Betti numbers of random hypersurface arrangements

Saugata Basu\(^1\) | Antonio Lerario\(^2\) | Abhiram Natarajan\(^1\)

\(^1\)Purdue University, West Lafayette, Indiana, USA
\(^2\)SISSA (Trieste) and Florida Atlantic University, Trieste, Italy

**Correspondence**
Abhiram Natarajan, Purdue University, West Lafayette, IN 47906, USA.
Email: abhiram.natarajan@alumni.purdue.edu

**Funding information**
NSF, Grant/Award Numbers: CCF-1618981, DMS-1620271, CCF-1910441, CCF-1910441

**Abstract**
We study the expected behavior of the Betti numbers of arrangements of the zeros of random (distributed according to the Kostlan distribution) polynomials in \(\mathbb{RP}^n\). Using a random spectral sequence, we prove an asymptotically exact estimate on the expected number of connected components in the complement of such hypersurfaces in \(\mathbb{RP}^n\). We also investigate the same problem in the case where the hypersurfaces are defined by random quadratic polynomials. In this case, we establish a connection between the Betti numbers of such arrangements with the expected behavior of a certain model of a randomly defined geometric graph. While our general result implies that the average zeroth Betti number of the union of random hypersurface arrangements is bounded from above by a function that grows linearly in the number of polynomials in the arrangement, using the connection with random graphs, we show an upper bound on the expected zeroth Betti number of random quadrics arrangements that is sublinear in the number of polynomials in the arrangement. This bound is a consequence of a general result on the expected number of connected components in our random graph model which could be of independent interest.

**MSC 2020**
14F25 (primary)
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 [1] 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.

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 [4] 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 \( s^nO(d)^n \) 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, [12, 17, 18]). 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 \( d_1, \ldots, d_s \). 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) [15, 22]. Specifically, on the space of homogenous polynomials of degree \( d \) in \( n+1 \) variables, a Kostlan form is defined as

\[
P(x) = \sum_{(\alpha_0, \ldots, \alpha_n)} \xi_{\alpha} \cdot x_{\alpha_0} \cdots x_{\alpha_n},
\]

where \( \xi_{\alpha} \sim \mathcal{N}(0, \frac{d!}{\alpha_0! \cdots \alpha_n!}) \) 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 \$1.1\) treats the problem in full generality without any restriction on the degrees (cf. Theorem 1). We then study the case when all the degrees are assumed to be equal to 2 (outlined in \$1.2\). This is the first non-trivial case, since for an arrangement of hyperplanes (that is, 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 [3, 5, 8, 25] for the deterministic case and [24, 26] 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 §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 2).

In order to prove Theorem 2, we study the behavior of a special kind of geometrically defined graph from a random viewpoint (outlined in §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, [2]). 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 subset $P \subset \mathbb{R}P^N$ and sample independent points $q_1, \ldots, q_s$ 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 $P$. We give a tight estimate on the expected number of isolated points of such a graph (cf. Theorem 3), from which we can deduce Theorem 2. Finally, we conclude by proving a Ramsey-type result about the random graph of quadrics (cf. Corollary 17).

1.1 | Random hypersurface arrangements

We are given random homogenous polynomials $P_1, \ldots, P_s$, where each $P_i \in \mathbb{R}[x_0, \ldots, x_n]_{(d_i)}$, and we look at the random arrangement of hypersurfaces defined in the projective space by the zero sets of these polynomials, that is,

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

where each $\Gamma_j$ is the real algebraic hypersurface given by the zero set of $P_j$, that is,

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

The main problem that we want to address concerns understanding the topological complexity of $\Gamma$, which will be measured by its Betti numbers.$^\dagger$

We observe that there are three sets of parameters that will play a role in our study: the degree sequence $d_1, \ldots, d_s$ 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 $P_1, \ldots, P_s$ will be independent samples from the Kostlan ensemble.)

Our first result concerns the asymptotic when $n$ is kept fixed and $d_1, \ldots, d_s, s \to \infty$ and gives information on the number of cells of $\mathbb{R}P^n \setminus \Gamma$. There is clearly an analogous statement for the

$^\dagger$ For a semialgebraic set $S$, we denote by $b_i(S)$ its $i$th Betti number with coefficients in $\mathbb{Z}/2\mathbb{Z}$. 
spherical version of this problem, and the two cases can be related 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 1** \((n \text{ fixed})\). Let \(P_1, \ldots, P_s \in \mathbb{R}[x_0, \ldots, x_n]\) 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 \(d = \max(d_1, \ldots, d_s)\). Then:

\[
\mathbb{E} [b_0(\mathbb{R}P^n \setminus \Gamma)] = \sum_{I \subset [s], |I| = n} \sqrt{\prod_{i \in I} d_i} + O(d^{(n-1)/2}s^{n-1}). \tag{1.1}
\]

Moreover, if all the degrees are the same \(d_1 = \cdots = d_s = d\), we have:

\[
\mathbb{E} [b_0(\mathbb{R}P^n \setminus \Gamma)] = \binom{s}{n} d^{n/2} + O(d^{(n-1)/2}s^{n-1}). \tag{1.2}
\]

**Remark 1.** As we will prove in Corollary 7, 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 (1.2)). 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 [6]:

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

**Remark 2.** It is possible to produce estimates for the expected number of cells also for other invariant distributions (classified in [22]), 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:

\[
\mathbb{E} [b_0(\mathbb{R}P^n \setminus \Gamma)] = \Theta(d^n s^n). \tag{1.3}
\]

We sketch a proof of this estimate in Remark 6.

### 1.2 Arrangements of random quadrics

The next result deals instead with the asymptotic structure of \(\Gamma\) when \(d_1, \ldots, d_s = 2, n\) is fixed, and \(s \to \infty\). It turns out that in this case, the problem of understanding the number of connected components of \(\Gamma\), that is, \(b_0(\Gamma)\) (Betti numbers of \(\Gamma\) and \(\mathbb{R}P^n \setminus \Gamma\) are related by the Alexander–Pontryiagin 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 2 $(n \geq 2 \text{ fixed}, s \to \infty)$. Let $P_1, \ldots, P_s \in \mathbb{R}[x_0, \ldots, x_n]$ be homogeneous Kostlan quadrics, and $n \geq 2$. 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

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

Remark 3. In the case $n = 1$, the expected number of zeroes in $\mathbb{R}P^1$ of a random Kostlan quadric $P$ is $\sqrt{2}$. Since the probability that two independent quadrics have a zero in common is zero, it follows that, when $n = 1$, $\mathbb{E}[b_0(\Gamma)] = s \sqrt{2}$.

Remark 4. The topology of a random intersection of quadrics has been studied in [24, 26], 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 $n \to \infty$ 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 2 (cf. Corollary 17), we rule out the existence of linear sized cliques in the complement of the quadrics graph. This must be contrasted with a result in [2] who prove a Ramsey type result (cf. Theorem 16) about existence of sub-linear sized cliques in general semi-algebraic graphs.

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 $q_1, \ldots, q_s$ 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 $q_1, \ldots, q_s$ are the quadrics.) The vertices of the random graph are points $\{v_1, \ldots, v_s\}$ (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 $G(P, s)$. 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 $s \cdot \frac{\text{vol}(P)}{\text{vol}(\mathbb{R}P^N)}$ many isolated points (this is the expected number of points falling inside $\mathcal{P}$). In Theorem 3, we prove that to the leading order there are no other isolated points.

Theorem 3 ($P \subset \mathbb{R}P^N$ fixed, $s \to \infty$). The expected number of connected component of the obstacle random graph satisfies

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

The connection between Theorems 3 and 2 comes from an interesting result of Calabi (see Theorem 9): the common zero set of two quadrics in $\mathbb{R}P^n$ is non-empty if and only if the line joining these two quadrics (the projective pencil) does not intersect the set $P \subset \mathbb{R}P^N$ of positive
quadrics. Since nonempty quadrics in projective space are connected, the incidence graph of the random 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 $q_i \in P$.

## 2 A RANDOM SPECTRAL SEQUENCE

We will only consider semi-algebraic sets which are compact. For a compact semi-algebraic set $S$, we denote by $H^*(S)$ the cohomology groups of the constant sheaf $(\mathbb{Z}_2)_S$ on $S$. Since $S$ is compact, $H^*(S)$ is isomorphic to the cohomology group $H^*_c(S)$ (cohomology of $(\mathbb{Z}_2)_S$ with compact support). Moreover, since $S$ is semi-algebraic $H^*(S)$ is isomorphic to the singular cohomology groups of $S$ as well.

