

BETTI NUMBERS OF DETERMINISTIC AND RANDOM SETS IN
SEMI-ALGEBRAIC AND O-MINIMAL GEOMETRY

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Abhiram Natarajan

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2020

Purdue University

West Lafayette, Indiana

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. Saugata Basu, Chair

Department of Mathematics, Purdue University

Dr. Elena Grigorescu, Co-Chair

Department of Computer Science, Purdue University

Dr. Hemanta Maji

Department of Computer Science, Purdue University

Dr. Simina Branzei

Department of Computer Science, Purdue University

Approved by:

Dr. Clifton W. Bingham

Head of the Graduate Program, Purdue University

Dedicated to Tatu, M.B.

ACKNOWLEDGMENTS

Writing this section has been tremendously emotional, and also somewhat disconcerting. Disconcerting because of this – obviously, if there are n people in your life, each of whom deserve at least an ε fraction of credit for your thesis, your personal credit is at most $1 - n\varepsilon$. For any $\delta > 0$, I am able to find n_δ people in my life, each of whom deserve at least ε_δ fraction of credit for my thesis, such that $1 - n_\delta\varepsilon_\delta < \delta$, leaving me wondering – how much credit do I really deserve for all this?

My biggest debt of gratitude is to my advisor Prof. Saugata Basu. Although I began discussing research with him only in Fall 2015, I had made up my mind to work with him in Fall 2013 itself when he gave a talk about his research to incoming graduate students. Needless to say, I did not understand anything he spoke about; however, at the end of his talk, I noticed that he said “if anyone is **interested to learn about this**, let’s meet”. Most professors ended their talks by saying how they were looking for brilliant, mathematically mature, and *<other adjectives that I am never able to attribute to myself>* students. I was sure it was not an accident that he used those words, and this was enough for me to become intrigued. After five years under his supervision, I can say his judgement has been unassailable $\aleph_0 - 1$ number of times (the exception being his decision to give me a chance to do research with him). I am grateful for all the mathematics I have learned from him, for his refined taste in research which I hope I have acquired in a small measure at least, and for being a paragon of adherence to the highest standards of rigour, and of course for the generous funding which allowed me to focus on research. In addition to all this, I thank him most for his patience with me and for giving me unfettered time to learn things at my own pace.

Next, I am very lucky to have been co-advised by Prof. Elena Grigorescu. In fact, I have been advised by her the longest. I took on my first long-term research project

under her supervision and obtained my first lessons in rigour under her guidance. She taught me a lot about the patience required to execute a meaningful research agenda and bring it to completion. While I have not talked about our publication together, I learned a lot from the experience. To this and a lot more things left unsaid, I am very grateful for her mentorship.

I was very fortunate to have collaborated with Prof. Antonio Lerario. In fact, I can say that the main core of this thesis might not have been possible without him. He is an excellent teacher. He is eternally teeming with ideas and research directions; a meeting with him always leaves me optimistic and hopeful. I sincerely hope that I can learn from him again in the future. He had absolutely no obligation to spend time with me, but he did it anyway, and for this, I am most grateful.

Next, I'd like to thank Prof. Joshua Grochow. Discussing research with him has been an intellectual privilege. He is another person who has absolutely no obligation to spend his time with me, but does it for god knows what reason. There are many things about him that might be annoying – his ability to immediately find a secondary path upon hitting a dead end, his fearless attitude towards learning and exploring, his supersonic brilliance, the fact that he is nicer than he is brilliant, or, paradoxically, that it is impossible to be annoyed by any of the previous reasons. I look forward to bringing our current projects to completion and hopefully have many more future collaborations with him. I can't believe my luck. I pray he never realizes that he can do better.

I'd like to thank Dr. Yi Wu for his mentoring during my first year at Purdue. He gave me my first chance in mathematical research, and I can't imagine being this happy with anything else. Many thanks to Prof. Hemanta Maji and Prof. Simina Branzei for serving on my committee and giving me valuable feedback and encouragement. I'd also like to thank the many exemplary teachers I have learned from. I owe all I know to them. Special thanks to Prof. Eugene Charniak, Prof. N C Naveen, Prof. Vasantha Ramaswamy, Prof. N K Srinath, and Prof. Eli Upfal for writing me recommendation letters for admission to graduate school. Finally, I'd

like to thank my collaborators Saugata Basu, Ilias Diakonikolas, Elena Grigorescu, Antonio Lerario, Jerry Li, Krzysztof Onak, Ludwig Schmidt, and Yi Wu for being fantastic people to work with.

A big thank you to people who share knowledge for free. I've benefited from innumerable video lectures, lecture notes, people responding to questions over email, stackexchange answers, etc. One of the greatest privileges of being a student today is the sheer volume of knowledge available for those who wish to learn.

The environment at Purdue was really a blessing. I highly recommend it to students who are looking for a balance between intellectual stimulation and freedom. Administrative activities were really a breeze due to the very co-operative staff. Most importantly, I've been very lucky to have had a lot of fantastic friends at Purdue and outside – Akash, Ashwin, Asish, Ganapathy, GV, Kaki, Kartik, Kaushal, Mayank, Negin, Omran, Onkar, Pramod, Rahul, Rohit, Sandeep, Sridhar, Vikhyat, Vikram, Vinit, Vivek, Warren – who made PhD life so much more bearable. Separate mention to (a) GV for finding me out of the blue and inviting me to the theory group (I might have very well not done theory otherwise) (b) Akash for teaching me a lot of things I would never have known about otherwise, and also for inspiring through his attitude towards learning (c) Vikram for *looking out* for me (Figure 2.2 is dedicated to you) (d) Asish for being there during both blue and red times. Please forgive me on two counts – (a) I would have liked to single out each one of the above and say something specific (b) I have left out other friends who have had a significant influence on my life. I am arbitrarily restricting myself to people I have interacted with most during the past five years because I am already embarrassed about the length of this section.

I owe a great lot to my family. They have always loved me unconditionally, been there for me, and have reposed blind trust in me. Specifically, I'd like to thank my late grandfather Krishnaswamy for giving me my first lessons in mathematics. He is one of the two people I dedicate my thesis to. I thank my father K Natarajan for sacrificing his evenings to patiently teach mathematics to a teenager whose only concern was Tendulkar scoring centuries; this thesis wouldn't exist otherwise. To

this date, he remains the best mathematics teacher I have learned from. I thank my mother Jyothi Natarajan for teaching me nearly everything else I know. I am very grateful for her selflessness, and for being ‘mad’ in general. My parents gave foremost importance to their children’s education, and I am very humbled to have received this privilege. I thank my sister Sarayu Natarajan for being somewhat of a prescient presence in my life. I could say that algorithms and algebraic geometry are central to my academic life; she taught me precisely two things during my school days - programming, and co-ordinate geometry. Coincidence? I think not.

I’d like to thank my in-laws for their support. My brother-in-law Kartik was a tremendous comforting presence during a particularly difficult time in my life. My parents-in-law, Srinath K S and Nagarathna Ramaswamy, have always encouraged me, trusted my decisions, and have never directly or indirectly placed any pressure on me during my PhD. When we needed their help, they flew to the USA with just two weeks notice, just to facilitate me completing my final semester without trouble. 100% support for education is a creed for them, and it is something my wife and I can only hope to emulate. I’d also like to thank my sister-in-law Poornima and brother-in-law Suhas for their company, and supportive presence in general. Importantly, I’d like to thank my dear little nephew Kanishka for being a tremendous source of entertainment and joy, and for overloading all of us with his cuteness.

It may be a cliché to say this, but you appreciate the value of people only during tough times. Many members of my family were tremendously supportive during the aforementioned difficult time. Special mention to the Nadig family (Ajja, Ajji, Nanni, Satish, Tejas, Nanda uncle, Roopa and Geetha aunty). They were there at every step helping me take one day at a time, whilst also unconditionally promising all their resources to help me. It appears that all this was based on some unfounded belief they had in me, and this trust has inspired me and made me aver to consciously nurture life as much as possible.

Last, definitely not the least, probably the most, I’d like to talk about what my wife Pavithra means to me. While it was my parents who hyped my capabilities

during my childhood, she has been solely responsible in keeping up the ruse over the majority of my adult life. She has assiduously stood beside me, given me support and encouragement, and has been a constant loving presence. She has been patient with all the vagaries of my academic life, which have sometimes been exacerbated by my idealistic tendencies. Not only has she provided emotional support, I've sometimes bounced ideas off her and learned things from/with her. It is hard to objectively state what she means to me – I either tend to resort to using extremely hackneyed phrases, or tongue-in-cheeks¹ remarks as a defense mechanism. My feelings for her are ineffable, so I will leave it at this. An observant reader will notice that I haven't actually said thanks; I find it impertinent to use a 'bounded' word to summarize my gratitude.

Also, fingers crossed, I'm excited about the imminent arrival into our family. I simply cannot wait to see you M.B. If you ever read this, I'll have you know that I used to be cool. This thesis is dedicated to her as well.

¹not a spelling error, insider's joke...

TABLE OF CONTENTS

| | Page |
|---|------|
| LIST OF TABLES | xi |
| LIST OF FIGURES | xii |
| ABSTRACT | xiv |
| 1 INTRODUCTION | 1 |
| 1.1 Semi-algebraic geometry | 1 |
| 1.2 O-minimal Geometry | 2 |
| 1.2.1 A soupçon of Model Theory | 2 |
| 1.2.2 O-minimal Structures | 4 |
| 1.3 Random Algebraic Geometry | 7 |
| 1.3.1 Some basic results in random algebraic geometry | 9 |
| 2 TOPOLOGICAL COMPLEXITY OF SEMI-ALGEBRAIC AND DEFIN- ABLE SETS | 12 |
| 2.1 Betti Numbers | 13 |
| 2.1.1 Betti numbers of semi-algebraic sets | 13 |
| 2.2 Applications of Bounds on Betti Numbers | 14 |
| 2.2.1 Discrete geometry applications | 16 |
| 2.3 Betti numbers of definable sets | 19 |
| 3 ZEROS OF POLYNOMIALS ON DEFINABLE HYPERSURFACES | 20 |
| 3.1 Introduction | 20 |
| 3.1.1 Existence of pathologies | 20 |
| 3.1.2 Pathologies are rare | 23 |
| 3.2 Pathological examples: Proof of Theorem 3.1.1 | 27 |
| 3.2.1 Construction of Gwoździewicz et al. | 27 |
| 3.2.2 Some basic facts | 30 |
| 3.3 Estimates on the size of pathological examples: proof of Theorem 3.1.2 | 34 |
| 3.4 Toward an O-minimal Polynomial Partitioning Theorem? | 40 |
| 3.4.1 Why do we not have an o-minimal polynomial partitioning theorem? | 40 |
| 3.4.2 An o-minimal polynomial partitioning theorem using the prob- abilistic method? | 42 |
| 4 BETTI NUMBERS OF RANDOM HYPERSURFACE ARRANGEMENTS | 44 |
| 4.1 Introduction | 44 |

| | Page |
|---|------|
| 4.1.1 Random hypersurface arrangements | 47 |
| 4.1.2 Arrangements of random quadrics | 49 |
| 4.1.3 A random graph model | 50 |
| 4.2 A Random Spectral Sequence | 51 |
| 4.2.1 Preliminaries on spectral sequences | 51 |
| 4.2.2 Random Mayer-Vietoris Spectral Sequence | 58 |
| 4.2.3 Average Betti numbers of hypersurface arrangements | 60 |
| 4.3 Obstacle Random Graphs and an Application to Arrangement of Quadrics | 63 |
| 4.3.1 The ‘Obstacle’ random graph model | 64 |
| 4.3.2 Average number of connected components of obstacle random graphs | 66 |
| 4.3.3 b_0 of arrangement of quadrics | 81 |
| 4.3.4 A Ramsey-type result | 83 |
| 4.4 More studies of the topology of random arrangements | 84 |
| REFERENCES | 87 |

LIST OF TABLES

| Table | Page |
|---|------|
| 1.1 Examples of real-algebraic sets and semi-algebraic sets | 1 |
| 1.2 Semi-algebraic sets vs Definable sets | 6 |
| 2.1 Betti numbers of semi-algebraic sets – examples | 14 |

LIST OF FIGURES

| Figure | Page |
|---|------|
| 2.1 “...a topologist cannot differentiate between a coffee mug and a donut because they are homotopy equivalent...” – illustration of smooth transformation from a coffee mug to a donut | 12 |
| 2.2 Just to belabor the point, these two koDubaLes are the same to me | 12 |
| 2.3 An example of an algebraic computation tree for testing membership in $\{(X, Y) \in \mathbb{R}^2 \mid 2X^2 - Y^2 \neq 0\}$ | 15 |
| 2.4 Example illustration of polynomial partitioning. The zero set of P breaks \mathbb{R}^2 into five connected components – $\mathcal{C}_1, \dots, \mathcal{C}_5$. Each \mathcal{C}_i is intersected by a subset of varieties in Γ | 17 |
| 3.1 Illustration of an ambient diffeotopy. The manifold pairs (M, X) are ambient diffeotopic to (N, Y) | 21 |
| 3.2 An illustration of the Thom isotopy lemma – if the perturbation is <i>small</i> , then the zero set is topologically the same | 31 |
| 3.3 Illustration of desirable situation to prove an o-minimal polynomial partitioning theorem | 42 |
| 4.1 For us, an arrangement is just the union of a finite number of algebraic sets | 44 |
| 4.2 Illustration of obstacle random graph. The thick lines denote edges of the graph, while the dotted lines denote non-edges, i.e. edges that were not included in the random graph because their geodesic completion intersected \mathcal{P} | 65 |
| 4.3 Illustration of $g_q(\mathcal{P})$, the <i>good cone</i> of a point q w.r.t. \mathcal{P} . The dashed lines are geodesics which are tangent to \mathcal{P} and incident on q . The shaded region is $g_q(\mathcal{P})$. Recall that in $\mathcal{G}(N, \mathcal{P}, s)$, by definition, if q is sampled and any point in $g_q(\mathcal{P})$ is sampled, these points would be connected to each other by any edge. | 71 |

| Figure | Page |
|--|------|
| 4.4 Illustration of the good cone of q w.r.t. $\tilde{\mathcal{P}}(\varepsilon)$. $\tilde{\mathcal{P}}(\varepsilon)$ is an approximation of \mathcal{P} which is convex and has a smooth boundary, such that $\mathcal{P} \subseteq \tilde{\mathcal{P}}(\varepsilon) \subseteq \mathcal{P}(\varepsilon)$. The dashed lines are geodesics which are tangent to \mathcal{P} and incident on q , and the dotted lines are geodesics which are tangent to $\tilde{\mathcal{P}}(\varepsilon)$ and incident on q . Observe that $g_q(\tilde{\mathcal{P}}(\varepsilon)) \subseteq g_q(\mathcal{P})$, and consequently, $\text{vol}\left(g_q(\tilde{\mathcal{P}}(\varepsilon))\right) \leq \text{vol}\left(g_q(\mathcal{P})\right)$ | 74 |
| 4.5 Illustration of the proof of Lemma 4.3.1 in a nutshell – cover the complement of the fattening of $\mathcal{P}(\varepsilon)$ with balls, show that each ball has positive probability of being <i>covered</i> , and then finish with a coupon-collector type argument. | 76 |

ABSTRACT

Natarajan, Abhiram Ph.D., Purdue University, May 2020. Betti numbers of deterministic and random sets in semi-algebraic and o-minimal geometry. Major Professors: Saugata Basu, Elena Grigorescu.

Studying properties of random polynomials has marked a shift in algebraic geometry. Instead of worst-case analysis, which often leads to overly pessimistic perspectives, randomness helps perform average-case analysis, and thus obtain a more realistic view. Also, via Erdős’ astonishing ‘probabilistic method’, one can potentially obtain deterministic results by introducing randomness into a question that apriori had nothing to do with randomness.

In this thesis, we study topological questions in real algebraic geometry, o-minimal geometry and random algebraic geometry, with motivation from incidence combinatorics. Specifically, we prove results along two different threads:

- (a) Topology of semi-algebraic and definable (over any o-minimal structure over \mathbb{R}) sets, in both deterministic and random settings.
- (b) Topology of random hypersurface arrangements. In this case, we also prove a result that could be of independent interest in random graph theory.

Towards the first thread, motivated by applications in o-minimal incidence combinatorics, we prove bounds (both deterministic and random) on the topological complexity (as measured by the Betti numbers) of general definable hypersurfaces restricted to algebraic sets (Basu et al., 2019b). Given any sequence of hypersurfaces, we show that there exists a definable hypersurface Γ , and a sequence of polynomials, such that each manifold in the sequence of hypersurfaces appears as a component of Γ restricted to the zero set of some polynomial in the sequence of polynomials. This shows that the topology of the intersection of a definable hypersurface and an alge-

braic set can be made *arbitrarily pathological*. On the other hand, we show that for random polynomials, the Betti numbers of the restriction of the zero set of a random polynomial to any definable set deviates from a Bézout-type bound with *bounded probability*.

Progress in o-minimal incidence combinatorics has lagged behind the developments in incidence combinatorics in the algebraic case due to the absence of an o-minimal version of the Guth-Katz *polynomial partitioning* theorem, and the first part of our work explains why this is so difficult. However, our average result shows that if we can prove that the measure of the set of polynomials which satisfy a certain property necessary for polynomial partitioning is suitably bounded from below, by the *probabilistic method*, we get an o-minimal polynomial partitioning theorem. This would be a tremendous breakthrough and would enable progress on multiple fronts in model theoretic combinatorics.

Along the second thread, we have studied the average Betti numbers of *random hypersurface arrangements* (Basu et al., 2019a). Specifically, we study how the average Betti numbers of a finite arrangement of random hypersurfaces grows in terms of the degrees of the polynomials in the arrangement, as well as the number of polynomials. This is proved using a random Mayer-Vietoris spectral sequence argument. We supplement this result with a better bound on the average Betti numbers when one considers an *arrangement of quadrics*. This question turns out to be equivalent to studying the expected number of connected components of a certain *random graph model*, which has not been studied before, and thus could be of independent interest. While our motivation once again was incidence combinatorics, we obtained the first bounds on the topology of arrangements of random hypersurfaces, with an unexpected bonus of a result in random graphs.





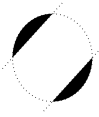
1 INTRODUCTION

1.1 Semi-algebraic geometry

Real algebraic geometry is algebraic geometry over the real numbers \mathbb{R} , or more generally, over real closed fields. The primary focus of real algebraic geometry is semi-algebraic sets, defined as elements of the boolean algebra over sets of the form $\{(x_1, \dots, x_n) \in \mathbb{R}^n \mid P(x_1, \dots, x_n) \leq 0, P \in \mathbb{R}[X_1, \dots, X_n]\}$. In other words, semi-algebraic sets are made of a finite number of conjunctions, disjunctions and negations of the locus of polynomial inequalities.

Real algebraic geometry has a long history, beginning from early works such as the Fourier-Motzkin elimination method and Sturm's counting theorem. Classical algebraic geometry, usually done over algebraically closed fields, enjoys the property that when an affine variety is projected down, the image is constructible. However, real varieties don't have this property; projections of real varieties can be semi-algebraic sets. For example, the projection of $Z(x^2 + y^2 - 1)$ is the interval $[-1, 1]$. However, due to the Tarski-Seidenberg theorem, we know that semi-algebraic sets project down to semi-algebraic sets as well, thus making semi-algebraic sets central to real-algebraic geometry. (Benedetti and Risler, 1991; Bochnak et al., 2013; Coste, 2000b) are excellent introductions to the subject.

Table 1.1.: Examples of real-algebraic sets and semi-algebraic sets

| real-algebraic sets | | semi-algebraic sets | | |
|---|---|---|---|---|
| $Z(x^2 + y^2 - 1)$ | $Z(y - x^2)$ | $\{-(x^2 + y^2 - 1) \geq 0\}$ | $\{y \geq x\} \wedge \{x \geq y\}$ | $\{x^2 + y^2 \leq 2\} \wedge ((y - x \geq 4) \vee \neg\{x - y \leq 4\})$ |
|  |  |  |  |  |

1.2 O-minimal Geometry

Semi-algebraic sets are known to possess many ‘tameness’ properties in a topological sense, such as stratifiability, finite triangulability, etc., which makes their study feasible. Often, in some theorems, it can be seen that it is only these tameness properties that are utilized. This in turn leads one to wonder if there are other classes of sets that semi-algebraic geometry can be generalized to. For an obvious example, observe that the graph of $y = e^x$, at least on say $[-1, 1]$ is isotopic to $[-1, 1]$ itself, so it is topologically no different. This question was articulated by Grothendieck in his *Esquisse d’un Programme* (Grothendieck, 1997):

“...investigate classes of sets with the tame topological properties of semi-algebraic sets...”

The answer to the above question is o-minimal geometry. O-minimal geometry, whose genesis was in model theory, is an axiomatic generalization of semi-algebraic geometry, in so much as, many results about semi-algebraic sets are actually corollaries of results in o-minimal geometry.

1.2.1 A soupçon of Model Theory

We refer the reader to references such as (Marker, 2006) to get a complete understanding of the basics of model theory. Below we shall present just a few definitions that are meant to serve as a hack to know enough model theory so as to make sense of the definition of o-minimality. We begin with the definition of a language.

Definition 1.2.1 *A language \mathcal{L} is given by:*

1. *A set of function symbols \mathcal{F} and $\{n_f \mid n_f \in \mathbb{N}\}_{f \in \mathcal{F}}$*
2. *A set of relation symbols \mathcal{R} and $\{n_R \mid n_R \in \mathbb{N}\}_{R \in \mathcal{R}}$*
3. *A set of constant symbols \mathcal{C}*

The numbers n_f and n_R are the arities of the respective functions and relations. Next we define what it means to have a *structure* based on a *language*.

Definition 1.2.2 An \mathcal{L} -structure \mathcal{S} is given by:

1. A non-empty set S called the universe, domain, or underlying set of \mathcal{S}
2. One function $f^{\mathcal{S}} : S^{n_f} \rightarrow S$ for each $f \in \mathcal{F}$
3. One set $R^{\mathcal{S}} \subseteq S^{n_R}$ for each $R \in \mathcal{R}$
4. One element $c^{\mathcal{S}} \in S$ for each $c \in \mathcal{C}$

For example, to study groups, we could use the language $\mathcal{L}_g = \{\cdot, e\}$, where \cdot denotes binary function symbol and e is a constant symbol. The set of relation symbols for \mathcal{L}_g is empty. An \mathcal{L}_g -structure $\mathcal{G} = (G, \cdot^{\mathcal{G}}, e^{\mathcal{G}})$ denotes a universe G equipped with a binary composition $\cdot^{\mathcal{G}}$ and a distinguished element $e^{\mathcal{G}}$. An example of an \mathcal{L}_g -structure is $\mathcal{G} = (\mathbb{R}, \cdot, 1)$, where we interpret \cdot as multiplication and e as 1.

The language L will be used to create formulas that define properties of \mathcal{L} -structures.

Definition 1.2.3 An \mathcal{L} -formula is a string created using the symbols $(\mathcal{F}, \mathcal{R}, \mathcal{C})$ of \mathcal{L} , variables v_1, v_2, \dots , equality $=$, boolean operators \wedge, \vee, \neg , and quantifiers \forall, \exists .

Finally, we define the notion of a *definable* set.

Definition 1.2.4 Let $\mathcal{S} = (S, \dots)$ be an \mathcal{L} -structure. A set $X \subseteq S^n$ definable if and only if there is an \mathcal{L} -formula ϕ such that X is exactly the set of all points of S^n that satisfy ϕ .

We skip defining the notion of ‘satisfy’ in a rigorous manner with the rationale that the reader has an intuitive understanding of what it means for a point to satisfy a formula.

Example 1.2.1 Here we provide an example to illustrate the efficacy of the above notions. Let $\mathcal{M} = (\mathbb{Q}, +, -, \cdot, 0, 1)$ be the field of rational numbers. Let $\phi(x, y, z)$ be the formula

$$\exists a \exists b \exists c \, xyz^2 + 2 = a^2 + xy^2 - yc^2,$$

and let $\psi(x)$ be the formula

$$\forall y \forall z ([\phi(y, z, 0) \wedge (\forall w (\phi(y, z, w) \implies \phi(y, z, w + 1)))] \implies \phi(y, z, x)).$$

By the result of (Robinson, 1949), the set of all points that satisfy $\psi(x)$ defines \mathbb{N} in \mathbb{Q} .

1.2.2 O-minimal Structures

Let's recall that our initial goal was to study structures where the definable sets are 'topologically tame'. Obviously not all structures have such definable sets. Consider, for instance, the structure $\mathcal{S}_{\mathbb{Z}}$, which is the smallest structure containing the semialgebraic sets with the set of integers $\mathbb{Z} \in \mathcal{S}_1$. Every Borel subset of \mathbb{R}^n is in $\mathcal{S}_{\mathbb{Z}}$. Thus even innocuous looking structures can have complicated definable sets.

To study structures where the definable sets are 'tame', we begin with a characterization of definable sets.

Proposition 1.2.1 (see Proposition 1.3.4 in (Marker, 2006)) *Let \mathcal{M} be an \mathcal{L} -structure. Suppose that D_n is a collection of subsets of M^n for all $n \in \mathbb{N}$ and $\mathcal{D} = (D_n)_{n \in \mathbb{N}}$ is the smallest collection such that:*

- i) $M^n \in D_n$
- ii) For all function symbols f of \mathcal{L} of arity exactly n , the graph of $f^{\mathcal{M}}$ is in D_{n+1}
- iii) For all relation symbols R of \mathcal{L} of arity exactly n , $R^{\mathcal{M}} \in D_n$
- iv) For all $i, j \leq n$, $\{(x_1, \dots, x_n) \in M^n \mid x_i = x_j\} \in D_n$

v) If $X \in D_n$, then $M \times X \in D_{n+1}$

vi) Each D_n is closed under complement, union, and intersection

vii) If $X \in D_{n+1}$, and $\pi : M^{n+1} \rightarrow M^n$ is the projection map which takes

$$(x_1, \dots, x_{n+1}) \mapsto (x_1, \dots, x_n),$$

then $\pi(X) \in D_n$

viii) If $X \in D_{n+m}$ and $b \in M^m$, then $\{a \in M^n \mid (a, b) \in X\} \in D_n$

Then, $X \subseteq M^n$ is definable if and only if $X \in D_n$.

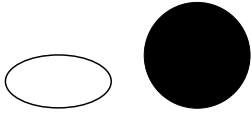

The above proposition gives a strong characterization of definable sets. Motivated by this, we define o-minimal structures, which are the principal objects of study in o-minimal geometry.

Definition 1.2.5 $\mathcal{S} = (\mathcal{S}_n)_{n \in \mathbb{N}}$, with $\mathcal{S}_n \subset \mathcal{P}(\mathbb{R}^n)$, is an o-minimal structure if:

- All algebraic subsets of \mathbb{R}^n are in \mathcal{S}_n
- \mathcal{S}_n is closed under complementation, finite unions & intersections
- If $A \in \mathcal{S}_n$, $B \in \mathcal{S}_m$, then $A \times B \in \mathcal{S}_{n+m}$
- If $A \in \mathcal{S}_{n+1}$, then $\Pi(A) \in \mathcal{S}_n$, where $\Pi : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$ is the projection on the first n coordinates
- Elements of \mathcal{S}_1 are precisely finite unions of points and intervals

The first four axioms make \mathcal{S} a structure. (van den Dries, 1984; Pillay and Steinhorn, 1986; Knight et al., 1986; Pillay and Steinhorn, 1988) noted that adding the fifth axiom rendered the definable sets to be tame. Thus the fifth axiom is what makes \mathcal{S} o-minimal.

