right)$. By the sampling condition in the definition of P^*, each ball of radius $n^{\frac{1}{1-d}}$ centered on $\partial\mathbb{B}$ contains $\Theta(1)$ points of P^*. In total there are thus $\Theta\left(nh^{\frac{d-1}{2}}\right)$ points $p^* \in P^*$ such that $(p^* + \delta\mathbb{B}) \cap H \neq \emptyset$. For the rest of this proof call these points *relevant*. How much a relevant point contributes to $\mathbb{E}\left[\operatorname{card} H \cap P\right]$ depends on how h compares to δ.

If $h \leq \delta$ then H intersects any ball $p^* + \delta\mathbb{B}$ with depth at most δ, and Claim 2.3.2 bounds the contribution of any relevant point p^* to $\mathbb{E}\left[\operatorname{card} H \cap P\right]$ by

$$\frac{\operatorname{vol}\left(H \cap (p^* + \delta\mathbb{B})\right)}{\operatorname{vol}\left(\delta\mathbb{B}\right)} \leq \frac{f(h,\delta)}{f(2\delta,\delta)} = O\left(\frac{h^{\frac{d+1}{2}}\delta^{\frac{d-1}{2}}}{\delta^d}\right) = O\left(\left(\frac{h}{\delta}\right)^{\frac{d+1}{2}}\right).$$

Shrinking h by a factor two, we obtain that a constant fraction (depending only on d) of the relevant points contribute for at least $\frac{f(h/2,\delta)}{f(2\delta,\delta)} = \Omega\left(\left(\frac{h}{\delta}\right)^{\frac{d+1}{2}}\right)$ to $\mathbb{E}\left[\operatorname{card}\left(H \cap P\right)\right]$, hence

$$\Theta(j) = \mathbb{E}\left[\operatorname{card}\left(H \cap P\right)\right] = \Theta\left(nh^{\frac{d-1}{2}}\left(\frac{h}{\delta}\right)^{\frac{d+1}{2}}\right) = \Theta\left(n\delta^{-\frac{d+1}{2}}h^d\right)$$

and $h = \Theta\left(\left(\frac{j}{n}\right)^{\frac{1}{d}}\delta^{\frac{d+1}{2d}}\right)$. The condition $h \leq \delta$ thus amounts to $\delta = \Omega\left(\left(\frac{j}{n}\right)^{\frac{2}{d-1}}\right)$, giving the second regime.

If $h > \delta$ then a constant fraction of the relevant points p^* are such that H intersects $p^* + \delta\mathbb{B}$ with depth at least $\delta/2$, thus containing a constant fraction of each of these balls (and the rest of the relevant points contribute less). It follows that $\Theta(j) = \Theta\left(nh^{\frac{d-1}{2}}\right)$ and $h = \Theta\left(\left(\frac{j}{n}\right)^{\frac{2}{d-1}}\right)$. The condition $h > \delta$ amounts to $\delta = O\left(\left(\frac{j}{n}\right)^{\frac{2}{d-1}}\right)$, giving the first regime.

Observe that in either case, the number of points in $H \cap P$ depends polynomially on h. Thus, multiplying the depth by 9 multiplies the expected number of points by a constant (depending only on d) and statement (ii) follows. □

Proof of Theorem 15. We use our construction of Section 2.3.2 with $\rho = 1 + \delta$. We fix h_j such that each W_i^j contains $\Theta(j)$ points of P; the values of h_j are given by Claim 2.3.9(i). By Claim 2.3.5, C_i^j is contained in a half-space that intersects $(1+\delta)\mathbb{B}$ with depth at most $9h_j$. Claim 2.3.9(ii) thus ensures that

$$\mathbb{E}\left[\operatorname{card}\left(C_i^j \cap P\right)\right] = O\left(\mathbb{E}\left[\operatorname{card}\left(W_i^j \cap P\right)\right]\right) = O(j)$$

and we can apply Theorem 10 (i). Lemma 2.3.4 and Claim 2.3.6 further guarantee that we can apply Theorem 10 (ii). By Claim 2.3.3, $m = \Theta\left(\left(\frac{1+\delta}{h_1}\right)^{\frac{d-1}{2}}\right)$ and we have three regimes.

If $\delta = O\left(\frac{1}{n}\right)^{\frac{2}{d-1}}$ then Claim 2.3.9(i) yields $h_1 = \Theta\left(\left(\frac{1}{n}\right)^{\frac{2}{d-1}}\right)$ and

$$m = \Theta\left(\left(\frac{1+\delta}{h_1}\right)^{\frac{d-1}{2}}\right) = \Theta\left(\left(\frac{1}{\left(\frac{1}{n}\right)^{\frac{2}{d-1}}}\right)^{\frac{d-1}{2}}\right) = \Theta\left(n\right).$$

If $\Omega\left(\frac{j}{n}\right)^{\frac{2}{d-1}} \leq \delta \leq O\left(n^{\frac{2}{d+1}}\right)$ then Claim 2.3.9(i) yields $h_1 = \Theta\left(\left(\frac{1}{n}\right)^{\frac{1}{d}} \delta^{\frac{d+1}{2d}}\right)$. If $\delta \leq 1$ then

$$m = \Theta\left(\left(\frac{1+\delta}{h_1}\right)^{\frac{d-1}{2}}\right) = \Theta\left(\left(\frac{1}{\left(\frac{1}{n}\right)^{\frac{1}{d}} \delta^{\frac{d+1}{2d}}}\right)^{\frac{d-1}{2}}\right) = \Theta\left(n^{\frac{d-1}{2d}} \delta^{\frac{1-d^2}{4d}}\right)$$

and if $\delta \geq 1$ then

$$m = \Theta\left(\left(\frac{1+\delta}{h_1}\right)^{\frac{d-1}{2}}\right) = \Theta\left(\left(\frac{\delta}{\left(\frac{1}{n}\right)^{\frac{1}{d}} \delta^{\frac{d+1}{2d}}}\right)^{\frac{d-1}{2}}\right) = \Theta\left(n^{\frac{d-1}{2d}} \delta^{\frac{(1-d)^2}{4d}}\right)$$

Up to multiplicative constants, the boundaries between the regimes can be set as in the statement of the Theorem. □

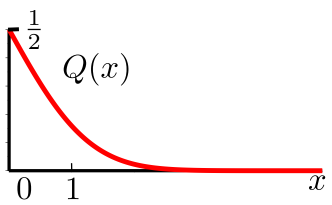
2.4 Gaussian Perturbation

The Gaussian model raises two difficulties compared to the Euclidean model: the computations are more technical and the fact that the perturbations have non-compact support requires to adapt the witness-collector construction. We expect some of the results to extend to arbitrary dimension *mutatis mutandis*, but for the sake of the presentation only spell out the analysis in the two-dimensional case.

2.4.1 Preliminaries

Recall that if $X \sim \mathcal{N}(\mu, \sigma^2)$ then for any $t \geq 0$ we have $\mathbb{P}[X \geq \mu + t\sigma] = Q(t)$, where Q is the tail probability of the standard Gaussian distribution:

$$\forall x \in \mathbb{R}, \quad Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt.$$



The solution to the functional equation

$$f(x)e^{f(x)} = x$$

is called the Lambert function $\mathcal{W}_0$ [15, Equation (3.1)]. For $x \geq 0$ the definition of $\mathcal{W}_0(x)$ is non-ambiguous and satisfies

$$\forall x \geq 1.01, \qquad \mathcal{W}_0(x) = \Theta(\ln x). \qquad (2.5)$$

This essentially follows from [15, Equations (4.6) and (4.9)]; note that the constant 1.01 is arbitrary and any constant strictly larger than 1 would do (the constants in the $\Theta()$ would change but we do not care). The following inequalities will be useful:

Lemma 2.4.1.

(i) For $x > 0$, $\quad \dfrac{e^{-\frac{x^2}{2}}}{x + \frac{1}{x}} < \sqrt{2\pi} Q(x) < \dfrac{e^{-\frac{x^2}{2}}}{x}.$

(ii) For $x > 1/4$, $\quad Q\left(x + \dfrac{1}{x}\right) = \Theta\left(Q(x)\right).$

(iii)

(a) $\sum_{i=0}^{n} e^{-i^2 x} = O\left(1 + \frac{1}{\sqrt{x}}\right)$

(b) *For any constant* $\gamma > 0$, *for* $x > \frac{\gamma}{n^2}$, $\sum_{i=0}^{n} e^{-i^2 x} = \Omega\left(\frac{1}{\sqrt{x}}\right)$

(c) *For any constant* $\gamma > 0$, *for* $x < \frac{\gamma}{n^2}$, $\sum_{i=0}^{n} e^{-i^2 x} = \Omega\left(n\right)$

Proof. The upper bound of statement (i) comes from

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt < \int_x^\infty \frac{t}{\sqrt{2\pi}x} e^{-\frac{t^2}{2}} dt = \int_{\frac{x^2}{2}}^\infty \frac{e^{-t}}{x\sqrt{2\pi}} dt = \frac{1}{\sqrt{2\pi}x} e^{-\frac{x^2}{2}}$$

and the lower bound comes from the fact that

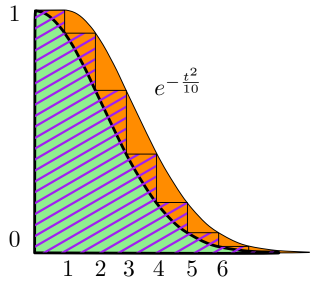
$$\begin{aligned}\left(1 + \frac{1}{x^2}\right) Q(x) &= \int_x^\infty \left(1 + \frac{1}{x^2}\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\ &> \int_x^\infty \left(1 + \frac{1}{t^2}\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}}.\end{aligned}$$

Now, for statement (ii), we have $Q(x) \geq Q(x + \frac{1}{x})$ since Q is a decreasing function. Moreover, from statement (i) we have

$$\begin{aligned}&\sqrt{2\pi} Q(x + \tfrac{1}{x}) > \frac{x + \frac{1}{x}}{1 + \left(x + \frac{1}{x}\right)^2} e^{-\frac{\left(x + \frac{1}{x}\right)^2}{2}} \\ &= \left(\frac{x^4 + x^2}{x^4 + 3x^2 + 1} e^{-1-\frac{1}{2x^2}}\right) \left(\frac{e^{-\frac{x^2}{2}}}{x}\right) > \left(\frac{x^4 + x^2}{x^4 + 3x^2 + 1} e^{-1-\frac{1}{2x^2}}\right) \sqrt{2\pi} Q(x)\end{aligned}$$

Statement (ii) then follows from noting that the image of $[1/4, +\infty)$ under the function $x \mapsto \frac{x^4+x^2}{x^4+3x^2+1} e^{-1-\frac{1}{2x^2}}$ is contained in some closed interval of $(0, +\infty)$.

The proof of Statements (iii-a) and (iii-b) follows from a standard comparison between series and integrals:



if $x > \frac{\gamma}{n^2}$,

$$\sum_{i=0}^{n} e^{-i^2 x} \geq \int_0^{n+1} e^{-t^2 x} dt \geq \int_0^{n\sqrt{x}} e^{-u^2} \frac{du}{\sqrt{x}}$$
$$\geq \frac{\int_0^{\sqrt{\gamma}} e^{-u^2} du}{\sqrt{x}} \geq \Omega\left(\frac{1}{\sqrt{x}}\right)$$

and for any $x > 0$,

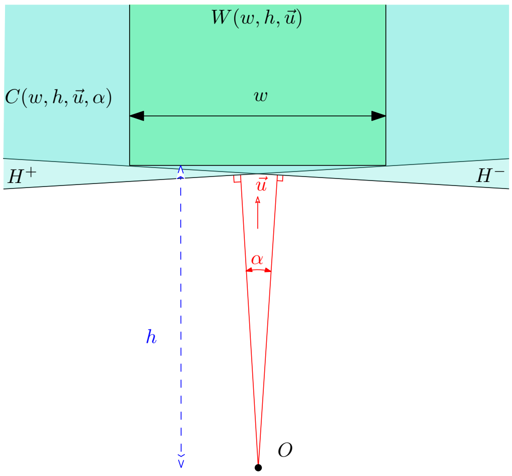
$$\sum_{i=0}^{n} e^{-i^2 x} \leq 1 + \int_0^{n} e^{-t^2 x} dt \leq 1 + \int_0^{n\sqrt{x}} e^{-u^2} \frac{du}{\sqrt{x}} \leq 1 + \int_0^{\infty} e^{-u^2} \frac{du}{\sqrt{x}}$$
$$= O\left(1 + \frac{1}{\sqrt{x}}\right).$$

Statement (iii-c) is trivial since, when $x < \frac{\gamma}{n^2}$, $\sum_{i=0}^{n} e^{-i^2 x} \geq n \cdot e^{-\gamma} = \Omega(n)$. □

2.4.2 Witness-Collector Construction

One Witness-Collector Pair. The pairs witness-collectors that we use to analyze Gaussian perturbations are based on the following basic construction. Let w, h and α be positive reals and $\vec{u}$ some vector in the plane.

- We define $R(\vec{u}, \alpha)$ as the set of half-planes whose inner normal makes an angle at most $\frac{\alpha}{2}$ with $\vec{u}$.
- We define $W(w, h, \vec{u})$ as the semi-infinite half strip with axis of symmetry $O + \mathbb{R}\vec{u}$, with width w and distance h to the origin. To save breath we define the *height* of a semi-infinite half strip as its distance to the origin – so $W(w, h, \vec{u})$ has height h.
- We define $C(w, h, \vec{u}, \alpha)$ as the union of the half-planes in $R(\vec{u}, \alpha)$ that do not contain $W(w, h, \vec{u})$.



The following more explicit description of $C(w, h, \vec{u}, \alpha)$ will be convenient:

Claim 2.4.2. *$C(w, h, \vec{u}, \alpha) = H^- \cup H^+$ where H^- and H^+ are the half-planes whose inner normals make an angle of $\pm\frac{\alpha}{2}$ with $\vec{u}$, that contain $W(w, h, \vec{u})$ and have one of the corners of $W(w, h, \vec{u})$ on their boundary.*

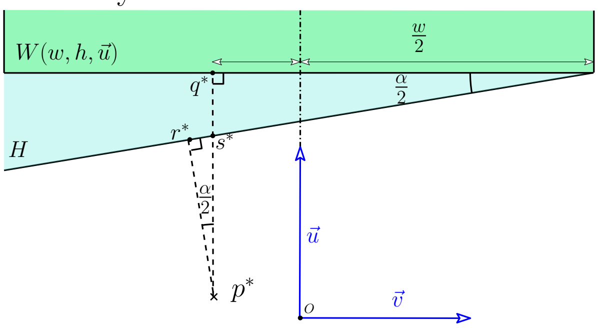
Proof. This follows from observing that any halfplane through $\partial H^+ \cap \partial H^-$ and contained in $C(w,h,\vec{u},\alpha)$ also contains $W(w,h,\vec{u})$. □

This construction has the following properties:
(a') Any halfplane whose inner normal makes an angle at most $\frac{\alpha}{2}$ with $\vec{u}$ contains $W(w,h,\vec{u})$ or is contained in $C(w,h,\vec{u},\alpha)$.
(b') If $h_j > h_{j+1}$ and $w_j < w_{j+1}$ then $W(w_j,h_j,\vec{u}) \subseteq W(w_{j+1},h_{j+1},\vec{u})$.
(c') $W(w,h,\vec{u}) \subseteq C(w,h,\vec{u},\alpha)$.
Families of pairs $(W(w,h,\vec{u}), C(w,h,\vec{u},\alpha))$ therefore combine easily into systems of witnesses and collectors. We will control the expected number of points in a witness by setting w and h adequately and tune α accordingly thanks to the next fact. We say that a point p^* is *in the slab of* $W(w,h,\vec{u})$ if the ray $p^* + \mathbb{R}^+\vec{u}$ intersects $W(w,h,\vec{u})$.

Claim 2.4.3. *Let $\vec{u}$ be arbitrary and let $\vec{v}$ denote a unit vector orthogonal to $\vec{u}$. If $p^* \in \mathbb{R}^2$ is in the slab of $W(w,h,\vec{u})$ and outside the interior of $C(w,h,\vec{u},\alpha)$ then*

$$d(p^*, C(w,h,\vec{u},\alpha)) = d(p^*, W(w,h,\vec{u}))\cos\frac{\alpha}{2} - \left(\frac{w}{2} + |\overrightarrow{Op^*}\cdot\vec{v}|\right)\sin\frac{\alpha}{2}.$$

Proof. Let H denote the half-plane contained in $C(w,h,\vec{u},\alpha)$ and whose distance to p^* is minimal. Let q^* and r^* denote respectively the orthogonal projections of p^* on $W(w,h,\vec{u})$ and $C(w,h,\vec{u},\alpha)$. Let s^* denote the intersection of p^*q^* with the boundary of H.



The assumptions ensure that

$$|p^*q^*| = d(p^*, W(w,h,\vec{u}))$$

and

$$|p^*r^*| = d(p^*, C(w,h,\vec{u},\alpha)).$$

With $\vec{v} \in \mathbb{S}^1, \vec{v} \perp \vec{u}$, we have

$$|q^*s^*| = \left(\frac{w}{2} + |\overrightarrow{Op^*}\cdot\vec{v}|\right)\tan\frac{\alpha}{2} \quad \text{and} \quad |p^*r^*| = |p^*s^*|\cos\frac{\alpha}{2} = (|p^*q^*| - |q^*s^*|)\cos\frac{\alpha}{2}$$

and the statement follows. □

System of Witnesses and Collectors. Our construction is parameterized by some positive real α and two sequences of positive reals $h_1 > h_2 > \cdots > h_\ell$ and $w_1 \le w_2 \le \cdots \le w_\ell$. We choose an inclusion-minimal cover of $\partial\mathbb{B}$ by half-planes $H_1, H_2, \ldots, H_m$ each intersecting $\partial\mathbb{B}$ in a circular arc of angle α; we let $\vec{u_i}$ denote the center of $H_i \cap \partial\mathbb{B}$ and note that $m = \Theta\left(\frac{1}{\alpha}\right)$. We define R_i as the set of half-planes whose inner normal is parallel to a vector from the origin to a point of

$H_i \cap \partial\mathbb{B}$ and let

$$W_i^j = W(w_j, h_j, \vec{u}_i) \quad \text{and} \quad C_i^j = C(w_j, h_j, \vec{u}_i, \alpha).$$

Lemma 2.4.4. *$R_1 \cup R_2 \cup \ldots \cup R_m$ covers the set of half-planes and $\{(W_i^j, C_i^j)\}_{\substack{1\le i\le m\\ 1\le j\le \ell}}$ is a system of witnesses and collectors for that covering. Moreover, some $\Omega\left(\frac{h_1}{w_1}\right)$ of the W_i^1 are pairwise disjoint.*

Proof. The definition readily ensures that the union of the R_i is the set of all half-planes and that Condition (a) holds. The monotonicity of the h_i and the w_i implies that Condition (b) is also satisfied. Claim 2.4.2 implies that each W_i^j is contained in the corresponding C_i^j, so Condition (c) holds. Each W_i^1 is contained in a wedge with apex the origin and opening angle $\Theta\left(\frac{w_1}{h_1}\right)$. Some $\Omega\left(\frac{h_1}{w_1}\right)$ of these wedges are disjoint (except in the origin), so the corresponding W_i^1's are pairwise disjoint. □

2.4.3 Warm-up: Gaussian Polygons Made Easy

To illustrate our construction, we revisit the classical problem of computing the expected number of faces of the convex hull from a Gaussian distribution:

Theorem 16 (Rényi and Sulanke [44])**.** *Let $P = \{p_1, p_2, \ldots, p_n\}$ be a set of random points chosen independently from $\mathcal{N}(0, I_2)$. The expected number of vertices of the convex hull of P is $\Theta\left(\sqrt{\ln n}\right)$.*