Suppose that $\Gamma_1, \ldots, \Gamma_s$ are compact semi-algebraic sets in real projective space, $\mathbb{R}P^n$. We want to study the cohomology of the union $\Gamma = \Gamma_1 \cup \cdots \cup \Gamma_s$. The following theorem follows from the Mayer–Vietoris exact sequence in cohomology for closed subspaces of a topological space (see, for example, [21, page 148]).

**Theorem 4 (MayerVietoris spectral sequence).** There exists a first quadrant cohomological spectral sequence $(E_r^{p,q}, \delta_r : E_r^{p,q} \to E_r^{p+r,q-r+1})_{r \in \mathbb{Z}_r}$, with

$$E_1^{p,q} = \bigoplus_{\alpha_0 < \cdots < \alpha_p} H^q(\Gamma_{\alpha_0, \ldots, \alpha_p}),$$

where $\Gamma_{\alpha_0, \ldots, \alpha_p} = \Gamma_{\alpha_0} \cap \cdots \cap \Gamma_{\alpha_p}$.

This spectral sequence converges to the cohomology of $\Gamma$, that is,

$$E_{\infty}^{p,q} \Rightarrow H^{p+q}(\Gamma),$$

and consequently

$$\text{rank } H^i(\Gamma) = \sum_{p+q=i} \text{rank } E_{\infty}^{p,q}. \quad (2.1)$$

Also, this spectral sequence collapses at $E_{\infty}^{1,0}$, and hence

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

**Remark 5.** An alternative way to obtain the same spectral sequence in Theorem 4 but using open covers instead of closed covers is as follows. It follows from the conic structure theorem at infinity of semi-algebraic sets (see, for example, [7, Proposition 5.49]), that is, that there exist open semi-algebraic neighborhoods $\Gamma_1^+, \ldots, \Gamma_s^+$ of $\Gamma_1, \ldots, \Gamma_s$, respectively, such that for any $1 \leq \alpha_0 < \cdots < \alpha_p \leq s$, $\Gamma_{\alpha_0, \ldots, \alpha_p}^+$ is homotopy equivalent to $\Gamma_{\alpha_0}^+ \cap \cdots \cap \Gamma_{\alpha_p}^+$, and $\Gamma^+ = \Gamma_1^+ \cup \cdots \cup \Gamma_s^+$ is homotopy equivalent to $\Gamma = \Gamma_1 \cup \cdots \cup \Gamma_s$. Then the spectral sequence in Theorem 4 is isomorphic to the spectral sequence associated to the open cover of $\Gamma^+$ by the subsets $\Gamma_1^+, \ldots, \Gamma_s^+$ (see, for example, [9, Example 15.7.1]).
Corollary 5 (of Theorem 4). Let $\Gamma_1, \ldots, \Gamma_s$ be randomly chosen closed subspaces of $\mathbb{R}P^n$. Consider the same definitions as in Theorem 4. For every $r \in \mathbb{Z}_{\geq 0}$, define $e_r^{p,q} := \mathbb{E}[\text{rank } E_r^{p,q}]$. We have

$$e_{r+1}^{p,q} \leq e_r^{p,q}, \quad (2.2)$$

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

$$e_{r+1}^{p,q} \geq e_r^{p,q} - e_{r-r,q+r-1}^{p-r,q-r}. \quad (2.3)$$

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

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

\[ \square \]

2.1 Average Betti numbers of hypersurface arrangements

Proof of Theorem 1. We give the proof for the spherical case; the asymptotic for projective case needs to be divided by two. By Theorem 4,

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

In our case, we have random closed subsets $\Gamma_{\alpha_0, \ldots, \alpha_p}$, and we need two results.

First let us recall the Edelman–Kostlan–Shub–Smale Theorem [15, 22, 29], on the expected number of common solutions in $\mathbb{R}P^n$ of a random system of equations: if $P_1, \ldots, P_n$ are random, independent, homogeneous polynomials of degrees $d_1, \ldots, d_n$, Kostlan distributed, then the expected number of common zeros in $\mathbb{R}P^n$ of these polynomials equals $\sqrt{d_1 \ldots d_n}$.

In particular, since $\Gamma_{\alpha_0, \ldots, \alpha_{n-1}}$ is the intersection in $\mathbb{R}P^n$ of $n$ random, independent, Kostlan distributed hypersurfaces of degree $d_{\alpha_0}, \ldots, d_{\alpha_{n-1}}$, the previous result gives the precise value of the expected rank of $\bigoplus_{\alpha_0 < \ldots < \alpha_{n-1}} H^0(\Gamma_{\alpha_0, \ldots, \alpha_{n-1}})$:

$$e_1^{n-1,0} = 2 \sum_{|I| \leq n} \sqrt{\prod_{i \in I} d_i}, \quad (2.4)$$

Next, we need a result of Gayet–Welschinger [18], which states that for every $0 \leq i \leq m$ the expectation of the $i$th Betti number of a random submanifold $X \subset \mathbb{R}P^n$ defined by $n - m$ independent Kostlan polynomials, is bounded by $\mathbb{E}[b_i(X)] \leq O(\sqrt{d^m})$.

This implies immediately that

$$e_1^{p,q} \leq \binom{s}{p+1} O(d^{(n-p-1)/2}), \quad (2.5)$$

for any $p < n-1$. In fact, each summand in $e_1^{p,q}$ is the expected $q$th Betti number of a random intersection of $p + 1$ random Kostlan hypersurfaces and, since there are a total of $\binom{s}{p+1}$ such summands, then (2.5) follows.
Denoting the reduced Betti numbers of a manifold by \( \tilde{b}_n(\cdot) \), by the Alexander–Pontryagin duality, we have
\[
b_{n-1}(\Gamma) = \tilde{b}_0(S^n \setminus \Gamma),
\]
thus when \( n > 1 \),
\[
b_{n-1}(\Gamma) = \tilde{b}_0(S^n \setminus \Gamma) = b_0(S^n \setminus \Gamma) - 1, \tag{2.6}
\]
and when \( n \leq 1 \),
\[
b_{n-1}(\Gamma) = \tilde{b}_0(S^n \setminus \Gamma) = b_0(S^n \setminus \Gamma). \tag{2.7}
\]
Consequently, when \( n > 1 \),
\[
\mathbb{E}[b_0(S^n \setminus \Gamma)] = \mathbb{E}[b_{n-1}(\Gamma)] + 1 \quad \text{(by (2.6))}
\]
\[
= \sum_{k=1}^{n} e_{n-k, k-1}^\infty + 1 \quad \text{(by (2.1) and linearity of expectation).} \tag{2.8}
\]
and when \( n \leq 1 \),
\[
\mathbb{E}[b_0(S^n \setminus \Gamma)] = \mathbb{E}[b_{n-1}(\Gamma)] \quad \text{(by (2.7))}
\]
\[
= \sum_{k=1}^{n} e_{n-k, k-1}^\infty. \quad \text{(by (2.1) and linearity of expectation).} \tag{2.9}
\]
First, observe that
\[
\sum_{k=2}^{n} e_{n-k, k-1}^\infty \leq \sum_{k=2}^{n} e_{1}^{n-k, k-1} \quad \text{(by (2.2))}
\]
\[
\leq \sum_{k=2}^{n} \binom{s}{n-k+1} O(d^{(k-1)/2}) \quad \text{(by (2.5))}
\]
\[
\leq s^{n-1} O(d^{(n-1)/2}). \tag{2.10}
\]
Now it remains to give precise bounds on \( e_{\infty}^{n-1,0} \), which is the same as obtaining precise bounds on \( e_{n}^{n-1,0} \), given that the spectral sequence collapses at \( E_n \) (cf. Theorem 4). Thus,
\[
e_{\infty}^{n-1,0} = e_n^{n-1,0} \leq e_1^{n-1,0} = 2 \sum_{I \subset [s], |I| = n} \sqrt{\prod_{i \in I} d_i},
\]
where the second step is due to (2.2) and the final equality is by (2.4). This provides an upper bound on \( e_{\infty}^{n-1,0} \). To lower bound \( e_{\infty}^{n-1,0} \), we make repeated use of inequality (2.3). Note that (2.3) 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 (2.3) holds for all $e_r^{p,0}$ when $r > 1$. Telescoping, we get

$$e_{\infty}^{n-1,0} = e_n^{n-1,0} \geq e_{n-1}^{n-1,0} - e_{n-1}^{0,n-2} \quad \text{(by (2.3))}$$

$$\geq e_{n-1}^{n-1,0} - e_1^{0,n-2} \quad \text{(by (2.2))}$$

$$\geq e_{n-2}^{n-1,0} - e_1^{1,n-3} - e_1^{0,n-2} \quad \text{(by (2.3))}$$

$$\vdots$$

$$\geq e_1^{n-1,0} - \left( \sum_{i=0}^{n-2} \left( \sum_{i=0}^{n-2} e_1^{i,n-2-i} \right) \right)$$

$$\geq 2 \sum_{I \subseteq [s]} |I| = n \sqrt{\prod_{i \in I} d_i} - \left( \sum_{i=0}^{n-2} \left( \sum_{i=0}^{n-2} e_1^{s \frac{i}{i+1}} O\left(d^{(n-1)/2}\right) \right) \right) \quad \text{(by (2.4) and (2.5))}.$$ 