Table 1.2.: Semi-algebraic sets vs Definable sets

| semi-algebraic sets | definable sets |
|---|--|
|  |  |

The smallest structure containing all semi-algebraic sets (denoted \mathcal{S}_{sa}) is known to be o-minimal. Note that the Tarski-Seidenberg theorem is required to prove that this is indeed the case. The interesting point is that this is not the only one. There are now many more structures which have been proved to be o-minimal. - e.g. smallest structure containing sets defined by the exponential function \mathcal{S}_{\exp} (Wilkie, 1996; Khovanskiĭ, 1991), restricted analytic functions \mathcal{S}_{an} (Van den Dries, 1986), Pfaffian functions $\mathcal{S}_{\mathcal{F}}$ (Wilkie, 1999), etc. Two resources to learn more about o-minimal structures are (van den Dries, 1998; Coste, 2000a).

Once again, we stress that while sets definable over any arbitrary o-minimal over \mathbb{R} naturally include sets far more general (for one, transcendental functions, albeit with restrictions sometimes, are allowed in defining definable sets) than semi-algebraic sets, they often share some of the 'topological tameness' properties that semi-algebraic sets share, thus making their study feasible. O-minimality is an extremely active topic of research. In pure mathematics, there have been striking applications of o-minimality in diophantine geometry, for e.g. in the resolution of the Andre-Oort conjecture (Pila, 2011). Besides pure mathematics, it is also important in applied mathematical areas. For instance, in neural networks, the activations functions are transcendental, so concepts represented by neural networks are definable, but not semi-algebraic (Tressl, 2010). Note however that while the ambit of o-minimal geometry is certainly bigger than that of semi-algebraic geometry, examples such as the topologist's sine curve, the cantor set, etc. are not admissible.

1.3 Random Algebraic Geometry

Gauss' fundamental theorem of algebra states that a complex polynomial of degree d has exactly d complex zeros (counted with multiplicity). However, when we consider a real polynomial of degree d , it can have any of $0, 2, \dots, 2\lfloor d/2 \rfloor$ number of real zeros. Given a quadratic polynomial $ax^2 + bx + c$, there is a test to check how many real zeros it has: if $b^2 - 4ac < 0$, it has no real zeros, and it has two (counted with multiplicity) otherwise. However, for a number of such enumerative questions, there often aren't algebraic tests of the above form. An alternative perspective here is to consider the question - how many real zeros does a polynomial of degree *usually* have?

As suggested by the term 'usually', instead of understanding the deterministic picture, random algebraic geometry aims to understand the problem from a statistical perspective. Specifically, by applying a probability distribution, we would like to study the statistical properties of polynomials. This approach has had a long history beginning with the works (Kac, 1943; Kac, 1949; Littlewood and Offord, 1938) where they considered random polynomials with standard Gaussian coefficients. The seminal paper (Edelman and Kostlan, 1995) studied the same question but in different settings, and more importantly introduced, what is commonly called the *Edelman-Kostlan* measure, or just Kostlan measure for short.

Definition 1.3.1 *The Edelman-Kostlan measure on $\mathbb{R}[X_0, \dots, X_n](d)$, i.e. the space of homogeneous polynomials of degree d in $n+1$ variables, is defined by choosing each coefficient of*

$$P = \sum_{|\alpha|=d} \xi_\alpha x_0^{\alpha_0} \cdots x_n^{\alpha_n}$$

independently from a centered Gaussian distribution, where,

$$\xi_\alpha \sim \mathcal{N}\left(0, \frac{d!}{\alpha_0! \cdots \alpha_n!}\right).$$

This measure is the restriction of the Fubini-Study measure to the space of real polynomials. The variances of the gaussian random variables $\frac{d!}{\alpha_0! \cdots \alpha_n!}$ are chosen in

such a way that the resulting probability distribution is invariant under orthogonal change of variables (there are no preferred points or direction in $\mathbb{R}P^n$, where zeroes of P are naturally defined).

Let us see some partial justification of this by looking at the two variable degree two case. Consider the Kostlan form with indeterminants (X_0, X_1) .

$$P(X_0, X_1) = \mathcal{N}(0, 1) X_0^2 + \mathcal{N}(0, 2) X_0 X_1 + \mathcal{N}(0, 1) X_1^2.$$

It is well know that finite subgroups of $O(2, \mathbb{R})$ are either C_n , the cyclic group of order n , or D_n , the dihedral group of order $2n$.

Case 1: Let $\begin{pmatrix} Y_0 \\ Y_1 \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} X_0 \\ X_1 \end{pmatrix}$. We have

$$\begin{aligned} P(Y_0, Y_1) &= \mathcal{N}(0, 1) Y_0^2 + \mathcal{N}(0, 2) Y_0 Y_1 + \mathcal{N}(0, 1) Y_1^2 \\ &= \mathcal{N}(0, 1) (X_0 \cos \theta - X_1 \sin \theta)^2 + \mathcal{N}(0, 1) (X_0 \sin \theta + X_1 \cos \theta)^2 \\ &\quad + \mathcal{N}(0, 2) (X_0 \cos \theta - X_1 \sin \theta)(X_0 \sin \theta + X_1 \cos \theta) \\ &= \mathcal{N}(0, (X_0 \cos \theta - X_1 \sin \theta)^4 + (X_0 \sin \theta + X_1 \cos \theta)^4) \\ &\quad + \mathcal{N}(0, 2(X_0 \cos \theta - X_1 \sin \theta)^2(X_0 \sin \theta + X_1 \cos \theta)^2) \\ &= \mathcal{N}(0, X_0^4(\cos^4 \theta + \sin^4 \theta + 2 \sin^2 \theta \cos^2 \theta)) \\ &\quad + \mathcal{N}(0, X_1^4(\cos^4 \theta + \sin^4 \theta + 2 \sin^2 \theta \cos^2 \theta)) \\ &\quad + \mathcal{N}(0, 4X_0^3 X_1(-\cos^3 \theta \sin \theta + \sin^3 \theta \cos \theta + \cos^3 \theta \sin \theta - \cos \theta \sin^3 \theta)) \\ &\quad + \mathcal{N}(0, 4X_0 X_1^3(-\cos \theta \sin^3 \theta + \sin \theta \cos^3 \theta + \cos \theta \sin^3 \theta - \cos^3 \theta \sin \theta)) \\ &\quad + \mathcal{N}(0, 2X_0^2 X_1^2(6 \cos^2 \theta \sin^2 \theta + \cos^4 \theta + \sin^4 \theta - 4 \cos^2 \theta \sin^2 \theta)) \\ &= \mathcal{N}(0, X_0^4) + \mathcal{N}(0, X_1^4) + \mathcal{N}(0, 2X_0^2 X_1^2) \\ &= P(X_0, X_1). \end{aligned}$$

Case 2: The case of a reflection is left as an exercise.

We see above that the distribution is invariant under transformations from finite subgroups of $O(2, \mathbb{R})$. This generalizes, i.e. for a homogenous Kostlan form P in n variables,

$$P \equiv_{\text{equiv.}} LP \quad \text{for any } L \in O(n, \mathbb{R}).$$

Additionally, if we consider zeros in projective space, where the zeros of homogenous polynomials are naturally defined, we can say that *no points or directions are preferred* in projective space. Moreover, if we extend this probability distribution to the whole space of complex polynomials, by replacing *real* with *complex* Gaussian variables, it can be shown that this extension is the *unique* Gaussian measure which is invariant under unitary change of variables. This makes real Kostlan polynomials a natural object of study.

This model for random polynomials received a lot of attention since pioneer works of Edelman, Kostlan, Shub and Smale (Edelman and Kostlan, 1995; Shub and Smale, 1993b; Edelman et al., 1994; Kostlan, 2002; Shub and Smale, 1993c; Shub and Smale, 1993a) on random polynomial systems solving. A nice recent textbook is (Breiding and Lerario, 2019).

1.3.1 Some basic results in random algebraic geometry

We shall now briefly review some results about random polynomials. The first question that was considered was the average number of real zeros of univariate polynomials with standard Gaussian co-efficients.

Theorem 1.3.1 ((Kac, 1943)) *Let $Z_P(d)$ be a random variable that denotes the number of real zeros of the random polynomial $P \in \mathbb{R}[X]_d$ defined as*

$$P = \mathcal{N}(0, 1) X^d + \mathcal{N}(0, 1) X^{d-1} + \dots + \mathcal{N}(0, 1).$$

Then

$$\lim_{d \rightarrow \infty} \mathbb{E}[Z_P(d)] = \frac{2}{\pi} \log d.$$

For Edelman-Kostlan forms, we have the following result.

Theorem 1.3.2 ((Kostlan, 1993; Shub and Smale, 1993b)) *Let $Z_P(d)$ be a random variable that denotes the number of real zeros of the random polynomial $P \in \mathbb{R}[X]_d$ defined as*

$$P = \mathcal{N}\left(0, \binom{d}{d}\right) X^d + \mathcal{N}\left(0, \binom{d}{d-1}\right) X^{d-1} + \dots + \mathcal{N}\left(0, \binom{d}{1}\right) X + \mathcal{N}\left(0, \binom{d}{0}\right).$$

Then

$$\mathbb{E}[Z_P(d)] = \sqrt{d}.$$

Let us now consider the case of polynomials in several variables. Obviously, for $P \in \mathbb{R}[X_1, \dots, X_n]_d$, then set of real zeros is no longer finite. In fact, it is now a real algebraic hypersurface. This hypersurface is not compact in general, however, there is a standard compactification. By taking the isomorphism

$$X_1^{\alpha_1} \dots X_n^{\alpha_n} \mapsto X_0^{d - \sum_{i=1}^n \alpha_i} X_1^{\alpha_1} \dots X_n^{\alpha_n},$$

we can instead just consider homogenous polynomials. In other words, we can consider polynomials in $\mathbb{R}[X_1, \dots, X_n]_{(d)}$ and look at zeros in real projective space. This zero set will be compact and smooth for generic polynomials, and more importantly, will contain the previous zero set as a dense subset. We can now consider the expected Betti numbers of these projective hypersurfaces.

Theorem 1.3.3 ((Gayet and Welschinger, 2016)) *Let $H_p(d)$ denote a Kostlan hypersurface which is the set of real zeros of a degree d homogenous Kostlan form in $n + 1$ variables. Then there exist universal constants a, b such that*

$$a \leq \lim_{d \rightarrow \infty} \frac{\mathbb{E}[b_i(H_p(d), \mathbb{Z}/2\mathbb{Z})]}{\text{Vol}(\mathbb{R}P^n) \sqrt{d^n}} \leq b,$$

where $\text{Vol}(\mathbb{R}P^n)$ denotes the total volume of the real projective space for the Fubini-Study metric.

More information about the topology of random hypersurfaces can be found in the survey (Welschinger, 2015).

2 TOPOLOGICAL COMPLEXITY OF SEMI-ALGEBRAIC AND DEFINABLE SETS

Colloquially speaking, topology studies properties of sets which are invariant under continuous transformations (stretching, bending, but not tearing). At a very high level, topology asks the following question – given two objects, is there a continuous transformation that transforms one into the other?

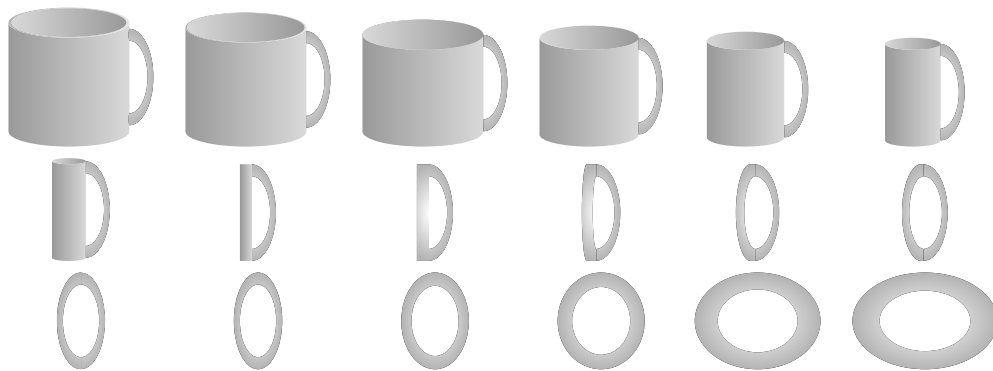


Figure 2.1.: “...a topologist cannot differentiate between a coffee mug and a donut because they are homotopy equivalent...” – illustration of smooth transformation from a coffee mug to a donut



Figure 2.2.: Just to belabor the point, these two koDubaLes are the same to me

2.1 Betti Numbers

There is a long history of research on topological complexity of sets arising in semi-algebraic geometry and o-minimal geometry. An important measure of the topological complexity are the Betti numbers. They have been studied for pure mathematical interest as well as for effecting fundamental advances in real algebraic geometry, discrete geometry, statistical learning theory, convex optimization, complexity theory, as well as applied areas such as robot motion planning, computer graphics. The reader is referred to surveys such as (Gabrielov and Vorobjov, 2004; Basu et al., 2005a; Basu, 2017), as well as the definitive book (Basu et al., 2006), and references therein, for an overview.

2.1.1 Betti numbers of semi-algebraic sets

The i -th Betti number of a semi-algebraic set S defined over \mathbb{R} , denoted $b_i(S)$, is the rank of the singular (co)homology group of S with integer coefficients, i.e. the rank of $H^i(S, \mathbb{Z})$. Informally, the i -th Betti number measures the number of i -dimensional holes in S . Specifically, $b_0(\cdot)$ measures the number of connected components, $b_1(\cdot)$ measures the number of one-dimensional/circular holes, $b_2(\cdot)$ measures the number of two-dimensional voids/cavities, etc. Table 2.1 shows some semi-algebraic sets and their Betti numbers.



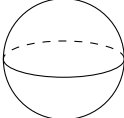
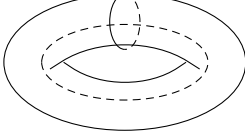
Given a semi-algebraic set $S \subset \mathbb{R}^n$, defined by at most m equations, each of degree at most d , a prototypical topological question is to bound the Betti numbers of S in terms of m, d, n . The first results along this line were obtained by (Oleinik and Petrovsky, 1949), and later by (Thom, 1965) and (Milnor, 1964).

Theorem 2.1.1 ((Oleinik and Petrovsky, 1949; Thom, 1965; Milnor, 1964))

Let $S \subseteq \mathbb{R}^n$ be defined by the conjunction of s inequalities,

$$P_1 \geq 0, \dots, P_s \geq 0, \quad P_i \in \mathbb{R}[X_1, \dots, X_n],$$

Table 2.1.: Betti numbers of semi-algebraic sets – examples

| Object | b_0 | b_1 | b_2 | $b_{i \geq 3}$ |
|---|-------|-------|-------|----------------|
| \cdot | 1 | 0 | 0 | 0 |
| $\dots\dots$ | 5 | 0 | 0 | 0 |
|  | 1 | 1 | 0 | 0 |
|  | 1 | 0 | 0 | 0 |
|  | 1 | 0 | 1 | 0 |
|  | 1 | 2 | 1 | 0 |

where $\deg(P_i) \leq d$ for all $1 \leq i \leq s$. Then

$$\sum_{i \geq 0} b_i(S) = O(sd)^n$$

This has been generalized to other types of semi-algebraic sets in several different ways, for e.g. (Basu et al., 2005b; Basu et al., 1996; Barone and Basu, 2012; Basu and Rizzie, 2018). Once again, (Basu et al., 2005a; Basu, 2017) are good surveys on this topic.

2.2 Applications of Bounds on Betti Numbers

As mentioned earlier, Betti numbers quantify the *topological complexity*. Heuristically, larger the Betti numbers, more *complex* an object is. This is underscored in one of the first applications of the bounds of the form in Theorem 2.1.1 in proving lower bounds in theoretical computer science. Specifically, it is in regard to proving lower bounds on the height of algebraic computation trees.

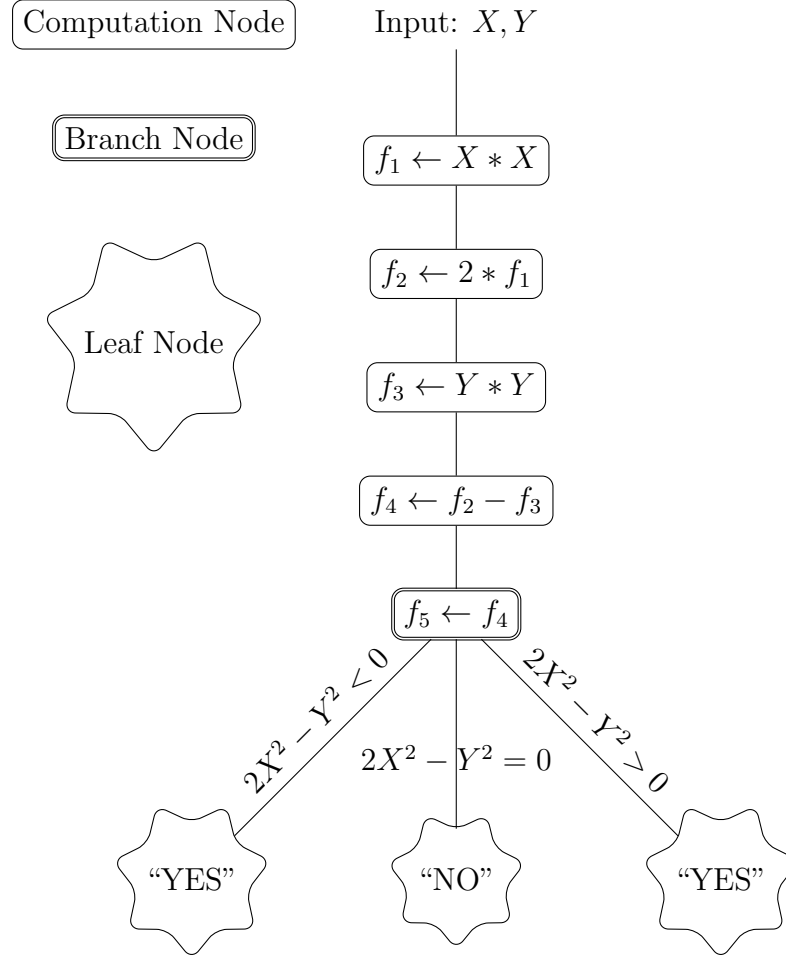


Figure 2.3.: An example of an algebraic computation tree for testing membership in $\{(X, Y) \in \mathbb{R}^2 \mid 2X^2 - Y^2 \neq 0\}$

An algebraic computation tree is a computational model that represents the steps a Turing machine might take. An example is depicted in Figure 2.3. Consider the following problem: given an input point $x \in \mathbb{R}^n$, determine if $x \in S \subseteq \mathbb{R}^n$, where S is semi-algebraic. The algebraic computation tree for this problem will be such that on the input x , the tree accepts x if and only if the computation terminates at a leaf node that is an accepting node, and $x \in S$ if and only if the tree accepts x .

In (Ben-Or, 1983) it was proved that the depth of an algebraic computation tree testing membership in S must be $\Omega(\log b_0(S))$. Subsequently, this result was extended in (Yao, 1997) where a lower bound was given based on any Betti number, not just

the 0-th. However, Yao’s result was in terms of the Borel-Moore Betti numbers. A survey of the results along this direction is available in (Bürgisser and Cucker, 2004). Relatively recently, the following bound was proved in (Gabrielov and Vorobjov, 2017), which proved a bound based on the individual singular betti numbers.

Theorem 2.2.1 ((Gabrielov and Vorobjov, 2017)) *The height of any algebraic computation tree for deciding membership in a semi-algebraic set $S \subseteq \mathbb{R}^n$ is bounded from below by*

$$\frac{c_1 \log b_i(S)}{i+1} - c_2 n,$$

where b_i is the i -th Betti number w.r.t. singular homology, and c_1, c_2 are some positive constants.

The intuition behind these types of bounds is that if S is topologically complicated, then the algebraic computation tree working with S must have larger depth.

In addition to lower bounds in computational complexity theory, bounds on Betti numbers have historically had applications in a number of other areas as well (Goodman and Pollack, 1986a; Goodman and Pollack, 1986b). Since about a decade ago, these bounds have been crucial in a tremendous number of problems in discrete geometry.

2.2.1 Discrete geometry applications

The seminal paper (Guth and Katz, 2015) introduced algebraic geometry techniques to solve two fundamental open questions in discrete geometry – the distinct distances problem proposed by Erdős, and the joints problem proposed by Bernard Chazelle. One of their techniques, called the polynomial partitioning technique, has been very influential. Below, we shall state a generalization of the polynomial partitioning technique, proved in (Guth, 2015).

Theorem 2.2.2 ((Guth and Katz, 2015; Guth, 2015)) *Let Γ be a finite set of k -dimensional varieties in \mathbb{R}^n , each defined by at most m polynomial equations, of*

degree at most d . For any $D \geq 1$, there is a non-zero polynomial P of degree at most D , so that for each connected component \mathcal{C} of $\mathbb{R}^n \setminus Z(P)$,

$$|\Gamma \cap \mathcal{C}| \leq C_{d,m,n} \frac{|\Gamma|}{D^{n-k}}.$$

At a high level, the polynomial partitioning technique (see Figure 2.4 for an example illustration) really gives us a divide and conquer technique – it allows you to break your space into pieces and solve a problem on each piece, and then put together the local solutions to get the global answer. While there already were older partitioning techniques in discrete geometry called cuttings and simplicial partitioning, polynomial partitioning is simpler and more powerful in higher dimensions. The polynomial partitioning technique has been a panacea for a huge number of problems in discrete geometry (Guth and Katz, 2015; Kaplan et al., 2012a; Solymosi and Tao, 2012; Kaplan et al., 2012b), and continues to be at the core of very recent fundamental advances (Aronov et al., 2019; Agarwal et al., 2019). To interpret Theorem 2.2.2, we need the following theorem proved in (Barone and Basu, 2016).

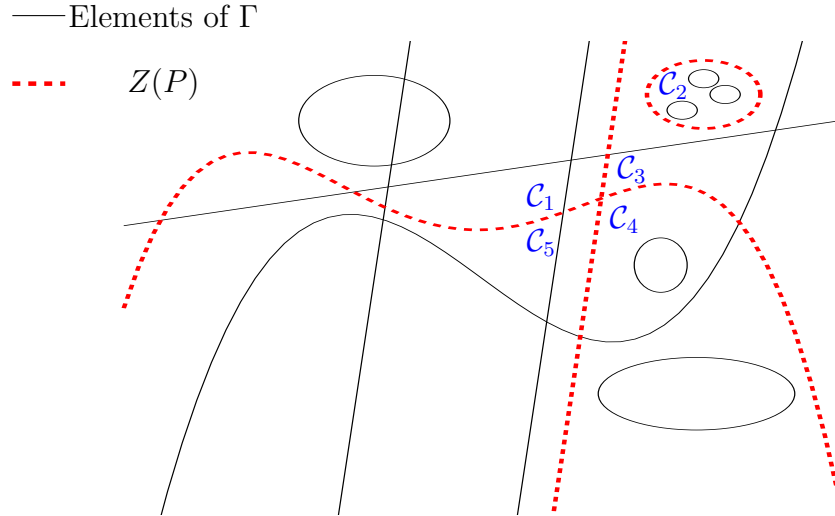


Figure 2.4.: Example illustration of polynomial partitioning. The zero set of P breaks \mathbb{R}^2 into five connected components – $\mathcal{C}_1, \dots, \mathcal{C}_5$. Each \mathcal{C}_i is intersected by a subset of varieties in Γ .

Theorem 2.2.3 ((Barone and Basu, 2016)) *Let $P_1, \dots, P_s \in \mathbb{R}[X_1, \dots, X_n]$ such that for all $1 \leq i \leq s$, $\deg(Q_i) \leq d_i$. Let k_i be an upper bound on the real dimension of $Z(\{Q_1, \dots, Q_i\})$ (by convention $k_{i \leq 0} = n$). Suppose that*

$$2 \leq d_1 \leq d_2 \leq \frac{1}{n+1}d_3 \leq \frac{1}{(n+1)^2}d_4 \leq \dots \leq \frac{1}{(n+1)^{s-2}}d_s.$$

Then,

$$b_0(Z(Q_1, \dots, Q_s)) \leq O(1)^s O(n)^{2n} \left(\prod_{1 \leq j \leq s} d_j^{k_{j-1} - k_j} \right) d_s^{k_{s-1}}.$$

A corollary of the above theorem (called a ‘Real-analogue’ of Bezout’s inequality), which was also re-proved using different techniques in (Solymosi and Tao, 2012) is below.

Corollary 2.2.1 ((Barone and Basu, 2016; Solymosi and Tao, 2012)) *Let γ be a k -dimensional real algebraic set in \mathbb{R}^n defined by at most m polynomial equations, each of degree at most d . If P is a polynomial of degree at most D , then γ intersects at most $C_{d,m,n} D^k$ different connected components of $\mathbb{R}^n \setminus Z(P)$.*

We are now in a position to interpret Theorem 2.2.2. We know that given a polynomial P of degree D in \mathbb{R}^n , $\mathbb{R}^n \setminus Z(P)$ can have at most $\sim D^n$ connected components. Also, Corollary 2.2.1 proves that of these D^n connected components, each k -dimensional γ intersects at most D^k of them. The polynomial partitioning theorem says that given *any* finite set Γ of k -dimensional varieties, *there exists* a polynomial such that the variety-connected-component intersections are equidistributed. In other words, each $\gamma \in \Gamma$ can intersect at most D^k connected components of $\mathbb{R}^n \setminus Z(P)$, so there are at most a total of $|\Gamma|D^k$ such intersections possible, and that P ensures that these intersections are equidistributed amongst the D^n connected components of $\mathbb{R}^n \setminus Z(P)$, i.e. there are at most $\frac{|\Gamma|D^k}{D^n}$ varieties intersecting each connected component. While not stated this way, Theorem 2.2.2 quite obviously holds if Γ is a finite set of semi-algebraic sets as well, not just real algebraic varieties.

Thus we see that bounds of the type in Theorem 2.2.3 and Corollary 2.2.1 are a crucial ingredient for polynomial partitioning. Motivated by this, and many other applications in discrete geometry (Matoušek and Patáková, 2015), studying bounds on the Betti numbers of semi-algebraic sets has remained an active field of study.

2.3 Betti numbers of definable sets

Parallel to the thrust to study incidences between algebraic and semi-algebraic sets, incidences between definable sets over arbitrary o-minimal expansions of \mathbb{R} has become an active research area as well, for example (Basu and Raz, 2017a; Chernikov and Starchenko, 2018; Chernikov et al., 2020; Chernikov et al., 2016). The progress along this direction has been significantly slower; each of these results use idiosyncratic techniques which don't really suggest methods of attack for other problems.

One matter that has stymied progress is the unavailability of a polynomial partitioning type result for definable sets. Needless to say, an o-minimal polynomial partitioning theorem would enable progress on a lot of different fronts, and would potentially provide greatly simplified proofs of already proved results. As explained in Section 2.2.1, we need bounds on the Betti numbers of certain kinds of semi-algebraic sets to prove polynomial partitioning theorems. The next chapter answers precisely this question – given a definable set γ , provide bounds on the Betti numbers of γ restricted to the zero sets of polynomials with growing degree in terms of the dimension of γ , the ambient dimension and the degree of the polynomial.

There has also been some previous work on Betti number bounds in o-minimal geometry. For instance, (Basu, 2009) generalizes many quantitative bounds already known for semi-algebraic sets to the case of definable sets. The premise of this work is that when one has a finite set of polynomials each of different degree, the dependence of the quantitative bounds on the cardinality of the set of polynomials is more crucial than the degrees of the polynomials.

3 ZEROS OF POLYNOMIALS ON DEFINABLE HYPERSURFACES

3.1 Introduction

3.1.1 Existence of pathologies

A classical fact from algebraic geometry states that given two real algebraic curves Γ and Z , if their intersection is transversal, it consists of at most $\deg(\Gamma) \cdot \deg(Z)$ many points. In particular, if we fix the first curve, we can say that there is a function $\beta_{\Gamma,0} : \mathbb{N} \rightarrow \mathbb{N}$ such that for every polynomial p of degree d , if Γ and $Z(p) = \{p = 0\}$ intersect transversally, then:

$$\#(\Gamma \cap Z(p)) \leq \beta_{\Gamma,0}(d) = \deg(\Gamma) \cdot d. \quad (3.1)$$

If we leave the *semialgebraic* world, but still remain in the *definable* setting, still such a function $\beta_{\Gamma,0}$ exists, but in general nothing can be said about its behavior. Here by definable we mean the class of definable sets in an o-minimal expansion of the real numbers, for example the o-minimal structure generated by semianalytic functions. (We refer the reader who is unfamiliar with o-minimal geometry to (van den Dries, 1998; Coste, 2000a) for easy to read introductions to the topic.)