Proof. We use the construction of Section 2.4.2 with $\ell = \ln^2 n$ and the values of α, w_j and h_j set as specified on the right. Lemma 2.4.4 ensures that we obtain a system of witnesses and collectors, so it only remains to analyze the expected number of points in W_i^j and C_i^j. We complete each vector $\vec{u}_i$ into a direct, orthonormal frame $(O, \vec{v}_i, \vec{u}_i)$; in that frame, the coordinates of any point $p \in P$ write (x_i, y_i) where x_i, y_i are independent random variables chosen from $\mathcal{N}(0, 1)$.

$$\alpha = \frac{1}{h_1} = \Theta\left(\frac{1}{\sqrt{\ln n}}\right)$$
$$w_j = 2$$
$$h_j = \sqrt{\mathcal{W}_0\left(\frac{n^2}{j^2}\right)}$$

The probability for p to be in W_i^j therefore writes

$$\mathbb{P}\left[p \in W_i^j\right] = \mathbb{P}\left[y_i > h_j\right] \mathbb{P}\left[|x_i| < 1\right] = \Theta\left(Q(h_j)\right).$$

Lemma 2.4.1 (i) yields $Q(x) = \Theta\left(\frac{1}{x}e^{-\frac{x^2}{2}}\right)$ for $x > 1$ so, since $j \le \ln^2 n$,

$$Q(h_j) = \Theta\left(\frac{e^{-\frac{1}{2}\mathcal{W}_0\left(\frac{n^2}{j^2}\right)}}{\sqrt{\mathcal{W}_0\left(\frac{n^2}{j^2}\right)}}\right) = \Theta\left(\frac{1}{\sqrt{\mathcal{W}_0\left(\frac{n^2}{j^2}\right)e^{\mathcal{W}_0\left(\frac{n^2}{j^2}\right)}}}\right) = \Theta\left(\frac{j}{n}\right) \tag{2.6}$$

and $\mathbb{E}\left[\text{card}\left(W_i^j \cap P\right)\right] = n\Theta\left(Q(h_j)\right) = \Theta(j)$. Since for $n \ge 3$, $\alpha < \frac{1}{\sqrt{\mathcal{W}_0(3^2)}} < \frac{\pi}{4}$, $\tan\frac{\alpha}{2} < 0.5$ and $\frac{2h_j}{w_j} \ge \frac{2h_\ell}{w_j} = \sqrt{\mathcal{W}_0\left(\frac{n^2}{\ln^4 n}\right)} \ge 1$. This means that $\tan\frac{\alpha}{2} < \frac{2h_j}{w_j}$, so the origin is not in C_i^j. By Claims 2.4.2 and 2.4.3, C_i^j is contained in the union of two half-planes with height $\tilde{h_j} = h_j\cos\frac{\alpha}{2} - \sin\frac{\alpha}{2} = h_j - O(\alpha)$. Thus,

$$\mathbb{E}\left[\text{card}\left(C_i^j \cap P\right)\right] \le 2nQ\left(\tilde{h_j}\right) = 2\left(nQ\left(h_j\right)\right)\left(\frac{Q\left(\tilde{h_j}\right)}{Q\left(h_j\right)}\right)$$

We already observed that $nQ\left(h_j\right) = \Theta(j)$. By Lemma 2.4.1 (i) we have

$$\frac{Q\left(\tilde{h_j}\right)}{Q\left(h_j\right)} = \frac{e^{-\frac{1}{2}\left(\tilde{h_j}^2 - h_j^2\right)}h_j}{\tilde{h_j}} = \frac{h_j}{\tilde{h_j}}e^{O\left(h_j\alpha + \alpha^2\right)}$$

and with Equation (2.5) we finally obtain

$$\mathbb{E}\left[\text{card}\left(C_i^j \cap P\right)\right] = O\left(je^{O(h_j\alpha)}\right) = O\left(je^{O\left(\sqrt{1-\frac{\ln j}{\ln n}}\right)}\right) = O(j),$$

and Property (b') holds. Theorem 10 (i) then yields that

$$\mathbb{E}\left[\text{card}\,\mathcal{H}^{(1)}\right] = O(m) = O(\sqrt{\ln n}).$$

Let H_i denote the halfplane with same height and inner normal as W_i^1 and $p_{\vec{u}_i}$ be the point of P extremal in direction $\vec{u}_i$. By construction $p_{\vec{u}_i}$ belongs to $\mathcal{H}^{(1)}$, thus

$$\mathbb{E}\left[\text{card}\left(W_i^1 \cap \mathcal{H}^{(1)}\right)\right] \ge \mathbb{P}\left[p_{\vec{u}_i} \in W_i^1\right] = \mathbb{P}\left[p_{\vec{u}_i} \in H_i\right]\mathbb{P}\left[p_{\vec{u}_i} \in W_i^1 \mid p_{\vec{u}_i} \in H_i\right]$$

We have

$$\mathbb{P}\left[p_{\vec{u}_i} \in H_i\right] = \mathbb{P}\left[P \cap H_i \neq \emptyset\right] \ge \mathbb{P}\left[P \cap W_i^1 \neq \emptyset\right] \ge 1 - \frac{1}{e} > \frac{1}{2}$$

by Lemma 2.2.1 (a). Gaussian noise perturbs points independently in directions x and y of frame $(O, \vec{v}_i, \vec{u}_i)$. The choice of $p_{\vec{u}_i}$ in P depends only on the y perturbation, thus knowing that $p_{\vec{u}_i} \in H_i$, deciding if it is in W_i^1 or in $H_i \setminus W_i^1$ depends only on the coordinate along direction v_i, thus

$$\begin{aligned}
\mathbb{P}\left[p_{\vec{u}_i} \in W_i^1 \mid p_{\vec{u}_i} \in H_i\right] &= \sum_{p \in H_i \cap P} \mathbb{P}\left[p \in W_i^1 \mid p_{\vec{u}_i} = p\right] \mathbb{P}\left[p_{\vec{u}_i} = p \mid p_{\vec{u}_i} \in H_i\right] \\
&= \sum_{p \in H_i \cap P} \mathbb{P}\left[|x_p| \leq 1\right] \mathbb{P}\left[p_{\vec{u}_i} = p \mid p_{\vec{u}_i} \in H_i\right] \\
&= \mathbb{P}\left[|x_{p_1}| \leq 1\right] \sum_{p \in H_i \cap P} \mathbb{P}\left[p_{\vec{u}_i} = p \mid p_{\vec{u}_i} \in H_i\right] \\
&= \mathbb{P}\left[|x_{p_1}| \leq 1\right] \quad = 1 - 2Q(1) > \frac{1}{2} \qquad (2.7)
\end{aligned}$$

Together we get that Lemma 2.4.4 ensures that we can also apply Theorem 10 (ii) and get that $\mathbb{E}\left[\operatorname{card} \mathcal{H}^{(1)}\right] = \Omega(\sqrt{\ln n})$ as well. □

2.4.4 Upper Bound on the Smoothed Complexity

As in the Euclidean case, for large Gaussian perturbation the smooth complexity is identical to the i.i.d. case. It is possible, as done in Section 2.4.4.1, to obtain a Gaussian analogue of Lemma 2.3.8 and apply the rescaling argument to get a smoothws complexity for any scale of perturbation. This bound is, however, worse than what we can obtain by a charging argument in the spirit of Claim 2.3.7 and Proposition 12, and is presented in Section 2.4.4.2.

2.4.4.1 Large Perturbation and Rescaling for Gaussian perturbation

Lemma 2.4.5. *Let $P^* = \{p_1^*, p_2^* \dots, p_n^*\}$ be a set of points in $\mathbb{R}^2$ with diameter $2r$. Let P be the set of $p_i = p_i^* + \eta_i$, where $\eta_i \sim \mathcal{N}(0, \sigma^2 I_2)$ are independent, for $\sigma \geq 3r\sqrt{\ln n}$. Then, the expected number of vertices of the convex hull of P is $O\left(\sqrt{\ln n}\right)$.*

Proof. We can suppose that P^* is included in $r\mathbb{B}$. We set up the Gaussian witness-collector construction of Section 2.4.2, with $\ell = \ln^2 n$ and the values of α, w_j and h_j set as specified on the right.

$$\alpha = \frac{\sigma}{h_1 - r} = \Theta\left(\frac{1}{\sqrt{\ln n}}\right) \qquad w_j = w = \frac{7}{2}\sigma \qquad h_j = r + \sigma\sqrt{\mathcal{W}_0\left(\frac{n^2}{j^2}\right)}$$

Every point $p \in P$ writes $p = x_i\vec{v}_i + y_i\vec{u}_i$ with $x_i, y_i \sim \mathcal{N}(0, \sigma^2)$ and $\vec{u}_i, \vec{v}_i \in \mathbb{S}^1, \vec{v}_i \perp \vec{u}_i$. Thus, the probability for p to be in W_i^j is upper-bounded by $\mathbb{P}\left[y_i > h_j\right]$ and is lower-bounded by $\mathbb{P}\left[y_i > h_j\right] \mathbb{P}\left[|x_i| \leq \frac{w-r}{2}\right]$.

It's easy to see that $\mathbb{P}\left[|x_i| \leq \frac{w-r}{2}\right]$ is lower-bounded from below by $\mathbb{P}\left[|x_i| \leq \frac{5w}{14}\right]$ (since $r \leq \sigma = \frac{2w}{7}$), which is a constant.

Now, $\mathbb{P}\left[y_i > h_j\right] \in \left[Q\left(\frac{h_j+r}{\sigma}\right), Q\left(\frac{h_j-r}{\sigma}\right)\right]$ so

$$\mathbb{P}\left[y_i > h_j\right] < Q\left(\sqrt{\mathcal{W}_0\left(\frac{n^2}{j^2}\right)}\right) = \Theta\left(\frac{j}{n}\right).$$

As

$$\begin{aligned}\sigma &\geq 3r\sqrt{\ln n} \geq 2r\sqrt{\ln(n^2)}\\ &\geq 2r\sqrt{\mathcal{W}_0\left(n^2\right)} \geq 2r\sqrt{\mathcal{W}_0\left(\frac{n^2}{j^2}\right)},\end{aligned}$$

we get

$$\frac{2r}{\sigma} \leq \frac{1}{\sqrt{\mathcal{W}_0\left(\frac{n^2}{j^2}\right)}} = \frac{\sigma}{h_j - r}$$

and

$$Q\left(\frac{h_j + r}{\sigma}\right) = Q\left(\frac{h_j - r}{\sigma} + \frac{2r}{\sigma}\right) \geq Q\left(\frac{h_j - r}{\sigma} + \frac{\sigma}{h_j - r}\right).$$

Now, using Lemma 2.4.1 (ii) , as long as $\frac{h_j - r}{\sigma} > \frac{1}{4}$ (which is true for $n \geq 3$ and $j \leq \ln^2 n$) we get that

$$Q\left(\frac{h_j - r}{\sigma} + \frac{\sigma}{h_j - r}\right) = \Omega\left(Q\left(\frac{h_j - r}{\sigma}\right)\right) = \Omega\left(\frac{j}{n}\right)$$

using Equation (2.6).

We obtain that for any $p^* \in P^*$, $\mathbb{P}\left[p \in W_i^j\right] = \Theta\left(\frac{j}{n}\right)$, and so

$$\mathbb{E}\left[\text{card}\left(P \cap W_i^j\right)\right] = \Theta\left(j\right).$$

Condition (a') therefore holds. Since $n \geq 3$, $\alpha < \frac{1}{\sqrt{\mathcal{W}_0(3^2)}} < \frac{\pi}{4}$, $\tan\frac{\alpha}{2} < 0.5$ and

$$\frac{2h_j}{w} \geq 2\frac{h_\ell - r}{w} = \frac{4}{7}\sqrt{\mathcal{W}_0\left(\frac{n^2}{\ln^4 n}\right)} \geq 0.5.$$

Thus, $\tan\frac{\alpha}{2} < \frac{2h_j}{w}$, so the origin is not in C_i^j. Using Claim 2.4.3 with p^* at the origin, the collector C_i^j is included in the union of two half-spaces with height $\tilde{h_j} = h_j \cos\frac{\alpha}{2} - \frac{w}{2}\sin\frac{\alpha}{2} \geq h_j \cos\frac{\alpha}{2} - \frac{7}{8}\alpha\sigma$, since $\sin(x) < x$ for $x > 0$.

Now,

$$\begin{aligned} h_j \cos\tfrac{\alpha}{2} &= h_j + (\cos\tfrac{\alpha}{2} - 1)h_j \\ &\geq h_j - \frac{1}{2}\left(\tfrac{\alpha}{2}\right)^2 h_j = h_j - \frac{1}{8}\alpha\sigma\left(\frac{h_j}{h_1 - r}\right) \\ &\geq h_j - \frac{1}{8}\alpha\sigma\left(\frac{h_j}{h_j - r}\right) \\ &\geq h_j - \frac{1}{8}\alpha\sigma \end{aligned}$$

since $\cos x - 1 \geq -\frac{1}{2}x^2$. Thus, we obtain $\tilde{h_j} \geq h_j - \alpha\sigma$.

Note that $Q\left(\frac{h_j - r}{\sigma}\right) \geq Q\left(\frac{\tilde{h_j} - r}{\sigma} + \alpha\right) = Q\left(\frac{\tilde{h_j} - r}{\sigma} + \frac{\sigma}{h_1 - r}\right)$.

Since $\tilde{h_j} \leq h_1$, $\frac{\sigma}{h_1 - r} \leq \frac{\sigma}{\tilde{h_j} - r}$ and

$$\begin{aligned} Q\left(\frac{h_j - r}{\sigma}\right) &> Q\left(\frac{\tilde{h_j} - r}{\sigma} + \frac{\sigma}{h_1 - r}\right) \\ &> Q\left(\frac{\tilde{h_j} - r}{\sigma} + \frac{\sigma}{\tilde{h_j} - r}\right) \\ &= \Omega\left(Q\left(\frac{\tilde{h_j} - r}{\sigma}\right)\right) \end{aligned}$$

using Lemma 2.4.1 (ii) (we have $\frac{\tilde{h_j} - r}{\sigma} > \frac{1}{4}$ for $n \geq 3$ and $j \leq \ln^2 n$), we get $Q\left(\frac{\tilde{h_j} - r}{\sigma}\right) = O\left(\frac{j}{n}\right)$ by Equation (2.6).

Thus, $\mathbb{E}\left[\text{card}\left(P \cap C_i^j\right)\right] = O(j)$, Theorem 10 (i) can be applied and we obtain $\mathbb{E}\left[\text{card}\,\mathcal{H}^{(1)}\right] = \mathbb{E}\left[\text{card}\,\text{CH}(P)\right] = O\left(\frac{1}{\alpha}\right) = O\left(\sqrt{\ln n}\right)$.

□

Corollary 17.

$$\mathcal{S}\left(n, \mathcal{N}(0, \sigma^2)\right) = O\left(\frac{\ln(n)}{\sigma^2}\sqrt{\ln\left(n\sigma^2\right)} + \sqrt{\ln n}\right).$$

Proof. Let P^* be some point of diameter at most 1. Without loss of generality, we assume that $P^* \subset \mathbb{B}$. We cover $\mathbb{B}$ with $m'' = \Theta\left(1 + \frac{1}{r^2}\right)$ disjoint cells of size $r = \frac{\sigma}{3\sqrt{\ln n}}$ We break down P^* into $P_1^* \cup P_2^* \cup \cdots \cup P_{m''}^*$ by taking its intersection with each covering cells; we let P_i denote the perturbation of P_i^* and $n_i = \text{card}\, P_i$. Every vertex of $\text{CH}(P)$ is a vertex of some $\text{CH}(P_i)$, and we can apply Lemma 2.4.5 to bound the number of vertices of $\text{CH}(P_i)$ from above by $\sqrt{\ln n_i}$. If $m'' > 1$, the sum is maximized with $\forall i, n_i = \frac{n}{m''}$, which bounds from above the number of vertices of CH(P) by

$$m''O\left(\sqrt{\ln\left(\frac{n}{m''}\right)}\right) = O\left(\frac{1}{r^2}\sqrt{\ln\left(nr^2\right)}\right) = O\left(\frac{\ln(n)}{\sigma^2}\sqrt{\ln\left(n\sigma^2\right)}\right)$$

using Equality (2.5). □

2.4.4.2 A Better Smoothed Upper Bound for Gaussian Perturbation

Theorem 18. $\mathcal{S}(n, \mathcal{N}(0, \sigma^2 I_2)) = O\left(\frac{\sqrt{\ln n}}{\sigma} + \sqrt{\ln n}\right)$.

Let P^* be some point set of diameter at most 1 in the plane and, without loss of generality, assume that P^* is contained in $\mathbb{B}$. We use a system of witnesses and collectors similar to the one presented in Section 2.4.2 with $\ell = \ln^2 n$. As in the Euclidean case, a key difference is that the depth of intersection of each witness W_i^j depends on i, and is adapted to P^*. Specifically, we set w and α to the values on the right, choose the $\vec{u_i}$ regularly spaced on $\mathbb{S}^1$ with $\widehat{\vec{u_i}\vec{u}_{i+1}} = \Theta(\alpha)$. We then define $R_i = R(\vec{u_i}, \alpha)$, $W_i^j = W(h_i^j, w, \vec{u_i})$ and $C_i^j = C(h_i^j, w, \vec{u_i}, \alpha)$ where h_i^j depends on P^* and is tuned so that the expected number of points in the witnesses are what they should be.

$$w = 2(1+\sigma)$$

$$\alpha = \frac{\sqrt{(2+\sigma)^2 + \frac{2\sqrt{2}\sigma}{\sqrt{\ln n}} + 2\sqrt{2}\sigma^2} - (2+\sigma)}{(1+\sigma\sqrt{\ln n})}$$

$$h_i^j \text{ s. t. } \mathbb{E}\left[\operatorname{card}\left(P \cap W(h_i^j, w, \vec{u_i})\right)\right] = j$$

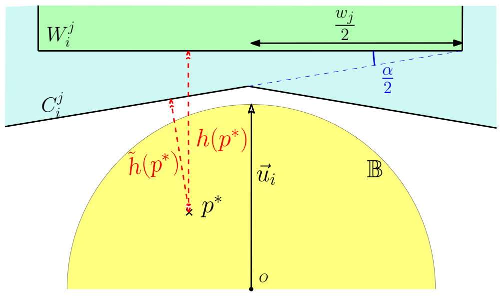
We first relate the distances from a point p^* of P^* to a witness W_i^j and a collector C_i^j:

Claim 2.4.6. *If $p^* \in \mathbb{R}^2$ is in the slab of W_i^j and outside the interior of C_i^j, then*

$$d(p^*, W_i^j) - d(p^*, C_i^j) \leq \frac{\sigma}{\sqrt{2\ln n}}.$$

Proof. Let $h(p^*)$ and $\tilde{h}(p^*)$ denote, respectively, $d(p^*, W_i^j)$ and $d(p^*, C_i^j)$. By Claim 2.4.3,

$$\tilde{h}(p^*) = h(p^*)\cos\frac{\alpha}{2} - \left(\frac{w}{2} + |\overrightarrow{Op^*}\cdot\vec{v}_i|\right)\sin\frac{\alpha}{2}$$



and with $1 - \cos x \leq \frac{x^2}{2}$, $\sin|x| < |x|$, $\frac{w}{2} = 1+\sigma$, and $|\overrightarrow{Op^*}\cdot\vec{v}_i| \leq 1$, this becomes

$$\begin{aligned} & h(p^*) - \tilde{h}(p^*) \\ & \leq h(p^*) - h(p^*)\cos\frac{\alpha}{2} + (2+\sigma)\sin\frac{\alpha}{2} \\ & \leq h(p^*)\frac{\alpha^2}{8} + (2+\sigma)\frac{\alpha}{2} \end{aligned}$$

The distance from p^* to $W_i^j = W(h_i^j, w, \vec{u_i})$ is maximized when p^* is located at the point of $\partial\mathbb{B}$ with outer normal $-\vec{u_i}$ and all other points of P^* are at the symmetric position, at the point of $\partial\mathbb{B}$ with normal $\vec{u_i}$. The same argument as in Equation (2.6) and the observation that $\ln(x) > \mathcal{W}_0(x)$ for $x \geq 3$ yield the upper bound

$$h(p^*) \leq 2 + \sigma\sqrt{\mathcal{W}_0(n^2)} \leq 2\left(1 + \sigma\sqrt{\ln n}\right).$$

Injecting this in the above inequality we get

$$h(p^*) - \tilde{h}(p^*) \leq \left(1 + \sigma\sqrt{\ln n}\right)\frac{\alpha^2}{4} + (2+\sigma)\frac{\alpha}{2}$$

The polynomial

$$P(\alpha) = \left(\left(1 + \sigma\sqrt{\ln n}\right)\frac{\alpha^2}{4} + (2+\sigma)\frac{\alpha}{2} - \frac{\sigma}{\sqrt{2\ln n}}\right)$$

can be checked to be negative for

$$0 \leq \alpha \leq \frac{\sqrt{(2+\sigma)^2 + \frac{2\sqrt{2}\sigma}{\sqrt{\ln n}} + 2\sqrt{2}\sigma^2} - (2+\sigma)}{(1 + \sigma\sqrt{\ln n})},$$

and that concludes the proof. □

The distance from a point p^* to W_i^j and C_i^j determines the probability that the perturbation of p^* belongs to either of these sets.