Altogether,

$$2 \sum_{I \subseteq [s]} |I| = n \sqrt{\prod_{i \in I} d_i} \geq e_{\infty}^{n-1,0} \geq 2 \sum_{I \subseteq [s]} |I| = n \sqrt{\prod_{i \in I} d_i} - s^{n-1} O\left(d^{(n-1)/2}\right). \quad (2.11)$$

From equations (2.11) and (2.10), we get that

$$\sum_{k=1}^{n} e_{\infty}^{n-k,k-1} = 2 \sum_{I \subseteq [s]} |I| = n \sqrt{\prod_{i \in I} d_i} + s^{n-1} O\left(d^{(n-1)/2}\right). \quad (2.12)$$

Putting (2.12) in equations (2.8) and (2.9) gives us that for any $n$,

$$\mathbb{E}[b_0(S^n \setminus \Gamma)] = 2 \sum_{I \subseteq [s]} |I| = n \sqrt{\prod_{i \in I} d_i} + s^{n-1} O\left(d^{(n-1)/2}\right),$$

completing the proof of the theorem. □

Remark 6 (Sketch of the proof of (1.3)). Following the same lines of the previous proof, in the case of random harmonic polynomials, we have

$$e_1^{n-1,0} = 2 \sum_{I \subseteq [s]} |I| = n \Theta(d^n) = \Theta(d^n s^n). \quad (2.13)$$

In fact in this case one can use a result of Bürgisser [11] for estimating the expectation of the Euler characteristic of a random intersection $\Gamma_{\alpha_0,\ldots,\alpha_{n-1}}$ in $\mathbb{R}P^n$ of $n$ random, independent, harmonic hypersurfaces of degree $d_{\alpha_0}, \ldots, d_{\alpha_{n-1}}$. In this case, the intersection is almost surely zero dimen-
ional, and the Euler characteristic coincides with the number of points. Using the notation of [11], the parameter δ of a random harmonic polynomial of degree d is δ = Θ(d) (see [17, Lemma 16] and [17, Example 2]) and [11, Theorem 1.1] implies (2.13). For the analog of (2.5) in the harmonic case, we can use Thom-Milnor's bound \( b(X) \leq O(d^n) \) for the smooth zero set \( X \) of an algebraic set in \( \mathbb{RP}^n \). This gives now

\[
e_p^q \leq \left( \frac{s}{p+1} \right) O(d^n). \tag{2.14}
\]

Since in (2.14) we have \( p \geq 1 \), the estimate (2.13) dominates (2.14) and (1.3) follows.

Strictly speaking we observe that the bound that we have used \( b(X) \leq O(d^n) \), as in Milnor’s proof, would be of the order \( O(d^n) \) in the case of an affine algebraic set defined by polynomials of degree bounded by \( d \), and of the order \( O(d^{n+1}) \) in the case of a projective algebraic set (which is the case of our interest), but Milnor’s proof can be easily improved as follows.

**Lemma 6 (Sketch of the proof of (1.3)).** For every \( n \geq 1 \), there exists \( c_n \) such that for every algebraic set \( X \subset \mathbb{RP}^n \) defined by polynomials of degree bounded by \( d \), we have \( b(X) \leq c_n d^n \).

**Proof.** We show it by induction on \( n \).

In the case \( n = 1 \), if \( X \neq \mathbb{RP}^1 \), then \( X \) consists of finitely many points and, up to a change of coordinates, we can assume all this points are contained in the open set \( \{ x_0 \neq 0 \} \simeq \mathbb{R} \). Therefore, \( b(X) = \# X \leq d \).

Let now \( n > 1 \). Consider in \( \mathbb{RP}^n \) the open sets \( A_\varepsilon = \{ x_0^2 < \varepsilon (x_0^2 + \cdots + x_n^2) \} \) and \( B = \{ x_0 \neq 0 \} \).

Then, by Mayer–Vietoris,

\[
b(X) \leq b(A_\varepsilon \cap X) + b(B \cap X) + b(A_\varepsilon \cap B \cap X).
\]

For \( \varepsilon > 0 \) small enough, we have the following.

(A) Set \( A_\varepsilon \cap X \) deformation retracts onto \( X \cap \{ x_0 = 0 \} = X \cap \mathbb{RP}^{n-1} \). Therefore, by the inductive hypothesis, \( b(X \cap A_\varepsilon) \leq c_{n-1} d^{n-1} \);

(B) The set \( A_\varepsilon \cap B \cap X \) deformation retracts onto \( X \cap \{ x_0^2 = \varepsilon (x_0^2 + \cdots + x_n^2) \} \). This is an algebraic set in \( \{ x_0 \neq 0 \} \simeq \mathbb{R}^n \) defined by polynomials of degree bounded by \( d \) and, by the affine Milnor’s bound, \( b(A_\varepsilon \cap B \cap X) \leq O(d^n) \);

(C) The set \( B \cap X \) is an algebraic set in \( \{ x_0 \neq 0 \} \simeq \mathbb{R}^n \) defined by polynomials of degree bounded by \( d \) and, again by the affine Milnor’s bound, \( b(B \cap X) \leq O(d^n) \).

Putting these three estimates together, we get the statement. \( \square \)

Below we give a corollary of Theorem 1 which gives a bound on the sum of the Betti numbers of \( \mathbb{RP}^n \setminus \Gamma \) (we prove the corollary for the spherical case, again one has to divide the asymptotics by two in the projective case).

**Corollary 7** \((n \text{ fixed})\). Let \( \Gamma \) be defined as in Theorem 1. Then, for all \( k > 0 \),

\[
\mathbb{E}[b_k(S^n \setminus \Gamma)] = O(d^{(n-1)/2} s^{n-k}). \tag{2.15}
\]
Consequently,
\[
\mathbb{E} \left[ \sum_{i=0}^{n-1} b_i(S^n \setminus \Gamma) \right] = 2 \sum_{I \subset [s]} \sqrt{\prod_{i \in I} d_i} + O(d^{(n-1)/2}s^{n-1}).
\] (2.16)

**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
\[
\mathbb{E}[b_k(S^n \setminus \Gamma)] \leq \mathbb{E}[b_{n-k-1}(\Gamma)]
\]
\[
= \sum_{i=0}^{n-k-1} e_{\infty}^{i,n-k-i-1}
\]
\[
\leq \sum_{i=0}^{n-k-1} e_I^{i,n-k-i-1}
\]
\[
\leq \sum_{i=0}^{n-k-1} \left( \frac{s}{i+1} \right) O\left(d^{(n-i-1)/2}\right) \quad \text{(by (2.5))}
\]
\[
\leq s^{n-k} O\left(d^{(n-1)/2}\right),
\]
proving equation (2.15). Using this, equation (1.1) of Theorem 1, and linearity of expectation, equation (2.16) follows immediately. \(\square\)

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

## 3 Obstacle Random Graphs and an Application to Arrangement of Quadrics

In this section, we study the top Betti number of \( \mathbb{RP}^n \setminus \Gamma \), when \( \Gamma \) 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 \( \Gamma \) 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 1 and Corollary 7 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 3.1, we introduce our random graph model which we call ‘Obstacle’ random graphs. In Section 3.3, we prove a theorem (Theorem 8) about the average number of connected components in this random graph model. Then, in Section 3.2, using a theorem of Calabi (Theorem 9), we obtain a result on the average zeroth Betti number of $\Gamma$ (Theorem 2), when $\Gamma$ is a finite union of the zero sets of quadrics.

3.1 The ‘Obstacle’ random graph model

In this section, we introduce the obstacle random graph model. Before, we need a definition of convexity in $\mathbb{RP}^N$.