In this direction Gwoździewicz, Kurdyka and Parusiński (Gwoździewicz et al., 1999) have proved that for every sequence $\{a_d \geq 0\}_{d \in \mathbb{N}}$ of natural numbers there exists a definable curve Γ , a subsequence $\{a_{d_m}\}_{m \in \mathbb{N}}$ and a sequence $\{p_m\}_{m \in \mathbb{N}}$ of polynomials of degree $\deg(p_m) = d_m$ such that:

$$\#(\Gamma \cap Z(p_m)) \geq a_{d_m}.$$

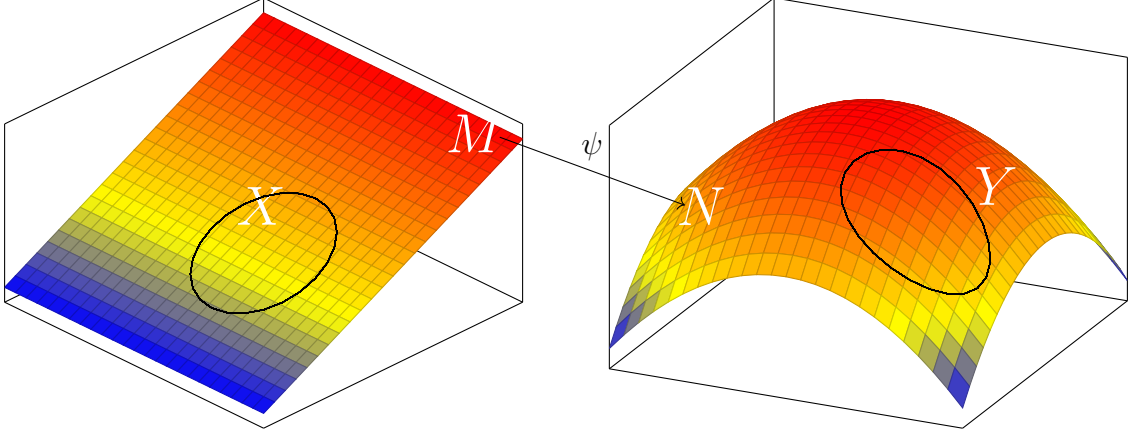


Figure 3.1.: Illustration of an ambient diffeotopy. The manifold pairs (M, X) are ambient diffeotopic to (N, Y) .

(In this paper we will show that the curve $\Gamma \subset \mathbb{R}P^2$ can be taken to be regular, definable and compact and that the polynomials p_{d_m} can be chosen in such a way that the intersection $\Gamma \cap Z(p_{d_m})$ is transversal, i.e. stable under small perturbations of the polynomial.)

In particular this shows that, for a fixed definable $\Gamma \subset \mathbb{R}P^2$, there is in general no upper bound on the number of zeroes of a polynomial p on Γ which is polynomial in $\deg(p)$. Generalizing this we will show that in higher dimensions the situation is even more interesting.

To state our first result, we will say that two manifold pairs (M, X) and (N, Y) are ambient-diffeotopic if there exists a diffeomorphism $\psi : M \rightarrow N$ such that $\psi(X) = Y$; in this case we write $(M, X) \sim (N, Y)$. This notion essentially says that X and Y are diffeomorphic and, up to a diffeomorphism, they are embedded in their ambient spaces in the same way. See Figure 3.1 for an illustration.

Of course, when Γ is an algebraic hypersurface and p is a polynomial, there are restrictions on the possible pairs $(\Gamma, Z(p) \cap \Gamma)$ (for example Betti numbers of $Z(p) \cap \Gamma$ grow at most as a polynomial in $\deg(p)$). Pick now a sequence of smooth and compact hypersurfaces $Z_1, Z_2, \dots \subset \mathbb{R}^{n-1}$. Our first Theorem says that (up to extracting subsequences) there exists a regular definable hypersurface $\Gamma \subset \mathbb{R}P^n$ such that each

manifold Z_d is diffeomorphic to a component of the zero set on Γ of some polynomial of degree d . Here (and in the rest of the paper) Γ will be *semianalytic* in $\mathbb{R}P^n$. More precisely, we will prove the following.

Theorem 3.1.1 (Existence of pathologies) *Let $\{Z_d\}_{d \in \mathbb{N}}$ be a sequence of smooth, compact hypersurfaces embedded in \mathbb{R}^{n-1} . There exist a regular¹, compact, semianalytic hypersurface $\Gamma \subset \mathbb{R}P^n$, a disk $D \subset \Gamma$ and a sequence $\{p_m\}_{m \in \mathbb{N}}$ of homogeneous polynomials of degree $\deg(p_m) = d_m$ such that the intersection $Z(p_m) \cap \Gamma$ is transversal and:*

$$(D, Z(p_m) \cap D) \sim (\mathbb{R}^{n-1}, Z_{d_m}) \quad \text{for all } m \in \mathbb{N}.$$

Remark 3.1.1 *Note that in the case $n = 2$ this implies the statement of (Gwoździwicz et al., 1999). In fact, we can take for $Z_d = \{x_1, \dots, x_{a_d}\} \subset \mathbb{R}$ a set consisting of a_d many points. Then we find a smooth definable curve $\Gamma \subset \mathbb{R}P^2$, an interval $I \subset \Gamma$ and a sequence of polynomials p_m of degree d_m such that the manifold pairs $(I, Z(p_m) \cap I)$ and $(\mathbb{R}, \{x_1, \dots, x_{a_{d_m}}\})$ are diffeomorphic, in particular $Z(p_m) \cap \Gamma$ consists of at least a_{d_m} many points.*

In higher dimensions we can measure the complexity of a manifold by its Betti numbers. If $\Gamma \subset \mathbb{R}P^n$ is a regular, compact, definable hypersurface, for every $0 \leq k \leq n-2$ let $\beta_{\Gamma,k} : \mathbb{N} \rightarrow \mathbb{N}$ be the function:

$$\beta_{\Gamma,k}(d) = \max_{\deg(p)=d} b_k(\Gamma \cap Z(p))$$

(here b_k denotes the k -th Betti number). When Γ is semialgebraic, we have

$$\beta_{\Gamma,k}(d) \leq c_{\Gamma} \cdot d^{n-1} \quad (\text{semialgebraic case}) \quad (3.2)$$

for some constant depending on Γ (this estimate actually requires some nontrivial work if Γ is singular, and it is proved in (Basu and Rizzie, 2018, Theorem 6.4)). On

¹Throughout the paper the word “regular” will mean “of regularity class C^k for some fixed $k \geq 2$ ”.

the other hand, as for the case of curves, there is no way to control the behavior of this function for a general definable Γ : in fact, given a sequence $\{a_d\}_{d \in \mathbb{N}}$, if we chose a sequence of hypersurfaces $\{Z_d\}$ with $b_k(Z_d) \geq a_d$, for the hypersurface Γ provided by Theorem 3.1.1 the function $\beta_{\Gamma,k}$ grows at least as fast as a_{d_m} .

Remark 3.1.2 *Estimates like (3.1) are basic building blocks in recent advances in incidence problems in the area of discrete geometry driven by the polynomial partitioning method (Guth and Katz, 2015) (see for example (Solymosi and Tao, 2012, Theorem A.2)). Recently, using different techniques such incidence results have been generalized from the semi-algebraic case to more general situations – namely, incidences between definable sets over arbitrary o-minimal expansions of \mathbb{R} , see (Basu and Raz, 2017a; Chernikov et al., 2020). In order to extend the polynomial partitioning technique to the o-minimal situation (as noted in (Basu and Raz, 2017b)) it is important to study the function $\beta_{\Gamma,k}$ where Γ is now an arbitrary definable hypersurface in an o-minimal structure (rather than just semi-algebraic). On one hand Theorem 3.1.1 seems to rule out the use of polynomial partitioning for incidence problems involving definable sets in arbitrary o-minimal structures, but on the other hand we also prove (see Theorem 3.1.2 below) that the pathological behavior exhibited in Theorem 3.1.1 is very rare, and this gives some hope that a modified version of the technique can still be applicable to incidence questions.*

3.1.2 Pathologies are rare

Given Γ , it is natural to ask how “stable” are the polynomials having the “pathological” behaviour of Theorem 3.1.1? In other words, if it is certainly true that nothing can be said on the function $\beta_{\Gamma,k}$ that bounds the Betti numbers of transversal intersection between a definable hypersurface Γ and the zero set of a polynomial in terms of the degree of the polynomial, is it possible to say that for “most polynomials” a polynomial upper bound still holds true for the Betti numbers? Our second result gives

an affirmative answer to this question, after the naive idea of “most polynomials” is made precise.

To make these questions precise, on the space $W_{n,d}$ of homogeneous polynomials of degree d in $n + 1$ variables we introduce a natural Gaussian measure, called the *Kostlan* measure, defined by choosing each coefficient of

$$p = \sum_{|\alpha|=d} \xi_\alpha \binom{d}{\alpha}^{1/2} x_0^{\alpha_0} \cdots x_n^{\alpha_n}$$

independently from a standard Gaussian distribution (i.e. $\xi_\alpha \sim N(0, 1)$). This measure is the restriction, to the space of real polynomials, of the Fubini-Study measure.

The scaling coefficients $\binom{d}{\alpha}^{1/2}$ are chosen in such a way that the resulting probability distribution is invariant under orthogonal change of variables (there are no preferred points or direction in $\mathbb{R}P^n$, where zeroes of p are naturally defined). Moreover, if we extend this probability distribution to the whole space of complex polynomials, by replacing *real* with *complex* Gaussian variables, it can be shown that this extension is the *unique* Gaussian measure which is invariant under unitary change of variables. This makes real Kostlan polynomials a natural object of study. (This model for random polynomials received a lot of attention since pioneer works of Edelman, Kostlan, Shub and Smale (Edelman and Kostlan, 1995; Shub and Smale, 1993b; Edelman et al., 1994; Kostlan, 2002; Shub and Smale, 1993c; Shub and Smale, 1993a) on random polynomial systems solving.)

The next Theorem estimates the size of the set of polynomials whose restriction to a definable hypersurface $\Gamma \subset \mathbb{R}P^n$ have a behaviour that deviates from the semi-algebraic case estimate (3.2). This result could be potentially useful in the study of incidence questions over o-minimal structures (cf. Remark 3.1.2). We observe, however, that for this theorem we do not need Γ to be definable (in fact it is enough that it is regular).

Theorem 3.1.2 *Let $\Gamma \subset \mathbb{RP}^n$ be a regular and compact hypersurface, and let p be a random Kostlan polynomial of degree d . Then there exists a constant c_Γ such that for every $0 \leq k \leq n-2$ and for every $t > 0$*

$$\mathbb{P} [b_k(\Gamma \cap Z(p)) \geq td^{n-1}] \leq \frac{c_\Gamma}{td^{\frac{n-1}{2}}}.$$

Combining this result with the construction of Theorem 3.1.1 we obtain the following estimate for the Gaussian volume of the set of “pathological” polynomials. The lower bounds follows from the fact that the intersection $Z(p_m) \cap \Gamma$ produced in Theorem 3.1.1 is transversal (hence stable under small perturbations of the polynomial p_m).

Corollary 3.1.1 (Pathologies are rare) *Given a sequence of natural numbers, i.e. $\{a_d\}_{d \in \mathbb{N}}$, let $\{Z_d\}_{d \in \mathbb{N}}$ be a sequence of hypersurfaces with $b_k(Z_d) \geq a_d$ for all $d \in \mathbb{N}$. Consider the hypersurface $\Gamma \subset \mathbb{RP}^n$ provided by Theorem 3.1.1. Then, for some constant $c_\Gamma > 0$:*

$$0 < \mathbb{P} [b_k(\Gamma \cap Z(p)) \geq a_{d_m}] \leq \frac{c_\Gamma d_m^{\frac{n-1}{2}}}{a_{d_m}}.$$

Remark 3.1.3 *Markov’s inequality gives an upper bound on the probability that a non-negative random variable takes values in the tail. Specifically, for a non-negative random variable X , we have that $\mathbb{P}[X > a] \leq \frac{\mathbb{E}X}{a}$. The conclusion of Theorem 3.1.2 follows after combining Markov’s inequality with the following fact (proved in Proposition 3.3.2): there exists a universal constant $c_{k,n} > 0$ such that for every $\Gamma \subset \mathbb{RP}^n$ regular, definable, compact hypersurface*

$$\mathbb{E}b_k(\Gamma \cap Z(p)) \leq |\Gamma|c_{k,n}d^{\frac{n-1}{2}} + O(d^{\frac{n-2}{2}}), \quad (3.3)$$

where $|\Gamma|$ denotes the volume of Γ , induced by restricting the Riemannian metric of \mathbb{RP}^n , and the implied constants in the $O(d^{\frac{n-2}{2}})$ depends on Γ .

Remark 3.1.4 *The content of (3.3) reveals an interesting and surprising property of the space of polynomials: by Theorem 3.1.1 there is a priori no upper bound on the*

homological complexity of $\Gamma \cap Z(p)$ (as a function of $d = \deg(p)$), but on average we cannot exceed a polynomial bound. Here is an example from (Khazhgali Kozhasov, 2017) of a similar phenomenon that appears in the study of random enumerative geometry. If X_1, \dots, X_4 are boundaries of smooth convex bodies in \mathbb{RP}^3 , one can ask for the number $\ell(X_1, \dots, X_4)$ of lines that are simultaneously tangent to all of them. This number is finite if the convex bodies are in general position in the projective space, but it can be arbitrarily large: for every $m > 0$ one can find $X_1, \dots, X_4 \subset \mathbb{RP}^3$ in general position such that there are at least m lines tangent to all of them. On the other hand (here is the surprising thing) there exists a constant $c > 0$, independent of the convex bodies, such that if we now average over all their possible configurations using the action of the orthogonal group $O(4)$ on \mathbb{RP}^3 , we get $\mathbb{E}_{g_1, \dots, g_4 \in O(4)} \ell(g_1 X_1, \dots, g_4 X_4) = c$. Here again there is no a priori upper bound, but there is an upper bound on average.

Remark 3.1.5 (The zero-dimensional case) Another case of interest, on which we can say more, is the case when $\Gamma \subset \mathbb{RP}^n$ is k -dimensional and we consider the common zero set of k polynomials on it. In this case we do not have to restrict to Kostlan polynomials and we can work with the more general class of random invariant polynomials: these are centered Gaussian probability measure on $W_{n,d}$ which are invariant under the action of the orthogonal group by change of variables (of course the Kostlan measure is one of them). These measures have been classified by Kostlan (Kostlan, 1993) and depend on $\lfloor \frac{d}{2} \rfloor$ many parameters. Consider now the common zero set X of independent random invariant polynomials p_1, \dots, p_k on Γ :

$$X = \Gamma \cap Z(p_1) \cap \dots \cap Z(p_k).$$

With probability one X is zero-dimensional and we can use integral geometry (see (Howard, 1993) or the appendix of (Burgisser and Lerario, 2018)) to deduce that:

$$\mathbb{E} \# (\Gamma \cap Z(p_1) \cap \dots \cap Z(p_k)) = \frac{|\Gamma|}{|\mathbb{RP}^k|} \prod_{j=1}^k \mathbb{E} \frac{|Z(p_j)|}{|\mathbb{RP}^{n-1}|}. \quad (3.4)$$

The quantity $\mathbb{E}|Z(p)|$ appearing in (3.4) can be evaluated using the definition of the invariant distribution in terms of its weights (see (Kostlan, 1993; Fyodorov et al., 2015)); when p is a Kostlan polynomial of degree d , then $\mathbb{E}|Z(p)| = \sqrt{d}|\mathbb{R}P^{n-1}|$. More generally (again by Integral Geometry) this expectation is bounded by $\mathbb{E}|Z(p)| \leq d|\mathbb{R}P^{n-1}|$. If each p_i has now degree d , we can apply Markov's inequality again and deduce that there exists $c_\Gamma > 0$ such that for any invariant Gaussian measure on the space of polynomials:

$$\mathbb{P}\{\#(\Gamma \cap Z(p_1) \cap \dots \cap Z(p_k)) \geq td^{k-1}\} \leq \frac{c_\Gamma}{t},$$

i.e., the probability of deviating from a Bézout-type bound is small.

3.2 Pathological examples: Proof of Theorem 3.1.1

3.2.1 Construction of Gwoździewicz et al.

Theorem 3.1.1 is a generalization of a result proved in (Gwoździewicz et al., 1999). Below we state the theorem and describe the proof.

Theorem 3.2.1 *For analytic $f : (a, \infty) \rightarrow \mathbb{R}$, let $A(d)$ denote the number of isolated solutions to the system $P(x, y) = 0, y = f(x), x > a$. If we are given a sequence $\mathbb{N} \ni d \rightarrow a(d) \in \mathbb{N}$, then there exists an analytic function $f : (a, \infty) \rightarrow \mathbb{R}$, subanalytic at infinity, and an increasing sequence $k \rightarrow d_k$ of integers such that*

$$a(d_k) \leq A(d_k),$$

for all $k \in \mathbb{N}$.

Proof One can easily construct by induction: a sequence $b_k \in \mathbb{N}$, two sequences $\varepsilon_k > 0, \eta_k > 0$, and a sequence of Polynomials

$$P_k = c_{1+b_k}t^{1+b_k} + \dots + c_{b_{k+1}}t^{b_{k+1}}$$

such that

1. For all $k \in \mathbb{N}$,

$$\|P_k\| \leq \varepsilon_k,$$

where $\|\cdot\|$ is the sum of absolute value of coefficients.

2. if $r : (0, 1) \rightarrow \mathbb{R}$ is continuous, $\sup_{t \in (0, 1)} |r(t)| \leq \eta_k$, then

$$|\{t \in (0, 1) : P_k(t) + r(t) = 0\}| \geq a(4b_k).$$

3. For all $n \in \mathbb{N}$,

$$\sum_{k > n} \varepsilon_k < \eta_n.$$

Below is how we can construct our sequences:

(Step 1) Choose any $\varepsilon_1 > 0, b_1 \in \mathbb{N}$. Initialize $\varepsilon_2 = \varepsilon_3 = \dots = \infty$. Let $i = 1$.

(Step 2) Choose any $b_{i+1} \geq a(4b_i) + 2$, and constants $c_{1+b_i}, \dots, c_{b_{i+1}}$ such that the polynomial $p_1(t) = c'_{1+b_i} t^{1+b_i} + \dots + c'_{b_{i+1}} t^{b_{i+1}}$ has at least $a(4b_i) + 2$ zeros in $(0, 1)$. Scale the co-efficients to form $c_{1+b_i}, \dots, c_{b_{i+1}}$ such that $|c_{1+b_i}| + \dots + |c_{b_{i+1}}| \leq \varepsilon_i$. Let

$$P_i(t) = c_{1+b_i} t^{1+b_i} + \dots + c_{b_{i+1}} t^{b_{i+1}}.$$

(Step 3) Define

$$\nu_i = \inf_{v \in \text{critical-points-in-}(0,1) \text{ of } P_i} P_i(v),$$

and set

$$\eta_i = \min(\nu_i, e^{-i}).$$

The value ν_i can be found by considering the system $\{y - f(x) = 0; f'(x) = 0\}$ and obtaining a lower bound on the absolute value of the non-zero roots of $\text{Res}_t(y - P_i(t), P'_i(t))$.

(Step 4) For each $j \geq 1$, update

$$\varepsilon_{i+j} = \min \left(\varepsilon_{i+j}, \frac{\eta_i}{2^{j+1}} \right).$$

(Step 5) $i = i + 1$. GOTO Step 2.

Now let

$$g(t) = \sum_{k=1}^{\infty} P_k(t).$$

Because $\varepsilon_{i+j} \leq \frac{\eta_i}{2^{j+1}}$, we have that

$$\sum_{j \geq 1} \varepsilon_{i+j} \leq \frac{\eta_i}{2} < \eta_i.$$

Also, we have that

$$\limsup_{n \rightarrow \infty} |c_n|^{1/n} < \lim_{n \rightarrow \infty} \eta_n^{1/n} \leq (e^{-n})^{1/n} < 1,$$

which ensures that the radius of convergence of $g(t)$ is > 1 . Finally put

$$f(x) = g \left(\frac{x}{\sqrt{x^2 + 1}} \right), x > 0,$$

and let

$$q_k(t, y) = y - \sum_{n=1}^{k-1} P_n(t), \quad k > 2.$$

q_k is of degree $\leq b_k$.

Clearly, every $t \in (0, 1)$ has a corresponding y_t such that $q_k(t, y_t) = 0$. $g(t) = P_k(t) + \text{continuous function}$, so there are at least $a(4b_k)$ zeros for $t \in (0, 1)$. Thus we can say that $q_k(t, y)$ has at least $a(4b_k)$ zeros on the graph of $g(t)$, for $t \in (0, 1)$.

$q_k\left(\frac{x}{\sqrt{x^2+1}}, y\right)$ is a rational function that looks like this:

$$y - a_0 - a_1 \frac{x}{\sqrt{x^2+1}} - a_2 \left(\frac{x}{\sqrt{x^2+1}}\right)^2 - a_3 \left(\frac{x}{\sqrt{x^2+1}}\right)^3 - \dots - a_{b_k} \left(\frac{x}{\sqrt{x^2+1}}\right)^{b_k}.$$

Now, multiply through by $(\sqrt{x^2+1})^{b_k}$ to get

$$y(\sqrt{x^2+1})^{b_k} - a_0(\sqrt{x^2+1})^{b_k} - a_1 x(\sqrt{x^2+1})^{b_k-1} - a_2 x^2(\sqrt{x^2+1})^{b_k-2} - \dots - a_{b_k} x^{b_k}.$$

Now, whenever the power of $\sqrt{x^2+1}$ is odd, we will have a leftover $\sqrt{x^2+1}$. To eliminate the square root term, we just collect all the square root terms on one side and square. Thus the degree of the polynomial will be the degree of the monomial $(y(\sqrt{x^2+1})^{b_k})^2$ which is $2 + 2b_k \leq 4b_k$. Thus, it is easy to find a polynomial $Q_k(x, y)$ of degree $\leq d_k = 4b_k$ which vanishes on the zeros of $q_k\left(\frac{x}{\sqrt{x^2+1}}, y\right)$.

Since Q_k has at least $a(d_k)$ zeros on the graph of f , it follows that $a(d_k) \leq A(d_k)$, as desired. ■

3.2.2 Some basic facts

For the next proof we will need a few elementary facts from differential topology and real algebraic geometry. First, if $D \subset \mathbb{R}^{n-1}$ is a disk and $f : \overline{D} \rightarrow \mathbb{R}$ is a regular function, we define:

$$\|f\|_{C^1(D, \mathbb{R})} = \sup_{z \in D} \|f(z)\| + \sup_{z \in D} \|\nabla f(z)\|.$$

If “zero” is a regular value of f , then $Z(f)$ is a regular hypersurface in D . If $Z \subset \mathbb{R}^{n-1}$ is a regular compact hypersurface we will write

$$(D, Z(f)) \sim (\mathbb{R}^{n-1}, Z)$$

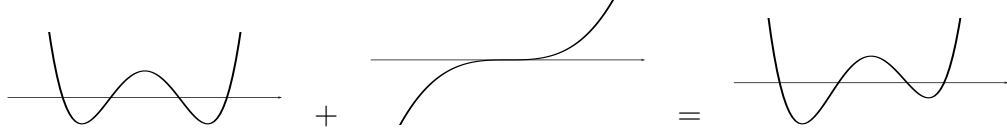


Figure 3.2.: An illustration of the Thom isotopy lemma – if the perturbation is *small*, then the zero set is topologically the same

to denote that the two pairs $(D, Z(f))$ and (\mathbb{R}^{n-1}, Z) are diffeomorphic. In this setting there exists $\delta > 0$ (depending on f) such that given any regular function $h : \overline{D} \rightarrow \mathbb{R}$ with $\|h\|_{C^1(D, \mathbb{R})} \leq \delta$, “zero” is a regular value of $f + h$ and:

$$(D, Z(f + h)) \sim (\mathbb{R}^{n-1}, Z)$$

(in particular the zero sets of f and h are diffeomorphic). We will (loosely) refer to this fact as *Thom’s isotopy Lemma*. Figure 3.2 contains an illustration in the 2-dimensional case.

We will also need the following classical approximation result from real algebraic geometry, due to Seifert (Seifert, 1936). Given a regular, compact hypersurface $Z \subset D \subset \mathbb{R}^{n-1}$, there exists a polynomial $q : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ such that “zero” is a regular value of q and

$$(D, Z(q)) \sim (\mathbb{R}^{n-1}, Z).$$

This follows from Weierstrass’ approximation Theorem; the reader can see (Kollár, 2017, Special case 5) for an elementary proof of Seifert’s result.

Proof [Proof of Theorem 3.1.1] Let $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^{n-1}$ and consider the two disks $D_1 = D(e_1, \frac{1}{2})$ and $D_2 = D(e_1, \frac{2}{3})$.

Pick Z_1 and consider a polynomial² q_2 such that:

$$(D_1, Z(q_2) \cap D_1) \sim (\mathbb{R}^{n-1}, Z_1).$$

²We start with q_2 and not q_1 , but the shift of the indices will be convenient to simplify the notation later.

Observe that, since $\|x\|^2$ does not vanish on D_1 , “zero” is also a regular value for $Q_2 = c_2\|x\|^2 q_2|_{D_1}$ for every positive constant $c_2 > 0$, and:

$$(D_1, Z(Q_2) \cap D_1) \sim (\mathbb{R}^{n-1}, Z_1).$$

(In the course of the proof we will pick a sequence of constants $\{c_k > 0\}_{k \in \mathbb{N}}$ that will only be specified later.) Call d_2 the degree of Q_2 and observe that Q_2 only contains monomials $x_1^{\alpha_1} \cdots x_{n-1}^{\alpha_{n-1}}$ with $2 \leq |\alpha| \leq d_2$. (We set $d_1 = 1$.)

By Thom’s isotopy Lemma, associated to the function $Q_2 : D_1 \rightarrow \mathbb{R}$ there is a $\delta_2 > 0$ such that for any other continuously differentiable function $h : D_1 \rightarrow \mathbb{R}$ with $\|h\|_{C^1(D, \mathbb{R})} \leq \delta_2$ we have that the equation $Q_2 + h = 0$ is regular on D_1 and the pair $(D_1, Z(Q_2 + h) \cap D_1)$ is isotopic to the pair $(D_1, Z(Q_2) \cap D_1)$.

Let now $k \geq 2$ and consider Z_{d_k} . Pick a polynomial q_{k+1} such that “zero” is a regular value for $q_{k+1}|_{D_1}$ and:

$$(D_1, Z(q_{k+1}) \cap D_1) \sim (\mathbb{R}^{n-1}, Z_{d_k}).$$

As before, observe that “zero” is also a regular value for $Q_{k+1} = c_{k+1}\|x\|^{2d_k} q_{k+1}|_{D_1}$, for any positive constant $c_{k+1} > 0$ and:

$$(D_1, Z(Q_{k+1}) \cap D_1) \sim (\mathbb{R}^{n-1}, Z_{d_k}).$$

Again, as before by Thom’s isotopy Lemma, associated to the function $Q_{k+1} : D_1 \rightarrow \mathbb{R}$ there is a $\delta_{k+1} > 0$ such that for any other continuously differentiable function $h : D_1 \rightarrow \mathbb{R}$ with $\|h\|_{C^1(D, \mathbb{R})} \leq \delta_{k+1}$ we have that the equation $Q_{k+1} + h = 0$ is regular on D_1 and the pair $(D_1, Z(Q_{k+1} + h) \cap D_1)$ is isotopic to the pair $(D_1, Z(Q_{k+1}) \cap D_1)$. Moreover, calling $d_{k+1} = \deg(Q_{k+1})$, we have that Q_{k+1} only contains monomials with total degree $2d_k \leq |\alpha| \leq d_{k+1}$.