Claim 2.4.7. $\mathbb{P}\left[p \in W_i^j\right] = \Theta\left(Q\left(\frac{d(p^*, W_i^j)}{\sigma}\right)\right)$ *and* $\mathbb{P}\left[p \in C_i^j\right] = O\left(Q\left(\frac{d(p^*, C_i^j)}{\sigma}\right)\right)$.

Proof. A perturbed point p is in W_i^j if it satisfies two conditions: (α) its displacement from p^* along $\vec{u_i}$ should be greater than $d(p^*, W_i^j)$, and (β) its displacement in the orthogonal direction is in the slab of width w_j. The conditions are independent, (α) is true with probability $Q\left(\frac{d(p^*, W_i^j)}{\sigma}\right)$ and (β) is true with constant probability since $w = 2 + 2\sigma$ ensures that the allowed orthogonal displacement for p^* is larger than σ. The statement for W_i^j follows. As for the collectors, the probability that a perturbed point p is in C_i^j is bounded from above by the sum of the probabilities to be in H^+ and to be in H^-, which are both $Q\left(\frac{d(p^*,C_i^j)}{\sigma}\right)$. $\square$

Combining the two previous claims we now get that witness and collector get, on average, essentially the same number of points.

Claim 2.4.8. *For any $p^* \in P^*$, we have $\mathbb{P}\left[p \in C_i^j\right] = O\left(\frac{1}{n} + \mathbb{P}\left[p \in W_i^j\right]\right)$.*

Proof. Let $h(p^*)$ and $\tilde{h}(p^*)$ denote, respectively, $d(p^*, W_i^j)$ and $d(p^*, C_i^j)$. Since $w \geq 2$ any point in P^* is in the slab of W_i^j.

First assume that p^* is not in C_i^j. Claim 2.4.6 then ensures that $\tilde{h}(p^*) \geq h(p^*) - \frac{\sigma}{\sqrt{2\ln n}}$. If $h(p^*) > \sigma\sqrt{2\ln n} + \frac{\sigma}{\sqrt{2\ln n}}$ then by Claim 2.4.7, Lemma 2.4.1 (i), and the fact that Q is decreasing, we have

$$\begin{aligned}\mathbb{P}\left[p \in C_i^j\right] = O\left(Q\left(\frac{\tilde{h}(p^*)}{\sigma}\right)\right) &= O\left(Q\left(\frac{h(p^*)}{\sigma} - \frac{1}{\sqrt{2\ln n}}\right)\right)\\ &= O\left(Q\left(\sqrt{2\ln n} + \frac{1}{\sqrt{2\ln n}} - \frac{1}{\sqrt{2\ln n}}\right)\right)\\ &= O\left(Q\left(\sqrt{2\ln n}\right)\right) = O\left(\frac{1}{n}\right)\end{aligned}$$

and the statement follows. If $h(p^*) \leq \sigma\sqrt{2\ln n} + \frac{\sigma}{\sqrt{2\ln n}}$ then we have

$$\mathbb{P}\left[p \in C_i^j\right] = O\left(Q\left(\frac{\tilde{h}(p^*)}{\sigma}\right)\right) = O\left(Q\left(\frac{h(p^*)}{\sigma} - \frac{1}{\sqrt{2\ln n}}\right)\right)$$

If $h(p^*) \leq \frac{\sigma}{4} + \frac{\sigma}{\sqrt{2\ln n}} \leq \sigma\left(\frac{1}{4} + \frac{1}{\sqrt{2\ln 3}}\right)$ then $h(p^*)$ is bounded from above by 2σ and

$$\mathbb{P}\left[p \in W_i^j\right] = \Omega\left(Q\left(2\right)\right) = \Omega(1).$$

Then, $\mathbb{P}\left[p \in C_i^j\right] \leq 1 = O\left(\mathbb{P}\left[p \in W_i^j\right]\right)$ and the statement also holds. Thus, we can suppose that $h(p^*) \geq \frac{\sigma}{4} + \frac{\sigma}{\sqrt{2\ln n}}$ and use Lemma 2.4.1 (ii) to get:

$$\mathbb{P}\left[p \in C_i^j\right] = O\left(Q\left(\frac{h(p^*)}{\sigma} - \frac{1}{\sqrt{2\ln n}} + \frac{1}{\frac{h(p^*)}{\sigma} - \frac{1}{\sqrt{2\ln n}}}\right)\right).$$

Since

$$\frac{1}{\frac{h(p^*)}{\sigma} - \frac{1}{\sqrt{2\ln n}}} \geq \frac{1}{\sqrt{2\ln n} + \frac{1}{\sqrt{2\ln n}} - \frac{1}{\sqrt{2\ln n}}} = \frac{1}{\sqrt{2\ln n}}$$

we get

$$\mathbb{P}\left[p \in C_i^j\right] = O\left(Q\left(\frac{h(p^*)}{\sigma}\right)\right) = O\left(\mathbb{P}\left[p \in W_i^j\right]\right)$$

and the statement also holds.

Finally assume that $p^* \in C_i^j$. In such a case Claims 2.4.3 and 2.4.6 do not apply directly, but we have $\frac{1}{2} \leq \mathbb{P}\left[p \in C_i^j\right] \leq 1$ so we have to argue that $\mathbb{P}\left[p \in W_i^j\right] = \Omega(1)$. Let us move from p^* in the direction $-\vec{u}_i$ until we reach some point $\bar{p}^*$ on the boundary of C_i^j; observe that $\mathbb{P}\left[\bar{p}^* + \eta \in C_i^j\right] \geq \frac{1}{2}$ where $\eta \sim \mathcal{N}(0, \sigma^2 I_2)$. Now $\bar{p}^*$ satisfies the hypotheses of Claim 2.4.6 and the above analysis implies that $\mathbb{P}\left[\bar{p}^* + \eta \in W_i^j\right] = \Omega\left(\mathbb{P}\left[\bar{p}^* + \eta \in C_i^j\right]\right) = \Omega(1)$. Moving from p^* to $\bar{p}^*$ only increased the distance to W_i^j, so we also have $\mathbb{P}\left[p \in W_i^j\right] \geq \mathbb{P}\left[\bar{p}^* + \eta \in W_i^j\right] = \Omega(1)$. □

We now have all the ingredients to prove our upper bound on the smoothed complexity under Gaussian noise.

Proof of Theorem 18. We set up our witnesses and collectors as described above. Since the parameter w is fixed and each sequence $\{h_i^j\}_j$ is decreasing, Lemma 2.4.4 yields that $\{(W_i^j, C_i^j)\}_{\substack{1 \leq i \leq m \\ 1 \leq j \leq \ell}}$ is a system of witnesses and collectors for the covering $R_1 \cup R_2 \cup \ldots \cup R_m$ of the set of half-planes. Each parameter h_i^j is set so that $\mathbb{E}\left[\text{card}\left(W_i^j \cap P\right)\right] = j$ and Claim 2.4.8 implies that $\mathbb{E}\left[\text{card}\left(C_i^j \cap P\right)\right] = O(j)$. Theorem 10 (i) thus implies that

$$\mathcal{S}(n, \mathcal{N}(0, \sigma^2 I_2)) = O\left(\frac{1}{\alpha}\right) = O\left(\frac{\left(1 + \sigma\sqrt{\ln n}\right)}{(2+\sigma)\left(\sqrt{1 + \frac{2\sqrt{2}}{(2+\sigma)^2}\left(\sigma^2 + \frac{\sigma}{\sqrt{\ln n}}\right)} - 1\right)}\right). \tag{2.8}$$

If $\sigma \leq \frac{1}{\sqrt{\ln n}}$ then Equation (2.8) simplifies into

$$\mathcal{S}(n, \mathcal{N}(0, \sigma^2 I_2)) = O\left(\frac{1}{\sqrt{1 + \frac{1}{(1+\sigma)^2}\left(\sigma^2 + \frac{\sigma}{\sqrt{\ln n}}\right)} - 1}\right).$$

Notice that in this case, $\frac{1}{(1+\sigma)^2}\left(\sigma^2 + \frac{\sigma}{\sqrt{\ln n}}\right)$ is bounded by some constant C and since for $0 < x < C$, $\sqrt{1+x} - 1 = \Theta(x)$,

$$\mathcal{S}(n, \mathcal{N}(0, \sigma^2 I_2)) = O\left(\frac{1}{\frac{1}{(1+\sigma)^2}\left(\sigma^2 + \frac{\sigma}{\sqrt{\ln n}}\right)}\right) = O\left(\frac{\sqrt{\ln n}}{\sigma}\right).$$

If $\frac{1}{\sqrt{\ln n}} \leq \sigma \leq 1$ then Equation (2.8) simplifies into

$$\mathcal{S}(n, \mathcal{N}(0, \sigma^2 I_2)) = O\left(\frac{\sigma\sqrt{\ln n}}{\sqrt{1+\Theta(\sigma^2)}-1}\right) = O\left(\frac{\sigma\sqrt{\ln n}}{\sigma^2}\right) = O\left(\frac{\sqrt{\ln n}}{\sigma}\right)$$

If $1 \leq \sigma$ then Equation (2.8) simplifies into

$$\mathcal{S}(n, \mathcal{N}(0, \sigma^2 I_2)) = O\left(\frac{\sigma\sqrt{\ln n}}{\sigma\Theta(1)}\right) = O(\sqrt{\ln n}).$$

In each case we therefore have $\mathcal{S}(n, \mathcal{N}(0, \sigma^2 I_2)) = O\left(\frac{\sqrt{\ln n}}{\sigma} + \sqrt{\ln n}\right)$. □

2.4.5 Lower Bound on Smoothed Complexity: Points in Convex Position

We finally investigate lower bounds on the smoothed complexity by analyzing the size of the convex hull of a Gaussian perturbation of points in convex position, as in Section 2.3.5.

Theorem 19. *Let $P^* = \{p_i^*, 1 \leq i \leq n\}$ be the set of vertices of a regular n-gon of radius 1 in $\mathbb{R}^2$. Let $P = \{p_i = p_i^* + \eta_i\}$ where $\eta_1, \eta_2, \ldots, \eta_n$ are random vectors in $\mathbb{R}^2$ chosen independently from $\mathcal{N}(0, \sigma^2 I_2)$. The expected number of vertices of the convex hull of P is*

Range of σ	$\left[0, \frac{1}{n^2}\right]$	$\left[\frac{1}{n^2}, \frac{1}{\sqrt{\ln n}}\right]$	$\left[\frac{1}{\sqrt{\ln n}}, +\infty\right)$
$\mathbb{E}\left[\text{card}\,\mathcal{H}^{(1)}\right]$	$\Omega(n)$	$\Omega\left(\frac{\sqrt[4]{\ln(n\sqrt{\sigma})}}{\sqrt{\sigma}}\right)$	$\Omega\left(\sqrt{\ln n}\right)$

We use the witness-collector construction presented in Section 2.4.2. We only care about the lower-bound, so, shortening W_i^1 into W_i, we need only define one level of witnesses $\{W_i\}_{1 \leq i \leq m}$ to apply Theorem 10 (ii).

Parameters Setting. We set h_1 and w_1 depending on σ and n as summarized below. We let $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m$ denote a family of vectors in $\mathbb{S}^1$ such that $\vec{u}_i$ is aligned with $p^*_{\left\lfloor \frac{2\pi i}{m} \right\rfloor}$, so these vectors are more or less equally spaced on $\mathbb{S}^1$. The witnesses are defined as $W_i = W(w_1, h_1, \vec{u}_i)$. We choose m maximal so that the $\{W_i\}$ are pairwise disjoint; Lemma 2.4.4 ensures that we can set $m = \Omega\left(\min\left(n, \frac{h_1}{w_1}\right)\right)$.

	$0 \le \sigma < \frac{2}{n^2}$	$\frac{2}{n^2} \le \sigma < \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)}$	$\frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)} \le \sigma$
w_1	2σ	$2\sigma + 2\sqrt{\sigma}\left(\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}(n\sqrt{\sigma})^{\frac{4}{3}}\right)\right)^{-1/4}$	$2\sigma + 2$
h_1	1	$1 + \sigma\sqrt{\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}(n\sqrt{\sigma})^{\frac{4}{3}}\right)}$	$1 + \sigma\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)}$

Preparation. Let $i \in \{1, 2 \ldots, m\}$. As in Section 2.4.3, we let $(O, \vec{v}_i, \vec{u}_i)$ denote some orthonormal frame and let H_i be the halfplane supporting W_i with inner normal $\vec{u}_i$. We renumber the points of P^* with indices in $\{-\frac{n-1}{2}, \ldots, \frac{n-1}{2}\}$ so that p_0^* is the point in direction $\vec{u}_i$. For the sake of the presentation we assume that n is odd (the case of even n follows with trivial modifications). We write (x_i, y_i) for the coordinates of p_i in $(O, \vec{v}_i, \vec{u}_i)$ and denote by $p_{\vec{u}_i} \in \mathcal{H}^{(1)}$ the point of P extremal in direction $\vec{u}_i$. Our goal is to prove that $\mathbb{E}\left[\text{card}\left(W_i \cap \mathcal{H}^{(1)}\right)\right]$ is $\Omega(1)$ in order to apply Theorem 10 (ii); in the light of

$$\mathbb{E}\left[\text{card}\left(W_i \cap \mathcal{H}^{(1)}\right)\right] \ge \mathbb{P}\left[p_{\vec{u}_i} \in W_i\right]$$

we set out to bound from below the probability that $p_{\vec{u}_i}$ lies in W_i. We write z_t for the distance from p_t^* to H_i and note that

$$z_0 = h_1 - 1 \qquad \text{and} \qquad z_t = h_1 - 1 + 1 - \cos\frac{2\pi t}{n}$$

For $x \in [-\frac{1}{2}, \frac{1}{2}]$ we have $8x^2 \le 1 - \cos(2\pi x) \le 20x^2$, hence

$$\begin{aligned} h_1 - 1 + 8\frac{t^2}{n^2} \quad &\le z_t \le \quad h_1 - 1 + 20\frac{t^2}{n^2} \\ \frac{8t^2}{n^2} \quad &\le z_t - z_0 \le \quad \frac{20t^2}{n^2}. \end{aligned} \tag{2.9}$$

Analysis for Small σ. We start with the case $\sigma < \frac{2}{n^2}$, where the analysis is simpler but already uses the main ingredients of the general case. Since $h_1 = 1$, we have $z_0 = 0$ and therefore p_0^* lies on the boundary of H_i. We condition on the event $\{p_0 \in H_i \text{ and } p_{\vec{u}_i} = p_0\}$ and obtain:

$$\mathbb{P}\left[p_{\vec{u}_i} \in W_i\right] \ge \mathbb{P}\left[p_0 \in W_i \mid p_0 \in H_i \text{ and } p_{\vec{u}_i} = p_0\right] \mathbb{P}\left[p_0 \in H_i \text{ and } p_{\vec{u}_i} = p_0\right] \tag{2.10}$$

We bound each of these terms in turn.

Claim 2.4.9. *When $\sigma < \frac{2}{n^2}$, $\mathbb{P}\left[p_0 \in H_i \text{ and } p_{\vec{u}_i} = p_0\right] = \Omega(1)$.*

Proof. Using the independence of the random variables $\{y_t\}_t$ we write

$$\begin{aligned}\mathbb{P}\left[p_0 \in H_i \text{ and } p_{\vec{u}_i} = p_0\right] &\geq \mathbb{P}\left[y_0 \geq h_1 \text{ and } \forall t \neq 0, y_t \leq h_1\right] \\ &= \mathbb{P}\left[y_0 \geq h_1\right] \prod_{t \neq 0} \mathbb{P}\left[y_t \leq h_1\right]\end{aligned}$$

As $p_0^* \in H_i$, the point p_0 has probability at least $\frac{1}{2}$ of remaining in the half-plane H_i after a Gaussian perturbation, so $\mathbb{P}\left[y_0 \geq h_1\right] \geq \frac{1}{2}$. Moreover, $y_t \sim \mathcal{N}(h_1 - z_t, \sigma^2)$ so Lemma 2.4.1 (i) and the bounds on z_t and σ lead to:

$$\mathbb{P}\left[y_t \geq h_1\right] = \mathbb{P}\left[y_t - \mathbb{E}\left[y_t\right] \geq z_t\right] = Q\left(\tfrac{z_t}{\sigma}\right) \leq Q\left(\frac{8t^2}{n^2\sigma}\right) \leq Q\left(4t^2\right) \leq e^{-2t^2},$$

and $\mathbb{P}\left[y_t \leq h_1\right] \geq 1 - e^{-2t^2}$. Taking the logarithm we obtain

$$\begin{aligned}\ln \mathbb{P}\left[p_0 \in H_i \text{ and } p_{\vec{u}_i} = p_0\right] &= \ln \mathbb{P}\left[y_0 \geq h_1\right] + 2\sum_{t=1}^{\frac{n-1}{2}} \ln \mathbb{P}\left[y_t \leq h_1\right] \\ &\geq \ln \tfrac{1}{2} + 2\sum_{t=1}^{\frac{n-1}{2}} \ln\left(1 - e^{-2t^2}\right)\end{aligned}$$

Then, using that for $x \in (0, \frac{1}{2}]$ we have $\ln(1-x) > -2x$ and Lemma 2.4.1 (iii-a) we get

$$-\ln \mathbb{P}\left[p_0 \in H_i \text{ and } p_{\vec{u}_i} = p_0\right] \leq \ln 2 - 2\sum_{t=1}^{\frac{n-1}{2}} -2e^{-2t^2} \leq \ln 2 + 4\sum_{t=0}^{\frac{n-1}{2}} e^{-2t^2} = O(1),$$

and

$$\mathbb{P}\left[p_{\vec{u}_i} = p_0\right] = e^{-O(1)} = \Omega(1).$$

□

Equation (2.10) finally implies that $\mathbb{P}\left[p_{\vec{u}_i} \in W_i\right]$ is $\Omega(1)$, so $\mathbb{E}\left[\text{card}\left(W_i \cap \mathcal{H}^{(1)}\right)\right]$ is indeed $\Omega(1)$ for this range of σ.

Relevant Points. The contribution of the tth point to $\mathbb{E}\left[\text{card}\left(H_i \cap P\right)\right]$ is $Q\left(\frac{z_t}{\sigma}\right)$. The gist of our analysis for larger σ is to split the points into two parts, the relevant points where $Q\left(\frac{z_t}{\sigma}\right) = \Theta\left(Q\left(\frac{z_0}{\sigma}\right)\right)$ and the irrelevant ones. The expected number of points in H_i is (up to a constant multiplicative factor) at least the number of relevant points times $Q\left(\frac{z_0}{\sigma}\right)$; fine tuning z_0 so that this product is $\Omega(1)$ then amounts to solving some functional equation. Specifically, we call a point p_t *relevant* if $|t| \leq t_m = \min\left(\left\lfloor \frac{n\sigma}{\sqrt{z_0}} \right\rfloor, \frac{n-1}{2}\right)$. We denote by P_r the relevant points.