**Definition 1** (Convex set in $\mathbb{RP}^N$). Let $\mathcal{P} \subset \mathbb{RP}^N$ be a measurable set and $f : S^N \to \mathbb{RP}^N$ the double cover map. We will say that $\mathcal{P}$ is convex if it is entirely contained in one single affine chart and each one of the two components of $f^{-1}(\mathcal{P})$ is geodesically convex in $S^N$.

Note that convex sets in $\mathbb{RP}^N$ are contractible and that for every pair of points $q_0, q_1 \in \mathcal{P}$ there is a ‘segment’ joining these two points and entirely contained in $\mathcal{P}$. This segment is build as follows: first lift $q_0, q_1$ to two points $\tilde{q}_0, \tilde{q}_1 \in S^N$ both belonging to the same component $\tilde{\mathcal{P}}$ of $p^{-1}(\mathcal{P}) = \tilde{\mathcal{P}} \sqcup -\tilde{\mathcal{P}}$. Then consider the spherical arc

$$\tilde{q}_t := \frac{(1-t)\tilde{q}_0 + t\tilde{q}_1}{\| (1-t)\tilde{q}_0 + t\tilde{q}_1 \|}, \quad t \in [0, 1].$$

This spherical arc is a (reparametrized) geodesic joining $\tilde{q}_0$ to $\tilde{q}_1$ and therefore, by assumption, it lies in $\tilde{\mathcal{P}}$. The arc joining $q_0$ with $q_1$ is $q_t := p(\tilde{q}_t), t \in [0, 1]$.

Note that with this definition a ball $B_{\mathbb{R}P^N}(x, \frac{\pi}{2} - \epsilon)$, for $\epsilon > 0$ small enough, is convex but not geodesically convex in $\mathbb{RP}^N$.

**Definition 2** (‘Obstacle’ random graph). Let $\{q_1, \ldots, q_s\} \subset \mathbb{RP}^N$ be a sample from the uniform distribution on $\mathbb{RP}^N$, and let $\mathcal{P} \subset \mathbb{RP}^N$ (the ‘obstacle’) be a measurable convex set. We define the obstacle random graph model $\mathbb{G}(N, \mathcal{P}, s)$ as follows.

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

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

In other words, it is an undirected graph where the vertices are $\{q_1, \ldots, q_s\}$, 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 [14]. See Figure 3.1 for an example illustration.

Remark 7. Two commonly studied random graph models are the Erdös–Rényi model (proposed in [16, 19]), and the geometric random graph model (proposed in [20]).

- 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 Erdös–Rényi model.
- Define the metric \( d : \mathbb{RP}^N \times \mathbb{RP}^N \to \mathbb{R} \), where

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

where \( \mathcal{L}(q_1, q_2) \), as defined earlier, is the projective line containing the points \( q_1 \) and \( q_2 \). 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.

3.2 | \( b_0 \) of arrangement of random quadrics

Our theorem on the expected \( b_0 \) of an arrangement of random quadrics, that is, Theorem 2, follows from a general theorem that we prove about the average number of connected components in the obstacle random graph model \( \mathcal{Q}(N, \mathcal{P}, s) \) as \( s \to \infty \).
Theorem 8 \((N \text{ fixed})\). Consider the obstacle random graph model \(G(N, P, s)\) as per Definition 2. Then

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

The proof of Theorem 8 is deferred to Section 3.3.

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 \(\mathbb{R}[x_0, \ldots, x_n]_{(2)}\), 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\(^\dagger\) measure under this linear isomorphism (see, for example, [26] for a discussion about this), that is:

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

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 \(P_n \subset \mathbb{R}P^N\) which is the projectivization of the set of positive definite matrices (equivalently of the set of positive quadratic forms):

\[P_n = \{[Q] \in \mathbb{R}P^N | Q > 0\}.\]

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 \(p : S^N \to \mathbb{R}P^N\) is a local isometry), with corresponding volume density. In this way, if \(q\) is a random Kostlan quadric, we have

\[\mathbb{P}\{q\text{ is a positive form}\} = \frac{\text{vol}(P_n)}{\text{vol}(\mathbb{R}P^N)}.\] (3.1)

Remark 8. The relative volume of \(P_n\) in \(\mathbb{R}P^N\) is known (see, for example, [27]) to decay exponentially fast when \(n\) increases

\[\lim_{n \to \infty} \frac{1}{n^2} \log \left(\frac{\text{vol}(P_n)}{\text{vol}(\mathbb{R}P^N)}\right) = -\frac{\log 3}{4}.\] (3.2)

The following result, which is due to Calabi [13], gives a geometric criterion for two quadrics intersecting in projective space.

Theorem 9 (Calabi, 1964). For \(n \geq 2\) let \(q_1, q_2 \in \mathbb{R}[x_0, \ldots, x_n]_{(2)}\) 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:

\[\Gamma_1 \cap \Gamma_2 \neq \emptyset \iff \ell \cap P_n = \emptyset.\]

One can refer to [23] for a proof of this using spectral sequences. As a consequence of Calabi’s Theorem (Theorem 9), studying the average zeroth Betti number of \(\Gamma\) reduces to studying the

\(^\dagger\) Stands for Gaussian Orthogonal Ensemble (see [30] for a description).
average number of connected components in the obstacle random graph model, that is, studying the average number of connected components of $\mathcal{G}(N, P_n, s)$.

Specifically, non-empty quadrics in projective space are connected, therefore the number of connected components of $\Gamma$ 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—to form the subgraph, we must discard the points that fall inside $P_n$ because the zero sets of quadrics in $P_n$ is empty.

**Definition 3.** We define the ‘quadrics graph’ as the graph form by taking an instance of $\mathcal{G}(N, P_n, s)$ and forming a subgraph of it by discarding the vertices that fall inside $P_n$. We shall use $\mathcal{H}(N, P_n, s)$ to denote this random subgraph model.

Relying on Calabi’s Theorem, and using Theorem 8, we shall now prove Theorem 2.

**Proof of Theorem 2.**

\[
\lim_{s \to \infty} \frac{\mathbb{E}[b_0(\Gamma)]}{s} = \lim_{s \to \infty} \frac{\mathbb{E}[b_0(\mathcal{H}(N, P_n, s))]}{s}
= \lim_{s \to \infty} \frac{\mathbb{E}[b_0(\mathcal{G}(N, P_n, s))] - \sum_{i=1}^{s} 1\{q_i \in P_n\}}{s}
\leq \frac{\text{vol}(P_n)}{\text{vol}(\mathbb{R}P^N)} - \frac{\text{vol}(P_n)}{\text{vol}(\mathbb{R}P^N)}
= 0.
\]

Equation (3.3) together with the fact that $\lim_{s \to \infty} \frac{\mathbb{E}[b_0(\Gamma)]}{s}$ is obviously non-negative completes the proof.

\[\square\]

### 3.3 Average number of connected components of obstacle random graphs

In this section, we shall prove Theorem 8. Below is a synopsis of the proof.

**A** An important definition is the definition of the good cone of a point. The good cone of a point $p$ with respect to $P$ is defined as the set of all points in $\mathbb{R}P^N \setminus P$, which if sampled, would be connected by an edge to $p$ (see Definition 4). Thus, after sampling a point $p$, if we were to sample another point $q$ which happens to be in the good cone of $p$, then $p$ and $q$ would be connected by an edge.

**B** The general strategy is to prove that, with high probability (w.h.p.), after a constant number of samples (say $\kappa$), the entire sphere is ‘covered’. Here covered is defined as the state where the union of the good cones of the already sampled points contains $\mathbb{R}P^N \setminus P$. By definition, this means that if we were to sample any additional points, they will be connected to at least one of the already sampled $\kappa$ vertices by an edge. Consequently, additional points cannot create
a new connected component, which in turn means that w.h.p., there are at most $x$ connected components.

(C) To prepare the ground for showing the above, a simple first step is Proposition 10, where we understand the distribution of the number of vertices in various regions in $\mathbb{RP}^N$. Specifically, Proposition 10 gives tail bounds on the number of vertices in $\mathcal{P}$, $\mathcal{P}(\varepsilon) \setminus \mathcal{P}$ (where $\mathcal{P}(\varepsilon)$ is the $\varepsilon$-neighborhood of $\mathcal{P}$) and $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$.

(D) Lemma 11 proves that the subgraph of $\mathcal{G}(N, \mathcal{P}, s)$ restricted to $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$ has number of connected components constant with respect to $s$. The proof of Lemma 11 involves the following sub-steps.