We choose the sequence of constants $\{c_k > 0\}$ at every step in such a way that

$$\|Q_{k+1}\|_{C^1(D_1, \mathbb{R})} \leq \min\{\delta_1, \dots, \delta_k\} 2^{-(k+1)}$$

and that the power series $\sum_{k \geq 2} Q_k$ converges on the disk D_2 .

Let now $\rho : \mathbb{R}^{n-1} \rightarrow [0, \infty)$ be a definable, regular, cut-off function such that $\rho|_{D_1} \equiv 1$ and $\rho|_{D_2^c} \equiv 0$ and define the function $g : D_2 \rightarrow \mathbb{R}$ by:

$$g(x) = \left(\sum_{k \geq 2} Q_k(x) \right) \cdot \rho(x).$$

We set $\hat{\Gamma} = \text{graph}(g) \subset \mathbb{R}^n$ and extend this to a regular, compact definable manifold $\Gamma \subset \mathbb{R}^n$. Note that the function ρ can be taken to be a restricted analytic function, and this will make Γ semianalytic in \mathbb{R}^n . The set $D \subset \text{graph}(g) \subset \Gamma$ will be the homeomorphic image of D_1 under the “graph” map $x \mapsto (x, g(x))$.

Let $P_1(x, y) = y$ and for every $k \geq 2$ define $P_k(x, y) = y - \sum_{j=2}^k Q_j(x)$. Observe that the degree of P_k is d_k . For every $k \geq 1$ we consider now the (equivalent) systems of equations:

$$\{y - g(x) = 0 = P_k(x, y)\} \iff \left\{ y - g(x) = 0 = Q_{k+1}(x) + \sum_{j \geq k+2} Q_j(x) = 0 \right\}$$

(the equivalence is obtained by eliminating y from the second equation using the first one). The set of solutions to these systems in D coincides with $Z(P_k) \cap D$.

Observe now that:

$$\left\| \sum_{j \geq k+2} Q_j \right\|_{C^1(D, \mathbb{R})} \leq \sum_{j \geq k+2} \frac{\delta_k}{2^j} \leq \delta_k.$$

In particular, since the equation $Q_{k+1} = 0$ was regular on D_1 , also the equation $Q_{k+1} + \sum_{j \geq k+2} Q_j = 0$ is regular on D_1 and we have:

$$\left(D_1, Z\left(Q_{k+1} + \sum_{j \geq k+2} Q_j\right) \cap D_1 \right) \sim (D_1, Z(Q_{k+1}) \cap D_1) \sim (\mathbb{R}^{n-1}, Z_{d_k}).$$

As a consequence the system $\{y - g(x) = 0 = P_k(x, y)\}$ is regular on $D_1 \times \mathbb{R}$ and under the graph map we have:

$$(D, Z(P_k) \cap D) \sim (\mathbb{R}^{n-1}, Z_{d_k}).$$

Finally, let $p_k = {}^h P_k + R_k$ be a homogeneous polynomial (here ${}^h P_k$ denotes the homogenization) whose zero set is transverse to Γ and with $\|R_k\|_{C^1(D, \mathbb{R})}$ small enough such that

$$(D, Z(p_k) \cap D) \sim (\mathbb{R}^{n-1}, Z_{d_k}).$$

(The existence of such R_k follows from the fact that the set of homogeneous polynomials of a given degree whose zero set intersect Γ transversely is dense).

■

3.3 Estimates on the size of pathological examples: proof of Theorem 3.1.2

Theorem 3.1.2 follows immediately from Proposition 3.3.2 (proved below) after applying Markov's inequality. In order to proceed we will need the following technical result. In the case Γ is a real algebraic set this was proved by Gayet and Welschinger (Gayet and Welschinger, 2016). Our strategy of proof is also very similar, and it essentially uses the same ideas, just adapted to the non-algebraic setting.

Proposition 3.3.1 *Let $\Gamma \subset \mathbb{RP}^n$ be a regular, compact hypersurface and $f : \Gamma \rightarrow \mathbb{R}$ be a Morse function. Let p be a random Kostlan distributed polynomial on \mathbb{RP}^n of degree d . Then, denoting by Q_{n-2} a $\text{GOE}(n-2)^3$ matrix, we have:*

$$\mathbb{E} \# \{ \text{critical points of } f|_{\Gamma \cap Z(p)} \} = \frac{|\Gamma|}{\pi} \frac{d^{\frac{n-1}{2}}}{(2\pi)^{\frac{n-2}{2}}} \cdot \mathbb{E} |\det Q_{n-2}| + O(d^{\frac{n-2}{2}}).$$

Remark 3.3.1 *Note that in the case $\dim \Gamma = 1$ this can be obtained by a simple application of integral geometry.*

Proof We will use the Kac-Rice formula for Riemannian manifolds. Since the involution $x \mapsto -x$ on the sphere with the round metric is an isometry, the quotient map $q : S^n \rightarrow \mathbb{RP}^n$ induces a Riemannian metric on \mathbb{RP}^n for which q is a Riemannian submersion. In this way $\Gamma \subset \mathbb{RP}^n$ inherits a Riemannian metric as well. For every point $y \in \Gamma$ such that $d_y f \neq 0$ (since Γ is compact and $f : \Gamma \rightarrow \mathbb{R}$ is Morse, there are only finitely many points where $d_y f$ vanishes) we consider an *orthonormal frame field* $\{v_1, \dots, v_{n-1}\}$ on a neighborhood $V \subset \Gamma$ of y such that for all $x \in V$

$$\ker d_x f = \text{span}\{v_2(x), \dots, v_{n-1}(x)\}.$$

Let us take now an open set $V \subset \Gamma$ which is contained in the open set $\{x_0 \neq 0\} \subset \mathbb{RP}^n$ (this is true after possibly shrinking V and relabeling the homogeneous coordinates $[x_0, \dots, x_n]$ in \mathbb{RP}^n). Let $\tilde{p} : \{x_0 \neq 0\} \rightarrow \mathbb{R}$ be the random function defined by $\tilde{p}(x_0, x_1, \dots, x_n) = p(1, x_1/x_0, \dots, x_n/x_0)$ and denote by \hat{p} its restriction to V : $\hat{p} = \tilde{p}|_V$ (thus \hat{p} is a random function on the Riemannian manifold $V \subset \Gamma \subset \mathbb{RP}^n$). Define the random map $F : V \rightarrow \mathbb{R}^{n-1}$ by:

$$F(x) = (\hat{p}(x), d_x \hat{p} v_2(x), \dots, d_x \hat{p} v_{n-1}(x)).$$

³GOE(m) stands for *Gaussian Orthogonal Ensemble*, an ensemble of random symmetric matrices constructed as follows: $X \in \text{GOE}(m)$ is a $m \times m$ random matrix where $X_{i,j} \sim \mathcal{N}(0, 1)$, and $X_{i,i} \sim \mathcal{N}(0, 2)$, see (Tao, 2012).

If the gradient of \hat{p} does not vanish on $\{p = 0\} \cap V$ (this happens with probability one), then $\{p = 0\} \cap V = \{\hat{p} = 0\}$ is a smooth submanifold of Γ . We claim that, with probability one, the number of critical points of $f|_{\{p=0\} \cap \Gamma}$ in V equals the number of zeroes of F . In fact, with probability one, none of the critical points of f lies on $\{p = 0\}$ and in this case a point $x \in V$ is critical for $f|_{\{p=0\} \cap \Gamma}$ if and only if $\hat{p}(x) = 0$ and the gradients of \hat{p} and f are collinear at x , i.e. $\hat{p}(x) = 0$ and $\ker d_x \hat{p} = \ker d_x f$, which is equivalent to $F(x) = 0$.

Let us denote by ω the volume density of Γ . Then the Kac-Rice formula for the random field F on the Riemannian manifold $\Gamma \cap V$ (Adler and Taylor, 2009) tells that for any open set $W \subset V$

$$\begin{aligned} \mathbb{E} \# \{F = 0\} \cap W &= \int_W \mathbb{E} \left\{ |\det J(x)| \mid F(x) = 0 \right\} \rho_{F(x)}(0) \omega(x) dx \\ &= \int_W \rho(x) \omega(x) dx. \end{aligned}$$

where the matrix $J(x)$ is the matrix of the derivatives at x of the components of F with respect to an orthonormal frame (in our case the chosen frame v_1, \dots, v_{n-1} and $\rho_{F(x)}(0)$ is the density at zero of the random vector $F(x)$).

We use now the fact that the Kostlan polynomial p is invariant by an orthogonal change of variable in $\mathbb{R}P^n$, hence for every $x \in V$ for the evaluation of

$$\rho(x) = \mathbb{E} \left\{ |\det J(x)| \mid F(x) = 0 \right\} \rho_{F(x)}(0)$$

we can assume $x = [1, 0, \dots, 0] = \bar{x}$. For simplicity let us also denote by $t_1, \dots, t_n : \{x_0 \neq 0\} \rightarrow \mathbb{R}$ the functions $t_i = x_i/x_0$. Then, since the stabilizer $O(n)$ of \bar{x} acts transitively on the set of frames at \bar{x} , we can also assume that $\{v_1(\bar{x}), \dots, v_{n-1}(\bar{x})\} = \{\partial_1(\bar{x}), \dots, \partial_{n-1}(\bar{x})\}$, where we have denoted by ∂_i the vector field $\partial/\partial t_i$.

For the calculation of the value of $\rho(\bar{x})$ we use local coordinates on $\Gamma \cap V$. Note that (t_1, \dots, t_{n-1}) are coordinates on $\Gamma \cap V$ (this is because the tangent space of Γ at \bar{x} equals $\text{span}\{\partial_1(\bar{x}), \dots, \partial_{n-1}(\bar{x})\}$). We denote by $\psi^{-1} : \{x_0 \neq 0\} \rightarrow \mathbb{R}^n$ the coordinate

chart on $\{x_0 \neq 0\} \subset \mathbb{R}P^n$ given by (t_1, \dots, t_n) . In this chart $\psi^{-1}(\Gamma \cap V)$, for a small enough V containing \bar{x} , can be seen as the graph of a function $g : \mathbb{R}^{n-1} \rightarrow \mathbb{R}$. Since the tangent space of $\psi^{-1}(\Gamma \cap V)$ at zero equals $\text{span}\{\partial_1, \dots, \partial_{n-1}\}$, the function g vanishes at zero, together with its differential. In this way we get a map $\varphi : \mathbb{R}^{n-1} \rightarrow \mathbb{R}P^n$ parametrizing V given by:

$$\varphi(t_1, \dots, t_{n-1}) = \psi(t_1, \dots, t_{n-1}, g(t_1, \dots, t_{n-1})).$$

Observe now that the frame $\{v_1, \dots, v_{n-1}\}$ coincides with $\{\partial_1, \dots, \partial_{n-1}\}$ only at zero; nevertheless, it is easy to verify that we could pick the frame $\{v_1, \dots, v_{n-1}\}$ such that in these coordinates:

$$v_i(t) = (1 + t_i)\partial_i + O(\|t\|^2) \quad i = 1, \dots, n-1.$$

In particular, denoting by $t = (t_1, \dots, t_{n-1})$, we have:

$$(d\hat{p}v_i)(t) = (1 + t_i)\partial_i\hat{p}(t, g(t)) + (1 + t_i)\partial_n\hat{p}(t, g(t))\partial_i g(t) + O(\|t\|^2).$$

From this it is immediate to see that:

$$\begin{aligned} F(\bar{x}) &= (\hat{p}(0), (d\hat{p}v_1)(0), \dots, (d\hat{p}v_{n-1})(0)) \\ &= (p_{d,0,\dots,0}, p_{d-1,0,10,\dots,0}, \dots, p_{d-1,0,\dots,0,1,0,\dots,0}) \end{aligned}$$

(in the multi-index of the i -th entry of this vector the 1 is in position $i+1$). In particular:

$$\rho_{F(\bar{x})}(0) = \frac{1}{(2\pi)^{\frac{n-1}{2}} d^{\frac{n-2}{2}}}$$

Let us evaluate now the matrix $J(\bar{x})$. For the first row $r_1(\bar{x})$ of $J(\bar{x})$ we immediately obtain:

$$r_1(\bar{x}) = (d_0\hat{p}v_1(0), \dots, d_0\hat{p}v_{n-1}(0)) = (\xi_{d-1,1,\dots,0}, \xi_{d-1,0,10,\dots,0}, \dots, \xi_{d-1,0,\dots,0,1})$$

Note that, except for the first entry, $r_1(\bar{x})$ coincides with $F(\bar{x})$; we denote by

$$w = (\xi_{d-1,0,10,\dots,0}, \dots, \xi_{d-1,0,\dots,0,1})$$

(i.e. the vector consisting of the last entries of the first row $r_1(\bar{x})$).

Let us now look at the $(n-2) \times (n-2)$ submatrix $\hat{J}(\bar{x})$ of $J(\bar{x})$, obtained by removing the first row and the first column. Observe that $\hat{J}(\bar{x}) = B + \xi_{d-1,1,0,\dots,0}M(\bar{x})$, where B is the matrix:

$$B = \begin{pmatrix} 2\xi_{d-2,0,2,0,\dots,0} & \xi_{d-2,0,1,1,0,\dots,0} & \cdots & \xi_{d-2,0,1,0,\dots,0,1} \\ \xi_{d-2,0,1,1,0,\dots,0} & 2\xi_{d-2,0,0,2,0,\dots,0} & \cdots & \xi_{d-2,0,0,1,0,\dots,0,1} \\ \vdots & & & \\ \xi_{d-2,0,1,0,\dots,0,1} & \xi_{d-2,0,1,0,\dots,0,1} & \cdots & 2\xi_{d-2,0,\dots,0,2} \end{pmatrix}, \quad (3.5)$$

and $M(\bar{x}) = (\partial_i \partial_j g(0))$. From (3.5) it is immediate to see that the matrix B is a random matrix distributed as:

$$B = \sqrt{d(d-1)}Q_{n-2}$$

where Q_{n-2} is a random $\text{GOE}(n-2)$ matrix. Hence

$$J(\bar{x}) = \left(\begin{array}{c|c} \xi_{d-1,1,0,\dots,0} & w \\ \hline * & B + \xi_{d-1,1,0,\dots,0}M(\bar{x}) \end{array} \right)$$

From this it follows that:

$$\begin{aligned}
& \mathbb{E} \left\{ \left| \det J(\bar{x}) \right| \middle| F(\bar{x}) = 0 \right\} \\
&= \mathbb{E} \left\{ \left| \det J(\bar{x}) \right| \middle| w = 0 \right\} \\
&= (d(d-1))^{\frac{n-2}{2}} \mathbb{E} \left\{ \left| \xi_{d-1,1,0,\dots,0} \right| \cdot \left| \det \left(Q_{n-2} + \frac{M(\bar{x})}{\sqrt{d-1}} \right) \right| \middle| w = 0 \right\} \\
&= (d(d-1))^{\frac{n-2}{2}} \cdot d^{\frac{1}{2}} \sqrt{\frac{2}{\pi}} \mathbb{E} \left| \det \left(Q_{n-2} + \frac{M(\bar{x})}{\sqrt{d-1}} \right) \right| = (*),
\end{aligned}$$

where in the last step we have used the fact that the random variables w , $\xi_{d-1,1,0,\dots,0}$ and Q_{n-2} and $\xi_{d-1,1,0,\dots,0}$ are independent. Note now that, by construction, the matrix $M(\bar{x})$ depends continuously on $\bar{x} \in \Gamma$, because we have assumed that Γ is of regularity class C^k with $k \geq 2$, and since Γ is compact:

$$(*) = (d(d-1))^{\frac{n-2}{2}} \cdot d^{\frac{1}{2}} \sqrt{\frac{2}{\pi}} (\mathbb{E} |\det(Q_{n-2})| + O(d^{-1/2})).$$

Putting all this together we obtain:

$$\begin{aligned}
\mathbb{E} \# \{F = 0\} \cap W &= \int_W \mathbb{E} \left\{ \left| \det J(x) \right| \middle| F(x) = 0 \right\} \rho_{F(x)}(0) \omega(x) dx \\
&= \int_W \frac{(d(d-1))^{\frac{n-2}{2}}}{(2\pi)^{\frac{n-1}{2}} d^{\frac{n-2}{2}}} \cdot d^{\frac{1}{2}} \sqrt{\frac{2}{\pi}} (\mathbb{E} |\det(Q_{n-2})| + O(d^{-1/2})) \omega(x) dx \\
&= \frac{|W|}{\pi} \frac{d^{\frac{n-1}{2}}}{(2\pi)^{\frac{n-2}{2}}} \cdot \mathbb{E} |\det Q_{n-2}| + O(d^{\frac{n-2}{2}}).
\end{aligned}$$

From this the conclusion follows. ■

In particular, since $f|_{\Gamma \cap \{p=0\}}$ is Morse with probability one (using standard arguments from differential topology it is not difficult to show that the set of such polynomials for which $f|_{\Gamma \cap \{p=0\}}$ is Morse has full measure), applying Morse's inequalities⁴ we get the following corollary.

⁴Given a compact, regular manifold Γ , and a Morse function $f : \Gamma \rightarrow \mathbb{R}$, the k -th Betti number of Γ is bounded by the number of critical points of f of index k .

Proposition 3.3.2 *There exists a universal constant $c_{k,n} > 0$ such that*

$$\mathbb{E}b_k(\Gamma \cap Z(p)) \leq |\Gamma|c_{k,n}d^{\frac{n-1}{2}} + O(d^{\frac{n-2}{2}})$$

(the implied constants in the $O(d^{\frac{n-2}{2}})$ depends on Γ).

3.4 Toward an O-minimal Polynomial Partitioning Theorem?

3.4.1 Why do we not have an o-minimal polynomial partitioning theorem?

To understand this, we need to have a working understanding of Guth's (Guth, 2015) proof of polynomial partitioning for any set of varieties (c.f. Theorem 2.2.2). The first step is to establish a cousin of the Borsuk-Ulam theorem.

Theorem 3.4.1 ((Guth, 2015)) *Define*

$$X_s = \prod_{j=1}^s S^{2^j-1}.$$

Any point $x \in X_s$ is going to be denoted in co-ordinates as (x_1, \dots, x_s) , where $x_j \in S^{2^j-1}$. Also define

$$Fl_j : X_s \rightarrow X_s,$$

which takes

$$(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_s) \mapsto (x_1, \dots, x_{j-1}, -x_j, x_{j+1}, \dots, x_s).$$

Suppose that you have a family of continuous functions $f_v : X_s \rightarrow \mathbb{R}$, for each $v \in \mathbb{Z}_2^s \setminus \{0\}$, that all obey the following antipodal-type condition:

$$f_v(Fl_j x) = (-1)^{v_j} f_v(x) \quad \text{for all } j = 1, \dots, s.$$

Then there exists a point $x \in X_s$ where $f_v(x) = 0$ for all $v \in \mathbb{Z}_2^s \setminus \{0\}$.

Let $\mathbb{R}[X_1, \dots, X_n]_{\leq D}$ denote the vector space of polynomials in n variables of degree at most D . For fixed n , we know that the dimension of $\mathbb{R}[X_1, \dots, X_n]_{\leq D}$ is $\sim_n D^n$. For each j , let D_j be such that the dimension of $\mathbb{R}[X_1, \dots, X_n]_{\leq D_j}$ is greater than 2^{j-1} . We have that $D_j \lesssim 2^{j/n}$. Next, pick a subspace of $\mathbb{R}[X_1, \dots, X_n]_{\leq D_j}$ with dimension $2^{j-1} + 1$, and identify $S^{2^{j-1}}$ with the unit sphere in this subspace. In this way, we get an embedding

$$X_s \subseteq \prod_{j=1}^s \mathbb{R}[X_1, \dots, X_n]_{\leq D_j}.$$

Let $D = \sum_j D_j \lesssim 2^{s/n}$. For $\vec{P} = (P_1, \dots, P_s) \in X_s \subseteq \prod_{j=1}^s \mathbb{R}[X_1, \dots, X_n]_{\leq D_j}$, then for any $w \in \mathbb{Z}_2^s$, define the cell

$$O(\vec{P}, w) = \{x \in \mathbb{R}^n \mid \text{Sign} P_j(x) = (-1)^{w_j} \text{ for all } j \in \{1, \dots, s\}\}.$$

Define the variables $G_v(\vec{P})$

$$G_v(\vec{P}) = \sum_{\substack{w \in \mathbb{Z}_2^s \\ w \cdot v = 0}} \sum_{\gamma \in \Gamma} \mathbb{1} \left\{ O(\vec{P}, w) \cap \gamma \neq \emptyset \right\} - \sum_{\substack{w \in \mathbb{Z}_2^s \\ w \cdot v = 1}} \sum_{\gamma \in \Gamma} \mathbb{1} \left\{ O(\vec{P}, w) \cap \gamma \neq \emptyset \right\}.$$

Note that the function G_v obeys the antipodal-type condition (Equation 3.4.1), i.e. $G_v(Fl_j \vec{P}) = (-1)^{v_j} G_v(\vec{P})$. Of course the functions $G_v : X_s \rightarrow \mathbb{R}$ are not continuous, because the indicator functions aren't. However, let's pretend for a moment that they are. Then Theorem 3.4.1 implies that there is a $\vec{P}_* \in X_s$ so that $G_v(\vec{P}_*) = 0$ for all $v \in \mathbb{Z}_2^s \setminus \{0\}$.

Note that $\sum_{\gamma \in \Gamma} \mathbb{1} \left\{ O(\vec{P}_*, w) \cap \gamma \neq \emptyset \right\}$ counts the number of varieties $\gamma \in \Gamma$ that intersect $O(\vec{P}_*, w)$. It can be proved that the system of $\{G_v = 0\}_{v \in \mathbb{Z}_2^s}$ is row equivalent to the system

$$\left\{ \sum_{\gamma \in \Gamma} \mathbb{1} \left\{ O(\vec{P}_*, v) \cap \gamma \neq \emptyset \right\} - \sum_{\gamma \in \Gamma} \mathbb{1} \left\{ O(\vec{P}_*, \vec{0}) \cap \gamma \neq \emptyset \right\} = 0 \right\}_{v \in \mathbb{Z}_2^s},$$

thus we have that $\sum_{\gamma \in \Gamma} \mathbb{1} \left\{ O(\vec{P}_*, w) \cap \gamma \neq \emptyset \right\}$ is the same for all $w \in \mathbb{Z}_2^s$. In other words, each of the 2^s cells $O(\vec{P}_*, w)$ would intersect the same number of varieties $\gamma \in \Gamma$. Since by Corollary 2.2.1 we know that each γ can enter at most $C_{d,m,n} D^k$ cells, the number of varieties intersecting each cell would be at most $2^{-s} C_{d,m,n} D^k |\Gamma| \leq C_n C_{d,m,n} D^{k-n} |\Gamma|$ proving Theorem 2.2.2. Obviously, we have to deal with the fact that the G_v are not actually continuous, but this is the core of the proof.

The proof works completely at all steps if Γ contains definable sets, except for the fact that Corollary 2.2.1 doesn't hold anymore. In fact, as per Theorem 3.1.1, no uniform bound even exists. This is why it has been difficult to establish an o-minimal version of the polynomial partitioning theorem.

3.4.2 An o-minimal polynomial partitioning theorem using the probabilistic method?

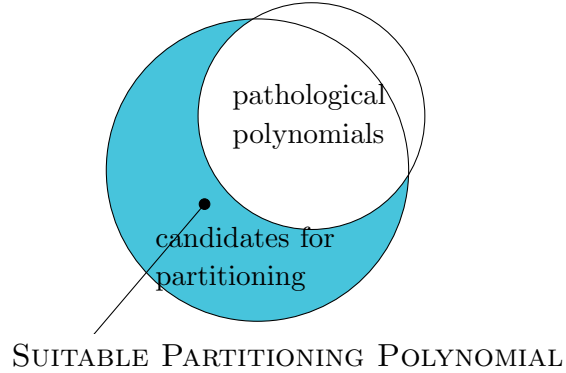


Figure 3.3.: Illustration of desirable situation to prove an o-minimal polynomial partitioning theorem

While Theorem 3.1.1 is a dampener on our hopes, on the other hand, Corollary 3.1.1 shows that for *most polynomials*, a Bezout-type bound holds. Specifically, we prove that for any definable hypersurface $\Gamma \subset \mathbb{R}P^n$, if p is a random Kostlan homogeneous polynomial of degree d in $n + 1$ variables,

$$\mathbb{P} [b_0(\Gamma \cap Z(p)) \gtrsim d^n] \lesssim \frac{1}{\sqrt{d^n}}.$$

This means that the measure of bad polynomials grows smaller with increasing degree. If we are able to prove that the measure of the set of polynomials that satisfy $\{G_v(P) = 0\}_{v \in \mathbb{Z}_2^s}$ is strictly $\gtrsim \frac{1}{\sqrt{d^n}}$ (maybe under some restrictions on Γ), by the probabilistic method, we would have proved that there exists a polynomial which has desirable properties, which also does not have pathological topological complexity on restriction to any definable hypersurface Γ , giving us a much needed o-minimal polynomial partitioning theorem.

Question 1 *Leveraging the results of (Basu et al., 2019b), prove an o-minimal polynomial partitioning theorem using the probabilistic method.*

4 BETTI NUMBERS OF RANDOM HYPERSURFACE ARRANGEMENTS

4.1 Introduction

The quantitative study of the ‘complexity’ of arrangements of hypersurfaces in some finite dimensional real space has a fairly long history in the area of discrete and computational geometry (see (Agarwal and Sharir, 2000) for a survey). The main mathematical results concern the combinatorial, as well as topological, complexities of the so called ‘cells’ of the arrangement. A cell of an arrangement refers to a connected component of any set obtained as the intersection of a subset of the given hypersurfaces with the complements of the remaining hypersurfaces (so by definition a cell is always locally closed and a full dimensional cell is open). It is worth recalling some of these results.

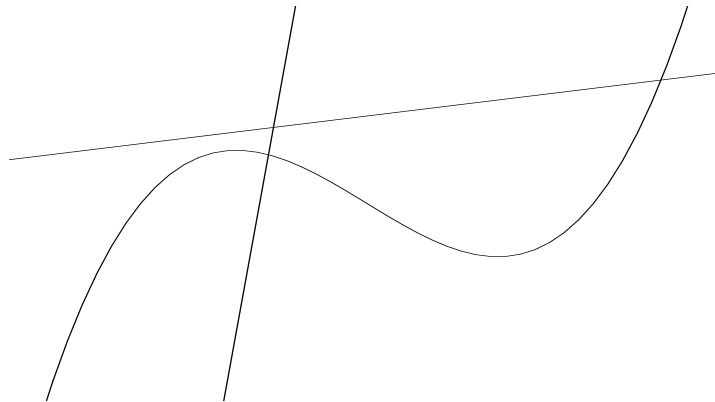


Figure 4.1.: For us, an arrangement is just the union of a finite number of algebraic sets

Given a set of s real algebraic hypersurfaces in \mathbb{R}^n each defined by a polynomial of degree at most d , it was proved in (Basu, 2003) that for each $i, 0 \leq i < n$, the sum over all cells of the arrangement of the i -th Betti number of the cells is bounded from

above by $s^{n-i}O(d)^n$. Taking $i = 0$, one obtains an upper bound of $s^n O(d)^n$ on the number of cells of the arrangement.