The same conditioning as in Equation (2.10) yields

$$\mathbb{P}\left[p_{\vec{u}_i} \in W_i\right] \geq \mathbb{P}\left[p_{\vec{u}_i} \in W_i \mid p_{\vec{u}_i} \in H_i \cap P_r\right] \mathbb{P}\left[p_{\vec{u}_i} \in H_i \cap P_r\right]. \tag{2.11}$$

One of the terms can be bounded as easily as for small σ.

Claim 2.4.10. *When $\sigma \geq \frac{2}{n^2}$, $\mathbb{P}\left[p_{\vec{u}_i} \in W_i \mid p_{\vec{u}_i} \in H_i \cap P_r\right] \geq \frac{1}{2}$.*

Proof. First, note that the parameter w_1 is set so that in the orthogonal projection on the $\vec{v}_i$-axis, the image of the witness contains the image of the ball $B(p_t, \sigma)$ whenever p_t is relevant. This ensures that

$$\mathbb{P}\left[p_{\vec{u}_i} \in W_i \mid p_{\vec{u}_i} \in H_i \text{ and } p_{\vec{u}_i} \text{ is relevant}\right] \geq 1 - 2Q(1) \geq \frac{1}{2}.$$

□

Counting Relevant Points in $\mathbf{H_i}$. Bounding the remaining probability requires different quantitative analysis depending on the range of σ but are based on the same principle: counting the expected number of relevant points in H_i. Since H_i has inner normal $\vec{u}_i$, we have

$$\mathbb{P}\left[p_{\vec{u}_i} \in H_i \mid p_{\vec{u}_i} \in P_r\right] = \mathbb{P}\left[H_i \cap P_r \neq \emptyset\right].$$

Thus, by the Chernoff bound of Lemma 2.2.1 (a), to show that the right-hand term is $\Omega(1)$ it suffices to show that H_i contains on average $\Omega(1)$ relevant points. Notice that

$$\mathbb{P}\left[p_t \in H_i\right] = \mathbb{P}\left[y_t - \mathbb{E}\left[y_t\right] > z_t\right] = \Omega\left(Q\left(\frac{z_t}{\sigma}\right)\right),$$

so the expected number of relevant points in H_i writes

$$\Omega\left(\sum_{t=-t_m}^{t_m} Q\left(\frac{z_t}{\sigma}\right)\right) = \Omega\left(Q\left(\frac{z_0}{\sigma}\right) \sum_{t=0}^{t_m} \frac{1}{\frac{z_t}{\sigma} + \frac{\sigma}{z_t}} \frac{z_0}{\sigma} e^{-\frac{1}{2\sigma^2}(z_t^2 - z_0^2)}\right). \tag{2.12}$$

Recall that $z_t = z_0 + \Theta\left(\frac{t^2}{n^2}\right)$. How we evaluate Equation (2.12) depends on the range of σ.

Large σ. When $\sigma \geq \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)}$, every point is relevant, *ie.* $t_m = \frac{n-1}{2}$, since

$$z_0 = \sigma \sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)} \quad \text{implies} \quad \frac{n\sigma}{\sqrt{z_0}} = n \sqrt{\frac{\sigma}{\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)}}} \geq \frac{n}{2}.$$

Claim 2.4.11. *When $\sigma \geq \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)}$, $\mathbb{P}\left[p_{\vec{u}_i} \in H_i \cap P_r\right] = \Omega(1)$.*

Proof. Since every point is relevant, this probability equals the probability that $H_i \cap P$ is non-empty. Computations similar to that of Equation (2.6) yield that $Q\left(\frac{z_0}{\sigma}\right) = \Theta\left(\frac{1}{n}\right)$. Moreover, $z_t \geq \frac{\sigma}{2}$ so $\frac{1}{\frac{z_t}{\sigma}+\frac{\sigma}{z_t}} = \Theta\left(\frac{\sigma}{z_t}\right)$. Also, $z_t = \Theta(z_0)$ and $z_t^2 - z_0^2 = \Theta\left(\frac{t^2 z_0}{n^2}\right)$. Injecting these three relations in Equation (2.12) we obtain that the expected number of (relevant) points in H_i writes

$$\mathbb{E}\left[\operatorname{card}\left(H_i \cap P\right)\right] = \Omega\left(\frac{1}{n}\sum_{t=0}^{\frac{n-1}{2}} \frac{z_0}{z_t} e^{-t^2\Theta\left(\frac{z_0}{n^2\sigma^2}\right)}\right) = \Omega\left(\frac{1}{n}\sum_{t=0}^{\frac{n-1}{2}} e^{-j^2\Theta\left(\frac{z_0}{n^2\sigma^2}\right)}\right)$$

Since $\frac{z_0}{n^2\sigma^2} < \frac{4}{n^2}$, Lemma 2.4.1 (iii-c) implies that

$$\sum_{t=0}^{\frac{n-1}{2}} e^{-t^2\Theta\left(\frac{z_0}{n^2\sigma^2}\right)} = \Omega(n),$$

so we finally get that H_i contains $\Omega(1)$ (relevant) points on average. The Chernoff bound of Lemma 2.2.1 (a) yields that $\mathbb{P}\left[H_i \cap P \neq \emptyset\right]$ is $\Omega(1)$, and so is

$$\mathbb{P}\left[p_{\vec{u}_i} \in H_i \cap P_r\right] = \Omega(1).$$

□

Intermediate σ. When $\frac{2}{n^2} \leq \sigma < \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)}$ we have $z_0 = \sigma\sqrt{\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}\left(n\sqrt{\sigma}\right)^{\frac{4}{3}}\right)}$. The function $x \mapsto \frac{x}{\sqrt{\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}\left(n\sqrt{x}\right)^{\frac{4}{3}}\right)}}$ is increasing. Let us define σ_0 as the solution of

$$\sigma_0 = \frac{1}{4}\sqrt{\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}\left(n\sqrt{\sigma_0}\right)^{\frac{4}{3}}\right)}$$

Using that $\mathcal{W}_0$ is the solution to $f(x)e^{f(x)} = x$, we obtain that

$$\sigma_0 = \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)}.$$

Then, for $\sigma < \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)}$ we have

$$\frac{n\sigma}{\sqrt{z_0}} \leq n\sqrt{\frac{\sigma}{\sqrt{\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}\left(n\sqrt{\sigma}\right)^{\frac{4}{3}}\right)}}} < n\sqrt{\frac{\sigma_0}{\sqrt{\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}\left(n\sqrt{\sigma_0}\right)^{\frac{4}{3}}\right)}}} = \frac{n}{2}$$

and $t_m = \left\lfloor \frac{n\sigma}{\sqrt{z_0}} \right\rfloor$. Notice that

$$\mathbb{P}\left[p_{\vec{u}_i} \in H_i \cap P_r\right] \geq \mathbb{P}\left[H_i \cap P_r \neq \emptyset\right] \mathbb{P}\left[H_i \cap (P \setminus P_r) = \emptyset\right].$$

The two quantities on the right-hand side are independent.

Claim 2.4.12. *When* $\frac{2}{n^2} \leq \sigma < \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)}$, $\mathbb{P}\left[P_r \cap H_i \neq \emptyset\right] = \Omega(1)$.

Proof. Note that z_0 is set so that $Q\left(\frac{z_0}{\sigma}\right) = \Theta\left(\frac{1}{t_m}\right) = \Theta\left(\frac{\sqrt{z_0}}{n\sigma}\right)$. Indeed, using Lemma 2.4.1 (i) and the fact that $z_0 = \Omega(\sigma)$, $Q\left(\frac{z_0}{\sigma}\right) = \Theta\left(\frac{e^{-\frac{z_0^2}{2\sigma^2}}}{\frac{z_0}{\sigma}}\right)$. The choice for z_0 comes from the resolution of the equation $\frac{1}{x}e^{-\frac{x^2}{2}} = \frac{\sqrt{x}}{n\sqrt{\sigma}}$ using the definition of the function $\mathcal{W}_0$.

Moreover, for $|t| \leq t_m$ we have $z_t = \Theta(z_0)$ and $z_t^2 - z_0^2 = \Theta\left(\frac{t^2 z_0}{n^2}\right)$. Indeed, $\sigma \geq \frac{2}{n^2}$ implies that $z_0 = \Omega(\sigma)$ and $z_t < z_0 + O\left(\frac{t_m^2}{n^2}\right) = O\left(\frac{z_0^2+\sigma^2}{z_0}\right) = O(z_0)$ and $z_t = z_0 + \Theta\left(\frac{t^2}{n^2}\right) = \Omega(z_0)$. Also, $z_t = \Omega(\sigma)$ so $\frac{1}{\frac{z_t}{\sigma}+\frac{\sigma}{z_t}} = \Omega\left(\frac{\sigma}{z_t}\right)$. Injecting these relations into Equation (2.12) we obtain that the expected number of relevant points in H_i is

$$\Omega\left(\frac{1}{t_m}\sum_{t=-t_m}^{t_m}\frac{1}{\frac{z_t}{\sigma}+\frac{\sigma}{z_t}}\frac{z_0}{\sigma}e^{-\frac{1}{2\sigma^2}(z_t^2-z_0^2)}\right) = \Omega\left(\frac{1}{t_m}\sum_{t=0}^{t_m}e^{-t^2\Theta\left(\frac{z_0}{n^2\sigma^2}\right)}\right).$$

Again, Lemma 2.4.1 (iii-b) ensures that

$$\sum_{t=0}^{t_m}e^{-t^2\Theta\left(\frac{z_0}{n^2\sigma^2}\right)} = \Omega\left(\frac{n\sigma}{\sqrt{z_0}}\right)$$

and the expected number of relevant points in H_i is $\Omega(1)$. The Chernoff bound of Lemma 2.2.1 (a) yields that $\mathbb{P}\left[H_i \cap P \neq \emptyset\right]$ is $\Omega(1)$. □

It remains to bound the third quantity:

Claim 2.4.13. *When* $\frac{2}{n^2} \leq \sigma < \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)}$, $\mathbb{P}\left[H_i \cap (P \setminus P_r) = \emptyset\right] = \Omega(1)$.

Proof. Every irrelevant point p_t belongs to H_i with probability $Q\left(\frac{z_t}{\sigma}\right)$. The probability that H_i contains no irrelevant point is therefore at least

$$\left(\prod_{t=t_m+1}^{\frac{n-1}{2}} 1 - Q\left(\frac{z_t}{\sigma}\right)\right)^2$$

Lemma 2.4.1 (i) and the fact that $z_t = z_0 + \Theta\left(\frac{t^2}{n^2}\right)$ ensure that

$$1 - Q\left(\frac{z_t}{\sigma}\right) \geq 1 - Q\left(\frac{z_0}{\sigma}\right) e^{-\frac{1}{2\sigma^2}(z_t^2 - z_0^2)} = 1 - Q\left(\frac{z_0}{\sigma}\right) e^{-t^2\Theta\left(\frac{z_0}{n^2\sigma^2}\right)}$$

so the probability that H_i contains no irrelevant point is at least

$$\gamma = \left(\prod_{t=t_m+1}^{\frac{n-1}{2}} 1 - \frac{1}{t_m} e^{-t^2\Theta\left(\frac{1}{t_m^2}\right)}\right)^2.$$

Taking the logarithm, and using $\ln(1-x) \geq -2x$ for $x \in [0,1]$, we get

$$\begin{aligned} -\ln\gamma &= -2\sum_{t=t_m+1}^{\frac{n-1}{2}} \ln\left(1 - \frac{1}{t_m} e^{-t^2\Theta\left(\frac{1}{t_m^2}\right)}\right) \\ &\leq \frac{4}{t_m}\sum_{t=t_m+1}^{\frac{n-1}{2}} e^{-t^2\Theta\left(\frac{1}{t_m^2}\right)} \leq \frac{4}{t_m}\sum_{t=0}^{\frac{n-1}{2}} e^{-t^2\Theta\left(\frac{1}{t_m^2}\right)} \end{aligned}$$

and Lemma 2.4.1 (iii-a) yields $0 \leq -\ln\gamma \leq O(1)$. It follows that the probability that H_i contains no irrelevant point is at least $e^{-O(1)} = \Omega(1)$. □

Wrapping Up. We can now obtain our lower bound.

Proof of Theorem 19. Lemma 2.4.4 and the preceding analysis ensure that the assumptions of Theorem 10 (ii) are satisfied, and we thus have $\mathbb{E}[\text{card}\,\text{CH}(P)] = \Omega(\min(n, h_1/w_1))$. We treat separately the three regimes.

If $\sigma < \frac{2}{n^2}$ then

$$\mathbb{E}[\text{card}\,\text{CH}(P)] = \Omega\left(\min\left(n, \left(\frac{1}{O\left(\frac{1}{n^2}\right)}\right)\right)\right) = \Omega(n)$$

which is the first regime announced in Theorem 19. (Note that the boundaries between the regimes can be set up to a multiplicative constant.)

If $\frac{2}{n^2} \leq \sigma < \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)}$ then

$$\mathbb{E}[\text{card}\,\text{CH}(P)] = \Omega\left(\frac{1 + \sigma\sqrt{\ln(n\sqrt{\sigma})}}{\sqrt{\sigma}\left(\sqrt{\sigma} + \frac{1}{\sqrt[4]{\ln(n\sqrt{\sigma})}}\right)}\right)$$

We simplify this expression by comparing σ and $\frac{1}{\sqrt{\ln\left(n\sqrt{\sigma}\right)}}$. Specifically, if $\sigma \leq \frac{1}{\sqrt{\ln n}}$ then

$$\sigma = O\left(\frac{1}{\sqrt{\ln\left(n\sqrt{\sigma}\right)}}\right) \quad \text{and} \quad \mathbb{E}\left[\operatorname{card}\operatorname{CH}(P)\right] = \Omega\left(\frac{\sqrt[4]{\ln\left(n\sqrt{\sigma}\right)}}{\sqrt{\sigma}}\right)$$

which is the second regime announced in Theorem 19.

If $\frac{1}{\sqrt{\ln n}} \leq \sigma < \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)} = O\left(\sqrt{\ln n}\right)$ then $\frac{1}{\sqrt{\ln\left(n\sqrt{\sigma}\right)}} = O(\sigma)$ and

$$\mathbb{E}\left[\operatorname{card}\operatorname{CH}(P)\right] = \Omega\left(\frac{\sigma\sqrt{\ln\left(n\sqrt{\sigma}\right)}}{\sigma}\right) = \Omega\left(\sqrt{\ln\left(n\sqrt{\sigma}\right)}\right) = \Omega\left(\sqrt{\ln n}\right)$$

If $\sigma \geq \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)} = \Omega\left(\sqrt{\ln n}\right)$ then

$$\mathbb{E}\left[\operatorname{card}\operatorname{CH}(P)\right] = \Omega\left(\frac{1+\sigma\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)}}{\sigma+1}\right) = \Omega\left(\sqrt{\ln n}\right)$$

The lower bound is the same as in the case $\frac{1}{\sqrt{\ln n}} \leq \sigma < \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4}\right)}$. Merging the two conditions we obtain that

$$\sigma \geq \frac{1}{\sqrt{\ln n}} \Rightarrow \mathbb{E}\left[\operatorname{card}\operatorname{CH}(P)\right] = \Omega\left(\sqrt{\ln n}\right)$$

which is the third regime announced in Theorem 19. □

2.4.6 Gaussian Perturbation of a Regular Polygon: An Upper Bound

In Section 2.4.3, we proved a lower bound on the smoothed complexity of convex hull with Gaussian perturbations, see Theorem 19. In this section, we prove that this lower bound is *almost* tight for this specific configuration of points (a regular polygon). By almost, we mean that the lower bound is tight up to some small intervals of σ. We believe that theses gaps are just a technical issue due to our technique.

Theorem 20. *Let $P^* = \{p_i^*, 1 \leq i \leq n\}$ be the set of vertices of a regular n-gon of radius 1 in $\mathbb{R}^2$. Let $P = \{p_i = p_i^* + \eta_i\}$ where $\eta_1, \eta_2, \ldots, \eta_n$ are random vectors in $\mathbb{R}^2$ chosen independently from $\mathcal{N}(0, \sigma^2 I_2)$. The expected number of vertices of the convex hull of P is*

Range of σ	$\left[0, \frac{2\ln^4 n}{n^2}\right]$	$\left[\frac{2\ln^4 n}{n^2}, \frac{1}{\sqrt{\ln n}}\right]$	$\left[\frac{1}{\sqrt{\ln n}}, +\infty\right)$
$\mathbb{E}\left[\operatorname{card}\mathcal{H}^{(1)}\right]$	$O(n)$	$O\left(\frac{\sqrt[4]{\ln(n\sqrt{\sigma})}}{\sqrt{\sigma}}\right)$	$O\left(\sqrt{\ln n}\right)$

Using Theorem 19 and Theorem 20 we get the following estimation:

Range of σ	$\left[0, \frac{1}{n^2}\right]$	$\left[\frac{1}{n^2}, \frac{2\ln^4 n}{n^2}\right]$	$\left[\frac{2\ln^4 n}{n^2}, \frac{1}{\sqrt{\ln n}}\right]$	$\left[\frac{1}{\sqrt{\ln n}}, +\infty\right)$
$\mathbb{E}\left[\operatorname{card}\mathcal{H}^{(1)}\right]$	$\Theta(n)$	$O(n)$, $\Omega\left(\frac{\sqrt[4]{\ln(n\sqrt{\sigma})}}{\sqrt{\sigma}}\right)$	$\Theta\left(\frac{\sqrt[4]{\ln(n\sqrt{\sigma})}}{\sqrt{\sigma}}\right)$	$\Theta\left(\sqrt{\ln n}\right)$

Note that we can suppose $\sigma > \frac{2\ln^4 n}{n^2}$ since otherwise the upper bound is trivial. The chosen parameters are similar to Section 2.4.3 (and so are some arguments in the proof), but in this case we need to define $\ell = \ln^2 n$ levels of witnesses and collectors.

Parameter Setting. We set h_j and w_j depending on σ and n as summurized below. We let $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m$ denote a family of vectors in $\mathbb{S}^1$ such that $\vec{u}_i$ is aligned with $p^*_{\lfloor \frac{2\pi i}{m} \rfloor}$, so these vectors are more or less equally spaced on $\mathbb{S}^1$. The witnesses are defined as $W_i^j = W(w_i, h_i, \vec{u}_i)$ and the collectors are defined as $C_i^j = C(w_j, h_j, \vec{u}_i, \alpha)$.

	$\frac{2\ln^4 n}{n^2} \leq \sigma < \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4j^2}\right)}$	$\frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4j^2}\right)} \leq \sigma$
w_j	$2\sigma + 2\sqrt{\sigma}\left(\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}\left(\frac{n}{j}\sqrt{\sigma}\right)^{\frac{4}{3}}\right)\right)^{-1/4}$	$2\sigma + 2$
h_j	$1 + \sigma\sqrt{\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}\left(\frac{n}{j}\sqrt{\sigma}\right)^{\frac{4}{3}}\right)}$	$1 + \sigma\sqrt{\mathcal{W}_0\left(\frac{n^2}{4j^2}\right)}$
α	$\Theta\left(\frac{\sigma}{h_0 - 1 + \sqrt{h_0 - 1}}\right)$	$\Theta\left(\frac{\sigma}{h_0 - 1 + \sqrt{h_0 - 1}}\right)$

Preparation. Let $i \in \{1, \ldots, m\}$. As in Section 2.4.3, we let $(O, \vec{v}_i, \vec{u}_i)$ denote some orthonormal frame and let H_i^j the halfplane supporting W_i^j, with inner normal $\vec{u}_i$. We renumber the points of P^* with indices in $\{-\frac{n-1}{2}, \ldots, \frac{n-1}{2}\}$ so that p_0^* is the point in direction $\vec{u}_i$. For the sake of presentation, we assume that n is odd (the case of even n follows from trivial modification). Our goal is to show that $\mathbb{E}\left[\operatorname{card}\left(W_i^j \cap P\right)\right] = \Theta(j)$ and that $\mathbb{E}\left[\operatorname{card}\left(C_i^j \cap P\right)\right] = O(j)$.