1. Cover $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$ with balls of radius $r > 0$, $r$ to be chosen later.
2. Then for each $r$-ball $B$, we proceed to lower bound the probability (Lemma 12) of choosing a point in $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$ such that the good cone of the point contains $B$. This involves showing that the volume of the good cone of a point is a continuous function of the position of the point, thus it attains a minimum. We prove this by first considering a smooth approximation of $\mathcal{P}$ containing $\mathcal{P}$ and contained in $\mathcal{P}(\varepsilon)$ (Proposition 13), and then applying a stereographic projection and proving continuity in Euclidean space (Lemma 14).
3. Now, we have a finite number of balls covering $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$, and there is a lower bound on the probability of each of these balls being covered. To complete our proof, we just need to show that after a finite number of samples, all balls will be covered. The final step, which is similar to a coupon-collector type argument (Lemma 15), 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 $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$ will not add a new connected component to the graph.

Sample $s$ points $q_1, \ldots, q_s$ independent and identically distributed from the uniform distribution on $\mathbb{RP}^N$. Let $\mathcal{P}(\varepsilon)$ be the $\varepsilon$-neighborhood of $\mathcal{P}$ in $\mathbb{RP}^N$. Define the random variables

\[ s_c(\varepsilon) = \sum_{i=1}^{s} \mathbb{1}\{ q_i \in \mathbb{RP}^N \setminus \mathcal{P}(\varepsilon) \}, \]

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

\[ s_g(\varepsilon) = \sum_{i=1}^{s} \mathbb{1}\{ q_i \in \mathcal{P}(\varepsilon) \setminus \mathcal{P} \}, \]

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,

\[ s = s_c(\varepsilon) + s_g(\varepsilon) + s_p. \]
Fix any $\beta < 1/2$. Now, let $\Omega_1(\varepsilon), \Omega_2(\varepsilon), \Omega_3$ be the following defined events:

$$\Omega_1(\varepsilon) = \left\{ s_e(\varepsilon) = s \cdot \left(1 - \frac{\text{vol}(P(\varepsilon))}{\text{vol}(\mathbb{R}P^N)}\right) \pm s^{1/2+\beta} \cdot \left(1 - \frac{\text{vol}(P(\varepsilon))}{\text{vol}(\mathbb{R}P^N)}\right) \right\},$$

$$\Omega_2(\varepsilon) = \left\{ s_a(\varepsilon) = s \cdot \left(\frac{\text{vol}(P(\varepsilon) \setminus P)}{\text{vol}(\mathbb{R}P^N)}\right) \pm s^{1/2+\beta} \cdot \left(\frac{\text{vol}(P(\varepsilon) \setminus P)}{\text{vol}(\mathbb{R}P^N)}\right) \right\},$$

$$\Omega_3 = \left\{ s_p = s \cdot \left(\frac{\text{vol}(P)}{\text{vol}(\mathbb{R}P^N)}\right) \pm s^{1/2+\beta} \cdot \left(\frac{\text{vol}(P)}{\text{vol}(\mathbb{R}P^N)}\right) \right\}.$$

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

**Proposition 10.** Let $\beta$ be as chosen before. For all $0 < \delta < 1, \varepsilon > 0$, there exists $\bar{s}_1 = \bar{s}_1(\delta, \beta, \varepsilon, N)$, such that if $s > \bar{s}_1$,

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

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 $\varepsilon > 0$,

$$\lim_{s \to \infty} \mathbb{P}[\Omega_1(\varepsilon) \cap \Omega_2(\varepsilon) \cap \Omega_3] = 1.$$
Thus $g_q(P)$ is a random variable that denotes the set of points in $\mathbb{RP}^N$ which, if sampled, would be connected to $q$ by an edge in $\mathcal{Q}(N, P, s)$. See Figure 2 for an example illustration of the good cone. Next, for $B \subseteq \mathbb{RP}^N$, $\varepsilon > 0$, define

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

In words, $G_B(P)$ denotes all points $x \in \mathbb{RP}^N \setminus P(\varepsilon)$ such that the good cone of $x$ with respect to $P$ completely contains whatever part of $B$ is outside $P(\varepsilon)$. The following lemma, whose proof is also postponed, gives a lower bound on the relative volume of $g_q(P)$, when $q$ is outside $P(\varepsilon)$.

**Lemma 12.** For $p \in \mathbb{RP}^n$, and $r \geq 0$, define $B(p, r)$ to be a ball centered at $p$ of radius $r$. For all $\varepsilon > 0$, there exists $r = r(\varepsilon, N) > 0$, $\delta_2 = \delta_2(\varepsilon, N)$, such that for any $p \in \mathbb{RP}^N \setminus P(\varepsilon)$,

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

Later, we will choose an $r \leq \varepsilon/8$, so that

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

**Remark 9.** Observe that the convex set $P \subseteq \mathbb{RP}^N$ is contained in one single affine chart, and therefore if we denote by $f : S^N \to \mathbb{RP}^N$ the double cover map, the pre-image $f^{-1}(P)$ consists of two isometric components, which for simplicity we still denote by $P$ and $-P$. Both these components
are entirely contained in a open hemisphere, and we assume that $\mathcal{P}$ is contained in

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

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

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

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 $U$, that is, intersections between $U$ and great circles, to (unparametrized) geodesics in $\mathbb{R}^N$, that is, straight segments. In particular, $\sigma$ 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 $f|_U : U \to \mathbb{R}P^N$ 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 9.

**Proposition 13.** For all $\varepsilon > 0$, there exists a smooth convex set $\bar{\mathcal{P}}(\varepsilon)$ such that $\mathcal{P} \subseteq \bar{\mathcal{P}}(\varepsilon) \subseteq \mathcal{P}(\varepsilon)$.

**Proof.** Consider the $\varepsilon/2$-neighborhood of $\mathcal{P}$, that is, $\mathcal{P}(\varepsilon/2)$. Since the set of smooth convex bodies is dense in the Hausdorff distance induced topology on the space of convex bodies (see [28, Theorem 2.7.1.]), there exists a body $C_\varepsilon$ that is convex, smooth and also satisfies

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

where $d_H$ denotes Hausdorff distance with the underlying metric being the usual round metric on $S^N$. We shall now show that $C_\varepsilon$ itself is the smooth approximation we desire, that is, $\bar{\mathcal{P}}(\varepsilon)$. We know that $d_H(\mathcal{P}, \mathcal{P}(\varepsilon/2)) = \varepsilon/2$. Observe that if $\mathcal{P}$ was not completely contained in $C_\varepsilon$, then $d_H(\mathcal{P}(\varepsilon/2), C_\varepsilon) \geq \varepsilon/2$, which contradicts equation (3.4). Similarly, it can be shown that $C_\varepsilon$ 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 (3.4). $\square$

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 14.** 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),

$$\frac{\text{vol}(\sigma(G_{B(p,r')})(\mathcal{P})))}{\text{vol}(\sigma(S^N))} \geq \delta'_2.$$
FIGURE 3 Illustration of the good cone of \( q \) with respect to \( \tilde{\mathcal{P}}(\varepsilon) \). \( \mathcal{P}(\varepsilon) \) 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, \( \text{vol}(g_q(\tilde{\mathcal{P}}(\varepsilon))) \leq \text{vol}(g_q(\mathcal{P})) \).