The above results are deterministic. Recently, the study of the *expected* topology of real varieties or semi-algebraic sets defined by randomly chosen real polynomials has assumed significance (see for example, (Gayet and Welschinger, 2015; Fyodorov et al., 2015; Burgisser and Lerario, 2018)). In this paper we initiate the study of quantitative properties of arrangements of real hypersurfaces from a random viewpoint in the same spirit as in the papers referred to above. We study the topological complexity of arrangements of s randomly chosen hypersurfaces of degrees d_1, \dots, d_s . The probability measure on the space of polynomials, according to which the polynomials are chosen, is the well known Kostlan distribution, which is a Gaussian distribution on the real vector space of homogeneous polynomials of a fixed degree (equipped with an inner product) (Edelman and Kostlan, 1995; Kostlan, 1993). Specifically, on the space of homogenous polynomials of degree d in $n + 1$ variables, a Kostlan form is defined as

$$P(x) = \sum_{\substack{(\alpha_0, \dots, \alpha_n) \\ \sum_{i=0}^n \alpha_i = d}} \xi_\alpha x_0^{\alpha_0} \dots x_n^{\alpha_n},$$

where $\xi_\alpha \sim \mathcal{N}\left(0, \frac{d!}{\alpha_0! \dots \alpha_n!}\right)$ are independently chosen. The variances are chosen in such a way that the resulting probability distribution is invariant under an orthogonal change of variables, meaning that there are no preferred points or direction in $\mathbb{R}P^n$, where the zeros of p are naturally defined. Moreover, if we extend this distribution to the space of complex polynomials by replacing real with complex Gaussian variables, it can be shown that this extension is the unique (up to multiples) Gaussian measure which is invariant under unitary change of variables, thus making real Kostlan polynomials a natural object of study.

Here we deviate slightly from the usual convention in the literature in discrete and computational geometry, and consider arrangements of hypersurfaces in real projective space $\mathbb{R}P^n$ rather than in \mathbb{R}^n (since the orthogonal invariance of the Kostlan measure is meaningful only over the projective space). However, asymptotically it

does not make a difference, whether we consider arrangements over affine or projective spaces.

We consider two variants of the problem of bounding the topological complexity of an arrangement of random real algebraic hypersurfaces in $\mathbb{R}P^n$ with specified degrees. Our first result outlined in §4.1.1 treats the problem in full generality without any restriction on the degrees (cf. Theorem 4.1.1). We then study the case when all the degrees are assumed to be equal to 2 (outlined in §4.1.2). This is the first non-trivial case, since for an arrangement of hyperplanes (i.e. with all degrees equal to one), the expected value of the topological complexity will coincide with that of deterministic generic arrangements. Since, it is known that the growth of the Betti numbers of semi-algebraic sets defined by quadratic polynomials show different behavior compared to that of general semi-algebraic sets (see (Barvinok, 1997; Basu et al., 2010; Lerario, 2016; Basu and Rizzie, 2018) for the deterministic case and (Lerario, 2015; Lerario and Lundberg, 2016) in the random setting), it could be expected that the average topological complexity of arrangements consisting of quadric hypersurfaces would be smaller than in the general case (at least in the dependence on the number s of hypersurfaces). We have partial results (outlined in §4.1.2) showing that this is indeed the case. While the $(n-1)$ -dimensional Betti number of the complement of a union s hypersurfaces of degree $d \geq 2$ in $\mathbb{R}P^n$ grows proportionally with s in the deterministic case, we show that in the random case with $d = 2$ the expected value of the same is $o(s)$ (cf. Theorem 4.1.2).

In order to prove Theorem 4.1.2, we study the behavior of a special kind of geometrically defined graph from a random viewpoint (outlined in §4.1.3). The geometric graph that we study is a special case of the more general graphs defined by semi-algebraic relations which has been widely studied in combinatorics (see for example (Alon et al., 2005)). In our case the semi-algebraic relation defining the graph is particularly simple and geometric, and hence we believe that study of this model could be of interest by itself. We fix a convex semi-algebraic subset $\mathcal{P} \subset \mathbb{R}P^N$ and sample independent points q_1, \dots, q_s from the uniform distribution on $\mathbb{R}P^N$, and we

put an edge between v_i and v_j , if and only if $i \neq j$ and the line connecting q_i and q_j *does not* intersect \mathcal{P} . We give a tight estimate on the expected number of isolated points of such a graph (cf. Theorem 4.1.3), from which we can deduce Theorem 4.1.2. Finally, we conclude by proving a Ramsey-type result about the random graph of quadrics (cf. Corollary 4.3.1).

4.1.1 Random hypersurface arrangements

We are given random homogenous polynomials P_1, \dots, P_s , where each P_i is homogenous in $n + 1$ variables and is of degree d_i , i.e. $P \in \mathbb{R}[x_0, \dots, x_n]_{(d_i)}$, and we look at the *random* arrangement of hypersurfaces defined in the projective space by the zero sets of these polynomials, i.e.,

$$\Gamma = \bigcup_{j=1}^s \Gamma_j \subset \mathbb{R}P^n,$$

where each Γ_j is the real algebraic hypersurface given by the zero set of P_j , i.e.,

$$\Gamma_j = Z(P_j) = \{[x_0, \dots, x_n] \in \mathbb{R}P^n \mid P_j(x_0, \dots, x_n) = 0\}.$$

The main problem that we want to address concerns understanding the topological complexity of Γ , which will be measured by its Betti numbers¹.

We observe that there are three sets of parameters that will play a role in our study: the degree sequence d_1, \dots, d_n of the hypersurfaces, the dimension n of the ambient projective space and the number s of independent hypersurfaces. (Of course, the choice of what is meant by *random* will also play a role: for us the polynomials P_1, \dots, P_s will be independent samples from the Kostlan ensemble.)

Our first result concerns the asymptotic when n is kept fixed and $d_1, \dots, d_s, s \rightarrow \infty$ and gives information on the number of cells of $\mathbb{R}P^n \setminus \Gamma$. There is clearly an analogous statement for the spherical version of this problem, and the two cases can be related

¹For a semialgebraic set S we denote by $b_i(S)$ its i^{th} Betti number with coefficients in $\mathbb{Z}/2\mathbb{Z}$

using standard techniques from algebraic topology (the spherical arrangement double covers the projective one and the asymptotics, up to a factor of two, are the same).

Theorem 4.1.1 (*n fixed*) *Let $P_1, \dots, P_s \in \mathbb{R}[x_0, \dots, x_n]$ be random, independent, Kostlan polynomials, where P_i has degree d_i . Let $\Gamma_i \subset \mathbb{R}P^n$ be the zero set of P_i , and define $\Gamma = \bigcup_{i=1}^s \Gamma_i$. Also, let $d = \max(d_1, \dots, d_s)$. Then:*

$$\mathbb{E}[b_0(\mathbb{R}P^n \setminus \Gamma)] = \sum_{\substack{I \subset [s] \\ |I|=n}} \sqrt{\prod_{i \in I} d_i} + O(d^{(n-1)/2} s^{n-1}). \quad (4.1)$$

Moreover if all the degrees are the same $d_1 = \dots = d_s = d$ we have:

$$\mathbb{E}[b_0(\mathbb{R}P^n \setminus \Gamma)] = \binom{s}{n} d^{n/2} + O(d^{(n-1)/2} s^{n-1}). \quad (4.2)$$

Remark 4.1.1 *As we will prove in Corollary 4.2.2, the expectation of the total Betti number of $\mathbb{R}P^n \setminus \Gamma$ has the same order as that of the expected number of connected components (cf. Equation (4.2)). This suggests an interesting phenomenon: the total amount of topology in $\mathbb{R}P^n \setminus \Gamma$ is the same (to the leading order) as the total number of cells of $\mathbb{R}P^n \setminus \Gamma$ and it is therefore natural to conjecture that a random cell is on average homologically a point — but unfortunately we were not able to prove this result. It is also interesting to compare the previous statement with its worst possible deterministic bound from (Basu et al., 1996):*

$$b_0(\mathbb{R}P^n \setminus \Gamma) \leq \binom{s}{n} (O(d))^n.$$

Remark 4.1.2 *It is possible to produce estimates for the expected number of cells also for other invariant distributions (classified in (Kostlan, 1993)), and the answer is given in terms of the parameter of the distribution. In general it is no longer true that we obtain an estimate where the leading term in d is of the type $O(d^{n/2})$, for*

instance sampling random harmonic polynomials of degree d , we get an estimate of the type:

$$\mathbb{E}[b_0(\mathbb{R}P^n \setminus \Gamma)] = \Theta\left(\frac{d^n s^n}{n!}\right).$$

4.1.2 Arrangements of random quadrics

The next result deals instead with the asymptotic structure of Γ when $d_1, \dots, d_s = 2$, n is fixed, and $s \rightarrow \infty$. It turns out that in this case, the problem of understanding the number of connected components of Γ , i.e. $b_0(\Gamma)$ (Betti numbers of Γ and $\mathbb{R}P^n \setminus \Gamma$ are related by the Alexander-Pontryagin duality), is related to the connectivity of a certain random graph model, and can be studied in a precise way. Specifically, our second theorem gives an upper bound on the average number of connected components in a random arrangement of quadrics' zero sets.

Theorem 4.1.2 (n fixed, $s \rightarrow \infty$) *Let $P_1, \dots, P_s \in \mathbb{R}[X_0, \dots, X_n]$ be homogeneous Kostlan quadrics. Let $\Gamma_i \subset \mathbb{R}P^n$ be the zero set of P_i , and define $\Gamma = \bigcup_{i=1}^s \Gamma_i$. Then*

$$\lim_{s \rightarrow \infty} \frac{\mathbb{E}[b_0(\Gamma)]}{s} = 0.$$

Remark 4.1.3 *The topology of a random intersection of quadrics has been studied in (Lerario, 2015; Lerario and Lundberg, 2016), also using a random spectral sequence (different from the one of this paper). There the following statement is proved: if $X \subset \mathbb{R}P^n$ is an intersection of k random quadrics, then for every fixed $i \geq 0$ with probability that goes to one faster than any polynomial as $n \rightarrow \infty$ we have $b_i(X) = 1$. In fact this phenomenon follows from a sort of “rigidification” of the spectral sequence structure in the large n limit (a similar phenomenon can be observed in the context of this paper).*

As a corollary of Theorem 4.1.2 (cf. Corollary 4.3.1), we rule out the existence of linear sized cliques in the complement of the quadrics graph. This must be contrasted

with a result in (Alon et al., 2005) who prove a Ramsey type result (cf. Theorem 4.3.3) about existence of sub-linear sized cliques in general semi-algebraic graphs.

4.1.3 A random graph model

The result on random arrangements of quadrics unexpectedly follows from the statistic of the number of connected components of a certain random graph introduced as follows. We pick a semialgebraic convex subset $\mathcal{P} \subset \mathbb{R}P^N$ and we sample independent points q_1, \dots, q_s from the uniform distribution on $\mathbb{R}P^N$. (In the forthcoming connection with the previous problem, N plays the role of the dimension of the space of quadratic forms and the points q_1, \dots, q_s are the quadrics.) The vertices of the random graph are points $\{v_1, \dots, v_s\}$ (one for each sample) and we put an edge between v_i and v_j , if and only if $i \neq j$ and the line connecting q_i and q_j *does not* intersect \mathcal{P} . We call such a graph a *obstacle random graph* and denote it by $\mathcal{G}(\mathcal{P}, s)$. Of course the same definition makes sense in every compact Riemannian manifold, where the notion of convexity comes from geodesics. An obstacle random graph is expected to have at least $s \cdot \frac{\text{vol}(\mathcal{P})}{\text{vol}(\mathbb{R}P^N)}$ many isolated points (this is the expected number of points falling inside \mathcal{P}). In Theorem 4.1.3 below we prove that to the leading order there are no other isolated points.

Theorem 4.1.3 ($\mathcal{P} \subset \mathbb{R}P^N$ fixed, $s \rightarrow \infty$) *The expected number of connected component of the obstacle random graph satisfies*

$$\lim_{s \rightarrow \infty} \frac{\mathbb{E}[b_0(\mathcal{G}(N, \mathcal{P}, s))]}{s} \leq \frac{\text{vol}(\mathcal{P})}{\text{vol}(\mathbb{R}P^N)}.$$

The connection between Theorem 4.1.3 and Theorem 4.1.2 comes from an interesting result of Calabi (see Theorem 4.3.2 below): the common zero set of two quadrics in $\mathbb{R}P^n$ is nonempty if and only if the line joining these two quadrics (the projective pencil) *does not* intersect the set $\mathcal{P} \subset \mathbb{R}P^N$ of positive quadrics. Since nonempty quadrics in projective space are connected, the incidence graph of the ran-

dom arrangement $\Gamma = \bigcup_{j=1}^s Z(q_j)$ is the same as the obstacle random graph minus its isolated points coming from vertices v_i whose corresponding quadric $q_i \in \mathcal{P}$.

4.2 A Random Spectral Sequence

4.2.1 Preliminaries on spectral sequences

We direct the reader to references such as (McCleary, 2001) for an in-depth treatment of spectral sequences. Our semi-algebraic sets will be assumed to possess finite triangulations. We shall study the simplicial cohomology (in our case, the topology is tame, so various cohomology theories coincide). Specifically, our method of proving involves a Mayer-Vietoris spectral sequence argument over a *double complex* arising from the *Mayer-Vietoris exact sequence*. We shall begin by reviewing what a double complex of vector spaces is and then introduce the associated spectral sequence.

Double Complexes

A double complex C is a bigraded vector space

$$C = \bigoplus_{p,q \in \mathbb{Z}} C^{p,q},$$

with morphisms $(d_{\wedge}^{p,q}, d_{>}^{p,q})_{p,q \in \mathbb{Z}}$, where $d_{\wedge}^{p,q} : C^{p,q} \rightarrow C^{p,q+1}$ are called ‘upward’ morphisms and $d_{>}^{p,q} : C^{p,q} \rightarrow C^{p+1,q}$ are called ‘rightward’ morphisms (we shall omit the superscripts on the morphisms whenever they are clear from context), satisfying:

$$(I) \quad d_{\wedge}^2 = 0, \quad d_{>}^2 = 0,$$

$$(II) \quad d_{\wedge} d_{>} + d_{>} d_{\wedge} = 0 \text{ (i.e. they anticommute).}$$

As can be seen, the superscripts p, q , are in keeping with the convention for how the (x, y) -plane is labeled; consequently, a double complex is called a *first quadrant double complex* if $C^{p,q} = 0$ whenever either $p < 0$ or $q < 0$.

$$\begin{array}{ccccccc}
& \vdots & & \vdots & & \vdots & \\
& \uparrow d_{\wedge} & & \uparrow d_{\wedge} & & \uparrow d_{\wedge} & \\
\longrightarrow & C^{p-1,q+1} & \xrightarrow{d_{>}} & C^{p,q+1} & \xrightarrow{d_{>}} & C^{p+1,q+1} & \longrightarrow \\
& \uparrow d_{\wedge} & & \uparrow d_{\wedge} & & \uparrow d_{\wedge} & \\
\longrightarrow & C^{p-1,q} & \xrightarrow{d_{>}} & C^{p,q} & \xrightarrow{d_{>}} & C^{p+1,q} & \longrightarrow \\
& \uparrow d_{\wedge} & & \uparrow d_{\wedge} & & \uparrow d_{\wedge} & \\
\longrightarrow & C^{p-1,q-1} & \xrightarrow{d_{>}} & C^{p,q-1} & \xrightarrow{d_{>}} & C^{p+1,q-1} & \longrightarrow \\
& \uparrow d_{\wedge} & & \uparrow d_{\wedge} & & \uparrow d_{\wedge} &
\end{array}$$

From the double complex, we can form a co-chain complex of vector spaces C^{\bullet} , called the *associated total complex*, defined by $C^n = \bigoplus_{p+q=n} C^{p,q}$, with $D = d_{\wedge} + d_{>} : C^n \rightarrow C^{n+1}$ as the co-boundary operator. Note that $D^2 = (d_{\wedge} + d_{>})^2 = d_{\wedge}^2 + (d_{\wedge}d_{>} + d_{>}d_{\wedge}) + d_{>}^2 = 0$, verifying that C^{\bullet} is indeed a valid co-chain complex.

There is a natural structure called *filtration* that we can find in our co-chain complex C^{\bullet} . In fact, each co-chain group $C^n = \bigoplus_{p+q=n} C^{p,q}$ has two filtrations. One filtration, called the *vertical* filtration, is by restricting p to be greater than or equal to some k . The second filtration, called the *horizontal* filtration, is obtained by restricting q to be greater than or equal to k .

The vertical filtration is:

$$C^n = C_0^n \supseteq C_1^n \supseteq \dots \supseteq C_n^n \supseteq C_{n+1}^n = 0,$$

where

$$C_k^n = \bigoplus_{\substack{p+q=n \\ p \geq k}} C^{p,q}.$$

Also let Z_k^n be the co-cycles in C_k^n , i.e.

$$Z_k^n = \{z \in C_k^n \mid Dz = 0\},$$

and

$$H_k^n = Z_k^n / (Z_k^n \cap B^n),$$

where $B^n = DC_{n-1}$. Thus we have a corresponding filtration of $H^n(C)$, i.e. $\dots \supseteq H_{k-1}^n \supseteq H_k^n \supseteq H_{k+1}^n \subseteq \dots$. Also denote $H^{k,n-k}$ the quotients H_k^n / H_{k+1}^n .

Spectral sequence associated to a Double Complex

A *spectral sequence* is a sequence of complexes $(E_r, d_r)_{r \geq 0}$, where $E_{r+1} \cong H_{d_r}(E_r)$, and d_r are differentials

$$d_r : E_r^{p,q} \rightarrow E_r^{p+r, q-r+1}.$$

Recall that we have graded co-chain groups $C^n = \bigoplus_{p+q=n} C^{p,q}$. We say that an element of $C^{\bullet, \bullet}$ is a (p, q) -strip if it is an element of $\bigoplus_{l \geq 0} C^{p+1, q-l}$. Let $Z^{p,q}$ be the set of all (p, q) -strips that are co-cycles. We get the following characterization for $Z^{p,q}$. Let $a \in C^{p,q}$, $a^{(i)} \in C^{p+i, q-i}$, for $i \geq 1$; thus $a \oplus a^{(1)} \oplus a^{(2)} \oplus \dots$ is a (p, q) -strip. Since it is a co-cycle, by definition, $D(a \oplus a^{(1)} \oplus a^{(2)} \oplus \dots) = 0 = \eta$ (say). η has co-ordinates in all $C^{p+l, q+1-l}$, $l \geq 0$. The co-ordinate in $C^{p, q+1}$ is $d_{\wedge} a$, the co-ordinate in $C^{p+1, q}$ is $d_{>} a + d_{\wedge} a^{(1)}$, the co-ordinate in $C^{p+2, q-1}$ is $d_{>} a^{(1)} + d_{\wedge} a^{(2)}$, and so on. All co-ordinates must be 0, thus we can conclude that $Z^{p,q}$ denotes the set of all $a \in C^{p,q}$ such that the following system of equations has a solution:

$$\begin{aligned} d_{\wedge} a &= 0, \\ d_{>} a &= -d_{\wedge} a^{(1)}, \\ d_{>} a^{(1)} &= -d_{\wedge} a^{(2)}, \\ d_{>} a^{(2)} &= -d_{\wedge} a^{(3)}, \\ &\vdots \end{aligned} \tag{4.3}$$

Also, we define $B^{p,q} \subseteq C^{p,q}$ to be all elements b such that it has a pre-image in C^{p+q-1} . Using the same analysis as above, we get the characterization that $B^{p,q}$ is all b such that the following system of equations has a solution:

$$\begin{aligned} d_{\wedge} b^{(0)} + d_{>} b^{(-1)} &= 0, \\ d_{\wedge} b^{(-1)} + d_{>} b^{(-2)} &= 0, \\ d_{\wedge} b^{(-2)} + d_{>} b^{(-3)} &= 0, \\ &\vdots, \end{aligned} \tag{4.4}$$

where $b^{(-i)} \in C^{p-i, q+i-1}$. Recall that $H^{p,q}$ was defined to be the quotients $H_p^{p+q} / H_{p+1}^{p+q}$. It is easy to see that

$$H^{p,q} \cong Z^{p,q} / B^{p,q}.$$

Now, define

$$Z_r^{p,q} = \{a \in C^{p,q} \mid \exists (a^{(1)}, \dots, a^{(r-1)}) \text{ with } (a, a^{(1)}, \dots, a^{(r-1)}) \text{ satisfying (4.3)}\},$$

and

$$\begin{aligned} B_r^{p,q} &= \{b \in C^{p,q} \mid \exists (b^{(0)}, b^{(-1)}, \dots) \text{ with } (b, b^{(0)}, b^{(-1)}, \dots) \text{ satisfying (4.4)}, \\ &\quad \text{and } b^{(-r)} = b^{(-r-1)} = \dots = 0\}. \end{aligned}$$

Thus we observe that we have the following subspace structure of the vector space $C^{p,q}$:

$$B_1^{p,q} \subseteq B_2^{p,q} \subseteq \dots \subseteq B^{p,q} \subseteq Z^{p,q} \subseteq Z_1^{p,q} \subseteq \dots \subseteq C^{p,q}.$$

Define the (p, q) -th graded piece, i.e. $E_r^{p,q}$, of the r^{th} page, i.e. E_r , as $E_r^{p,q} = Z_r^{p,q} / B_r^{p,q}$. It should be seen as an approximation to $H^{p,q} = Z^{p,q} / B^{p,q}$.

We shall now define the differentials d_r . Let $[a] \in E_r^{p,q}$ for some $a \in Z_r^{p,q}$. Then there exists $a^{(1)}, \dots, a^{(r-1)}$ satisfying (4.3). Now, d_r is defined as follows

$$d_r : E_r^{p,q} \rightarrow E_r^{p+r,q-r+1}, \quad \text{taking } [a] \mapsto [d_{>} a^{(r-1)}] \in E_r^{p+r,q-r+1}.$$

We refer the reader to (McCleary, 2001) for a proof that this map is well-defined.

Mayer-Vietoris Spectral Sequence

Of interest to us is a particular double complex that arises from the Mayer-Vietoris exact sequence. Let A_1, \dots, A_s be sub-complexes of a finite simplicial complex $A = A_1 \cup \dots \cup A_s$. By the definition of a simplicial complex, any finite intersection $A_{\alpha_0} \cap \dots \cap A_{\alpha_p}$, denoted $A_{\alpha_0, \dots, \alpha_p}$ is a sub-complex of A . Let $C^i(A)$ denote the vector space over \mathbb{R} of i -co-chains of A , and $C^*(A) = \bigoplus_i C^i(A)$. A basic definition from simplicial homology is that of the singular co-boundary homomorphism $d : C^q(A) \rightarrow C^{q+1}(A)$, i.e.

$$(d\omega)([a_0, \dots, a_{q+1}]) = \sum_{0 \leq i \leq q+1} \omega([a_0, \dots, \hat{a}_i, \dots, a_{q+1}]),$$

extended linearly, where $w \in C^q(A)$, $[a_0, \dots, a_{q+1}]$ is a $q+1$ simplex in A , and $\hat{}$ denotes omission of a vertex.

The *Mayer-Vietoris sequence* is an exact sequence of vector spaces, where each vector space is of the form $\bigoplus_{\alpha_0 < \dots < \alpha_p} C^*(A_{\alpha_0, \dots, \alpha_p})$. Specifically, it is the following sequence:

$$\begin{aligned} 0 \longrightarrow C^*(A) &\xrightarrow{r} \bigoplus_{\alpha_0} C^*(A_{\alpha_0}) \xrightarrow{\delta} \bigoplus_{\alpha_0 < \alpha_1} C^*(A_{\alpha_0, \alpha_1}) \dots \\ &\xrightarrow{\delta} \bigoplus_{\alpha_0 < \alpha_1 < \dots < \alpha_p} C^*(A_{\alpha_0, \alpha_1, \dots, \alpha_p}) \xrightarrow{\delta} \dots \end{aligned} \quad (4.5)$$

Here r is just induced by restriction. We need to define δ , which is called the Čech differential. Let $\omega \in \bigoplus_{\alpha_0 < \dots < \alpha_p} C^q(A_{\alpha_0, \dots, \alpha_p})$. Then, $\delta(\omega) \in \bigoplus_{\alpha_0 < \dots < \alpha_{p+1}} C^q(A_{\alpha_0, \dots, \alpha_{p+1}})$.