Let $j \in \{1, \ldots, \ell\}$. We write z_t for the distance from p_t^* to H_i^j and note that

$$z_0 = h_j - 1 \qquad \text{and} \qquad z_t = h_j - 1 + 1 - \cos\frac{2\pi t}{n}$$

Same argument as Equation 2.9 gives $z_t - z_0 = \Theta\left(\frac{t^2}{n^2}\right)$.

Relevants Points. The contribution of the t-th point to $\mathbb{E}\left[\mathrm{card}\left(H_i^j \cap P\right)\right]$ is $Q\left(\frac{z_t}{\sigma}\right)$. The gist of our analysis for larger σ is to split the points into two parts, the relevant points where $Q\left(\frac{z_t}{\sigma}\right) = \Theta\left(Q\left(\frac{z_0}{\sigma}\right)\right)$ and the irrelevant ones. The expected number of points in H_i^j is (up to a constant multiplicative factor) at least the number of relevant points times $Q\left(\frac{z_0}{\sigma}\right)$; fine tuning z_0 so that this product is $\Omega(1)$ then amounts to solving some functional equation. Also, we need to show that the contribution of the irrelevant points are negligeable. Specifically, we call a point p_t *relevant* if $|t| \leq t_m = \min\left(\left\lfloor \frac{n\sigma}{\sqrt{z_0}} \right\rfloor, \frac{n-1}{2}\right)$. A relevant point in H_i^j is in W_i^j with constant probability, since the width of the witness is chosen to enclose $B(p^*, \sigma)$ in $\vec{v}_i$ direction, when p is relevant. Thus, $\mathbb{E}\left[\mathrm{card}\left(W_i^j \cap P\right)\right] = \Omega\left(\mathbb{E}\left[\mathrm{card}\left(W_i^j \cap P\right)\right]\right)$. Since $W_i^j \subset H_i^j$, we get $\mathbb{E}\left[\mathrm{card}\left(W_i^j \cap P\right)\right] = \Theta\left(\mathbb{E}\left[\mathrm{card}\left(H_i^j \cap P\right)\right]\right)$.

Counting points in H_i^j. A first step to compute the expected number of points in W_i^j is to compute the expected number of points in H_i^j. The expected number of points in H_i^j writes

$$\mathbb{E}\left[\mathrm{card}\left(H_i^j \cap P\right)\right] = \sum_{t=-\frac{n-1}{2}}^{\frac{n-1}{2}} Q\left(\frac{z_t}{\sigma}\right) = \Theta\left(\sum_{t=0}^{\frac{n-1}{2}} \frac{1}{\frac{z_t}{\sigma} + \frac{\sigma}{z_t}} \frac{z_0}{\sigma} e^{-\frac{1}{2\sigma^2}(z_t^2 - z_0^2)}\right) \tag{2.13}$$

Intermediate σ When $\frac{2\ln^4 n}{n^2} \leq \sigma < \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4j^2}\right)}$ we have

$$z_0 = \sigma\sqrt{\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}\left(\frac{n}{j}\sqrt{\sigma}\right)^{\frac{4}{3}}\right)}.$$

The function $x \mapsto \frac{x}{\sqrt{\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}\left(\frac{n}{j}\sqrt{x}\right)^{\frac{4}{3}}\right)}}$ is increasing. Let us define σ_0 as the solution of

$$\sigma_0 = \frac{1}{4}\sqrt{\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}\left(\frac{n}{j}\sqrt{\sigma_0}\right)^{\frac{4}{3}}\right)}$$

Using that $\mathcal{W}_0$ is the solution to $f(x)e^{f(x)} = x$, we obtain that

$$\sigma_0 = \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4j^2}\right)}.$$

Then, for $\sigma < \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4j^2}\right)}$ we have

$$\frac{n\sigma}{\sqrt{z_0}} \leq n\sqrt{\frac{\sigma}{\sqrt{\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}\left(\frac{n}{j}\sqrt{\sigma}\right)^{\frac{4}{3}}\right)}}} < n\sqrt{\frac{\sigma_0}{\sqrt{\frac{3}{2}\mathcal{W}_0\left(\frac{2}{3}\left(\frac{n}{j}\sqrt{\sigma_0}\right)^{\frac{4}{3}}\right)}}} = \frac{n}{2}$$

and $t_m = \left\lfloor \frac{n\sigma}{\sqrt{z_0}} \right\rfloor$.

Claim 2.4.14. *When* $\frac{2\ln^4 n}{n^2} < \sigma < \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4j^2}\right)}$, $\mathbb{E}\left[\text{card}\left(H_i^j \cap P\right)\right] = \Theta(j)$.

Proof. Note that z_0 is set so that $Q\left(\frac{z_0}{\sigma}\right) = \Theta\left(\frac{j}{t_m}\right) = \Theta\left(\frac{j\sqrt{z_0}}{n\sigma}\right)$. Indeed, using Lemma 2.4.1 (i) and the fact that $z_0 = \Omega(\sigma)$, $Q\left(\frac{z_0}{\sigma}\right) = \Theta\left(\frac{e^{-\frac{z_0^2}{2\sigma^2}}}{\frac{z_0}{\sigma}}\right)$. The choice for z_0 comes from the resolution of the equation $\frac{1}{x}e^{-\frac{x^2}{2}} = \frac{\sqrt{x}}{n\sqrt{\sigma}}$ using the definition of the function $\mathcal{W}_0$.

Moreover, for $|t| \leq t_m$ we have $z_t = \Theta(z_0)$ and $z_t^2 - z_0^2 = \Theta\left(\frac{t^2 z_0}{n^2}\right)$. Indeed, $\sigma \geq \frac{2\ln^4 n}{n^2} \geq \frac{2j^2}{n^2}$ implies that $z_0 = \Omega(\sigma)$ and $z_t < z_0 + O\left(\frac{t_m^2}{n^2}\right) = O\left(\frac{z_0^2+\sigma^2}{z_0}\right) = O(z_0)$ and $z_t = z_0 + \Theta\left(\frac{t^2}{n^2}\right) = \Omega(z_0)$. Also, $z_t = \Omega(\sigma)$ so $\frac{1}{\frac{z_t}{\sigma}+\frac{\sigma}{z_t}} = \Theta\left(\frac{\sigma}{z_t}\right)$. Injecting these relations into Equation (2.13) we obtain that the expected number of relevant points in H_i^j is

$$\Omega\left(\frac{j}{t_m}\sum_{t=0}^{t_m}\frac{1}{\frac{z_t}{\sigma}+\frac{\sigma}{z_t}}\frac{z_0}{\sigma}e^{-\frac{1}{2\sigma^2}(z_t^2-z_0^2)}\right) = \Omega\left(\frac{j}{t_m}\sum_{t=0}^{t_m}e^{-t^2\Theta\left(\frac{z_0}{n^2\sigma^2}\right)}\right).$$

Again, Lemma 2.4.1 (iii-b) ensures that

$$\sum_{t=0}^{t_m}e^{-t^2\Theta\left(\frac{z_0}{n^2\sigma^2}\right)} = \Omega\left(\frac{n\sigma}{\sqrt{z_0}}\right)$$

and the expected number of relevant points in H_i^j is $\Omega(j)$. Same arguments yields the upper bound

$$\mathbb{E}\left[\text{card}\left(H_i^j \cap P\right)\right] = O\left(\frac{j}{t_m}\sum_{t=0}^{n}e^{-t^2\Theta\left(\frac{z_0}{n^2\sigma^2}\right)}\right) = O\left(\frac{j}{t_m}\left(1+\frac{n\sigma}{\sqrt{z_0}}\right)\right) = O(j)$$

using Lemma 2.4.1 (iii-a), since $z_0 = \Omega(\sigma^2)$. ☐

Large σ When $\sigma \geq \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4j^2}\right)}$, every point is relevant, *ie.* $t_m = \frac{n-1}{2}$, since

$$z_0 = \sigma\sqrt{\mathcal{W}_0\left(\frac{n^2}{4j^2}\right)} \quad \text{implies} \quad \frac{n\sigma}{\sqrt{z_0}} = n\sqrt{\frac{\sigma}{\sqrt{\mathcal{W}_0\left(\frac{n^2}{4j^2}\right)}}} \geq \frac{n}{2}.$$

Claim 2.4.15. *When $\sigma \geq \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4j^2}\right)}$, $\mathbb{E}\left[\text{card}\left(H_i^j \cap P\right)\right] = \Theta(j)$.*

Proof. Computations similar to that of Equation (2.6) yield that $Q\left(\frac{z_0}{\sigma}\right) = \Theta\left(\frac{j}{n}\right)$. Moreover, $z_t \geq \frac{\sigma}{2}$ so $\frac{1}{\frac{z_j}{\sigma}+\frac{\sigma}{z_j}} = \Theta\left(\frac{\sigma}{z_j}\right)$. Also, $z_t = \Theta(z_0)$ and $z_t^2 - z_0^2 = \Theta\left(\frac{t^2 z_0}{n^2}\right)$. Injecting these three relations in Equation (2.12) we obtain that the expected number of (relevant) points in H_i writes

$$\mathbb{E}\left[\text{card}\left(H_i^j \cap P\right)\right] = \Theta\left(\frac{j}{n}\sum_{t=0}^{\frac{n-1}{2}} \frac{z_0}{z_t} e^{-t^2\Theta\left(\frac{z_0}{n^2\sigma^2}\right)}\right) = \Theta\left(\frac{j}{n}\sum_{t=0}^{\frac{n-1}{2}} e^{-t^2\Theta\left(\frac{z_0}{n^2\sigma^2}\right)}\right)$$

Since $\frac{z_0}{n^2\sigma^2} < \frac{4}{n^2}$, Lemma 2.4.1 (iii-c) implies that

$$\sum_{j=0}^{\frac{n-1}{2}} e^{-j^2\Theta\left(\frac{z_0}{n^2\sigma^2}\right)} = \Omega(n),$$

so we finally get that H_i contains $\Omega(j)$ (relevant) points on average. The upper bound comes trivially since $e^{-j^2\Theta\left(\frac{z_0}{n^2\sigma^2}\right)} = O(1)$. □

We then conclude that for $\sigma > \frac{2\ln^4 n}{n^2}$, $\mathbb{E}\left[\text{card}\left(W_i^j \cap P\right)\right] = \Theta(j)$. It remains to control the expected number of points in the collector.

Couting points in C_i^j.

Claim 2.4.16. *Let $\sigma > \frac{2\ln^4 n}{n^2}$. Then, $\mathbb{E}\left[\text{card}\left(C_i^j \cap P\right)\right] = O(j)$.*

Proof. The collector C_i^j is the union of two halfplanes of height $\tilde{h}_j$. The contribution of a halfplane depends only on its height and to compute the contribution to the collector, we compute the contribution to a halfplane $\tilde{H}_i^j$ parallel to H_i^j of height $\tilde{h}_j$. This contribution will be computed separately for relevant and irrelevant points.

As above z_t denote the distance of p_t^* to H_i^j and we denote $\tilde{z}_t$ its distance to $\tilde{H}$. Let's first consider the contribution of the relevant points. We suppose without loss of generality that $\tilde{H}_i^j$ corresponds to the half-plane on the left of $\vec{u}_i$, and we consider only the relevant points on the left side $\{p_0, \ldots, p_{t_m}\}$. The contribution of the relevants points on the right side will be trivially bounded by the contribution of the left-sided relevant points.

Using Claim 2.4.3, we get that for $0 \le t \le t_m$,

$$\tilde{z}_t = z_t \cos\frac{\alpha}{2} - \sin\left(\frac{\alpha}{2}\right)\frac{w_j + \Theta(\frac{t}{n})}{2}.$$

In the large σ settings, $w_j > 1$ so $\frac{t_m}{nw_j} < 1$. For the intermediate σ settings, $\frac{t_m}{n} = \frac{\sigma}{\sqrt{z_0}}$ and $\frac{t_m}{nw_j} = \frac{\sigma}{2\sigma\sqrt{z_0}\left(1+\frac{1}{\sqrt{z_0}}\right)} \le \frac{1}{1+\sqrt{z_0}} < 1$, since $w_j = 2\sigma\left(1 + \frac{1}{\sqrt{z_0}}\right)$.
Thus,

$$\begin{aligned}
z_t &\le \tilde{z}_t + z_t\left(1 - \cos\frac{\alpha}{2}\right) + \frac{\alpha(w_j + C_1\frac{t_m}{n})}{4} \\
&\le \tilde{z}_t + \frac{\alpha w_j}{4}\left(1 + \frac{(C_1\frac{t_m}{n})}{w_j} + \frac{z_t\alpha}{2w_j}\right) \\
&\le \tilde{z}_t + \frac{\alpha w_j}{4}(C_1 + 2)
\end{aligned}$$

if we choose $\alpha \le \frac{\sigma}{C(h_0-1))(1+\frac{1}{\sqrt{h_0-1}})}$ with $C > \frac{z_{t_m}}{z_0}$ (remember that $z_{t_m} = \Theta(z_0)$) and $C_1 > 0$ some constants.

In particular, since $w_j < 4\sigma\left(1 + \frac{1}{\sqrt{z_0}}\right)$ we have

$$\frac{z_t}{\sigma} \le \frac{\tilde{z}_t}{\sigma} + \alpha(C_1 + 2)\left(1 + \frac{1}{\sqrt{z_0}}\right).$$

We choose $\alpha = \frac{\sigma}{C_2(h_0-1)(1+\frac{1}{\sqrt{h_0-1}})}$ with $C_2 = C(C_1 + 2)$.
Since $\tilde{z}_t \le \tilde{z}_{t_m} < z_{t_m} < Cz_0 \le C(h_0 - 1)$,

$$\begin{aligned}
Q\left(\frac{z_t}{\sigma}\right) &\ge Q\left(\frac{\tilde{z}_t}{\sigma} + (C_1+2)\alpha\left(1 + \frac{1}{\sqrt{z_0}}\right)\right) \\
&\ge Q\left(\frac{\tilde{z}_t}{\sigma} + \frac{\sigma}{C(h_0-1)(1+\frac{1}{\sqrt{h_0-1}})}\left(1 + \frac{1}{\sqrt{z_0}}\right)\right) \\
&\ge Q\left(\frac{\tilde{z}_t}{\sigma} + \frac{\sigma}{Cz_0\left(1+\frac{1}{\sqrt{z_0}}\right)}\left(1 + \frac{1}{\sqrt{z_0}}\right)\right) \\
&= Q\left(\frac{\tilde{z}_t}{\sigma} + \frac{\sigma}{Cz_0}\right)
\end{aligned}$$

and since $\tilde{z}_t \ge \frac{1}{4}\sigma$ we conclude by

$$Q\left(\frac{\tilde{z}_t}{\sigma}+\frac{\sigma}{Cz_0}\right) \geq Q\left(\frac{\tilde{z}_t}{\sigma}+\frac{\sigma}{\tilde{z}_t}\right) = \Omega\left(Q\left(\frac{\tilde{z}_t}{\sigma}\right)\right)$$

using Lemma 2.4.1 (ii).
Thus,

$$Q\left(\frac{\tilde{z}_t}{\sigma}\right) = O\left(Q\left(\frac{z_t}{\sigma}\right)\right)$$

and the contribution of relevant points to the collector is bounded, up to a constant, by the contribution of relevant points to the witness.

Taking the above formula with $t = t_m$ gives that $Q\left(\frac{\tilde{z}_{t_m}}{\sigma}\right) = O\left(Q\left(\frac{z_{t_m}}{\sigma}\right)\right)$. Since we chose t_m so that $Q\left(\frac{z_{t_m}}{\sigma}\right) = \Theta\left(Q\left(\frac{z_0}{\sigma}\right)\right)$, we obtain $Q\left(\frac{\tilde{z}_{t_m}}{\sigma}\right) = O\left(Q\left(\frac{z_0}{\sigma}\right)\right)$. Basic circle geometry gives that $\tilde{z}_{i_m+j} - \tilde{z}_{i_m} > z_j - z_0$ and thus the contribution of irrelevant points to $\tilde{H}_i^j$ is bounded by the contribution of all points to H_i^j which is $O(j)$. □

Wrapping up. With the chosen parameter, Theorem 10 (i) applies and $\mathbb{E}\left[\mathcal{H}^{(1)}\right] = \mathbb{E}\left[\text{card}\,\text{CH}(P)\right] = O\left(\frac{1}{\alpha}\right)$. Notice that for $\frac{1}{\sqrt{\ln n}} < \sigma < \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4j^2}\right)}$, $\frac{1}{\alpha} = O\left(\frac{h_0}{\sigma}\right) = O\left(\sqrt{\ln n}\right)$, which is the same behavior as the case where $\sigma > \frac{1}{4}\sqrt{\mathcal{W}_0\left(\frac{n^2}{4j^2}\right)}$. Thus, for $\sigma > \frac{1}{\sqrt{\ln n}}$, we get $\frac{1}{\alpha} = O\left(\frac{1}{\sqrt{\ln n}}\right)$. For $\frac{2\ln^4 n}{n^2} < \sigma < \frac{1}{\sqrt{\ln n}}$, $\frac{1}{\alpha} = O\left(\frac{\sqrt{h_0}}{\sigma}\right) = O\left(\frac{\sqrt[4]{\ln\left(n\sqrt{\sigma}\right)}}{\sigma}\right)$. This finishes the proof of Theorem 20.

2.5 Concluding Remarks

2.5.1 Poisson Distribution

Theorem 10 is established for a set of n independent elements. Except for some technicalities in the presentation, nothing prevents making n a random variable to prove *eg.* analogs of Theorems 11 and 16 for Poisson distributions. (As this was not required for our application to smoothed complexity analysis, we opted for a simpler presentation where n is fixed.)

2.5.2 Silhouette of Polytopes

Glisse, Lazard, Michel and Pouget [33] used the witness and collector approach to study the expected size of the silhouette of a 3D random convex polytope defined as the convex hull of a Poisson point process of intensity n on the unit sphere. The silhouette of the polytope from a given viewpoint is the two dimensional convex hull of the projection of the points, thus the problem reduces to the size of the convex hull of i.i.d. points in a disk for the distribution corresponding to the projection of a Poisson point process. Glisse *et al.* analyzed the size of that convex hull using a system of witnesses and collectors adapted to that distribution and proved that the worst point of view yields a silhouette of expected size $\Theta\left(\sqrt{n}\right)$.

2.5.3 ℓ^∞ Perturbation and Snap-Rounding

Systems of witnesses and collectors can be designed for perturbations that are uniform in the ball for other metrics. In [6], denoting $\square$ the unit square in 2D, we prove the following theorem:

Theorem 21. *Let $P^* = \{p_i^* : 1 \leq i \leq n\}$ be an $(\Theta\left(n\right), \Theta(1))$-sample of the unit circle in $\mathbb{R}^2$ and let $P = \{p_i = p_i^* + \eta_i\}$ where $\eta_1, \eta_2, \dots, \eta_n$ are random variables chosen independently from $\mathcal{U}_{\delta\square}$. For any fixed k, and $\delta \in [n^{-2}, 1]$*

$$\mathbb{E}\left[\operatorname{card}\mathcal{H}^{(k)}\right] = \Theta\left(n^{\frac{1}{5}}\delta^{-\frac{2}{5}}\right)$$

As in the Euclidean case, the witnesses and collectors are parallel half planes, but the partition of ranges must be adapted to cope with the lack of rotational symmetry. The angle α_i of the set of directions covered by R_i is no longer constant and is much smaller when the ranges are almost horizontal or vertical than when they are oblique. The bound of Theorem 21 is confirmed experimentally (*cf.* the slopes of $-\frac{2}{5}$ in the plots of Figure 2.3a). Theorem 21 implies that for $\delta \in [n^{-2}, 1]$, $\mathcal{S}(n, \mathcal{U}_{\delta\square})$ is $\Omega\left(n^{\frac{1}{5}}\delta^{-\frac{2}{5}}\right)$. It is also known to be $O\left(\left(\frac{n \ln n}{\delta}\right)^{\frac{2}{3}}\right)$, for all ranges of δ, by the upper bound obtained by Damerow and Sohler for maximal points under ℓ^∞ noise [17].