Define the map

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

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

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

is bounded below by a continuous function.
Let \( q' \) be the point shortest to \( q \) on \( \partial \tilde{Q}_n(\varepsilon) \), the boundary of \( \tilde{Q}_n(\varepsilon) \). Let \( \Pi(\tilde{Q}_n(\varepsilon)) \) be the projection of \( \tilde{Q}_n(\varepsilon) \) onto \( T_{q'}(\partial \tilde{Q}_n(\varepsilon)) \), the tangent space of \( \tilde{Q}_n(\varepsilon) \) at \( q' \). Observe that

\[
\lambda(q) = \max_{v \in \Pi(\tilde{Q}_n(\varepsilon))} ||v||_2
\]
is continuous. Consequently, observe that

\[
\frac{\text{vol}(\sigma(g_{y^{-1}(q)(\tilde{P}_n(\varepsilon)))))}{\text{vol}(S^{N-1})} \geq 1 - \frac{\text{vol}(\text{spherical cap with angle } \tan^{-1} \frac{\lambda(q)}{2||q-q'||_2})}{\text{vol}(S^{N-1})}.
\]

\( \beta(q) \) is a continuous function, and thus attains a maximum on \( S^{N-1} \setminus \text{int}(Q_n(\varepsilon)) \) (remember that \( ||q-q'||_2 \) can never become 0 because \( q \) is always outside \( P(\varepsilon) \)) 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 \( p \in \mathbb{R}^P \setminus P(\varepsilon) \), we can find a direction \( \tilde{v}' \) in \( \mathbb{R}^N \) and an angle \( \hat{\theta} \) such that for all directions \( \hat{v} \) with \( \cos^{-1} \frac{\tilde{v}' \cdot \hat{v}}{||\tilde{v}'||_2 ||\hat{v}||_2} \leq \hat{\theta} \), we have that \( \ell_o(p, \tilde{v}') \subseteq \sigma(p, \tilde{P}(\varepsilon)) \), where \( \ell_o(p, \tilde{v}') \) denotes the line in \( \mathbb{R}^N \) through \( \sigma(p) \) in the direction \( \tilde{v}' \). Note that \( \tilde{v}' \) and \( \hat{\theta} \) depend on \( p \) continuously. Let \( p_{\tilde{v}'} \) be the point of intersection of the line \( \ell_o(p, \tilde{v}') \) and \( S^{N-1} \), and now let \( p_2 \) be the mid-point on the line joining \( p \) and \( p_{\tilde{v}'} \). Since \( \hat{\theta} \) depends on \( p \) continuously, it has a minimum on \( \mathbb{R}^P \setminus \text{int}(P(\varepsilon)) \), and thus we can pick \( \tilde{r}'' = \tilde{r}''(\varepsilon, N) > 0 \) such that \( B(p_2, \tilde{r}'') \subseteq \sigma(g_{p}(\tilde{P}(\varepsilon))) \).

Now, for the sake of contradiction, assume that for all \( r' > 0 \), \( \min_{q \in S^{N-1} \setminus \text{int}(Q_n(\varepsilon))} \tilde{\alpha}'(q) = 0 \), and let \( q \) be the point at which \( \tilde{\alpha}' \) attains the minimum. Then we can find a sequence \( (r_n) \), with \( r_n \to 0 \), and a sequence \( (q_n) \), with \( q_n \to q \), 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 \( S^{N-1} \setminus \text{int}(Q(\varepsilon)) \) is compact, and \( B(p_2, r'') \) is obviously compact as well, this means that \( (\lim_{n \to \infty} b_n) \notin (\lim_{n \to \infty} g_{q_n}) \), 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.

\[ \square \]

Proof of Lemma 12. Set \( r = \min(r', \varepsilon/s) \). The proof of the lemma follows by noting that since \( \sigma \) is smooth, bijective and angle-preserving (conformal)\(^1\), proving that there is a set of strictly positive measure that is good for all \( r \)-balls centered in \( \mathbb{R}^N \setminus P(\varepsilon) \) follows from Lemma 14. This is because since \( \delta_2' > 0 \), the pre-image under \( \sigma \) of any set of measure at least \( \delta_2' \) will be strictly positive \( (\delta_2 \text{ will be the measure of the pre-image, under } \sigma \text{, of the set in } \mathbb{R}^N \text{ which attains the minimum measure } \delta_2') \).

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

**Lemma 15.** Let \( q'_i \) denote the \( i \)th sampled point. For any \( \varepsilon > 0 \), define \( C = C(\varepsilon) \) to be a random variable that denotes the minimum index such that the union of the good cones of \( q'_1 ... q'_C \) covers

\[ ^1 \text{Note that the stereographic projection is not isometric, and thus does not preserve areas. However, angle-preservation is enough for us.} \]
$\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$, that is,

$$\bigcup_{i=1}^{C} g_{q_i'}(P_n) \supseteq \mathbb{RP}^N \setminus \mathcal{P}(\varepsilon).$$

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

$$P[C \leq \gamma] \geq 1 - \delta_3.$$

**Proof.** Take a covering of $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$ with $r$-balls (where $r$ is from Lemma 12) of size $Q = Q(\varepsilon, N)$, and let the $Q$ balls that cover $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$ be $B_1, \ldots, B_Q$. Remember that the conditional distribution of sampling from $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$ is uniform. Let $C_i$ denote the additional number of points needed to be sampled from $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$ such that $B_i$ is covered, given that balls $B_1, \ldots, B_{i-1}$ 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 $B_1, \ldots, B_{i-1}$ 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 12, each $C_i$ is a geometric random variable with parameter $\mu_i \geq \delta_2$. This means

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

where Geom($\mu$) refers to the geometric distribution with mean $\frac{1}{\mu}$. Thus

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

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

$$E[C] \leq \frac{Q}{\delta_2}. \quad (3.5)$$

Set $\gamma = \frac{Q}{\delta_2 \delta_3}$. Applying Markov’s inequality on $C$, and using equation (3.5), the lemma follows. \(\square\)

**Proof of Lemma 11.** Lemma 15 shows that we will have a covering of $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$ with *good sets*, with probability at least $1 - \delta_3$, if we have $s_e(\varepsilon) \geq \gamma$. To complete the proof of Lemma 11, we have to set $\bar{s}_2$ appropriately so that if $s \geq \bar{s}_2$, then $s_e(\varepsilon) \geq \gamma$. Conditioning on $\Omega_1(\varepsilon)$, it is clear that if $s \geq k \cdot \gamma(\frac{\text{vol}(\mathbb{RP}^N)}{\text{vol}(\mathbb{RP}^N) - \text{vol}(\mathcal{P}(\varepsilon))})$, for an appropriately chosen constant $k$, then $s_e(\varepsilon) \geq \gamma$. Thus, conditioned on $\Omega_1(\varepsilon)$, setting $\bar{s}_2 = k \cdot \gamma(\frac{\text{vol}(\mathbb{RP}^N)}{\text{vol}(\mathbb{RP}^N) - \text{vol}(\mathcal{P}(\varepsilon))})$ ensures we have a covering of $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$ with *good sets* with probability at least $1 - \delta_3$.

Since $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$ is covered, any new point that is added to $\mathbb{RP}^N \setminus \mathcal{P}(\varepsilon)$ will be connected to at least one of the existing $\gamma$ vertices, which in turn means that the number of connected components of the graph stays fixed as $\gamma$. The lemma follows by setting $a = k(\frac{\text{vol}(\mathbb{RP}^N)}{\text{vol}(\mathbb{RP}^N) - \text{vol}(\mathcal{P}(\varepsilon))}). \square$
Proof of Theorem 8. We shall prove that, for all $\epsilon, \delta, \delta_1$, \( \lim_{s \to \infty} \frac{\mathbb{E}[b_0(G(N,P,s))]}{s} \) is bounded from above by \( \frac{\text{vol}(P)}{\text{vol}(\mathbb{R}P^N)} \) plus some terms which depend on $\epsilon, \delta, \delta_1$. 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. Consider the decomposition of the graph from earlier, that is, $G(N,P,s)$ is decomposed into $G_1(N,P,s,\epsilon)$, which is the subgraph of $G(N,P,s)$ restricted to the vertices in $\mathbb{R}P^N \setminus P(\epsilon)$, $G_2(N,P,s,\epsilon)$, which is the subgraph of $G(N,P,s)$ restricted to the vertices in $P(\epsilon) \setminus P$, and $G_3(N,P,s)$, which is the subgraph of $G(N,P,s)$ restricted to the vertices in $P$. Thus,

\[
\mathbb{E}[b_0(G(N,P,s))] \leq \mathbb{E}[b_0(G_1(N,P,s,\epsilon))] + \mathbb{E}[b_0(G_2(N,P,s,\epsilon))] + \mathbb{E}[b_0(G_3(N,P,s))].
\] (3.6)

To bound $\mathbb{E}[b_0(G_1(N,P,s,\epsilon))]$, we break the sample space into three disjoint events, that is, $\Omega_1(\epsilon) \cap (b_0(G_1(N,P,s,\epsilon)) \leq \frac{\tilde{s}_2}{a})$, $\Omega_1(\epsilon) \cap (b_0(G_1(N,P,s,\epsilon)) \leq \frac{\tilde{s}_2}{a})^c$, and $\Omega_1(\epsilon)^c$, where $\tilde{s}_2$ and $a$ are from Lemma 11. Thus, for any $\epsilon > 0$, we have

\[
\mathbb{E}[b_0(G_1(N,P,s))] \leq \int_{\Omega_1(\epsilon) \cap (b_0(G_1(N,P,s,\epsilon)) \leq \frac{\tilde{s}_2}{a})} b_0(G_1(N,P,s,\epsilon)) \, d\omega + \int_{\Omega_1(\epsilon) \cap (b_0(G_1(N,P,s,\epsilon)) \leq \frac{\tilde{s}_2}{a})^c} b_0(G_1(N,P,s,\epsilon)) \, d\omega + \int_{\Omega_1(\epsilon)^c} b_0(G_1(N,P,s,\epsilon)) \, d\omega.
\] (3.7)