Each element of $\bigoplus_{\alpha_0 < \dots < \alpha_{p+1}} C^q(A_{\alpha_0, \dots, \alpha_{p+1}})$ is a tuple of size $\binom{s}{p+2}$, where each item in the tuple is a linear form on $C^q(A_{\alpha_0, \dots, \alpha_{p+1}})$, for a specific $(p+2)$ -tuple $0 \leq \alpha_0 < \dots < \alpha_{p+1} \leq s$. Thus, we can index $\delta(\omega)$ by $(p+2)$ -tuples, and consequently, it suffices to define $\delta(\omega)_{\alpha_0, \dots, \alpha_{p+1}}$. Let $s \in A_{\alpha_0, \dots, \alpha_{p+1}}$ be a q -simplex. $\delta(\omega)_{\alpha_0, \dots, \alpha_{p+1}}$ is a linear form on $C_q(A_{\alpha_0, \dots, \alpha_{p+1}})$, defined as

$$\delta(\omega)_{\alpha_0, \dots, \alpha_{p+1}}(s) = \sum_{0 \leq i \leq p+1} (-1)^i \omega_{\alpha_0, \dots, \hat{\alpha}_i, \dots, \alpha_{p+1}}(s).$$

It is known that the sequence (4.5) is exact. We now initialize the zeroth page of our spectral sequence, i.e. the double complex

$$E_0 = \bigoplus_{p, q \geq 0} E_0^{p, q}, \quad \text{where } E_0^{p, q} = \bigoplus_{\alpha_0 < \dots < \alpha_p} C^q(A_{\alpha_0, \dots, \alpha_p}),$$

and the total differential $D = \delta + (-1)^p d$ (verify that $D^2 = 0$). Thus we get that E_0 has the following shape:

$$\begin{array}{ccccccc} & & \vdots & & \vdots & & \vdots \\ & & \uparrow d & & \uparrow d & & \uparrow d \\ 0 & \longrightarrow & \bigoplus_{\alpha_0} C^3(A_{\alpha_0}) & \xrightarrow{\delta} & \bigoplus_{\alpha_0 < \alpha_1} C^3(A_{\alpha_0, \alpha_1}) & \xrightarrow{\delta} & \bigoplus_{\alpha_0 < \alpha_1 < \alpha_2} C^3(A_{\alpha_0, \alpha_1, \alpha_2}) \longrightarrow \\ & & \uparrow d & & \uparrow d & & \uparrow d \\ 0 & \longrightarrow & \bigoplus_{\alpha_0} C^2(A_{\alpha_0}) & \xrightarrow{\delta} & \bigoplus_{\alpha_0 < \alpha_1} C^2(A_{\alpha_0, \alpha_1}) & \xrightarrow{\delta} & \bigoplus_{\alpha_0 < \alpha_1 < \alpha_2} C^2(A_{\alpha_0, \alpha_1, \alpha_2}) \longrightarrow \\ & & \uparrow d & & \uparrow d & & \uparrow d \\ 0 & \longrightarrow & \bigoplus_{\alpha_0} C^1(A_{\alpha_0}) & \xrightarrow{\delta} & \bigoplus_{\alpha_0 < \alpha_1} C^1(A_{\alpha_0, \alpha_1}) & \xrightarrow{\delta} & \bigoplus_{\alpha_0 < \alpha_1 < \alpha_2} C^1(A_{\alpha_0, \alpha_1, \alpha_2}) \longrightarrow \\ & & \uparrow d & & \uparrow d & & \uparrow d \\ 0 & \longrightarrow & \bigoplus_{\alpha_0} C^0(A_{\alpha_0}) & \xrightarrow{\delta} & \bigoplus_{\alpha_0 < \alpha_1} C^0(A_{\alpha_0, \alpha_1}) & \xrightarrow{\delta} & \bigoplus_{\alpha_0 < \alpha_1 < \alpha_2} C^0(A_{\alpha_0, \alpha_1, \alpha_2}) \longrightarrow \\ & & \uparrow d & & \uparrow d & & \uparrow d \\ & & 0 & & 0 & & 0 \end{array}$$

There are two spectral sequences associated with $E_0^{p, q}$ both converging to $H_D^*(E_0)$, one corresponding to taking the horizontal filtration, and another corresponding to

taking the vertical filtration. The first two terms by taking the horizontal filtration are $E_1 = H_\delta(E_0)$ and $E_2 = H_d(H_\delta(E_0))$. Because of the exactness of the Mayer-Vietoris sequence (c.f. (4.5)), we have

$$E_1 = \left[\begin{array}{ccc} \vdots & \vdots & \vdots \\ \uparrow d & \uparrow 0 & \uparrow 0 \\ C^3(A) & 0 & 0 \\ \uparrow d & \uparrow 0 & \uparrow 0 \\ C^2(A) & 0 & 0 \\ \uparrow d & \uparrow 0 & \uparrow 0 \\ C^1(A) & 0 & 0 \\ \uparrow d & \uparrow 0 & \uparrow 0 \\ C^0(A) & 0 & 0 \end{array} \right],$$

and consequently,

$$E_2 = \left[\begin{array}{ccc} \vdots & \vdots & \vdots \\ H^3(A) & 0 & 0 \\ H^2(A) & 0 & 0 \\ H^1(A) & 0 & 0 \\ H^0(A) & 0 & 0 \end{array} \right].$$

The sequence converges at E_2 , showing that $H_D^*(E_0) \cong H^*(A)$. The first two terms by taking the vertical filtration are $E'_1 = H_d(E_0)$ and $E'_2 = H_\delta(H_d(E_0))$. Specifically,

$$E'_1 = \begin{array}{c} \vdots \\ \vdots \\ \vdots \\ \oplus_{\alpha_0} H^3(A_{\alpha_0}) \xrightarrow{\delta} \oplus_{\alpha_0 < \alpha_1} H^3(A_{\alpha_0, \alpha_1}) \xrightarrow{\delta} \oplus_{\alpha_0 < \alpha_1 < \alpha_2} H^3(A_{\alpha_0, \alpha_1, \alpha_2}) \longrightarrow \\ \oplus_{\alpha_0} H^2(A_{\alpha_0}) \xrightarrow{\delta} \oplus_{\alpha_0 < \alpha_1} H^2(A_{\alpha_0, \alpha_1}) \xrightarrow{\delta} \oplus_{\alpha_0 < \alpha_1 < \alpha_2} H^2(A_{\alpha_0, \alpha_1, \alpha_2}) \longrightarrow \\ \oplus_{\alpha_0} H^1(A_{\alpha_0}) \xrightarrow{\delta} \oplus_{\alpha_0 < \alpha_1} H^1(A_{\alpha_0, \alpha_1}) \xrightarrow{\delta} \oplus_{\alpha_0 < \alpha_1 < \alpha_2} H^1(A_{\alpha_0, \alpha_1, \alpha_2}) \longrightarrow \\ \oplus_{\alpha_0} H^0(A_{\alpha_0}) \xrightarrow{\delta} \oplus_{\alpha_0 < \alpha_1} H^0(A_{\alpha_0, \alpha_1}) \xrightarrow{\delta} \oplus_{\alpha_0 < \alpha_1 < \alpha_2} H^0(A_{\alpha_0, \alpha_1, \alpha_2}) \longrightarrow \end{array}.$$

Now, since this spectral sequence also converges to $H^*(A)$, we have that

$$\text{rank } H^i(A) = \sum_{p+q=i} \text{rank } E_\infty'^{p,q} \leq \sum_{p+q=i} \text{rank } E_1'^{p,q}.$$

4.2.2 Random Mayer-Vietoris Spectral Sequence

We have a finite family of closed semi-algebraic sets and we want to consider the cohomology of the union. Let A_1, \dots, A_s be triangulations of $\Gamma_1, \dots, \Gamma_s$, respectively. Thus we have a finite simplicial complex $A = A_1 \cup \dots \cup A_s$. By definition, any finite intersection $A_{\alpha_0} \cap \dots \cap A_{\alpha_p}$, denoted $A_{\alpha_0, \dots, \alpha_p}$, is a sub-complex of A . Let $C^i(A)$ denote the vector space over \mathbb{R} of i -co-chains of A , and $C^*(A) = \bigoplus_i C^i(A)$. We shall use the Mayer-Vietoris spectral sequence. From the above, we have the following theorem.

Theorem 4.2.1 (Mayer-Vietoris spectral sequence (see e.g. (Basu, 2003)))

There exists a first quadrant cohomological spectral sequence $(E_r, \delta_r)_{r \in \mathbb{Z}}$, where each E_r is a double complex

$$E_r = \bigoplus_{p,q \in \mathbb{Z}} E_r^{p,q},$$

and

$$E_0^{p,q} = \bigoplus_{\alpha_0 < \dots < \alpha_p} C^q(A_{\alpha_0, \dots, \alpha_p}),$$

with morphisms

$$\delta_r : E_r^{p,q} \rightarrow E_r^{p+r, q-r+1},$$

where

$$E_{r+1} \cong H_{\delta_r}(E_r).$$

This spectral sequence converges to the cohomology of the union, i.e.

$$E_r^{p,q} \Rightarrow H^{p+q}(A),$$

and consequently

$$\text{rank } H^i(A) = \sum_{p+q=i} \text{rank } E_{\infty}^{p,q}. \quad (4.6)$$

Also, this spectral sequence collapses at E_n , i.e.

$$E_{\infty}^{n-1,0} \cong E_n^{n-1,0}.$$

Corollary 4.2.1 (of Theorem 4.2.1) *Let A_1, \dots, A_s be random simplicial complexes.*

Consider the same definitions as in Theorem 4.2.1. For every $r \geq 0$, define $e_r^{a,b} := \mathbb{E}[\text{rank } E_r^{a,b}]$. We have

$$e_{r+1}^{p,q} \leq e_r^{p,q}, \quad (4.7)$$

and, when $E_r^{p+r, q-r+1} = 0$,

$$e_{r+1}^{p,q} \geq e_r^{p,q} - e_r^{p-r, q+r-1}. \quad (4.8)$$

Proof Follows immediately from the deterministic versions of the same statements, which in turn follow from the structure of the differentials, i.e., specifically the fact that

$$E_{r+1}^{p,q} \cong \text{Ker}(\delta_r : E_r^{p,q} \rightarrow E_r^{p+r,q-r+1}) / \text{Im}(\delta_r : E_r^{p-r,q+r-1} \rightarrow E_r^{p,q}).$$

■

4.2.3 Average Betti numbers of hypersurface arrangements

Proof [Proof of Theorem 4.1.1] We give the proof for the spherical case; the asymptotic for projective case needs to be divided by two. By Theorem 4.2.1,

$$E_1^{p,q} \cong \bigoplus_{\alpha_0 < \dots < \alpha_p} H^q(A_{\alpha_0, \dots, \alpha_p}).$$

In our case, we have random complexes $A_{\alpha_0, \dots, \alpha_p}$, and we need two results. First is the result by (Edelman and Kostlan, 1995; Kostlan, 1993; Shub and Smale, 1993b), which gives the precise value of the expected rank of $\bigoplus_{\alpha_0 < \dots < \alpha_{n-1}} H^0(A_{\alpha_0, \dots, \alpha_{n-1}})$:

$$e_1^{n-1,0} = 2 \sum_{\substack{I \subset [s] \\ |I|=n}} \sqrt{\prod_{i \in I} d_i}. \quad (4.9)$$

Next, we need a bound that follows immediately from (Gayet and Welschinger, 2015). They prove that for a smooth real projective manifold X , $\mathbb{E}[b_i(X)] \leq O(\sqrt{d^{\dim(X)}})$. Noting that there are a total of $\binom{s}{p+1}$ such manifolds, we have:

$$e_1^{p,q} \leq \binom{s}{p+1} O(d^{(n-p-1)/2}), \quad (4.10)$$

for any $p < n - 1$.

Denoting the reduced Betti numbers of a manifold by $\tilde{b}_*(\cdot)$, by the Alexander-Pontryagin duality, we have

$$b_{n-1}(\Gamma) = \tilde{b}_{n-1}(\Gamma) = \tilde{b}_0(S^n \setminus \Gamma) = b_0(S^n \setminus \Gamma) - 1,$$

thus

$$\begin{aligned} \mathbb{E}[b_0(S^n \setminus \Gamma)] &= \mathbb{E}[b_{n-1}(\Gamma)] + 1 && \text{(Alexander-Pontryagin duality)} \\ &= \sum_{k=1}^n e_{\infty}^{n-k, k-1} + 1 && \text{(by (4.6) and linearity of expectation).} \end{aligned} \quad (4.11)$$

First, observe that

$$\begin{aligned} \sum_{k=2}^n e_{\infty}^{n-k, k-1} &\leq \sum_{k=2}^n e_1^{n-k, k-1} && \text{(by (4.7))} \\ &\leq \sum_{k=2}^n \binom{s}{n-k+1} O(d^{(k-1)/2}) && \text{(by (4.10))} \\ &\leq s^{n-1} O(d^{(n-1)/2}). \end{aligned} \quad (4.12)$$

Now it remains to give precise bounds on $e_{\infty}^{n-1,0}$, which is the same as as obtaining precise bounds on $e_n^{n-1,0}$, given that the spectral sequence collapses at E_n (cf. Theorem 4.2.1). Clearly,

$$e_{\infty}^{n-1,0} = e_n^{n-1,0} \leq e_1^{n-1,0} = 2 \sum_{\substack{I \subset [s] \\ |I|=n}} \sqrt{\prod_{i \in I} d_i}.$$

For the lower bound, we make repeated use of Equation (4.8). Note that (4.8) is true only when $E_r^{p+r, q-r+1} = 0$, which happens when $r > q + 1$ since we have a first-

quadrant spectral sequence. Thus (4.8) holds for all $e_r^{p,0}$ when $r > 1$. Telescoping, we get

$$\begin{aligned}
e_\infty^{n-1,0} &= e_n^{n-1,0} \\
&\geq e_{n-1}^{n-1,0} - e_{n-1}^{0,n-2} && \text{(by (4.8))} \\
&\geq e_{n-1}^{n-1,0} - e_1^{0,n-2} && \text{(by (4.7))} \\
&\geq e_{n-2}^{n-1,0} - e_{n-2}^{1,n-3} - e_1^{0,n-2} && \text{(by (4.8))} \\
&\geq e_{n-2}^{n-1,0} - e_1^{1,n-3} - e_1^{0,n-2} && \text{(by (4.7))} \\
&\vdots \\
&\geq e_1^{n-1,0} - \left(\sum_{i=0}^{n-2} e_1^{i,n-2-i} \right) \\
&\geq 2 \sum_{\substack{I \subset [s] \\ |I|=n}} \sqrt{\prod_{i \in I} d_i} - \left(\sum_{i=0}^{n-2} \binom{s}{i+1} O(d^{(n-i-1)/2}) \right) && \text{(by (4.9), (4.10)).}
\end{aligned}$$

Thus,

$$2 \sum_{\substack{I \subset [s] \\ |I|=n}} \sqrt{\prod_{i \in I} d_i} \geq e_\infty^{n-1,0} \geq 2 \sum_{\substack{I \subset [s] \\ |I|=n}} \sqrt{\prod_{i \in I} d_i} - s^{n-1} O(d^{(n-1)/2}). \quad (4.13)$$

Putting Equations (4.13) and (4.12) in Equation (4.11) completes the proof of the theorem. ■

Below we give a corollary of Theorem 4.1.1 which gives a bound on the sum of the Betti numbers of $\mathbb{R}P^n \setminus \Gamma$ (we prove the corollary for the spherical case, again one has to divide the asymptotics by two in the projective case).

Corollary 4.2.2 (n fixed) *Let Γ be defined as in Theorem 4.1.1. Then, for all $k > 0$,*

$$\mathbb{E}[b_k(S^n \setminus \Gamma)] = O(d^{(n-1)/2} s^{n-k}). \quad (4.14)$$

Consequently,

$$\mathbb{E} \left[\sum_{i=0}^{n-1} b_i(S^n \setminus \Gamma) \right] = 2 \sum_{\substack{I \subset [s] \\ |I|=n}} \sqrt{\prod_{i \in I} d_i} + O(d^{(n-1)/2} s^{n-1}). \quad (4.15)$$

Proof By Alexander-Pontryagin duality, when $k > 0$,

$$b_k(S^n \setminus \Gamma) = \tilde{b}_k(S^n \setminus \Gamma) = \tilde{b}_{n-k-1}(\Gamma) \leq b_{n-k-1}(\Gamma),$$

thus

$$\begin{aligned} \mathbb{E} [b_k(S^n \setminus \Gamma)] &\leq \mathbb{E} [b_{n-k-1}(\Gamma)] \\ &= \sum_{i=0}^{n-k-1} e_{\infty}^{i, n-k-i-1} \\ &\leq \sum_{i=0}^{n-k-1} e_1^{i, n-k-i-1} \\ &\leq \sum_{i=0}^{n-k-1} \binom{s}{i+1} O(d^{(n-i-1)/2}) \quad (\text{by (4.10)}) \\ &\leq s^{n-k} O(d^{(n-1)/2}), \end{aligned}$$

proving Equation (4.14). Using this, Equation (4.1) of Theorem 4.1.1, and linearity of expectation, Equation (4.15) follows immediately. \blacksquare

Thus the expected total Betti number of $\mathbb{R}P^n \setminus \Gamma$ has the same order as that of its number of connected components.

4.3 Obstacle Random Graphs and an Application to Arrangement of Quadrics

In this section, we study the top Betti number of $\mathbb{R}P^n \setminus \Gamma$, when Γ is the union of a finite set of quadrics. It turns out that in this case, the problem of understanding the number of connected components of Γ is related to the connectivity of a certain random graph model.

In the study of the topological complexity of arrangements of hypersurfaces, there are two sets of parameters that play a part. First is the sequence of degrees of the polynomials defining the hypersurfaces. Second is the number of polynomials in the arrangement. The former is often called the ‘algebraic part’ and the latter is called the ‘combinatorial part’. While the algebraic part is indeed important, in several applications, for instance in discrete and computational geometry, it is the combinatorial part of the complexity that is of paramount interest. This is because one typically encounters arrangements of a large number of objects, where each object has “bounded complexity”.

Theorem 4.1.1 and Corollary 4.2.2 together suggest that in arrangements of s random hypersurfaces, the top Betti number of the complement of the union of the arrangement grows linearly in s . In line with many results where the growth of the Betti numbers of semi-algebraic sets defined by quadratic inequalities is shown to be different, in this section we prove a bound on the average top Betti number of the complement of the union of an arrangement of Kostlan quadrics that is sub-linear in s . In Section 4.3.1, we introduce our random graph model which we call “Obstacle” random graphs. In Section 4.3.2, we prove a theorem (Theorem 4.3.1) about the average number of connected components in this random graph model. Then, in Section 4.3.3, using a theorem of Calabi (Theorem 4.3.2), we obtain a result on the average zeroth Betti number of Γ (Theorem 4.1.2), when Γ is a finite union of the zero sets of quadrics.

4.3.1 The ‘Obstacle’ random graph model

In this section we introduce the obstacle random graph model.

Definition 4.3.1 (‘Obstacle’ random graph) *Let $\{q_1, \dots, q_s\} \subset \mathbb{R}P^N$ be a sample from the uniform distribution on $\mathbb{R}P^N$, and let $\mathcal{P} \subset \mathbb{R}P^N$ (the “obstacle”) be a measurable convex set. We define the obstacle random graph model $\mathcal{G}(N, \mathcal{P}, s)$ as follows:*

1. $\mathcal{G}(N, \mathcal{P}, s)$ has s vertices $\{q_1, \dots, q_s\}$.
2. Define $\ell(q_i, q_j) := \{[\lambda_a q_i + \lambda_b q_j]\}_{[\lambda_a, \lambda_b] \in \mathbb{R}P^1}$. The edge set is defined as the set of unordered pairs

$$\{(q_i, q_j) \mid 1 \leq i < j \leq s \text{ and } \ell(q_i, q_j) \cap \mathcal{P} = \emptyset\}.$$

In other words, it is an undirected graph where the vertices are $\{q_1, \dots, q_s\}$, and for every pair of distinct vertices q_i, q_j has an undirected edge if and only if the great circle connecting the vertices does not intersect \mathcal{P} .

This model bears some similarity to random visibility graphs (De Berg et al., 2000). See Figure 4.2 for an example illustration.

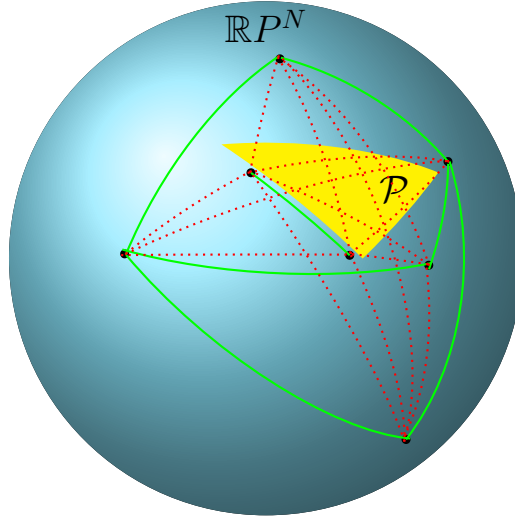


Figure 4.2.: Illustration of obstacle random graph. The thick lines denote edges of the graph, while the dotted lines denote non-edges, i.e. edges that were not included in the random graph because their geodesic completion intersected \mathcal{P} .

Remark 4.3.1 Two commonly studied random graph models are the Erdős-Rényi model (proposed in (Erdős and Rényi, 1959; Gilbert, 1959)), and the geometric random graph model (proposed in (Gilbert, 1961)).

- In the obstacle random graph, the edge probabilities are random variables, and the random variables are not independent. Thus this model is dissimilar to the Erdős-Rényi model.
- Define the metric $d : \mathbb{R}P^N \times \mathbb{R}P^N \rightarrow \mathbb{R}$, where

$$d(q, q') = \begin{cases} 0 & q_1 = q_2 \\ 1 & \ell(q_1, q_2) \cap \mathcal{P} \neq \emptyset \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

While our graph is a geometric random graph on s vertices with an edge appearing between two distinct vertices q_1, q_2 when $d(q_1, q_2) \leq \frac{1}{2}$, note that d is a non-continuous function that is difficult to work with, and thus standard results in the geometric random graph literature do not apply.

4.3.2 Average number of connected components of obstacle random graphs

We shall now study the average number of connected components in the obstacle random graph model $\mathcal{G}(N, \mathcal{P}, s)$ as $s \rightarrow \infty$. Specifically, we prove the following theorem.

Theorem 4.3.1 (N fixed, $s \rightarrow \infty$) *Consider $\mathcal{G}(N, \mathcal{P}, s)$, the obstacle random graph model as per Definition 4.3.1. Then*

$$\lim_{s \rightarrow \infty} \frac{\mathbb{E}[b_0(\mathcal{G}(N, \mathcal{P}, s))]}{s} \leq \frac{\text{vol}(\mathcal{P})}{\text{vol}(\mathbb{R}P^N)}.$$

Below is a synopsis of the proof of Theorem 4.3.1.

1. A simple first step is Proposition 4.3.1 where we understand the distribution of the number of vertices in various regions in $\mathbb{R}P^N$. Specifically, Proposition 4.3.1 gives tail bounds on the number of vertices in $\mathcal{P}, \mathcal{P}(\varepsilon) \setminus \mathcal{P}$ (where $\mathcal{P}(\varepsilon)$ is the ε -neighbourhood of \mathcal{P}) and $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$.

2. The second and final (and most involved) step is the proof of Lemma 4.3.1 which proves that the subgraph of $\mathcal{G}(N, \mathcal{P}, s)$ restricted to $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ has number of connected components constant w.r.t. s . The proof of Lemma 4.3.1 involves the following sub-steps.

- (a) Cover $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ with balls of radius $r > 0$, r to be chosen later.
- (b) We define the *good cone* of a point p w.r.t. \mathcal{P} as the set of all points in $\mathbb{R}P^N \setminus \mathcal{P}$ such that an edge would appear between the point and p . Then for each r -ball B , we proceed to lower bound the probability (Lemma 4.3.2) of choosing a point in $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ such that the good cone of the point contains B . This involves showing that the volume of the good cone (also to be defined later) of a point is a continuous function of the position of the point. We prove this by first considering a smooth approximation of \mathcal{P} containing \mathcal{P} and contained in $\mathcal{P}(\varepsilon)$ (Proposition 4.3.3), and then applying a stereographic projection and proving continuity in Euclidean space (Lemma 4.3.3).
- (c) Finally, a geometric coupon-collector type argument (Lemma 4.3.4) gives tail bounds on the number of points required for all r -balls to be contained in good cones. This ensures that any new point sampled in $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ will not add a new connected component to the graph.

Sample s points q_1, \dots, q_s i.i.d. from the uniform distribution on $\mathbb{R}P^N$. Let $\mathcal{P}(\varepsilon)$ be the ε -neighbourhood of \mathcal{P} in $\mathbb{R}P^N$. Define the random variables

$$s_e(\varepsilon) = \sum_{i=1}^s \mathbb{1} \{q_i \in \mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)\},$$

which is the number of points in $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$,

$$s_a(\varepsilon) = \sum_{i=1}^s \mathbb{1} \{q_i \in \mathcal{P}(\varepsilon) \setminus \mathcal{P}\},$$

which is the number of points in $\mathcal{P}(\varepsilon) \setminus \mathcal{P}$, and

$$s_p = \sum_{i=1}^s \mathbb{1}_{\{q_i \in \mathcal{P}\}},$$

which is the number of points in \mathcal{P} . Obviously,

$$s = s_e(\varepsilon) + s_a(\varepsilon) + s_p.$$

Now, let $\Omega_1(\varepsilon), \Omega_2(\varepsilon), \Omega_3$ be the following defined events:

$$\begin{aligned} \Omega_1(\varepsilon) &= \left\{ s_e(\varepsilon) = s \cdot \left(1 - \frac{\text{vol}(\mathcal{P}(\varepsilon))}{\text{vol}(\mathbb{R}P^N)} \right) \pm o(\sqrt{s}) \right\}, \\ \Omega_2(\varepsilon) &= \left\{ s_a(\varepsilon) = s \cdot \left(\frac{\text{vol}(\mathcal{P}(\varepsilon) \setminus \mathcal{P})}{\text{vol}(\mathbb{R}P^N)} \right) \pm o(\sqrt{s}) \right\}, \\ \Omega_3 &= \left\{ s_p = s \cdot \left(\frac{\text{vol}(\mathcal{P})}{\text{vol}(\mathbb{R}P^N)} \right) \pm o(\sqrt{s}) \right\}. \end{aligned}$$

Below we have a simple proposition that gives tail bounds on the random variables $s_e(\varepsilon)$, $s_a(\varepsilon)$ and s_p .

Proposition 4.3.1 *For all $0 < \delta < 1$, $\varepsilon > 0$, $\alpha > 0$, there exists $\tilde{s}_1 = \tilde{s}_1(\delta, \alpha) = \left(\frac{2}{\log 6/\delta} \right)^{2\alpha}$, such that if $s > \tilde{s}_1$,*

$$\mathbb{P}[\Omega_1(\varepsilon)^c], \mathbb{P}[\Omega_2(\varepsilon)^c], \mathbb{P}[\Omega_3^c] < \frac{\delta}{3},$$

and consequently, for all $\varepsilon > 0$,

$$\mathbb{P}[\Omega_1(\varepsilon) \cap \Omega_2(\varepsilon) \cap \Omega_3] > 1 - \delta.$$

This also implies that for all $\varepsilon > 0$,

$$\lim_{s \rightarrow \infty} \mathbb{P}[\Omega_1(\varepsilon) \cap \Omega_2(\varepsilon) \cap \Omega_3] = 1.$$

We will need the additive Chernoff-Hoeffding bound for Binomial random variables.

Proposition 4.3.2 (See for e.g. (Boucheron et al., 2013)) *For a random variable $X \sim \text{Binomial}(n, p)$,*

$$\mathbb{P}[X < \mathbb{E}[X] - t], \mathbb{P}[X > \mathbb{E}[X] + t] < e^{-2t^2/n}.$$

Consequently, if $n \geq \tilde{n} = \tilde{n}(t, \delta) = \frac{2t^2}{\log 2/\delta}$,

$$\mathbb{P}[|X - \mathbb{E}[X]| > t] < \delta.$$

Proof [Proof of Proposition 4.3.1] Obviously $s_e(\varepsilon)$, $s_a(\varepsilon)$ and s_p are Binomial random variables. Note that

$$\mathbb{E}[s_e(\varepsilon)] = s \cdot \left(1 - \frac{\text{vol}(\mathcal{P}(\varepsilon))}{\text{vol}(\mathbb{R}P^N)}\right).$$

By Proposition 4.3.2, we have that

$$\mathbb{P}[\Omega_1(\varepsilon)^c] \leq \mathbb{P}\left[\left|s_e(\varepsilon) - s \cdot \left(1 - \frac{\text{vol}(\mathcal{P}(\varepsilon))}{\text{vol}(\mathbb{R}P^N)}\right)\right| > s^{1/2-\alpha}\right] < \frac{\delta}{3}. \quad (4.16)$$

Similarly, by noting that

$$\mathbb{E}[s_a(\varepsilon)] = s \cdot \left(\frac{\text{vol}(\mathcal{P}(\varepsilon) \setminus \mathcal{P})}{\text{vol}(\mathbb{R}P^N)}\right),$$

and

$$\mathbb{E}[s_p] = s \cdot \left(\frac{\text{vol}(\mathcal{P})}{\text{vol}(\mathbb{R}P^N)}\right),$$

again by Proposition 4.3.2, we have that

$$\mathbb{P}[\Omega_2(\varepsilon)^c] \leq \mathbb{P}\left[\left|s_a(\varepsilon) - s \cdot \left(\frac{\text{vol}(\mathcal{P}(\varepsilon) \setminus \mathcal{P})}{\text{vol}(\mathbb{R}P^N)}\right)\right| > s^{1/2-\alpha}\right] < \frac{\delta}{3}, \quad (4.17)$$

and

$$\mathbb{P}[\Omega_3^c] \leq \mathbb{P}\left[\left|s_p - s \cdot \left(\frac{\text{vol}(\mathcal{P})}{\text{vol}(\mathbb{R}P^N)}\right)\right| > s^{1/2-\alpha}\right] < \frac{\delta}{3}. \quad (4.18)$$

The first part of the claim follows by (4.16), (4.17), and (4.18), and the second part follows by applying a union bound on the equations. \blacksquare

Recall that $\mathcal{G}(N, \mathcal{P}, s)$ is the graph over all the s points q_1, \dots, q_s . Let $\mathcal{G}_1(N, \mathcal{P}, s, \varepsilon)$ denote the subgraph of $\mathcal{G}(N, \mathcal{P}, s)$ restricted to the vertices in $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$, let $\mathcal{G}_2(N, \mathcal{P}, s, \varepsilon)$ denote the subgraph of $\mathcal{G}(N, \mathcal{P}, s)$ restricted to the vertices in $\mathcal{P}(\varepsilon) \setminus \mathcal{P}$, and let $\mathcal{G}_3(N, \mathcal{P}, s)$ denote the subgraph of $\mathcal{G}(N, \mathcal{P}, s)$ restricted to the vertices in \mathcal{P} . Note that $\mathcal{G}_3(N, \mathcal{P}, s)$ contains s_p vertices and no edges whatsoever. The following lemma gives us some information of the distribution of the zeroth Betti number of $\mathcal{G}_1(N, \mathcal{P}, s, \varepsilon)$.