Snap Rounding. Given a grid whose pixel have size δ, rounding points with real coordinates at the center of their pixel have some similarity with ℓ^∞ noise. Actually, for a single point, and if the origin of the grid is random, the two processus are identical, but when several points are involved things are different: clearly rounding creates collisions while noising separates identical points. However for the regular n-gon, provided that $\delta < \frac{1}{n}$ the two processes give convex hulls of similar size as confirmed by Figure 2.3b.

2.5.4 Delaunay Triangulation

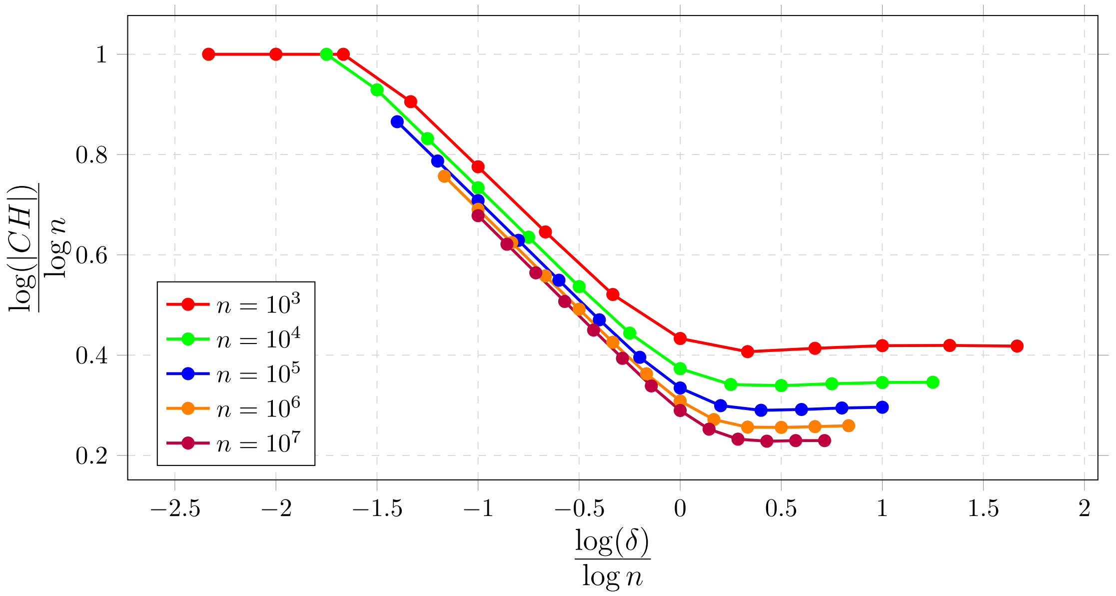
Systems of witnesses and collectors can also be used to prove the following well known result of Dwyer [30]:

Theorem 22 (Dwyer [30])**.** *The expected complexity of the Delaunay triangulation of n random points uniformly distributed in the unit ball $\mathbb{B}$ of dimension d is $\Theta(n)$.*

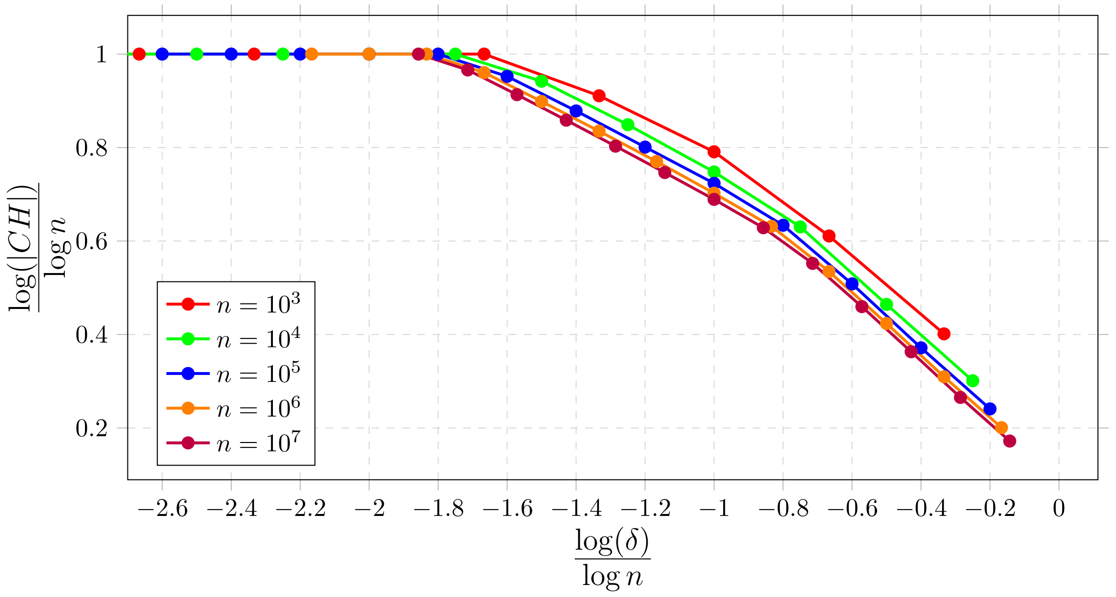
In [23], Devillers *et al.* gave a proof, considerably simpler than Dwyer's, of this result up to logarithmic factors; these factors can be removed thanks to Theorem 10 using a system of witnesses and collectors that we now outline.

The faces of dimension k of the Delaunay triangulation are hyperedges of size $k+1$ in the hypergraph where the ranges are balls in $\mathbb{R}^d$. More precisely, given a set P of n points in general position, $k+1$ points define a face of the Delaunay triangulation $DT(P)$ iff there exists a ball with the $k+1$ points on its boundary and no other points inside. Thus the hypergraph define using the balls as ranges may be a strict superset of the Delaunay faces. The proof splits the ranges in three subsets and builds a system of witnesses and collectors for each these subsets.

Balls Centered Deep Inside $\mathbb{B}$. Let $r_j = O\left(\left(\frac{j}{n}\right)^{\frac{1}{d}}\right)$ denote the radius of a ball completely contained in $\mathbb{B}$ and expected to contain j points. We use a minimal covering of $\mathbb{B}$ with balls of radius r_1 and keep the balls centered inside $(1-r_1 \ln^2 n)\mathbb{B}$ to define our first level witnesses W_i^1. We define W_i^j as the ball concentric with W_i^1 with radius r_j, and C_i^j as the ball concentric with W_i^j with radius $r_j + 2r_1$. We finally let R_i be the set of balls centered in W_i^1. This system of witnesses and collectors verifies the hypotheses of Theorem 10 (i), and a constant fraction of the first layer $\{(W_i^1, C_i^1)\}_i$ verifies the hypotheses of Theorem 10 (ii). Altogether, they allow to conclude that the number of Delaunay balls centered in $(1 - r_1 \ln^2 n)\mathbb{B}$ is $\Theta(n)$.



(a) Experimental results for the complexity of the convex hull of a ℓ^∞ perturbation of amplitude δ of the regular n-gon inscribed in the unit circle. Each data point corresponds to an average over 1000 experiments.



(b) Experimental results for the complexity of the convex hull of a rounding of the regular n-gon inscribed in the unit circle on a grid of pixel size δ.

Figure 2.3: Experimental results for the ℓ^∞ perturbation and rounding.

Balls Centered Near $\partial\mathbb{B}$. The Delaunay balls centered in an annulus of width $2r_1 \ln^2 n$ around $\partial\mathbb{B}$ can be counted easily since their number is sublinear. To this aim we can cover the above annulus by collectors of diameter $O(r_1 \ln^2 n)$ and use associated empty witnesses.

Balls Centered outside $\mathbb{B}$. Balls centered outside $\mathbb{B}$ are a bit more delicate, since they can have a large radius but, possibly, a small probability to be empty. A first remark is that ball of infinite radius are half-plane and are counted by Theorem 11. Actually, the construction of Theorem 11 can be adapted to count all balls of radii between α and 2α by using balls of radius α to define the witnesses and balls of radii 2α for the collectors. Then it is possible to sum on various values of α to cover all the possible radii. As a side result we get the expected size of the α-shape of points uniformly distributed in $\mathbb{B}$.

Chapter 3

A Chaotic Random Convex Hull

Outline

3.1 Introduction

Consider a sequence of points in a convex body in dimension d whose convex hull is dynamically maintained when the points are inserted one by one, the convex hull size may increase, decrease, or remain constant when a new point is added. Studying the expected size of the convex hull when the points are evenly distributed in the convex is a classical problem of probabilistic geometry that yields some surprising facts. For example, although it seems quite natural to think that the expected size of the convex hull is increasing with n the number of points, this fact is only formally proved for n large enough [24]. The asymptotic behavior of the expected size is known to be polylogarithmic for a polytopal body and polynomial for a smooth one. If for a polytope or a smooth body, the asymptotic behavior is *somehow* "nice", for "most" convex bodies the behavior is unpredictable [7, corollary 3]. It is possible to construct strange convex objects that have no such nice behaviors and this note exhibits a convex body, such that the behavior of the expected size of a random polytope oscillates between the polytopal and smooth behaviors when n increases.

More formally, let K be a convex body in $\mathbb{R}^d$ and $(x_1, \ldots, x_n)$ a sample of n points chosen uniformly and independently at random in K. Let K_n be the convex hull of these points and $f_0(K_n)$ the number of vertices of K_n.

It is well known [8, 43] that if P is a polytope, then

$$\mathbb{E}\left[f_0(P_n)\right] = c_{d,P} \ln^{d-1} n + o(\log^{d-1} n) \tag{3.1}$$

and if K is a smooth convex body (i.e with $\mathcal{C}^2$ boundary with a positive Gaussian curvature), then

$$\mathbb{E}\left[f_0(K_n)\right] = c_{d,K} n^{\frac{d-1}{d+1}} + o(n^{\frac{d-1}{d+1}}) \tag{3.2}$$

where $c_{d,P}$ and $c_{d,K}$ are constants depending only on d and on the convex body. These are the two extreme behaviors : every random polytope of a convex body in $\mathbb{R}^d$ has a behavior between (3.1) and (3.2) for n large enough [7, corollary 3]. For general convex bodies, we cannot expect such a beautiful formula.

Theorem 23 (Bàràny-Larman [7])**.** *For any function $G(n) \to_{n\to\infty} \infty$ and for most (in the Baire category sense) convex bodies K in $\mathbb{R}^d$,*

$$G(n) \ln^{d-1} n > \mathbb{E}\left[f_0(K_n)\right]$$

for infinitely many n and

$$G(n)^{-1} n^{\frac{d-1}{d+1}} < \mathbb{E}\left[f_0(K_n)\right]$$

for infinitely many n.

Note that this *"most"* does not contain convex polytopes and smooth convex bodies, which are the most used in practice.

In this chapter, we present an explicit example of a convex body which has this chaotic behavior.

Notations. Let's introduce some notations used in this chapter:

- $K \oplus L$ will denote the Minkowski sum of K and L, defined as

$$J \oplus K = \{x + y \mid x \in J, y \in K\};$$

- $d_H(J, K)$ will denote the Hausdorff distance of the convex bodies K and L, defined as

$$d_H(J, K) = \min\{r \in \mathbb{R}^+ \mid J \subset K \oplus \mathbb{B}, K \subset J \oplus r\mathbb{B}\}$$

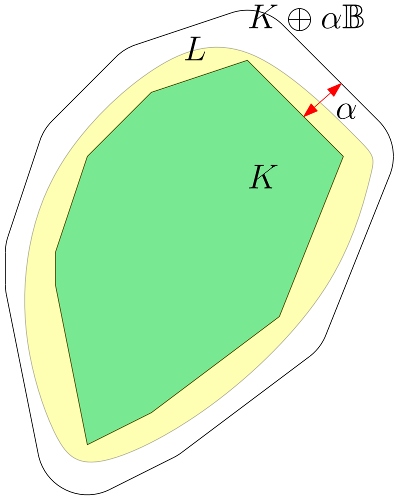
  where $\mathbb{B}$ is the Euclidean unit ball in $\mathbb{R}^d$.

3.2 Approximations of Convex Bodies

In this section, we present an intermediate lemma about random polytopes of close (in terms of Hausdorff distance) convex bodies. Let K be a convex body in $\mathbb{R}^d$ and

K_n be a random polytope in K. We want to show that if L is an approximation of K with small Hausdorff distance, L_n is approximating the asymptotic behavior of K_n for some value of n.
Let's assume that the expected size of K_n is in $c_{d,K}g(n,d) + o(g(n,d))$, where $c_{d,K}$ is a constant and g some function.

Then, for every close-enough compact set L containing K, L_n has an expected size as close as we want from $c_{d,K}g(n,d)$ for values of n as big as we want.
The idea of this lemma is very simple: if L is very close to K, the volume in $L \setminus K$ is small, and points chosen uniformly in L are very unlikely to be in $L \setminus K$. Then, even if asymptotically the expected size of L_n is different from the size of K_n, there exists some n where the expected size of L_n is as close as we want to $c_{d,K}g(n,d)$.



Lemma 3.2.1. *Let K be a convex body in $\mathbb{R}^d$ such that*

$$\mathbb{E}\left[f_0(K_n)\right] = c_{d,K}g(n,d) + o(g(n,d)). \tag{3.3}$$

Let $\varepsilon > 0$ and $N \in \mathbb{N}$.
Then, there exist $p > N$ and $\alpha > 0$ such that for any compact set L containing K with $d_H(K,L) < \alpha$,

$$\frac{\mathbb{E}\left[f_0(L_p)\right]}{c_{d,K}g(p,d)} \in [1-\varepsilon, 1+\varepsilon].$$

Proof. First, for all $n \in \mathbb{N}^*$

$$\mathbb{E}\left[f_0(L_n)\right] = \mathbb{P}\left[L_n \subset K\right]\mathbb{E}\left[f_0(L_n)|L_n \subset K\right] + \mathbb{P}\left[L_n \not\subset K\right]\mathbb{E}\left[f_0(L_N)|L_n \not\subset K\right].$$

As the points are uniformly distributed, $\mathbb{E}\left[f_0(L_n)|L_n \subset K\right] = \mathbb{E}\left[f_0(K_n)\right]$.
Using (3.3), let's choose p such that

$$\frac{\mathbb{E}\left[f_0(K_p)\right]}{c_{d,K}g(p,d)} \in \left[1-\frac{\varepsilon}{2}, 1+\frac{\varepsilon}{2}\right], \tag{3.4}$$

then

$$\frac{\mathbb{P}\left[L_p \subset K\right]\mathbb{E}\left[f_0(L_p)|\ L_p \subset K\right]}{c_{d,K}g(p,d)} \leq 1 + \frac{\varepsilon}{2}.$$

As

$$\mathbb{P}\left[L_p \not\subset K\right] = 1 - \mathbb{P}\left[L_p \subset K\right] = 1 - \left(\frac{\mathrm{vol}\,(K)}{\mathrm{vol}\,(L)}\right)^p$$

we get

$$\mathbb{P}\left[L_p \not\subset K\right] \mathbb{E}\left[f_0(L_p)|\; L_p \not\subset K\right] \leq \left(1 - \left(\frac{\mathrm{vol}\,(K)}{\mathrm{vol}\,(L)}\right)^p\right) p.$$

Now, as $1 \geq \frac{\mathrm{vol}(K)}{\mathrm{vol}(L)} \geq \frac{\mathrm{vol}(K)}{\mathrm{vol}(K \oplus B_\alpha)} \to_{\alpha\to 0} 1$ we can choose α such that

$$\left(\frac{\mathrm{vol}\,(K)}{\mathrm{vol}\,(L)}\right)^p \geq \max\left(1 - c_{d,K}\frac{g(p,d)}{p}\frac{\varepsilon}{2}, 1 - \frac{\varepsilon}{2}\right). \tag{3.5}$$

Finally,

$$\mathbb{E}\left[f_0(L_p)\right] \leq c_{d,K}g(p,d)\left(1 + \frac{\varepsilon}{2}\right) + c_{d,K}g(p,d)\frac{\varepsilon}{2} = c_{d,K}g(p,d)(1+\varepsilon). \tag{3.6}$$

For the lower bound, using Inequalities (3.4) and (3.5) we get

$$\begin{aligned}
\mathbb{E}\left[f_0(L_p)\right] &\geq \mathbb{P}\left[L_p \subset K\right] \mathbb{E}\left[f_0(L_n)|\; L_n \subset K\right] \\
&\geq c_{d,K}g(p,d)\left[\left(\frac{\mathrm{vol}\,(K)}{\mathrm{vol}\,(L)}\right)^p (1 - \frac{\varepsilon}{2})\right] \\
&\geq c_{d,K}g(p,d)\left[\left(\frac{\mathrm{vol}\,(K)}{\mathrm{vol}\,(L)}\right)^p - \frac{\varepsilon}{2}\right] \\
&\geq c_{d,K}g(p,d)\left(1 - \frac{\varepsilon}{2} - \frac{\varepsilon}{2}\right) \\
&= c_{d,K}g(p,d)(1-\varepsilon). \qquad (3.7)
\end{aligned}$$

Inequalities (3.7) and (3.6) prove the lemma. □

3.3 Construction of the Convex Body

Given an increasing function G, we want to construct a convex body in $\mathbb{R}^d$ where the size of a convex hull of random points has a chaotic behavior between $\ln^{d-1} n$ and $n^{\frac{d-1}{d+1}}$ on some values arbitrarily big. More formally,

Theorem 24. *Let $G : \mathbb{N}^* \to \mathbb{R}_+^*$ an increasing function such that $G(n) \to_{n\to\infty} \infty$. We can construct a convex body K such that:*
For all $N \in \mathbb{N}^*$, there exist $M_1, M_2 > N$, where

$$\mathbb{E}\left[f_0(K_{M_1})\right] < G(M_1)\ln^{d-1} M_1$$

and

$$\mathbb{E}\left[f_0(K_{M_2})\right] > G(M_2)^{-1}\; M_2^{\frac{d-1}{d+1}}.$$

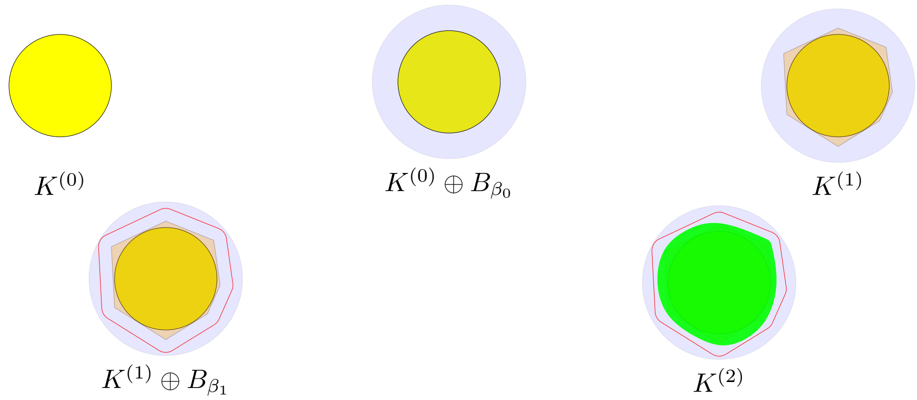


Figure 3.1: Iterations are made of polygonal and smooth approximations

Proof. The main idea of the proof is, starting from a convex body $K^{(0)}$, to iterate smooth and polytopal approximations. Lemma 3.2.1 will give us some number of points where the behavior of the random convex body will be very close to $n^{\frac{d-1}{d+1}}$ (which is the behavior for smooth convex bodies) or very close to $\ln^{d-1} n$ (which is the behavior for polytopes).

Iterations. We create an increasing sequence of convex bodies starting from the unit ball, made of polytopal and smooth approximations.