To bound $\mathbb{E}[b_0(G_2(N,P,s,\epsilon))]$, we break the sample space into two disjoint events, that is, $\Omega_2(\epsilon)$ and $\Omega_2(\epsilon)^c$. Thus, for any $\epsilon > 0$, we have

\[
\mathbb{E}[b_0(G_2(N,P,s))] \leq \int_{\Omega_2(\epsilon)} b_0(G_2(N,P,s,\epsilon)) \, d\omega + \int_{\Omega_2(\epsilon)^c} b_0(G_2(N,P,s,\epsilon)) \, d\omega.
\] (3.8)

Finally, to bound $\mathbb{E}[b_0(G_3(N,P,s))]$, we break the sample space into two disjoint events, that is, $\Omega_3$ and $\Omega_3^c$. Thus, we have

\[
\mathbb{E}[b_0(G_3(N,P,s))] \leq \int_{\Omega_3} b_0(G_3(N,P,s)) \, d\omega + \int_{\Omega_3^c} b_0(G_3(N,P,s)) \, d\omega.
\] (3.9)

Thus, from (3.7), (3.8), (3.9), and (3.6), we have that

\[
\mathbb{E}[b_0(G(N,P,s))] \leq A + B + C + D + E + F + G.
\] (3.10)

Let us now estimate $A, B, C, D, E, F, G$. Because we are integrating over the space where $b_0(G_1(N,P,s,\epsilon)) \leq \frac{\tilde{s}_2}{a}$, for all $\epsilon > 0$, obviously,

\[
A \leq \frac{\tilde{s}_2}{a}.
\] (3.11)
We apply the trivial bound of $s$ on $b_0(G_1(N, P, s, \varepsilon))$ to get that, for all $\delta > 0, \varepsilon > 0$,

$$B \leq \mathbb{P}[\Omega_1(\varepsilon)^c] s \leq \frac{\delta}{3} s,$$  \hspace{1cm} (3.12)

as long as $s \geq \bar{s}_1 = s_1(\delta, \beta, \varepsilon, N)$ (cf. Proposition 10). By Lemma 11, for all $\delta_1 > 0, \varepsilon > 0$, if $s > \bar{s}_2(\varepsilon, \delta_1, N)$, $\mathbb{P}[b_0(G_1(N, P, s, \varepsilon)) > \bar{s}_2/\delta_1 | \Omega_1(\varepsilon)] < \delta_1$, for some specific $a = a(\varepsilon, N)$. Thus, for $\delta_1 > 0, \varepsilon > 0$,

$$C \leq \mathbb{P}[\Omega_1(\varepsilon)] \cdot \mathbb{P}[b_0(G_1(N, P, s, \varepsilon)) > \bar{s}_2/a | \Omega_1(\varepsilon)] s \leq \mathbb{P}[\Omega_1(\varepsilon)] \delta_1 s \leq \delta_1 s.$$  \hspace{1cm} (3.13)

Trivially, $b_0$ of a graph is bounded from above by the number of vertices in the graph. Thus, for all $\varepsilon > 0$,

$$D \leq \mathbb{P}[\Omega_2(\varepsilon)] \left( s \cdot \left( \frac{\text{vol}(P(\varepsilon) \setminus P)}{\text{vol}(\mathbb{R}P^N)} \right) + s^{1/2+\beta} \cdot \left( \frac{\text{vol}(P(\varepsilon) \setminus P)}{\text{vol}(\mathbb{R}P^N)} \right) \right) \leq \left( s + s^{1/2+\beta} \right) \cdot \left( \frac{\text{vol}(P(\varepsilon) \setminus P)}{\text{vol}(\mathbb{R}P^N)} \right).$$  \hspace{1cm} (3.14)

At the same time, similar to (3.12), for all $\delta > 0, \varepsilon > 0$

$$E \leq s \mathbb{P}[\Omega_2(\varepsilon)] \leq s \frac{\delta}{3},$$  \hspace{1cm} (3.15)

if $s \geq \bar{s}_1 = \bar{s}_1(\delta, \beta, \varepsilon, N)$ (by Proposition 10). By equation (3.2), we have that

$$F \leq \mathbb{P}[\Omega_3] \left( s \cdot \left( \frac{\text{vol}(P)}{\text{vol}(\mathbb{R}P^N)} \right) + s^{1/2+\beta} \cdot \left( \frac{\text{vol}(P)}{\text{vol}(\mathbb{R}P^N)} \right) \right) \leq \left( s + s^{1/2+\beta} \right) \cdot \left( \frac{\text{vol}(P)}{\text{vol}(\mathbb{R}P^N)} \right).$$  \hspace{1cm} (3.16)

Finally, again, for all $\delta > 0, \varepsilon > 0$, if $s \geq \bar{s}_1 = \bar{s}_1(\delta, \beta, \varepsilon, N)$,

$$G \leq s \mathbb{P}[\Omega_3] \leq s \frac{\delta}{3}.$$  \hspace{1cm} (3.17)

Putting (3.11), (3.12), (3.13), (3.14), (3.15), (3.16), and (3.17) in (3.10), we have that for all $\varepsilon > 0, \delta > 0, \delta_1 > 0$,

$$\lim_{s \to \infty} \frac{\mathbb{E}[b_0(G(N, P, s))]}{s} \leq \frac{\delta}{3} + \delta_1 + \frac{\text{vol}(P(\varepsilon) \setminus P)}{\text{vol}(\mathbb{R}P^N)} + \frac{\delta}{3} + \frac{\text{vol}(P)}{\text{vol}(\mathbb{R}P^N)} + \frac{\delta}{3}. $$  \hspace{1cm} (3.18)

Since equation (3.18) is true for any choice of $\varepsilon, \delta, \delta_1$, we have that

$$\lim_{s \to \infty} \frac{\mathbb{E}[b_0(G(N, P, s))]}{s} \leq \frac{\text{vol}(P)}{\text{vol}(\mathbb{R}P^N)}.$$  \hspace{1cm} \Box
### 3.4 On the probability of having large cliques in the complement of the quadrics graph $\Gamma$

We first recall the definition (Definition 3) of the random quadrics graph $H(N, P_n, s)$ which is a subgraph of $G(N, P_n, s)$. Quadrics graphs are random objects but for any fixed value of $n$, those that occur with positive probability are examples of semi-algebraic graphs which we define below.

**Definition 5.** Let $R \subseteq \mathbb{R}^d \times \mathbb{R}^d$ be a symmetric semi-algebraic relation, that is, $R$ is a semi-algebraic subset of $\mathbb{R}^d \times \mathbb{R}^d$, and $(x, y) \in R \iff (y, x) \in R$ for every $x, y \in \mathbb{R}^d$. We say a (undirected) graph $G = (V, E)$ is a $R$-semi-algebraic graph, if there exists a map $\phi : V \to \mathbb{R}^d$, such that for any $u, v \in V, (u, v) \in E \iff (\phi(u), \phi(v)) \in R$.

The property of existence of large cliques in semi-algebraic graphs or their complements have been studied by various authors. For example, Alon et al. [2] prove the following Ramsey-theoretic theorem.

**Theorem 16 (Alon et al. [2]).** For any symmetric semi-algebraic relation $R$ the following is true. There exists a constant $\delta = \delta(R) > 0$, such that for any $R$-semi-algebraic graph with $n$ vertices one of the following is true.

(A) There exists a clique of size $n\delta$ in $G$.
(B) The complement of $G$ has a clique of size $n\delta$.

It is clear from its definition that for fixed $n$ each quadrics graph $H(N, P_n, s)$ is an $R$-semi-algebraic graph for a fixed appropriately defined symmetric semi-algebraic relation $R$. In view of Theorem 16, it is natural to ask the probability of a quadric graph or its complement to have a large clique.

The following result rules out with probability 1, in the limit as $s \to \infty$, the existence of large cliques in the complement graph of $\Gamma$ (where large means of size at least a constant fraction of $s$).