Lemma 4.3.1 *For all $\varepsilon > 0$, $\delta_1 > 0$, there exists $\tilde{s}_2 = \tilde{s}_2(\varepsilon, \delta_1, N)$, $a = a(\varepsilon, N)$, such that for all $s > \tilde{s}_2$*

$$\mathbb{P}\left[b_0(\mathcal{G}_1(N, \mathcal{P}, s, \varepsilon)) \leq \frac{\tilde{s}_2}{a} \mid \Omega_1(\varepsilon)\right] \geq 1 - \delta_1.$$

For any point $q \in \mathbb{R}P^N$, define

$$g_q(\mathcal{F}) = \{x \in \mathbb{R}P^N \mid \ell(q, x) \cap \mathcal{F} = \emptyset\}.$$

By definition, $g_q(\mathcal{P})$ is a random variable that denotes the set of points in $\mathbb{R}P^N$ which, if sampled, would be connected to q by an edge in $\mathcal{G}(N, \mathcal{P}, s)$. We will refer to $g_q(\mathcal{F})$ as the *good cone of q w.r.t. \mathcal{F}* , or just *good cone* if \mathcal{F} is clear from context (see Figure 4.3 for an example illustration of the good cone). The following lemma gives a lower bound on the relative volume of $g_q(\mathcal{P})$, when q is outside $\mathcal{P}(\varepsilon)$.

Lemma 4.3.2 *For $B \subseteq \mathbb{R}P^N$, $\varepsilon > 0$, define*

$$G_B(\mathcal{F}) = \{x \in \mathbb{R}P^N \setminus \mathcal{P}(\varepsilon) \mid g_x(\mathcal{F}) \supseteq B \cap (\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon))\}.$$

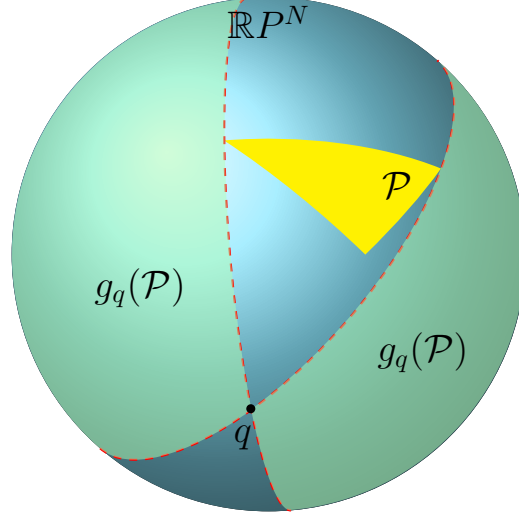


Figure 4.3.: Illustration of $g_q(\mathcal{P})$, the *good cone* of a point q w.r.t. \mathcal{P} . The dashed lines are geodesics which are tangent to \mathcal{P} and incident on q . The shaded region is $g_q(\mathcal{P})$. Recall that in $\mathcal{G}(N, \mathcal{P}, s)$, by definition, if q is sampled and any point in $g_q(\mathcal{P})$ is sampled, these points would be connected to each other by any edge.

For all $\varepsilon > 0$, there exists $r = r(\varepsilon, N) > 0$, $\delta_2 = \delta_2(\varepsilon, N)$, such that for any $p \in \mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$,

$$\frac{\text{vol}(G_{B(p,r)}(\mathcal{P}))}{\text{vol}(\mathbb{R}P^N)} \geq \delta_2.$$

Define

$$\alpha^r : \mathbb{R}P^N \setminus \text{int}(\mathcal{P}(\varepsilon)) \rightarrow [0, \infty), \quad \text{which takes } p \mapsto \frac{\text{vol}(G_{B(p,r)}(\mathcal{P}))}{\text{vol}(\mathbb{R}P^N)}.$$

where $r \leq \varepsilon/8$ is going to be chosen later. Note that since we are going to be choosing $r \leq \varepsilon/8$,

$$B(p, r) \subseteq \mathbb{R}P^N \setminus \mathcal{P}(\varepsilon/2), \quad \forall p \in \mathbb{R}P^N \setminus \mathcal{P}(\varepsilon).$$

Remark 4.3.2 Observe that the convex set $\mathcal{P} \subset \mathbb{R}P^N$ is contained in one single affine chart, and therefore if we denote by $f : S^N \rightarrow \mathbb{R}P^N$ the double cover map, the

preimage $f^{-1}(\mathcal{P}_n)$ (which for simplicity we still denote by \mathcal{P}) is entirely contained in a open hemisphere, which we assume it is

$$U = \text{int}B\left(e_0, \frac{\pi}{2}\right) \subset S^N$$

for some point $e_0 \in S^N$. Let us denote now by

$$\sigma : U \rightarrow \mathbb{R}^N$$

the stereographic projection constructed as follows: we identify \mathbb{R}^n with $T_{e_0}S^N$ and for every point $y \in U$ we take $\sigma(y)$ to be the point of intersection between $T_{e_0}S^N$ and the line from the origin to y . This stereographic projection has an interesting property that we will use: it maps (unparametrized) geodesics entirely contained in U , i.e. intersections between U and great circles, to (unparametrized) geodesics in \mathbb{R}^N , i.e. straight segments. In particular σ maps convex sets to convex sets, and the same is true for its inverse. In particular we can use results from convex geometry in \mathbb{R}^N to obtain results for the convex geometry of U . Since $f|_U : U \rightarrow \mathbb{R}P^N$ is a local isometry onto its image, the same is true for the geometry of convex sets in $\mathbb{R}P^N$.

The next proposition is an application of the idea explained in Remark 4.3.2.

Proposition 4.3.3 *For all $\varepsilon > 0$, there exists a smooth convex set $\tilde{\mathcal{P}}(\varepsilon)$ such that $\mathcal{P} \subseteq \tilde{\mathcal{P}}(\varepsilon) \subseteq \mathcal{P}(\varepsilon)$.*

Proof Consider the $\varepsilon/2$ -neighbourhood of \mathcal{P} , i.e. $\mathcal{P}(\varepsilon/2)$. Since the set of smooth convex bodies is dense in the Hausdorff distance induced topology on the space of convex bodies (see (Schneider, 2014, Theorem 2.7.1.)), there exists a body C_ε that is convex, smooth and also satisfies

$$d_H(C_\varepsilon, \mathcal{P}(\varepsilon/2)) \leq \varepsilon/3, \tag{4.19}$$

where d_H denotes Hausdorff distance with the underlying metric being the usual round metric on S^N . We shall now show that C_ε itself is the smooth approximation we desire, i.e. $\tilde{\mathcal{P}}(\varepsilon)$. We know that $d_H(\mathcal{P}, \mathcal{P}(\varepsilon/2)) = \varepsilon/2$. Observe that if \mathcal{P} was not completely contained in C_ε , then $d_H(\mathcal{P}(\varepsilon/2), C_\varepsilon) \geq \varepsilon/2$, which contradicts Equation (4.19). Similarly, it can be shown that C_ε is completely contained in $\mathcal{P}(\varepsilon)$ because otherwise, we would again have $d_H(\mathcal{P}(\varepsilon/2), C_\varepsilon) \geq \varepsilon/2$ (because $d_H(\mathcal{P}(\varepsilon/2), \mathcal{P}(\varepsilon)) = \varepsilon/2$) contradicting Equation (4.19). ■

The following lemma proves that for every r' -ball (where $r' > 0$ is appropriately chosen) contained in $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$, there is a set of positive measure such that the good cone of any point in this set contains the ball, which in turn implies that with each vertex sampled, there is a positive probability that a particular r' ball is *covered*.

Lemma 4.3.3 *For all $\varepsilon > 0$, there exists $r' = r'(\varepsilon, N)$, $\delta'_2 = \delta'_2(\varepsilon, N) > 0$ such that for any $p \in \mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$,*

$$\frac{\text{vol}(\sigma(G_{B(p, r')}(\mathcal{P})))}{\text{vol}(\sigma(S^N))} \geq \delta'_2.$$

Proof Let $\mathcal{Q}_n(\varepsilon) = \sigma(\mathcal{P}(\varepsilon)) \subseteq \mathbb{R}^N$, and $\tilde{\mathcal{Q}}(\varepsilon) = \sigma(\tilde{\mathcal{P}}(\varepsilon))$ (cf. Proposition 4.3.3). Note that for any $p \in \mathbb{R}P^N$, $g_p(\mathcal{P}) \supseteq g_p(\tilde{\mathcal{P}}(\varepsilon))$, and for any $B \subseteq \mathbb{R}P^N$, $G_B(\mathcal{P}) \supseteq G_B(\tilde{\mathcal{P}}(\varepsilon))$ (see Figure 4.4 for an illustration).

Define the map

$$\tilde{\alpha}^s : S^{N-1} \setminus \text{int}(\mathcal{Q}(\varepsilon)) \rightarrow [0, \infty), \quad \text{which takes } q \mapsto \frac{\text{vol}(\sigma(G_{B(\sigma^{-1}(q), s)}(\tilde{\mathcal{P}}(\varepsilon))))}{\text{vol}(S^{N-1})}.$$

To establish the lemma, we need to show that for an appropriately chosen r , $\tilde{\alpha}^r$ attains a minimum on its domain. As a first step, we shall show that the map

$$\tilde{\alpha}^0 : S^{N-1} \setminus \text{int}(\mathcal{Q}(\varepsilon)) \rightarrow [0, \infty), \quad \text{which takes } q \mapsto \frac{\text{vol}(\sigma(g_{\sigma^{-1}(q)}(\tilde{\mathcal{P}}(\varepsilon))))}{\text{vol}(S^{N-1})},$$

is bounded below by a continuous function.

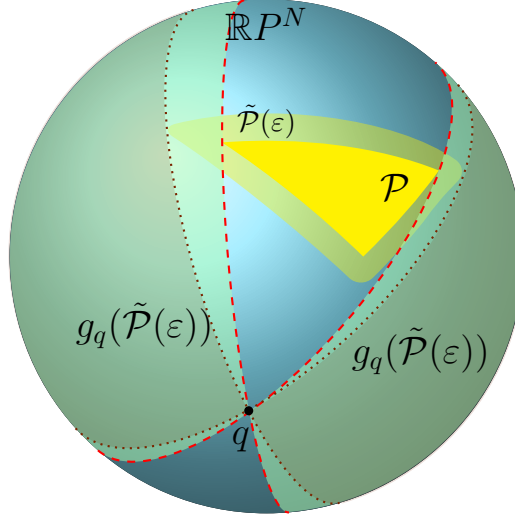
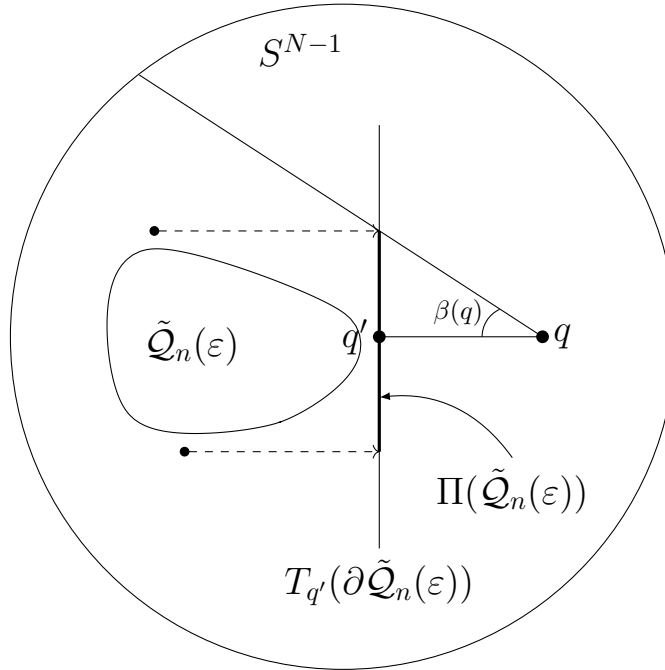


Figure 4.4.: Illustration of the good cone of q w.r.t. $\tilde{\mathcal{P}}(\varepsilon)$. $\tilde{\mathcal{P}}(\varepsilon)$ is an approximation of \mathcal{P} which is convex and has a smooth boundary, such that $\mathcal{P} \subseteq \tilde{\mathcal{P}}(\varepsilon) \subseteq \mathcal{P}(\varepsilon)$. The dashed lines are geodesics which are tangent to \mathcal{P} and incident on q , and the dotted lines are geodesics which are tangent to $\tilde{\mathcal{P}}(\varepsilon)$ and incident on q . Observe that $g_q(\tilde{\mathcal{P}}(\varepsilon)) \subseteq g_q(\mathcal{P})$, and consequently, $\text{vol}(g_q(\tilde{\mathcal{P}}(\varepsilon))) \leq \text{vol}(g_q(\mathcal{P}))$.



Let q' be the point shortest to q on $\partial\tilde{\mathcal{Q}}_n(\varepsilon)$, the boundary of $\tilde{\mathcal{Q}}_n(\varepsilon)$. Let $\Pi(\tilde{\mathcal{Q}}_n(\varepsilon))$ be the projection of $\tilde{\mathcal{Q}}_n(\varepsilon)$ onto $T_{q'}(\partial\tilde{\mathcal{Q}}_n(\varepsilon))$, the tangent space of $\tilde{\mathcal{Q}}_n(\varepsilon)$ at q' . Observe that

$$\lambda(q) = \max_{v \in \Pi(\tilde{\mathcal{Q}}_n(\varepsilon))} \|v\|_2$$

is continuous. Consequently, observe that

$$\frac{\text{vol}\left(\sigma\left(g_{\sigma^{-1}(q)}(\tilde{\mathcal{P}}_n(\varepsilon))\right)\right)}{\text{vol}(S^{N-1})} \geq 1 - \underbrace{\frac{\text{vol}\left(\text{spherical cap with angle } \tan^{-1}\left(\frac{\lambda(q)}{2\|q-q'\|_2}\right)\right)}{\text{vol}(S^{N-1})}}_{\beta(q)}.$$

$\beta(q)$ is a continuous function, and thus attains a maximum on $S^{N-1} \setminus \text{int}(\mathcal{Q}_n(\varepsilon))$ (remember that $\|q - q'\|_2$ can never become 0 because q is always outside $\mathcal{P}(\varepsilon)$) proving that $\tilde{\alpha}^0$ is bounded below by a continuous function that attains a minimum on its domain.

From this, we have that for every $p \in \mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$, we can find a direction \vec{v}' in \mathbb{R}^N and an angle θ such that for all directions \vec{v} with $\cos^{-1} \frac{\vec{v} \cdot \vec{v}'}{\|\vec{v}\|_2 \|\vec{v}'\|_2} \leq \theta$, we have that $\ell_\sigma(p, \vec{v}) \subseteq \sigma\left(g_p(\tilde{\mathcal{P}}(\varepsilon))\right)$, where $\ell_\sigma(p, \vec{v})$ denotes the line in \mathbb{R}^N through $\sigma(p)$ in the direction \vec{v} . Note that \vec{v}' and θ depend on p continuously. Let $p_{v'}$ be the point of intersection of the line $\ell_\sigma(p, \vec{v}')$ and S^{N-1} , and now let p_2 be the midpoint on the line joining p and $p_{v'}$. Since θ depends on p continuously, it has a minimum on $\mathbb{R}P^N \setminus \text{int}(\mathcal{P}(\varepsilon))$, and thus we can pick $r'' = r''(\varepsilon, N) > 0$ such that $B(p_2, r'') \subseteq \sigma\left(g_p(\tilde{\mathcal{P}}(\varepsilon))\right)$.

Now, for the sake of contradiction, assume that for all $r' > 0$,

$$\min_{q \in S^{N-1} \setminus \text{int}(\mathcal{Q}_n(\varepsilon))} \tilde{\alpha}^{r'}(q) = 0,$$

and let q be the point at which $\tilde{\alpha}^{r'}$ attains the minimum. Then we can find a sequence (r_n) , with $r_n \rightarrow 0$, and a sequence (q_n) , with $q_n \rightarrow q$, where $q_n \in B(p, r_n)$, such that for all $n < \infty$, there exists a point $b_n \in B(p_2, r'')$ with $b_n \notin g_{q_n}$. Since $S^{N-1} \setminus \text{int}(\mathcal{Q}(\varepsilon))$ is compact, and $B(p_2, r'')$ is obviously compact as well, this means

that $(\lim_{n \rightarrow \infty} b_n) \notin (\lim_{n \rightarrow \infty} g_{q_n})$, implying that there is a point in $B(p_2, r'')$ which does not belong to g_q , which gives us the contradiction we require. ■

Proof [Proof of Lemma 4.3.2] Set $r = \min(r', \varepsilon/8)$. The proof of the lemma follows by noting that since σ is smooth, bijective and angle-preserving (conformal)², proving that there is a set of strictly positive measure that is good for all r -balls centered in $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ follows from Lemma 4.3.3. This is because since $\delta'_2 > 0$, the pre-image under σ of any set of measure at least δ'_2 will be strictly positive (δ_2 will be the measure of the pre-image, under σ , of the set in \mathbb{R}^N which attains the minimum measure δ'_2). ■

The lemma below gives bounds on the number of samples from $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ required to cover all of $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ with good cones.

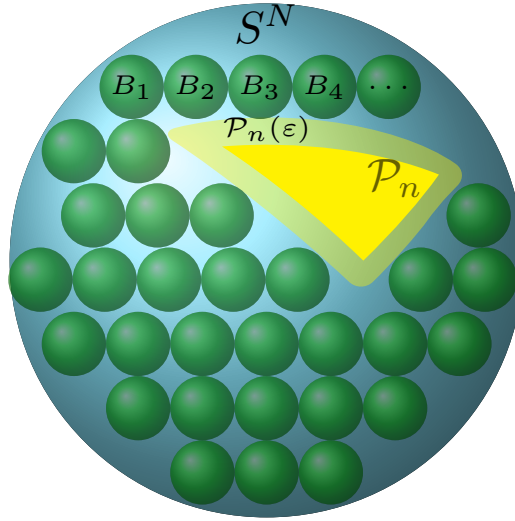


Figure 4.5.: Illustration of the proof of Lemma 4.3.1 in a nutshell – cover the complement of the fattening of $\mathcal{P}(\varepsilon)$ with balls, show that each ball has positive probability of being *covered*, and then finish with a coupon-collector type argument.

²Note that the stereographic projection is not isometric, and thus does not preserve areas. However, angle-preservation is enough for us.

Lemma 4.3.4 *For any $\varepsilon > 0$, define $C = C(\varepsilon)$ to be a random variable that denotes that number of points q'_1, \dots, q'_C needed outside $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ s.t.*

$$\bigcup_{i=1}^C g_{q'_i}(\mathcal{P}_n) \supseteq \mathbb{R}P^N \setminus \mathcal{P}(\varepsilon).$$

Then, for all $\delta_3 > 0$, there exists $\alpha = \alpha(\varepsilon, \delta_3, N)$ such that

$$\mathbb{P}[C \leq \alpha] \geq 1 - \delta_3.$$

Proof Take a covering of $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ with r -balls (where r is from Lemma 4.3.2) of size $Q = Q(\varepsilon, N)$, and let the Q balls that cover $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ be B_1, \dots, B_Q . Remember that the conditional distribution of sampling from $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ is uniform. Let C_i denote the additional number of points needed to be sampled from $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ such that B_i is covered, given that balls B_1, \dots, B_{i-1} are already covered by $\bigcup_{i=1}^{C_{i-1}} g_{q'_i}$. By definition,

$$C \leq \sum_{i=1}^Q C_i.$$

When balls B_1, \dots, B_{i-1} are already covered, B_i could already be covered. Let the probability that B_i is already covered be p_i . If not covered, by Lemma 4.3.2, each C_i is a geometric random variable with parameter $\mu_i \geq \delta_2$. This means

$$C_i = \begin{cases} 0 & \text{with prob. } p_i \\ \text{Geom}(\mu_i) & \text{with prob. } 1 - p_i \end{cases}.$$

Thus

$$\mathbb{E}[C_i] = 0 \cdot p_i + (1 - p_i) \cdot \frac{1}{\mu_i} \leq \frac{1}{\mu_i} \leq \frac{1}{\delta_2},$$

and by linearity of expectation, in turn, we get that

$$\mathbb{E}[C] \leq \frac{Q}{\delta_2}. \tag{4.20}$$

Set $\alpha = \frac{Q}{\delta_2 \delta_3}$. Applying Markov's inequality on C , and using Equation (4.20), the lemma follows. \blacksquare

Remark 4.3.3 *The proof of Lemma 4.3.4 is similar to the coupon collector problem (see for e.g. (Isaac, 1995) for a description). The balls represent coupons. When the good cone of a sampled point encapsulates a ball, this is equivalent to collecting a coupon. See Figure 4.5 for an illustration.*

Proof [of Lemma 4.3.1] The above lemma shows that we will have a covering of $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ with *good sets*, with probability at least $1 - \delta_3$, if we have $s_e(\varepsilon) \geq \alpha$. To complete the proof of Lemma 4.3.1, we have to set \tilde{s}_2 appropriately so that if $s \geq \tilde{s}_2$, then $s_e(\varepsilon) \geq \alpha$. Conditioning on $\Omega_1(\varepsilon)$, it is clear that if $s \geq k \cdot \alpha \left(\frac{\text{vol}(\mathbb{R}P^N)}{\text{vol}(\mathbb{R}P^N) - \text{vol}(\mathcal{P}(\varepsilon))} \right)$, for an appropriately chosen constant k , then $s_e(\varepsilon) \geq \alpha$. Thus, conditioned on $\Omega_1(\varepsilon)$, setting $\tilde{s}_2 = k \cdot \alpha \left(\frac{\text{vol}(\mathbb{R}P^N)}{\text{vol}(\mathbb{R}P^N) - \text{vol}(\mathcal{P}(\varepsilon))} \right)$ ensures we have a covering of $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ with *good sets* with probability at least $1 - \delta_3$.

Since, $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ is covered, any new point that is added to $\mathbb{R}P^N \setminus \mathcal{P}(\varepsilon)$ will be connected to at least one of the existing α vertices, which in turn means that the number of connected components of the graph stays fixed as α . The lemma follows by setting $a = k \left(\frac{\text{vol}(\mathbb{R}P^N)}{\text{vol}(\mathbb{R}P^N) - \text{vol}(\mathcal{P}(\varepsilon))} \right)$. \blacksquare

Proof [of Theorem 4.3.1] We shall prove that, for all $\varepsilon, \delta, \delta_1$, $\lim_{s \rightarrow \infty} \frac{\mathbb{E}[b_0(\mathcal{G}(N, \mathcal{P}, s))]}{s}$ is bounded from above by $\frac{\text{vol}(\mathcal{P}_n)}{\text{vol}(\mathbb{R}P^N)}$ plus some terms which depend on $\varepsilon, \delta, \delta_1$. We know that the number of connected components of a graph is bounded from above by the

sum of the number of connected components of subgraphs of the graph that form a decomposition of the original graph. Thus, for any $\varepsilon > 0$, we can estimate

$$\begin{aligned}
& \mathbb{E} [b_0(\mathcal{G}(N, \mathcal{P}, s))] \\
& \leq \mathbb{E} [b_0(\mathcal{G}_1(N, \mathcal{P}, s, \varepsilon))] + \mathbb{E} [b_0(\mathcal{G}_2(N, \mathcal{P}, s, \varepsilon))] + \mathbb{E} [b_0(\mathcal{G}_3(N, \mathcal{P}, s))] \\
& \leq \underbrace{\int_{\Omega_1(\varepsilon) \cap (b_0(\mathcal{G}_1(N, \mathcal{P}, s, \varepsilon)) \leq \tilde{s}_2/a)} b_0(\mathcal{G}_1(N, \mathcal{P}, s, \varepsilon)) d\omega}_A + \underbrace{\int_{\Omega_1(\varepsilon)^c} b_0(\mathcal{G}_1(N, \mathcal{P}, s, \varepsilon)) d\omega}_B \\
& + \underbrace{\int_{\Omega_1(\varepsilon) \cap (b_0(\mathcal{G}_1(N, \mathcal{P}, s, \varepsilon)) \leq \tilde{s}_2/a)^c} b_0(\mathcal{G}_1(N, \mathcal{P}, s, \varepsilon)) d\omega}_C + \underbrace{\int_{\Omega_2(\varepsilon)} b_0(\mathcal{G}_2(N, \mathcal{P}, s, \varepsilon)) d\omega}_D \\
& + \underbrace{\int_{\Omega_2(\varepsilon)^c} b_0(\mathcal{G}_2(N, \mathcal{P}, s, \varepsilon)) d\omega}_E + \underbrace{\int_{\Omega_3} b_0(\mathcal{G}_3(N, \mathcal{P}, s)) d\omega}_F \\
& + \underbrace{\int_{\Omega_3^c} b_0(\mathcal{G}_3(N, \mathcal{P}, s)) d\omega}_G, \tag{4.21}
\end{aligned}$$

where the \tilde{s}_2 and a are from Lemma 4.3.1. Because we are integrating over the space where $b_0(\mathcal{G}_1(N, \mathcal{P}, s, \varepsilon)) \leq \tilde{s}_2/a$, obviously,

$$A \leq \frac{\tilde{s}_2}{a}. \tag{4.22}$$

We apply the trivial bound of s on $b_0(\mathcal{G}_1(N, \mathcal{P}, s, \varepsilon))$ to get that, for all $\delta > 0$,

$$B \leq \mathbb{P} [\Omega_1(\varepsilon)^c] s \leq \frac{\delta}{3} s, \tag{4.23}$$

as long as $s \geq \tilde{s}_1 = \tilde{s}_1(\delta, \alpha)$, where $\alpha > 0$ is any constant (cf. Proposition 4.3.1). By Lemma 4.3.1, for all $\delta_1 > 0$, if $s > \tilde{s}_2 = \tilde{s}_2(\varepsilon, \delta_1, N)$,

$$\mathbb{P} [b_0(\mathcal{G}_1(N, \mathcal{P}, s, \varepsilon)) > \tilde{s}_2/a \mid \Omega_1(\varepsilon)] < \delta_1,$$

for some specific $a = a(\varepsilon, N)$. Thus,

$$C \leq \mathbb{P}[\Omega_1(\varepsilon)] \cdot \mathbb{P}[b_0(\mathcal{G}_1(N, \mathcal{P}, s, \varepsilon)) > \tilde{s}_2/a \mid \Omega_1(\varepsilon)] s \leq \mathbb{P}[\Omega_1(\varepsilon)] \delta_1 s \leq \delta_1 s. \quad (4.24)$$

Trivially, b_0 of a graph is bounded from above by the number of vertices in the graph.