Let's define $K^{(0)}$ as the unit ball. For all $n \in \mathbb{N}^*$, $K^{(n)}$ is an approximation of $K^{(n-1)}$ where $d_H(K^{(n-1)}, K^{(n)}) < \beta_{n-1}$, with $(\beta_i)_{i\in\mathbb{N}}$ some decreasing sequence, as shown in Figure 3.1.

- If n is odd, $K^{(n-1)}$ is a smooth convex body, so $K^{(n)}$ is a convex polytope. Let's choose $q_n > n$, such that

$$\frac{\mathbb{E}\left[f_0(K_{q_n}^{(n-1)})\right]}{q_n^{\frac{d-1}{d+1}}} > \frac{2}{G(q_n)}.$$

We define

$$\varepsilon_n = 1 - \frac{2}{c_{d,K^{(n-1)}} G(q_n)}. \tag{3.8}$$

Using Lemma 3.2.1,with $\varepsilon = \varepsilon_n$, there exist α_{n-1} and $p_n > q_n$ such that for every compact set L containing $K^{(n-1)}$ with $d_H(K^{(n-1)}, L) < \alpha_{n-1}$,

$$\mathbb{E}\left[f_0(L_{p_n})\right] > c_{d,K^{(n-1)}} p_n^{\frac{d-1}{d+1}} (1 - \varepsilon_n). \tag{3.9}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[f_0(L_{p_n})\right] &> c_{d,K^{(n-1)}} p_n^{\frac{d-1}{d+1}}(1-\varepsilon_n)\\ &> p_n^{\frac{d-1}{d+1}} G(q_n)^{-1}\\ &> p_n^{\frac{d-1}{d+1}} G(p_n)^{-1}. \qquad (3.10)\end{aligned}$$

Now let's define $\beta_{n-1} = \min(\frac{\alpha_{n-1}}{2}, \frac{\beta_{n-2}}{2})$ if $n > 1$ and $\beta_0 = \frac{\alpha_0}{2}$. We define $K^{(n)}$ as a convex polytope with $d_H(K^{(n-1)}, K^{(n)}) < \beta_{n-1}$, so (3.10) works for $L = K^{(n)}$.

- If n is even, $K^{(n-1)}$ is a convex polytope, so $K^{(n)}$ is a smooth approximation.

  Let's choose q_n such that

$$\frac{\mathbb{E}\left[f_0(K^{(n-1)}_{q_n})\right]}{\ln^{d-1} q_n} < \frac{G(q_n)}{2}$$

  and define

$$\varepsilon_n = \frac{G(q_n)}{2c_{d,K^{(n-1)}}} - 1.$$

  Using Lemma 3.2.1 with $\varepsilon = \varepsilon_n$, there exist α_{n-1} and $p_n > q_n$ such that for every compact set with $d_H(K^{(n-1)}, L) < \alpha_{n-1}$,

$$\mathbb{E}\left[f_0(L_{p_n})\right] < c_{d,K^{(n-1)}} \ln^{d-1}(p_n)(1+\varepsilon_n).$$

  Finally,

$$\begin{aligned}\mathbb{E}\left[f_0(L_{p_n})\right] &< G(q_n)\ln^{d-1}(p_n)\\ &< G(p_n)\ln^{d-1}(p_n). \qquad (3.11)\end{aligned}$$

  Again, we define $\beta_{n-1} = \min(\frac{\alpha_{n-1}}{2}, \frac{\beta_{n-2}}{2})$. We define $K^{(n)}$ as a smooth approximation of $K^{(n-1)}$ such that $d_H(K^{(n-1)}, K^{(n)}) < \beta_{n-1}$, so (3.11) works for $L = K^{(n)}$.

Note that for all $m > n \in \mathbb{N}$,

$$\begin{aligned}d_H(K^{(n)}, K^{(m)}) \leq \sum_{k=n}^{m-1} d_H(K^{(k)}, K^{(k+1)}) &< \sum_{k=n}^{m-1} \beta_k\\ &\leq \sum_{k=0}^{m-n-1} \frac{\beta_n}{2^k} \leq 2\beta_n \leq \alpha_n.\end{aligned}$$

That means for all $m > n$, the property (3.10) or (3.11) (depending on the evenness of n) are also true for $K^{(m)}$.

Now, defining $K = \overline{\cup_{i=0}^{\infty} K^{(i)}}$, the property (3.10) and (3.11) are true for arbitrary $n \in \mathbb{N}$ with $L = K$, by considering $K^{(n)}$ and $K^{(n+1)}$.
As we can choose q_n as big as we want for any n (it will just decrease α_{n-1}), we can choose this sequence to be increasing. As a result, $\mathbb{E}\left[f_0(K_n)\right]$ will have a chaotic behavior within $n^{\frac{d-1}{d+1}}/G(n)$ and $G(n)\ln^{d-1} n$, as shown in Figure 3.2.

□

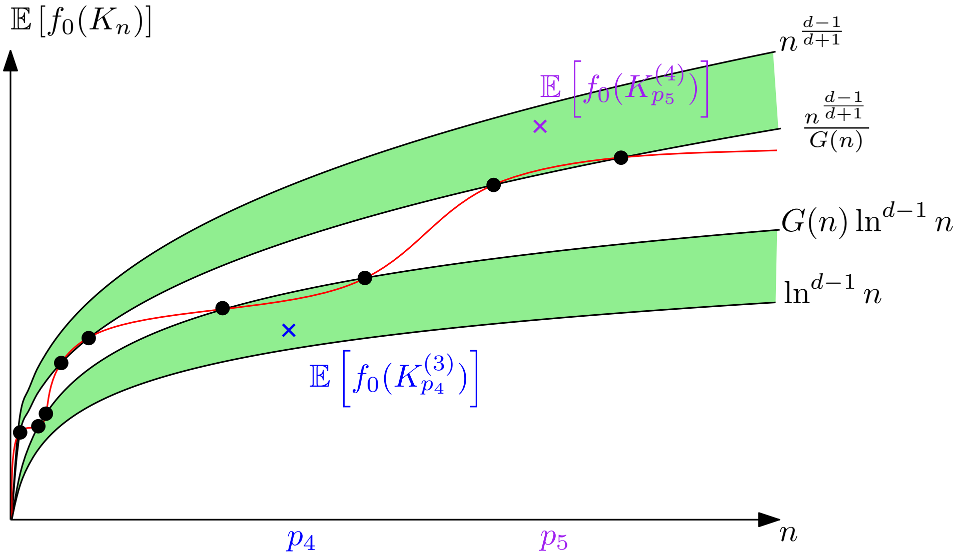


Figure 3.2: The expected size of the random polytope of K.

3.4 Concluding remarks

We have constructed a convex body K such that the expected size of the convex hull of a random polytope in K has a chaotic behavior. This construction is the limit of a sequence of bodies $(K^{(i)})$ that alternate polytopes and smooth shapes so it is difficult to provide an explicit description of K, in this note we just show that constructing such a sequence is possible by a repeated application of Lemma 3.2.1

but there is no obstacle, except long and painful computations, to a more constructive version with explicit description of the sequence. Notice that in such a case the complexity of $K^{(i)}$ will be increasing quite rapidly. Actually, since $K^{(i)}$ is constrained in a slab of width β_i around $K^{(i-1)}$, the size of $K^{(i)}$ can be lower bounded for polytopes, see [12]: $|K^{(i)}| = \Omega\left(\beta_i^{-\frac{d-1}{2}}\right)$ and since $\beta_i < \frac{\alpha_0}{2^i}$ we get, at least, an exponential behavior for the size of $K^{(i)}$. Even with a constructive description of the $K^{(i)}$, the description of K as the limit of the $K^{(i)}$ will remain quite abstract, but will allow to develop a membership test, given a point p, $p \in \mathrm{int}(K)$ can be decided by computing the sequence $K^{(i)}$ up to an index where $p \in K^{(i)}$ or $p \notin K^{(i)} \oplus \beta_i \mathbb{B}$.

Chapter 4

A Generator of Random Convex Polygons in a Disk

Outline

4.1 Introduction

Let $\mathcal{D}$ be a disk in $\mathbb{R}^2$ with radius R centered at $\mathfrak{o}$, and $X_1, \ldots, X_n$ be a sample of n points uniformly and independently distributed in $\mathcal{D}$. Let's define the polygon K_n as the convex hull of $X_1, \ldots, X_n$, and $f_0(K_n)$ its number of vertices. This kind of polygon has been well studied, and using [13, 43] one can easily check that

$$\mathbb{E}\left[f_0(K_n)\right] = c\, n^{\frac{1}{3}} + o(n^{\frac{1}{3}}) \tag{4.1}$$

where $c = 2^{\frac{4}{3}} 3^{-\frac{1}{3}} \Gamma(\frac{5}{3}) \pi^{\frac{5}{3}} \approx 3.383228964$ and Γ denotes the usual Gamma function. To generate such a polygon, one can explicitly generate n points uniformly in $\mathcal{D}$ and compute the convex hull. For a very large number of points, it could be interesting to generate fewer points to get the same polygon, for example to evaluate the constant that are not explicitly known for the asymptotic distribution of the perimeter, or other parameters such as the higher moments of the extremal points.

In this chapter, we propose an algorithm that generates far fewer points at random in order to get K_n, so that the time and the memory needed is reduced for n large.

4.2 Algorithm

We start with random polygon K_i in $\mathcal{D}$ where i is very small, and we increase the number of points until we get K_n .
The idea is that given the convex hull of small number of points, the number of points generated in $\mathcal{D}$ that are deeply inside (and so will not change the convex hull) is a random variable that we can easily simulate, so that we need to generate only the small number of points that are close to the convex hull.
The outline of the algorithm is the following:

- Generate a small number of points in $\mathcal{D}$ and compute its convex hull;
- Compute the radius of the largest disk centered at $\mathfrak{o}$ inscribed in the convex hull;
- Choose a number of points to simulate at this step;
- Simulate the number of points that fall in this inscribed disk at this step;
- Generate the rest of the random points in the annulus defined by these two discs and update the convex hull.
- continue this process until the sum of the simulated and generated points is equal to n.

To simplify the notations, we assume $\mathcal{D}$ to have radius 1.

Notations.

- n is the total number of points simulated and m_i is the total number of points from step 1 to step i;
- k_i is the number of generated points at step i, and $k = \sum_i k_i$ is the total number of generated points;
- h_i is size of the convex hull at step i, and h is the size of the final convex hull.
- p_i is the probability to fall in the annulus at step i.

Initialization. First, we have to generate a random polytope K_i with a small number of points in $\mathcal{D}$ such that $\mathfrak{o}$ is inside the random polytope. This is not too much to ask, as (see [3])

$$\mathbb{P}\left[\mathfrak{o} \notin K_i\right] = 2^{-(i-1)}i \tag{4.2}$$

We initialize the random polygon by generating 100 points in the disk. As the probability that $\mathfrak{o} \notin K_{100}$ is lower than 1.6×10^{-28}, it's very unlikely that this is not enough. Otherwise, we add another sample of 100 points, until $\mathfrak{o} \in P_{100.j}$ for some j.

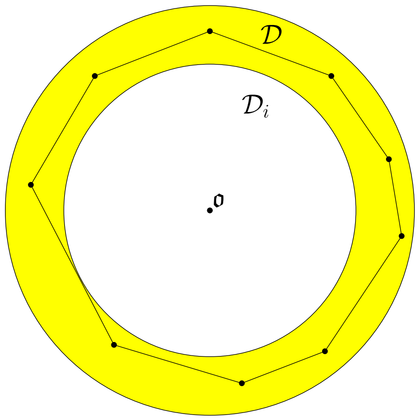


Figure 4.1: At step i, we simulate the number of points that falls in $\mathcal{D}_i$ and we generate points uniformly in the yellow annulus

Simulation of Points. At the beginning of each step i of the loop, we are given a polygon K_{m_i}, which is the convex hull of m_i points. Let s_i the number of new points simulated at step i. We choose $s_i = m_i$ if $m_i < n \ln^{-2} n$, and $s_i = n \ln^{-2} n$ otherwise. Let $\mathcal{D}_i$ be the largest inscribed disk in K_{m_i} centered in $\mathfrak{o}$, and r_i its radius.

Using a simulation of a binomial variable of parameter s_i and $1 - r_i^2$, we can evaluate the number of points that falls in $\mathcal{D} \setminus \mathcal{D}_i$. As the points in $\mathcal{D}_i$ will not change the convex hull, we do not need to generate them. Then, we generate the rest of the points uniformly in the annulus $\mathcal{D} \setminus \mathcal{D}_i$, and we update the convex hull using a Graham scan.

Generation of Random Points. To generate random points in an annulus with radii r_i and 1, one need to generate the polar angles uniformly in $[-\pi, \pi)$ and the squared radii uniformly in $[r_i^2, 1)$.

As we want to perform a Graham scan in linear time, the points have to be sorted by their polar angles. This can be done in expected linear time and size, using a bucket sort, as the angles are uniformly chosen, see [16].

Full Algorithm.

```
Data: integer n
Result: Convex hull of n uniformly chosen points in the disk D
Simulated_Points ← 0 ;
do
    Generate min(100,n − Simulated_Points) points in the disk of radius 1;
    Simulated_Points ← Simulated_Points+min(100,n − Simulated_Points);
while o is not in the convex hull and Simulated_Points < n;

while Simulated_Points< n do
    Compute inscribed_radius;
    p ← inscribed_radius^2;
    if Simulated_Points < n ln^{-2} n then
        k ← Simulated_Points;
    else
        k ← min(⌊n ln^{-2} n⌋,n − Simulated_points);
    end
    X ← Simulation of Binomial variable with parameters k, 1 − p;
    Generate X points uniformly and sorted in the annulus of radii
    inscribed_radius, 1;
    Simulated_Points ← Simulated_Points + k ;
    Merge the list of the convex hull and the new points;
    Update the convex hull with a Graham scan on the list;
end
return Convex hull
```

Algorithm 1: Algorithm of the Generator of Random Polygon in a disk

4.3 Complexity

Clearly the size complexity is $\max_i(h_i + k_i)$ and the time complexity is $\sum_i(h_i + k_i)$ since the Graham scan and the points generation are linear in the number of points [34].

For the initialization, as the probability that $\mathfrak{o} \notin K_n$ decreases exponentially, it is very unlikely that more than one loop is necessary. Let's call $\mathfrak{p}$ the minimal number of points such that $\mathfrak{o} \in K_{\mathfrak{p}}$.
Using formula (4.2), the expectation of $\mathfrak{p}$ is very small :

$$
\begin{aligned}
\mathbb{E}\left[\mathfrak{p}\right] &= \sum_{j=1}^{\infty} j\,\mathbb{P}\left[\mathfrak{p}=j\right] = \sum_{j=1}^{\infty} \mathbb{P}\left[\mathfrak{p}\geq j\right] \\
&= \sum_{i=1}^{3} \mathbb{P}\left[\mathfrak{p}\geq j\right] + \sum_{j=4}^{\infty} \mathbb{P}\left[\mathfrak{o}\notin K_{j-1}\right] \\
&= 3 + \sum_{j=3}^{\infty} \mathbb{P}\left[\mathfrak{o}\notin K_{j}\right] \\
&= 3 + \sum_{j=3}^{\infty} j 2^{-(j-1)} = 3+2 = 5. \qquad (4.3)
\end{aligned}
$$

Thus, the expected size and time complexity of the initialization is $O(1)$.

For i large enough, we have [7]:

$$
\mathbb{E}\left[d_H\left(K_{m_i}, \mathcal{D}\right)\right] = \Theta\left(\frac{\ln m_i}{m_i}\right)^{\frac{2}{3}} \qquad (4.4)
$$

where d_H, the Hausdorff distance, is the maximal distance between a point in K_i and the boundary of $\mathcal{D}$.
Recall that $\mathcal{D}_i$ is the annulus with radii $r_i = 1 - d_H(K_{m_i}, \mathcal{D})$ and 1 and let p_i be the probability that a random point in $\mathcal{D}$ falls in $\mathcal{D}_i$.

Using (4.4), there exist a constant $c_0 > 0$ such that, for i large ,

$$
\mathbb{E}\left[p_i\right] = 1 - r_i^2 < 2(1-r_i) < c_0\left(\frac{\ln m_i}{m_i}\right)^{\frac{2}{3}}.
$$

Let' s call i_τ the last step i where $m_i < \frac{n}{\ln^2 n}$.
At each step $i \leq i_\tau$, k_i is a binomial variable with parameter p_i and $m_i < \frac{n}{\ln^2 n}$.
Thus, for i large enough,

$$
\begin{aligned}
\mathbb{E}\left[k_i\right] &= \mathbb{E}\left[\ \mathbb{E}\left[k_i \mid p_i\ \right]\right] \\
&= m_i\,\mathbb{E}\left[p_i\right] \\
&= O(m_i^{\frac{1}{3}} \log^{\frac{2}{3}}(m_i)) \\
&= O\left(\left(\frac{n}{\log^2 n}\right)^{\frac{1}{3}} \log^{\frac{2}{3}} n\right) = O(n^{\frac{1}{3}}). \qquad (4.5)
\end{aligned}
$$

As we choose $m_{i+1} = 2m_i$, i_τ is bounded by $\log_2(n)$.
For $i > i_\tau$, k_i is a binomial variable with parameter p_i and $\frac{n}{\ln^2 n}$, so using Equation (4.5) and the fact that $m_i > \frac{n}{\ln^2 n}$, we get $\mathbb{E}\left[k_i\right] = O(n^{\frac{1}{3}})$.

As we simulate at each step $i > i_\tau$ at least $\frac{n}{\ln^2(n)}$, the number of step after i_τ is bounded by $\ln^2(n)$.

Now,

$$\begin{aligned}\mathbb{E}[k] &= \sum_{i=1}^{i_\tau} \mathbb{E}[k_i] + \sum_{i>i_\tau} \mathbb{E}[k_i] \\ &\leq O\left(\ln^{\frac{2}{3}} n \left(\sum_{i=1}^{\log_2 n} m_i^{\frac{1}{3}}\right)\right) + \ln^2 n\ O(n^{\frac{1}{3}}) \\ &\leq O\left(\ln^{\frac{2}{3}} n \left(\sum_{i=1}^{\log_2 n} (2^i)^{\frac{1}{3}}\right)\right) + O(n^{\frac{1}{3}} \ln^2 n) = O(n^{\frac{1}{3}} \ln^2 n) \end{aligned} \tag{4.6}$$

At each step i, the expected size of the convex hull is $O(m_i^{\frac{1}{3}}) = O(n^{\frac{1}{3}})$, and $\mathbb{E}[k_i] = O(n^{\frac{1}{3}})$. Thus, using Chernoff bound, the expected size of the largest list, that is the expected size complexity, is $O(n^{\frac{1}{3}})$.

As our points are sorted according to their polar angle, computing the convex hull with a Graham scan is done in linear time ($O(n^{\frac{1}{3}} \ln^2 n)$), the generation of the k points and the computation of the largest annulus as well ($O(n^{\frac{1}{3}})$). Thus, the expected time complexity is $O(n^{\frac{1}{3}} \log^2 n)$.

4.4 Experiments

This algorithm has been implemented in C++ and integrated into the CGAL library.

As expected, the distribution of $\mathbb{E}[f_0(K_n)]$ is asymptotically the same as the theoretical one, see Equation (4.1), with the same constant, see Figure 4.2. This estimation has been done on 1000 experiments for each value of n.

Complexities. To evaluate the expected size complexity, we estimate the average size of the largest list of points used in the loop, see Figure 4.3. The time complexity can be evaluated by estimating the total number of points generated, see Figure 4.4.

As a first application, we can estimate the distribution of the smallest and the largest edge of a random polygon, see Figure 4.5 and Figure 4.6. The average have been done on 100 experiments for each value of n. The expected minimal edge seems to be $O(n^{-\frac{1}{3}})$, and the maximal $O\left(\frac{\ln^2 n}{n}\right)^{\frac{1}{3}}$.

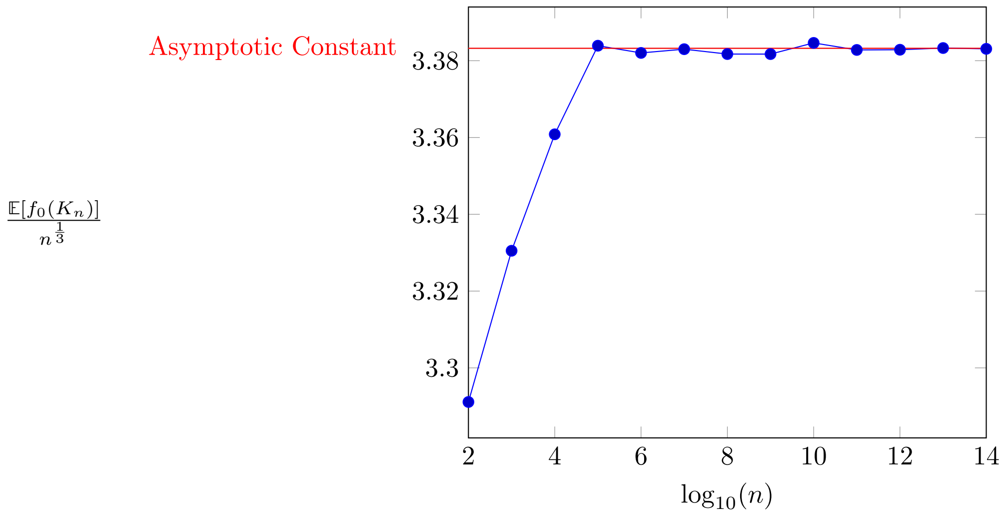


Figure 4.2: Average size of K_n divided by $n^{\frac{1}{3}}$

4.5 Conclusion

We propose an algorithm that generates random polygons given by the convex hull of random points, without generating all the points. There is no theoretical obstacle to generalize to higher dimension. The theoretical results used for the evaluation of the complexity are known in arbitrary dimension [7, 43], so the analysis can be done as well.

We can reduce the expected time complexity by a logarithmic factor if we allow to increase the expected size complexity by a logarithmic factor. Instead of bounding the simulated points at step i by $\frac{n}{\ln^2 n}$ when m_i becomes bigger than $\frac{n}{\ln^2 n}$, we can choose to always simulate m_i points. In this case,

$$\mathbb{E}\left[k_i\right] = O(m_i^{\frac{1}{3}} \ln^{\frac{2}{3}} m_i) = O(n^{\frac{1}{3}} \ln^{\frac{2}{3}} n) \tag{4.7}$$

so the expected size complexity becomes $O(n^{\frac{1}{3}} \ln^{\frac{2}{3}} n)$. On the other hand, the expected time complexity is reduced to $O(n^{\frac{1}{3}} \ln^{\frac{2}{3}} n)$, as

$$\begin{aligned}\mathbb{E}\left[k\right] &= \sum_i \mathbb{E}\left[k_i\right] \\ &= O\left(\ln^{\frac{2}{3}} n \left(\sum_{i=1}^{\log_2 n} m_i^{\frac{1}{3}}\right)\right) = O\left(n^{\frac{1}{3}} \ln^{\frac{2}{3}} n\right).\end{aligned}$$

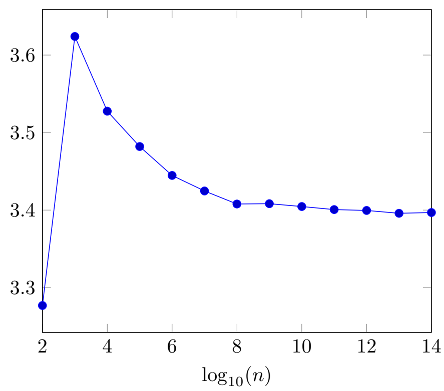


Figure 4.3: Average maximal size of the list divided by $n^{\frac{1}{3}}$

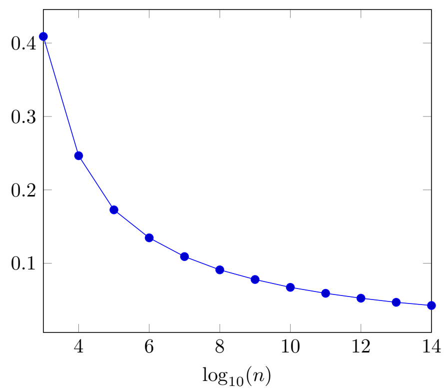


Figure 4.4: Average number of generated points divided by $n^{\frac{1}{3}} \ln^2 n$

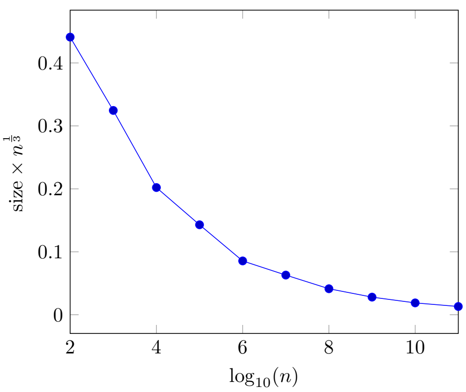


Figure 4.5: Average length of the minimal edge

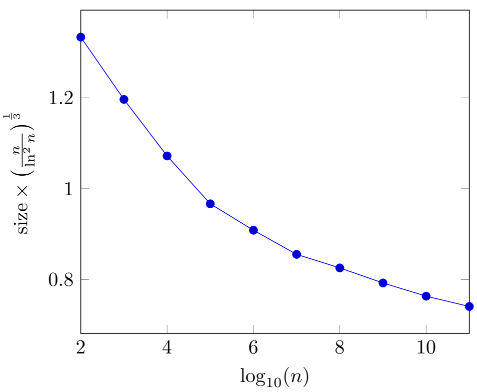


Figure 4.6: Average length of the maximal edge

Conclusion & Perspectives

We give new results on the complexity analysis of random convex hull. The Witness & Collector technique allows to capture the order of magnitude of the expected number of faces of a convex hull made of random points. These random points might be identically distributed: in this case we give an easier proof of classic results. They can also be a noisy version of an arbitrary set of points: this corresponds to the smoothed complexity.

We performed the smoothed analysis for two kind of perturbations. The first perturbation distribution is the uniform noise in a ball. This first analysis is new, and given in arbitrary dimension. The second one is the Gaussian perturbation: in this case we improve the results of the state of the art in dimension 2; we believe that this result can be generalized in higher dimensions.

For these two results, we provide an upper and a lower bound for the smoothed complexity. In the two cases, the lower bound does not match with the upper bound. The problem of having tight bounds is still open; however the case of large perturbations is tight, since it corresponds to the average-case magnitude.

The Witness & Collector technique can be applied for other structures based on hypergraphs; taking the ranges as balls of $\mathbb{R}^d$, the Delaunay triangulation average size can be estimated. An interesting question would be to bound the smoothed complexity under a given noise, for an arbitrary initial set of points. The technique works for a bad initial point set: if the points are on the moment curve $t \mapsto (t, t^2, t^3)$, it's possible to compute the noisy complexity under a uniform Euclidean noise. Is it possible to deal with an arbitrary initial set of points? If not, can we, at least, show that under large perturbations, the smoothed complexity corresponds to the average complexity?

Other geometric structures can be interesting to investigate. One example is the *visibility complex* of a set of n triangles in $\mathbb{R}^3$. The visibility complex encodes all the visibility informations of a 3D scene. In the worst case, its size is $\Theta(n^4)$, which makes the structure unusable. The worst-case seems, however, to be too pessimistic in practice; in particular for uniformly distributed unit balls the expected size is linear and for polygons or polyhedra of bounded aspect ratio [21]. This would justify a smoothed analysis.

The asymptotic expected average number of vertices of a random polytope is known with high precision when the points are uniformly and independently chosen on a well shaped convex body. If the convex body is badly shaped, then it can be as unpredictable as we want. Actually, *most* of the convex bodies are badly shaped, however this notion of *most* does not include the convex bodies used and studied *in practice*. We give a construction of such a set, where the magnitude of the expected size oscillates between two regimes. This is an illustration of the fact that we cannot hope a good asymptotic formula for every arbitrary convex body.

We also propose an algorithm generating efficiently a random convex hull in a disk. We reduced the number of points to be generated, using the observation that, after generating a small number of points, a large region of the disk will not contain any of the extreme points. Generating efficiently random geometric objects can be interesting to study limit behaviors. It can also provide test case for algorithms, if the average-case analysis of the algorithm is not known. An open question, for example, would be to generate efficiently the convex hull of n points uniformly chosen, given the fact that the n points are in convex position. Since this event happens with a very small probability, a simple rejection algorithm is inneficient so generating such a structure when n is large is challenging.

These problems can be seen as special cases of the problem of sampling an object efficiently according to a given distribution μ. Several algorithms using Markov chains can solve (at least approximately) this kind of problem. However, computing the mixing time (that is, the number of steps before reaching the steady state distribution) can be difficult, and several interesting cases are still open.

Bibliography

[1] A. V. Aho. Complexity theory. In E. K. Blum and A. V. Aho, editors, *Computer Science, The Hardware, Software and Heart of It*, pages 241–267. Springer, 2011. doi:10.1007/978-1-4614-1168-0_12.

[2] D. Aloise, A. Deshpande, P. Hansen, and P. Popat. Np-hardness of euclidean sum-of-squares clustering. *Machine Learning*, 75(2):245–248, 2009. ISSN 0885-6125. doi:10.1007/s10994-009-5103-0.

[3] N. Alon and J. H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., Hoboken, NJ, third edition, 2008. ISBN 978-0-470-17020-5. doi:10.1002/9780470277331. With an appendix on the life and work of Paul Erdős.

[4] E. Anderson, Z. Bai, C. Bischof, L. S. Blackford, J. Demmel, J. J. Dongarra, J. Du Croz, S. Hammarling, A. Greenbaum, A. McKenney, and D. Sorensen. *LAPACK Users' Guide (Third Ed.)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999. ISBN 0-89871-447-8.

[5] D. Arthur, B. Manthey, and H. Röglin. Smoothed analysis of the k-means method. *J. ACM*, 58(5):19:1–19:31, Oct. 2011. ISSN 0004-5411. doi:10.1145/2027216.2027217.

[6] D. Attali, O. Devillers, and X. Goaoc. The Effect of Noise on the Number of Extreme Points. Research Report RR-7134, Inria, 2009. URL `https://hal.inria.fr/inria-00438409`.

[7] I. Bárány. Intrinsic volumes and f-vectors of random polytopes. *Mathematische Annalen*, 285(4):671–699, 1989.

[8] I. Bárány. Random points and lattice points in convex bodies. *Bull. Amer. Math. Soc. (N.S.)*, 45(3):339–365, 2008. ISSN 0273-0979. doi:10.1090/S0273-0979-08-01210-X.

[9] I. Bárány and D. Larman. Convex bodies, economic cap coverings, random polytopes. *MATHEMATIKA*, 35:274–291, 1988. ISSN 2041-7942. doi:10.1112/S0025579300015266.

[10] I. Bárány and V. Vu. Central limit theorems for gaussian polytopes. *The Annals of Probability*, 35(4):1593–1621, 2007.

[11] W. Blaschke. *Vorlesungen über Differenzialgeometrie II. Affine Differenzialgeometrie*. Springer, 1923.

[12] K. Böröczky Jr. Polytopal approximation bounding the number of k-faces. *Journal of Approximation Theory*, 102(2):263 – 285, 2000. ISSN 0021-9045. doi:10.1006/jath.1999.3413.

[13] I. Bárány. Random polytopes in smooth convex bodies. *Mathematika*, 39: 81–92, 6 1992. ISSN 2041-7942. doi:10.1112/S0025579300006872.

[14] F. Cazals and J. Giesen. Delaunay triangulation based surface reconstruction. In J.-D. Boissonnat and M. Teillaud, editors, *Effective Computational Geometry for Curves and Surfaces*, pages 231–276. Springer-Verlag, Mathematics and Visualization, 2006. doi:10.1007/978-3-540-33259-6_6.

[15] R. M. Corless, G. H. Gonnet, D. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert W function. *Advances in Computational Mathematics*, 5(1):329–359, 1996. doi:10.1007/BF02124750.

[16] T. H. Cormen, C. Stein, R. L. Rivest, and C. E. Leiserson. *Introduction to Algorithms*. McGraw-Hill Higher Education, 2nd edition, 2001. ISBN 0070131511.

[17] V. Damerow and C. Sohler. Extreme points under random noise. In *Proc. 12th European Sympos. Algorithms*, pages 264–274, 2004. doi:10.1007/978-3-540-30140-0_25.

[18] G. Dantzig. Maximization of linear function of variables subject to linear inequalities. In T. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 339–347. 1951.

[19] M. de Berg. Improved bounds on the union complexity of fat objects. *Discrete & Computational Geometry*, 40:127–140, 2008. doi:10.1007/s00454-007-9029-7.

[20] M. de Berg, H. Haverkort, and C. P. Tsirogiannis. Visibility maps of realistic terrains have linear smoothed complexity. In *Proceedings of the Twenty-fifth Annual Symposium on Computational Geometry*, SCG '09, pages 163–168, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-501-7. doi:10.1145/1542362.1542397.

[21] O. Devillers, V. Dujmovic, H. Everett, X. Goaoc, S. Lazard, H.-S. Na, and S. Petitjean. The expected number of 3D visibility events is linear. *SIAM Journal on Computing*, 32(6):1586–1620, 2003. doi:10.1137/S0097539702419662. Article dans revue scientifique avec comité de lecture. internationale.

[22] O. Devillers, M. Glisse, and X. Goaoc. Complexity analysis of random geometric structures made simpler. Research Report RR-8168, INRIA, 2012. URL `https://hal.inria.fr/hal-00761171`.

[23] O. Devillers, M. Glisse, and X. Goaoc. Complexity analysis of random geometric structures made simpler. In *Symposium on Computational Geometry*, pages 167–176, 2013. doi:10.1145/2462356.2462362.

[24] O. Devillers, M. Glisse, X. Goaoc, G. Moroz, and M. Reitzner. The monotonicity of f-vectors of random polytopes. *Electron. Commun. Probab.*, 18:no. 23, 1–8, 2013. ISSN 1083-589X. doi:10.1214/ECP.v18-2469.

[25] O. Devillers, P. Duchon, and R. Thomasse. A generator of random convex polygons in a disc. AofA 2014- 25th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, June 2014. URL `https://hal.inria.fr/hal-01015603`. Poster.

[26] O. Devillers, M. Glisse, and R. Thomasse. A chaotic random convex hull. AofA 2014- 25th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, June 2014. URL `https://hal.inria.fr/hal-01015598`. Poster.

[27] O. Devillers, M. Glisse, X. Goaoc, and R. Thomasse. On the Smoothed Complexity of Convex Hulls. In L. Arge and J. Pach, editors, *31st International Symposium on Computational Geometry (SoCG 2015)*, volume 34 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 224–238, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-939897-83-5. doi:http://dx.doi.org/10.4230/LIPIcs.SOCG.2015.224.

[28] O. Devillers, M. Glisse, X. Goaoc, and R. Thomasse. On the smoothed complexity of convex hulls. In *Symposium on Computational Geometry*, pages 224–239, 2015. doi:10.4230/LIPIcs.SOCG.2015.224.

[29] O. Devillers, M. Glisse, X. Goaoc, and R. Thomasse. Smoothed complexity of convex hulls by witnesses and collectors. Research Report 8787, INRIA, Oct. 2015. URL `https://hal.inria.fr/hal-01214021`.

[30] R. Dwyer. The expected number of k-faces of a Voronoi diagram. *Internat. J. Comput. Math.*, 26(5):13–21, 1993. doi:10.1016/0898-1221(93)90068-7.

[31] B. Efron. The convex hull of a random set of points. *Biometrika*, 52:331–343, 1965. ISSN 0006-3444.

[32] J. Erickson. Nice point sets can have nasty Delaunay triangulations. *Discrete and Computational Geometry*, 30(1):109–132, 2003. doi:10.1007/s00454-003-2927-4.

[33] M. Glisse, S. Lazard, J. Michel, and M. Pouget. Silhouette of a random polytope. Research Report 8327, INRIA, 2013. URL https://hal.inria.fr/hal-00841374/.

[34] R. L. Graham. An efficient algorith for determining the convex hull of a finite planar set. *Information processing letters*, 1(4):132–133, 1972.

[35] M. Grotschel, L. Lovasz, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag. doi:10.1007/978-3-642-78240-4.

[36] M. Inaba, N. Katoh, and H. Imai. Applications of weighted voronoi diagrams and randomization to variance-based k-clustering: (extended abstract). In *Proceedings of the Tenth Annual Symposium on Computational Geometry*, SCG '94, pages 332–339, New York, NY, USA, 1994. ACM. ISBN 0-89791-648-4. doi:10.1145/177424.178042.

[37] V. Klee and G. J. Minty. How Good is the Simplex Algorithm? In O. Shisha, editor, *Inequalities III*, pages 159–175. Academic Press Inc., New York, 1972.

[38] M. Mahajan, P. Nimbhorkar, and K. Varadarajan. The planar k-means problem is np-hard. In S. Das and R. Uehara, editors, *WALCOM: Algorithms and Computation*, volume 5431 of *Lecture Notes in Computer Science*, pages 274–285. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-00201-4. doi:10.1007/978-3-642-00202-1_24.

[39] J.-F. Marckert. Probability that n random points in a disk are in convex position. 2014. URL http://arxiv.org/abs/1402.3512.

[40] V. Petrov. *Limit Theorems of Probability Theory. Sequence of Independent Random Variables.* Number 4 in Oxford studies in probability. Clarendon Press, 1995. URL http://www.citeulike.org/group/2854/article/1615442.

[41] H. Raynaud. Sur l'enveloppe convexe des nuages de points aleatoires dans R^n. *J. Appl. Probab.*, 7:35–48, 1970. URL http://www.jstor.org/stable/10.2307/3212146.

[42] M. Reitzner. The combinatorial structure of random polytopes. *Advances in Mathematics*, 191(1):178 – 208, 2005. ISSN 0001-8708. doi:http://dx.doi.org/10.1016/j.aim.2004.03.006.

[43] M. Reitzner. Random polytopes. In *New perspectives in stochastic geometry*, pages 45–76. Oxford Univ. Press, Oxford, 2010.

[44] A. Rényi and R. Sulanke. Über die konvexe Hülle von n zufällig gewählten Punkten I. *Z. Wahrsch. Verw. Gebiete*, 2:75–84, 1963. doi:10.1007/BF00535300.

[45] A. Rényi and R. Sulanke. Über die konvexe Hülle von n zufällig gewählten Punkten II. *Z. Wahrsch. Verw. Gebiete*, 3:138–147, 1964. doi:10.1007/BF00535973.

[46] A. Sankar, D. A. Spielman, and S.-H. Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM Journal on Matrix Analysis and Applications*, 28(2):446–476, 2006.

[47] Schütt, Carsten and Werner, Elisabeth. The convex floating body. *Math. Scand.*, 66(2):275–290, 1990.

[48] C. Schütt. The convex floating body and polyhedral approximation. *Israel Journal of Mathematics*, 73(1):65–77, 1991. ISSN 0021-2172. doi:10.1007/BF02773425.

[49] D. A. Spielman and S.-H. Teng. Smoothed analysis: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51:385 – 463, 2004. doi:10.1145/990308.990310.

[50] J. J. Sylvester. On a special class of questions on the theory of probabilities. *Birmingham British Association Report*, pages 8–9, 1865.

[51] L. N. Trefethen and D. Bau, III. *Numerical Linear Algebra.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997. ISBN 0-89871-361-7.

[52] P. Valtr. The probability that n random points in a triangle are in convex position. *Combinatorica*, 16(4):567–573, 1996. ISSN 0209-9683. doi:10.1007/BF01271274.

[53] A. Vattani. k-means requires exponentially many iterations even in the plane. *Discrete & Computational Geometry*, 45(4):596–616, 2011. ISSN 0179-5376. doi:10.1007/s00454-011-9340-1.

[54] J. Wilkinson. Error analysis of direct methods of matrix inversion. *J.Assoc.Comput.Mach.*, 8:261–330, 1961.