Thus,

$$D \leq \mathbb{P}[\Omega_2(\varepsilon)] \left(s \cdot \left(\frac{\text{vol}(\mathcal{P}(\varepsilon) \setminus \mathcal{P})}{\text{vol}(\mathbb{R}P^N)} \right) + o(\sqrt{s}) \right) \leq \varepsilon s + o(\sqrt{s}). \quad (4.25)$$

At the same time, as in case of (4.23), for all $\delta > 0$,

$$E \leq s \cdot \mathbb{P}[\Omega_2(\varepsilon)^c] \leq s \frac{\delta}{3}, \quad (4.26)$$

if $s \geq \tilde{s}_1 = \tilde{s}_1(\delta, \alpha)$, with $\alpha > 0$ any constant (by Proposition 4.3.1). By Equation (4.31), we have that

$$F \leq \mathbb{P}[\Omega_3] \left(s \cdot \left(\frac{\text{vol}(\mathcal{P})}{\text{vol}(\mathbb{R}P^N)} \right) + o(\sqrt{s}) \right) \leq s \cdot \left(\frac{\text{vol}(\mathcal{P})}{\text{vol}(\mathbb{R}P^N)} \right) + o(\sqrt{s}). \quad (4.27)$$

Finally, again, for all $\delta > 0$, if $s \geq \tilde{s}_1 = \tilde{s}_1(\delta, \alpha)$, $\alpha > 0$ any constant,

$$G \leq s \cdot \mathbb{P}[\Omega_3^c] \leq s \frac{\delta}{3}. \quad (4.28)$$

Putting equations (4.22), (4.23), (4.24), (4.25), (4.26), (4.27), (4.28) in (4.21), we have that for all $\varepsilon > 0$, $\delta > 0$, $\delta_1 > 0$,

$$\lim_{s \rightarrow \infty} \frac{\mathbb{E}[b_0(\mathcal{G}(N, \mathcal{P}, s))]}{s} \leq \underbrace{0}_{A/s} + \underbrace{\frac{\delta}{3}}_{B/s} + \underbrace{\delta_1}_{C/s} + \underbrace{\varepsilon}_{D/s} + \underbrace{\frac{\delta}{3}}_{E/s} + \underbrace{\frac{\text{vol}(\mathcal{P})}{\text{vol}(\mathbb{R}P^N)}}_{F/s} + \underbrace{\frac{\delta}{3}}_{G/s}. \quad (4.29)$$

Since Equation (4.29) is true for any choice of $\varepsilon, \delta, \delta_1$, we have that

$$\lim_{s \rightarrow \infty} \frac{\mathbb{E}[b_0(\mathcal{G}(N, \mathcal{P}, s))]}{s} \leq \frac{\text{vol}(\mathcal{P})}{\text{vol}(\mathbb{R}P^N)}.$$



4.3.3 b_0 of arrangement of quadrics

Once we fix a scalar product on \mathbb{R}^{n+1} , there is a natural isomorphism between the vector space $\text{Sym}(n+1, \mathbb{R})$ of real symmetric matrices and the space $\mathbb{R}[x_0, \dots, x_n]_{(2)}$, which is given by associating to a symmetric matrix Q the quadratic form defined by $q(x) = \langle x, Qx \rangle$. It turns out that the Kostlan measure is the pushforward of the GOE^3 measure under this linear isomorphism (see for e.g. (Lerario and Lundberg, 2016) for a discussion about this), i.e.:

$$Q \text{ is a GOE matrix} \iff q \text{ is a Kostlan polynomial.}$$

Let $\mathbb{R}P^N = P(\text{Sym}(n, \mathbb{R}))$ be the projectivization of the space of symmetric matrices (here $N = \binom{n+2}{2} - 1$) and consider the set $\mathcal{P}_n \subset \mathbb{R}P^N$ which is the projectivization of the set of positive definite matrices (equivalently of the set of *positive* quadratic forms):

$$\mathcal{P}_n = \{[Q] \in \mathbb{R}P^N \mid Q > 0\}.$$

We endow $\text{Sym}(n+1, \mathbb{R})$ with the Frobenius metric, which corresponds to the Bombieri-Weil metric under the above linear isomorphism; on the projective space $\mathbb{R}P^N$ we consider the quotient Riemannian metric (for this metric the quotient map $p : S^N \rightarrow \mathbb{R}P^N$ is a local isometry), with corresponding volume density. In this way, if q is a random Kostlan quadric, we have:

$$\mathbb{P}[\{q \text{ is a positive form}\}] = \frac{\text{vol}(\mathcal{P}_n)}{\text{vol}(\mathbb{R}P^N)}. \quad (4.30)$$

³Stands for Gaussian Orthogonal Ensemble (see (Tao, 2012) for a description).

Remark 4.3.4 *The relative volume of \mathcal{P}_n in $\mathbb{R}P^N$ is known (see e.g. (Majumdar et al., 2011)) to decay exponentially fast when n increases:*

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \log \left(\frac{\text{vol}(\mathcal{P}_n)}{\text{vol}(\mathbb{R}P^N)} \right) = -\frac{\log 3}{4}. \quad (4.31)$$

The following result, which is due to Calabi (Calabi, 1964), gives a geometric criterion for two quadrics intersecting in projective space.

Theorem 4.3.2 (Calabi, 1964) *For $n \geq 1$ let $q_1, q_2 \in \mathbb{R}[x_0, \dots, x_n]_{(2)}$ and denote by $\Gamma_1, \Gamma_2 \subset \mathbb{R}P^n$ their (possibly empty) zero sets. Define $\ell(q_1, q_2) \subset \mathbb{R}P^N$ to be the projective line $\ell(q_1, q_2) := \{[\lambda_1 q_1 + \lambda_2 q_2]\}_{[\lambda_1, \lambda_2] \in \mathbb{R}P^1}$ (a pencil of quadrics). Then:*

$$\Gamma_1 \cap \Gamma_2 \neq \emptyset \iff \ell \cap \mathcal{P}_n = \emptyset.$$

One can refer to (Lerario, 2012) for a proof of this using spectral sequences. Relying on Calabi's Theorem, and using Theorem 4.3.1, we shall now prove Theorem 4.1.2.

Proof [Proof of Theorem 4.1.2] As a consequence of Calabi's Theorem (Theorem 4.3.2), studying the average zeroth Betti number of Γ is equivalent to studying the average number of connected components in the obstacle random graph model, i.e. studying the average number of connected components of $\mathcal{G}(N, \mathcal{P}_n, s)$.

In fact nonempty quadrics in projective space are connected and therefore the number of connected components of Γ in this case equals the number of connected components of the incidence graph of the zero sets $Z(q_i)$ of the sampled quadrics. This incidence graph is a subgraph of the corresponding obstacle graph – we must discard the points that fall inside \mathcal{P}_n because the zero sets of quadrics in \mathcal{P}_n is empty. Thus

$$\begin{aligned}
\lim_{s \rightarrow \infty} \frac{\mathbb{E}[b_0(\Gamma)]}{s} &= \lim_{s \rightarrow \infty} \frac{\mathbb{E}[b_0(\mathcal{G}(N, \mathcal{P}_n, s)) - \sum_{i=1}^s \mathbb{1}\{q_i \in \mathcal{P}_n\}]}{s} \\
&= \lim_{s \rightarrow \infty} \frac{\mathbb{E}[b_0(\mathcal{G}(N, \mathcal{P}_n, s))]}{s} - \frac{s \cdot \mathbb{P}[q \in \mathcal{P}_n]}{s} \\
&\leq \frac{\text{vol}(\mathcal{P}_n)}{\text{vol}(\mathbb{R}P^N)} - \frac{\text{vol}(\mathcal{P}_n)}{\text{vol}(\mathbb{R}P^N)} \quad (\text{by Theorem 4.3.1 and (4.30)}) \\
&= 0.
\end{aligned} \tag{4.32}$$

Equation (4.32) together with the fact that $\lim_{s \rightarrow \infty} \frac{\mathbb{E}[b_0(\Gamma)]}{s}$ is obviously non-negative completes the proof. ■

4.3.4 A Ramsey-type result

Semi-algebraic graphs have been studied from the point of view of Ramsey theory. (Alon et al., 2005) prove the following theorem.

Theorem 4.3.3 ((Alon et al., 2005)) *For any semi-algebraic graph $G = (V, E)$, there exists a constant $\varepsilon > 0$, and two sets $V_1, V_2 \subset V$, each with size at least $\varepsilon|V|$, such that either $V_1 \times V_2 \subset E$, or $(V_1 \times V_2) \cap E = \emptyset$. Consequently, there exists another constant $\delta > 0$, and $V' \subset V$ of size at least $|V|^\delta$ such that $V' \times V' \subset E$ or $(V' \times V') \cap E = \emptyset$. In other words, one of the following is true:*

1. *There exists a clique of size n^δ in G .*
2. *The complement of G has a clique of size n^δ .*

The quadrics graph, i.e. Γ , is a subgraph of $\mathcal{G}(N, \mathcal{P}_n, s)$. It is formed by discarding the vertices that fall inside \mathcal{P}_n (because the zero sets of quadrics inside \mathcal{P}_n are empty). In Γ , an edge is placed between vertices if the corresponding quadrics intersect, thus it is clear that Γ is a semi-algebraic graph. The following result rules out the probability of large cliques in the complement graph of Γ .

Corollary 4.3.1 (of Theorem 4.1.2) *Let Γ be the graph of quadrics as defined in Theorem 4.1.2. Denote by Γ^c the complement of the graph Γ on the same set of vertices. Then, for any $\varepsilon > 0$,*

$$\lim_{s \rightarrow \infty} \mathbb{P}[\Gamma^c \text{ contains a clique of size } \varepsilon s] = 0.$$

Proof Let Ω_a denote the event that there exists a clique of size εs in Γ^c . Thus we have

$$\begin{aligned} 0 &= \lim_{s \rightarrow \infty} \frac{\mathbb{E}[b_0(\Gamma)]}{s} && \text{(by Theorem 4.1.2)} \\ &= \lim_{s \rightarrow \infty} \frac{\int_{\Omega_a} b_0(\Gamma) d\omega + \int_{\Omega_a^c} b_0(\Gamma) d\omega}{s} \\ &\geq \lim_{s \rightarrow \infty} \frac{\varepsilon s \cdot \mathbb{P}[\Omega_a] + 0}{s}. \end{aligned} \tag{4.33}$$

The final step follows by noting that if the complement of Γ contains a clique of size εs , it means that all εs vertices were isolated in Γ , in turn implying that Γ has at least εs connected components. The corollary follows by Equation (4.33) and by noting that $\lim_{s \rightarrow \infty} \mathbb{P}[\Omega_a]$ is obviously non-negative. ■

Juxtaposing with Theorem 4.3.3, Corollary 4.3.1 proves that, in the quadrics random graph, among the two conditions of Theorem 4.3.3, a condition stricter than (2) holds with probability 0 as the number of vertices tends to infinity.

4.4 More studies of the topology of random arrangements

The current study leaves a number of other open questions. First, we begin with the definition of a sign condition.

Definition 4.4.1 A sign condition σ on a tuple of polynomials $\vec{P} = (P_1, \dots, P_s)$, where each $P_i \in \mathbb{R}[X_1, \dots, X_n]$, is an element of $\{-1, 1\}^s$. The realization of $\sigma = (\sigma_1, \dots, \sigma_s)$ over \vec{P} , denoted $\mathcal{R}(\sigma, \vec{P})$ is defined as

$$\mathcal{R}(\sigma, \vec{P}) = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid \text{sign}(P_i(x_1, \dots, x_n)) = \sigma_i \text{ for all } 1 \leq i \leq s\}.$$

Given a tuple of polynomials, not all sign conditions are realizable. For instance, if we consider the tuple $(x - y, x - y - 5)$, the sign condition $\{-1, 1\}$ is not realizable. Trivially speaking, there could be 2^s different sign conditions. (Warren, 1968) gives an upper bound on the number of sign conditions which depends on parameters such as degrees of P_i , in addition to s . It would be interesting to study questions about sign conditions when the P_i are Kostlan distributed.

Question 2 Given a tuple of Kostlan distributed polynomials $\vec{P} = (P_1, \dots, P_s)$, $P_i \in \mathbb{R}[X_1, \dots, X_n]$, each of degree D_i , what is the probability that a sign condition is realizable, i.e.

$$\mathbb{P} \left[\mathcal{R}(\sigma, \vec{P}) \neq \emptyset \right] = ?$$

Note that the number of realizable sign conditions is a trivial lower bound on the number of connected components of the complement of the union of the zero sets of the polynomials, i.e. given $P = (P_1, \dots, P_s)$,

$$\left| \{ \sigma \in \{-1, 1\}^s \mid \mathcal{R}(\sigma, \vec{P}) \neq \emptyset \} \right| \leq b_0 \left(\mathbb{R}^n \setminus \left(\bigcup_{i=1}^s Z(P_i) \right) \right).$$

For Kostlan distributed $\vec{P} = (P_1, \dots, P_s)$, let R be a random variable that denotes the total number of realizable sign conditions. We have that

$$R = \sum_{\sigma \in \{-1, 1\}^s} \mathbb{1} \left\{ \mathcal{R}(\sigma, \vec{P}) \neq \emptyset \right\}. \quad (4.34)$$

If we know the answer to Question 2, or more realistically, if we know bounds on $\mathbb{P}[\mathcal{R}(\sigma, \vec{P}) \neq \emptyset]$, we will be able to obtain bounds on $\mathbb{E}[R]$ by just using linearity of expectation on Equation (4.34).

Next, by Equation 4.4, we have that

$$\mathbb{E}[R] \leq \mathbb{E} \left[b_0 \left(\mathbb{R}^n \setminus \left(\bigcup_{i=1}^s Z(P_i) \right) \right) \right],$$

and by Theorem 4.1.1, we know $\mathbb{E}[b_0(\mathbb{R}^n \setminus (\bigcup_{i=1}^s Z(P_i)))]$. Any ‘gap’ between $\mathbb{E}[R]$ and $\mathbb{E}[b_0(\mathbb{R}^n \setminus (\bigcup_{i=1}^s Z(P_i)))]$ can be studied further by studying the average Betti numbers of realizable sign conditions.

Question 3 *Given a tuple of Kostlan distributed polynomials $\vec{P} = (P_1, \dots, P_s)$, $P_i \in \mathbb{R}[X_1, \dots, X_n]$, each of degree D_i , obtain bounds on*

$$\mathbb{E} \left[b_i \left(\mathcal{R}(\sigma, \vec{P}) \right) \mid \mathcal{R}(\sigma, \vec{P}) \neq \emptyset \right]?$$

REFERENCES

- Adler, R. J. and Taylor, J. E. (2009). *Random fields and geometry*. Springer Science & Business Media.
- Agarwal, P. and Sharir, M. (2000). Arrangements and their applications. In J.R. Sack, J. U., editor, *Handbook of computational geometry*, pages 49–119. North-Holland, Amsterdam.
- Agarwal, P. K., Aronov, B., Ezra, E., and Zahl, J. (2019). An efficient algorithm for generalized polynomial partitioning and its applications. In *35th International Symposium on Computational Geometry, SoCG 2019, June 18-21, 2019, Portland, Oregon, USA*.
- Alon, N., Pach, J., Pinchasi, R., Radoičić, R., and Sharir, M. (2005). Crossing patterns of semi-algebraic sets. *J. Combin. Theory Ser. A*, 111(2):310–326.
- Aronov, B., Ezra, E., and Zahl, J. (2019). Constructive polynomial partitioning for algebraic curves in \mathbb{R}^3 with applications. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2636–2648. SIAM.
- Barone, S. and Basu, S. (2012). Refined bounds on the number of connected components of sign conditions on a variety. *Discrete & Computational Geometry*, 47(3):577–597.
- Barone, S. and Basu, S. (2016). On a real analog of bezout inequality and the number of connected components of sign conditions. *Proceedings of the London Mathematical Society*, 112(1):115–145.
- Barvinok, A. I. (1997). On the Betti numbers of semialgebraic sets defined by few quadratic inequalities. *Math. Z.*, 225(2):231–244.
- Basu, S. (2003). Different bounds on the different Betti numbers of semi-algebraic sets. *Discrete Comput. Geom.*, 30(1):65–85. ACM Symposium on Computational Geometry (Medford, MA, 2001).
- Basu, S. (2009). Combinatorial complexity in o-minimal geometry. *Proceedings of the London Mathematical Society*, 100(2):405–428.
- Basu, S. (2017). Algorithms in real algebraic geometry: a survey. In *Real algebraic geometry*, volume 51 of *Panor. Synthèses*, pages 107–153. Soc. Math. France, Paris.
- Basu, S., Lerario, A., and Natarajan, A. (2019a). Betti numbers of random hypersurface arrangements. *arXiv preprint arXiv:1911.13256*.
- Basu, S., Lerario, A., and Natarajan, A. (2019b). Zeroes of polynomials on definable hypersurfaces: pathologies exist, but they are rare. *The Quarterly Journal of Mathematics*, 70(4):1397–1409.

- Basu, S., Pasechnik, D. V., and Roy, M.-F. (2010). Bounding the Betti numbers and computing the Euler-Poincaré characteristic of semi-algebraic sets defined by partly quadratic systems of polynomials. *J. Eur. Math. Soc. (JEMS)*, 12(2):529–553.
- Basu, S., Pollack, R., and Roy, M.-F. (1996). On the number of cells defined by a family of polynomials on a variety. *Mathematika*, 43(1):120–126.
- Basu, S., Pollack, R., and Roy, M.-F. (2005a). Betti number bounds, applications and algorithms. *Current trends in combinatorial and computational geometry: papers from the special program at MSRI*, 52:87–97.
- Basu, S., Pollack, R., and Roy, M.-F. (2005b). On the betti numbers of sign conditions. *Proceedings of the American Mathematical Society*, 133(4):965–974.
- Basu, S., Pollack, R., and Roy, M. F. (2006). Algorithms in real algebraic geometry. *Algorithms and Computation in Mathematics*, 10.
- Basu, S. and Raz, O. E. (2017a). An o-minimal Szemerédi-Trotter theorem. *The Quarterly Journal of Mathematics*, 69(1):223–239.
- Basu, S. and Raz, O. E. (2017b). An o-minimal Szemerédi-Trotter theorem. *Oberwolfach Report 22/2017*, pages 1–17.
- Basu, S. and Rizzie, A. (2018). Multi-degree bounds on the betti numbers of real varieties and semi-algebraic sets and applications. *Discrete & Computational Geometry*, 59(3):553–620.
- Ben-Or, M. (1983). Lower bounds for algebraic computation trees. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, pages 80–86. ACM.
- Benedetti, R. and Risler, J.-J. (1991). *Real algebraic and semialgebraic sets*. Hermann.
- Bochnak, J., Coste, M., and Roy, M.-F. (2013). *Real algebraic geometry*, volume 36. Springer Science & Business Media.
- Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration inequalities*. Oxford University Press, Oxford. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.
- Breiding, P. and Lerario, A. (2019). Lectures on random algebraic geometry.
- Bürgisser, P. and Cucker, F. (2004). Variations by complexity theorists on three themes of Euler, Bézout, Betti, and Poincaré. In *Complexity of computations and proofs*, volume 13 of *Quad. Mat.*, pages 73–151. Dept. Math., Seconda Univ. Napoli, Caserta.
- Burgisser, P. and Lerario, A. (2018). Probabilistic schubert calculus. *Crelle’s journal*. DOI: <https://doi.org/10.1515/crelle-2018-0009>.
- Calabi, E. (1964). Linear systems of real quadratic forms. *Proc. Amer. Math. Soc.*, 15:844–846.
- Chernikov, A., Galvin, D., and Starchenko, S. (2020). Cutting lemma and Zarankiewicz’s problem in distal structures. *Selecta Math. (N.S.)*, 26(2):Paper No. 25.

- Chernikov, A. and Starchenko, S. (2018). Regularity lemma for distal structures. *J. Eur. Math. Soc. (JEMS)*, 20(10):2437–2466.
- Chernikov, A., Starchenko, S., and Thomas, M. E. (2016). Ramsey growth in some nip structures. *Journal of the Institute of Mathematics of Jussieu*, pages 1–29.
- Coste, M. (2000a). *An introduction to o-minimal geometry*. Istituti Editoriali e Poligrafici Internazionali, Pisa. Dip. Mat. Univ. Pisa, Dottorato di Ricerca in Matematica.
- Coste, M. (2000b). *An Introduction to Semialgebraic Geometry*. Istituti editoriali e poligrafici internazionali Pisa.
- De Berg, M., Van Kreveld, M., Overmars, M., and Schwarzkopf, O. C. (2000). Visibility graphs. In *Computational geometry*, pages 307–317. Springer.
- Edelman, A. and Kostlan, E. (1995). How many zeros of a random polynomial are real? *Bull. Amer. Math. Soc. (N.S.)*, 32(1):1–37.
- Edelman, A., Kostlan, E., and Shub, M. (1994). How many eigenvalues of a random matrix are real? *J. Amer. Math. Soc.*, 7(1):247–267.
- Erdős, P. and Rényi, A. (1959). On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297.
- Fyodorov, Y. V., Lerario, A., and Lundberg, E. (2015). On the number of connected components of random algebraic hypersurfaces. *J. Geom. Phys.*, 95:1–20.
- Gabrielov, A. and Vorobjov, N. (2004). Complexity of computations with pfaffian and noetherian functions. *Normal forms, bifurcations and finiteness problems in differential equations*, pages 211–250.
- Gabrielov, A. and Vorobjov, N. (2017). On topological lower bounds for algebraic computation trees. *Foundations of Computational Mathematics*, 17(1):61–72.
- Gayet, D. and Welschinger, J.-Y. (2015). Expected topology of random real algebraic submanifolds. *Journal of the Institute of Mathematics of Jussieu*, 14(4):673–702.
- Gayet, D. and Welschinger, J.-Y. (2016). Betti numbers of random real hypersurfaces and determinants of random symmetric matrices. *J. Eur. Math. Soc. (JEMS)*, 18(4):733–772.
- Gilbert, E. N. (1959). Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144.
- Gilbert, E. N. (1961). Random plane networks. *Journal of the Society for Industrial and Applied Mathematics*, 9(4):533–543.
- Goodman, J. E. and Pollack, R. (1986a). There are asymptotically far fewer polytopes than we thought. *Bull. Amer. Math. Soc. (N.S.)*, 14(1):127–129.
- Goodman, J. E. and Pollack, R. (1986b). Upper bounds for configurations and polytopes in \mathbf{R}^d . *Discrete Comput. Geom.*, 1(3):219–227.
- Grothendieck, A. (1997). Esquisse d’un programme. *London Mathematical Society Lecture Note Series*, pages 5–48.

- Guth, L. (2015). Polynomial partitioning for a set of varieties. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 159, pages 459–469. Cambridge University Press.
- Guth, L. and Katz, N. H. (2015). On the Erdős distinct distances problem in the plane. *Ann. of Math. (2)*, 181(1):155–190.
- Gwoździewicz, J., Kurdyka, K., and Parusiński, A. (1999). On the number of solutions of an algebraic equation on the curve $y = e^x + \sin x, x > 0$, and a consequence for o-minimal structures. *Proceedings of the American Mathematical Society*, 127(4):1057–1064.
- Howard, R. (1993). The kinematic formula in Riemannian homogeneous spaces. *Mem. Amer. Math. Soc.*, 106(509):vi+69.
- Isaac, R. (1995). *The pleasures of probability*. Undergraduate Texts in Mathematics. Springer-Verlag, New York. Readings in Mathematics.
- Kac, M. (1943). On the average number of real roots of a random algebraic equation. *Bull. Amer. Math. Soc.*, 49:314–320.
- Kac, M. (1949). On the average number of real roots of a random algebraic equation. II. *Proc. London Math. Soc. (2)*, 50:390–408.
- Kaplan, H., Matoušek, J., Safernová, Z., and Sharir, M. (2012a). Unit distances in three dimensions. *Combinatorics, Probability and Computing*, 21(4):597–610.
- Kaplan, H., Matoušek, J., and Sharir, M. (2012b). Simple proofs of classical theorems in discrete geometry via the guth–katz polynomial partitioning technique. *Discrete & Computational Geometry*, 48(3):499–517.
- Khazhgali Kozhasov, A. L. (2017). On the number of flats tangent to convex hyper-surfaces in random position. *arXiv:1702.06518*.
- Khovanskii, A. (1991). *Fewnomials*, volume 88. American Mathematical Soc.
- Knight, J. F., Pillay, A., and Steinhorn, C. (1986). Definable sets in ordered structures. ii. *Transactions of the American Mathematical Society*, 295(2):593–605.
- Kollár, J. (2017). Nash’s work in algebraic geometry. *Bull. Amer. Math. Soc. (N.S.)*, 54(2):307–324.
- Kostlan, E. (1993). On the distribution of roots of random polynomials. In *From Topology to Computation: Proceedings of the Smalefest (Berkeley, CA, 1990)*, pages 419–431. Springer, New York.
- Kostlan, E. (2002). On the expected number of real roots of a system of random polynomial equations. In *Foundations of computational mathematics (Hong Kong, 2000)*, pages 149–188. World Sci. Publ., River Edge, NJ.
- Lerario, A. (2012). Convex pencils of real quadratic forms. *Discrete & Computational Geometry*, 48(4):1025–1047.
- Lerario, A. (2015). Random matrices and the average topology of the intersection of two quadrics. *Proc. Amer. Math. Soc.*, 143(8):3239–3251.

- Lerario, A. (2016). Complexity of intersections of real quadrics and topology of symmetric determinantal varieties. *J. Eur. Math. Soc. (JEMS)*, 18(2):353–379.
- Lerario, A. and Lundberg, E. (2016). Gap probabilities and betti numbers of a random intersection of quadrics. *Discrete & Computational Geometry*, 55(2):462–496.
- Littlewood, J. E. and Offord, A. C. (1938). On the Number of Real Roots of a Random Algebraic Equation. *J. London Math. Soc.*, 13(4):288–295.
- Majumdar, S. N., Nadal, C., Scardicchio, A., and Vivo, P. (2011). How many eigenvalues of a gaussian random matrix are positive? *Physical Review E*, 83(4):041105.
- Marker, D. (2006). *Model theory: an introduction*, volume 217. Springer Science & Business Media.
- Matoušek, J. and Patáková, Z. (2015). Multilevel polynomial partitions and simplified range searching. *Discrete & Computational Geometry*, 54(1):22–41.
- McCleary, J. (2001). *A user’s guide to spectral sequences*. Number 58. Cambridge University Press.
- Milnor, J. (1964). On the betti numbers of real varieties. *Proceedings of the American Mathematical Society*, 15(2):275–280.
- Oleinik, O. and Petrovsky, I. (1949). On the topology of real algebraic hypersurfaces. *Izv. Acad. Nauk SSSR*, 13:389–402.
- Pila, J. (2011). O-minimality and the andré-oort conjecture for c^n . *Annals of mathematics*, pages 1779–1840.
- Pillay, A. and Steinhorn, C. (1986). Definable sets in ordered structures. i. *Transactions of the American Mathematical Society*, 295(2):565–592.
- Pillay, A. and Steinhorn, C. (1988). Definable sets in ordered structures. iii. *Transactions of the American Mathematical Society*, 309(2):469–476.
- Robinson, J. (1949). Definability and decision problems in arithmetic. *The Journal of Symbolic Logic*, 14(2):98–114.
- Schneider, R. (2014). *Convex bodies: the Brunn - Minkowski theory*. Number 151. Cambridge University Press.
- Seifert, H. (1936). Algebraische Approximation von Mannigfaltigkeiten. *Math. Z.*, 41(1):1–17.
- Shub, M. and Smale, S. (1993a). Complexity of Bézout’s theorem. I. Geometric aspects. *J. Amer. Math. Soc.*, 6(2):459–501.
- Shub, M. and Smale, S. (1993b). Complexity of Bezout’s theorem. II. Volumes and probabilities. In *Computational algebraic geometry (Nice, 1992)*, volume 109 of *Progr. Math.*, pages 267–285. Birkhäuser Boston, Boston, MA.
- Shub, M. and Smale, S. (1993c). Complexity of Bezout’s theorem. III. Condition number and packing. *J. Complexity*, 9(1):4–14. Festschrift for Joseph F. Traub, Part I.

- Solymosi, J. and Tao, T. (2012). An incidence theorem in higher dimensions. *Discrete Comput. Geom.*, 48(2):255–280.
- Tao, T. (2012). *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI.
- Thom, R. (1965). Sur l’homologie des variétés algébriques réelles. *Differential and combinatorial topology*, pages 255–265.
- Tressl, M. (2010). Introduction to o-minimal structures and an application to neural network learning. *Preprint*.
- van den Dries, L. (1984). Remarks on Tarski’s problem concerning $(\mathbf{R}, +, \cdot, \exp)$. In *Logic colloquium ’82 (Florence, 1982)*, volume 112 of *Stud. Logic Found. Math.*, pages 97–121. North-Holland, Amsterdam.
- Van den Dries, L. (1986). A generalization of the tarski-seidenberg theorem, and some nondefinability results. *Bulletin of the American Mathematical Society*, 15(2):189–193.
- van den Dries, L. (1998). *Tame topology and o-minimal structures*, volume 248 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge.
- Warren, H. E. (1968). Lower bounds for approximation by nonlinear manifolds. *Trans. Amer. Math. Soc.*, 133:167–178.
- Welschinger, J.-Y. (2015). Topology of random real hypersurfaces. *Rev. Colombiana Mat.*, 49(1):139–160.
- Wilkie, A. J. (1996). Model completeness results for expansions of the ordered field of real numbers by restricted pfaffian functions and the exponential function. *Journal of the American Mathematical Society*, 9(4):1051–1094.
- Wilkie, A. J. (1999). A theorem of the complement and some new o-minimal structures. *Selecta Mathematica, New Series*, 5(4):397–421.
- Yao, A. C.-C. (1997). Decision tree complexity and betti numbers. *Journal of computer and system sciences*, 55(1):36–43.