**Corollary 17 (of Theorem 2).** Let $H(N, P_n, s)$ be the random quadrics graph as per Definition 3. Let $H(N, P_n, s)^c$ be the complement of the graph on the same set of vertices. Then, for any $\varepsilon > 0$,$$
\lim_{s \to \infty} \mathbb{P}[H(N, P_n, s)^c \text{ contains a clique of size } \varepsilon s] = 0.
$$

**Proof.** Let $\Omega_a$ denote the event that there exists a clique of size $\varepsilon s$ in $H(N, P_n, s)^c$. Thus, we have

\[
0 = \lim_{s \to \infty} \mathbb{E}\left[\frac{\mathbb{b}_0(H(N, P_n, s))}{s}\right] \quad \text{(by Theorem 2)}
\]

\[
= \lim_{s \to \infty} \int_{\Omega_a} \mathbb{b}_0(H(N, P_n, s)) d\omega + \int_{\Omega_a^c} \mathbb{b}_0(\Gamma) d\omega
\]

\[
\geq \lim_{s \to \infty} \frac{\varepsilon s \cdot \mathbb{P}[\Omega_a] + 0}{s}.
\]

The last inequality follows by noting that if the complement of $H(N, P_n, s)$ contains a clique of size $\varepsilon s$, it means that all $\varepsilon s$ vertices were isolated in $H(N, P_n, s)$, in turn implying that $H(N, P_n, s)$...
has at least \( \varepsilon s \) connected components. By noting that \( \lim_{s \to \infty} \mathbb{P}[\Omega_a] \) is obviously non-negative, and from (3.19), we deduce that

\[
\lim_{s \to \infty} \mathbb{P}[\Omega_a] = 0,
\]
or in other words

\[
\lim_{s \to \infty} \mathbb{P}[\mathcal{H}(N, P_n, s)^c \text{ contains a clique of size } \varepsilon s] = 0. \tag*{□}
\]

Juxtaposing with Theorem 16, Corollary 17 shows that in the quadrics random graph the probability of having an independent set of size at least a fixed fraction of the number of vertices goes to 0 with \( s \). This is to be compared with condition (B) in Theorem 16, which posits the existence of an independent set of size at least some fixed polynomial in the number of vertices.

APPENDIX: PROOF OF PROPOSITION 10

We will need the additive Chernoff–Hoeffding bound for binomial random variables.

**Proposition A.1** (See, for example, [10]). Let random variable \( X \sim \text{Binomial}(n, p) \), and \( t \in [0, 1] \). Then

\[
\mathbb{P}[|X - \mathbb{E}[X]| > t \cdot \mathbb{E}[X]] < 2e^{-t^2 \mathbb{E}[X]/3}.
\]

Consequently, if \( n > \bar{n} = \bar{n}(t, \delta, p) = \frac{3 \log 2/\delta}{p t^2} \),

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

This also implies

\[
\lim_{n \to \infty} \mathbb{P}[|X - \mathbb{E}[X]| > t \cdot \mathbb{E}[X]] = 0. \tag{A.1}
\]

**Proof of Proposition 10.** Obviously \( s_c(\varepsilon) \), \( s_a(\varepsilon) \) and \( s_p \) are Binomial random variables. Note that

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

Putting \( t = s^{\beta-1/2} \) and substituting \( \delta \) with \( \delta/3 \) in Proposition A.1, we have that if \( s > \left(\frac{3 \log 6/\delta}{1 - \frac{\text{vol}(P(\varepsilon))}{\text{vol}(\mathbb{R}P^N)}}\right)^{1/2\beta} \),

\[
\mathbb{P}[\Omega_1(\varepsilon)^c] \leq \mathbb{P}\left[|s_c(\varepsilon) - s \cdot \left(1 - \frac{\text{vol}(P(\varepsilon))}{\text{vol}(\mathbb{R}P^N)}\right)| > s^{1/2 + \beta} \cdot \left(1 - \frac{\text{vol}(P(\varepsilon))}{\text{vol}(\mathbb{R}P^N)}\right)\right] < \frac{\delta}{3}. \tag{A.2}
\]

Similarly, by noting that

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

\[ E[s_P] = s \cdot \left( \frac{\text{vol}(P)}{\text{vol}(\mathbb{RP}^N)} \right), \]

again by Proposition A.1, we have that, if \( s > \left( \frac{3 \log \delta}{\text{vol}(\mathbb{RP}^N)} \right)^{1/2} \),

\[ \mathbb{P}[\Omega_2^c] \leq \mathbb{P} \left[ s_d(\varepsilon) - s \cdot \left( \frac{\text{vol}(P) \setminus P}{\text{vol}(\mathbb{RP}^N)} \right) > s^{1/2 + \beta} \cdot \left( \frac{\text{vol}(P) \setminus P}{\text{vol}(\mathbb{RP}^N)} \right) \right] < \frac{\delta}{3} \tag{A.3} \]

and, if \( s > \left( \frac{3 \log \delta}{\text{vol}(\mathbb{RP}^N)} \right)^{1/2} \),

\[ \mathbb{P}[\Omega_3^c] \leq \mathbb{P} \left[ s_p - s \cdot \left( \frac{\text{vol}(P)}{\text{vol}(\mathbb{RP}^N)} \right) > s^{1/2 + \beta} \cdot \left( \frac{\text{vol}(P)}{\text{vol}(\mathbb{RP}^N)} \right) \right] < \frac{\delta}{3} . \tag{A.4} \]

Setting \( \bar{s}_1 = \max \{ (\frac{3 \log \delta}{\text{vol}(\mathbb{RP}^N)} \right)^{1/2}, (\frac{3 \log \delta}{\text{vol}(P) \setminus P)} \right)^{1/2}, (\frac{3 \log \delta}{\text{vol}(\mathbb{RP}^N)} \right)^{1/2} \}, \) observe that if \( s \geq \bar{s}_1 \),

\[ \mathbb{P}[\Omega_1(\varepsilon) \cap \Omega_2(\varepsilon) \cap \Omega_3] = 1 - \mathbb{P} [\Omega_1(\varepsilon)^c \cup \Omega_2(\varepsilon)^c \cup \Omega_3^c] \]

\[ \geq 1 - \mathbb{P} [\Omega_1(\varepsilon)^c] + \mathbb{P} [\Omega_2(\varepsilon)^c] + \mathbb{P} [\Omega_3^c] \quad \text{(by a union bound)} \]

\[ \geq 1 - 3 \cdot \frac{\delta}{3} = 1 - \delta \quad \text{(by (A.2), (A.3), and (A.4)).} \]

Finally, from (A.1), we can easily deduce that

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

This completes the proof of Proposition 10.

\[ \square \]

ACKNOWLEDGEMENT

The authors would like to thank Arthur Renaudineau for pointing out a missing hypothesis in the statement of Corollary 5 in the first arXiv version. Saugata Basu was partially supported by NSF grants CCF-1618981, DMS-1620271 and CCF-1910441. Abhiram Natarajan was supported by NSF grant CCF-1910441.

JOURNAL INFORMATION

The Journal of the London Mathematical Society is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

REFERENCES

1. P. Agarwal and M. Sharir, Arrangements and their applications, in: J. Urrutia, and J. R. Sack (eds.), Handbook of computational geometry, North-Holland, Amsterdam, 2000, pp. 49–119. MR 1746675.
2. N. Alon, J. Pach, R. Pinchasi, R. Radoičić, and M. Sharir, Crossing patterns of semi-algebraic sets, J. Combin. Theory Ser. A 111 (2005), no. 2, 310–326. MR 1746675.

3. A. I. Barvinok, On the Betti numbers of semialgebraic sets defined by few quadratic inequalities, Math. Z. 225 (1997), no. 2, 231–244. MR 98f:14044.

4. S. Basu, Different bounds on the different Betti numbers of semialgebraic sets, Discrete Comput. Geom. 30 (2003), no. 1, 65–85. MR 1991587 (2004f:14082).

5. S. Basu, D. V. Pasechnik, and M.-F. Roy, 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 (2010), no. 2, 529–553. MR 2608951 (2011d:14102).

6. S. Basu, R. Pollack, and M.-F. Roy, On the number of cells defined by a family of polynomials on a variety, Mathematika 43 (1996), no. 1, 120–126. MR 1401711.

7. S. Basu, R. Pollack, and M.-F. Roy, Algorithms in real algebraic geometry, 2nd edn, Algorithms Comput. Math., vol. 10, Springer, Berlin, 2006 (second edition). MR 1998147 (2004g:14064).

8. S. Basu and A. Rizzie, Multi-degree bounds on the betti numbers of real varieties and semi-algebraic sets and applications, Discrete Comput. Geom. 59 (2018), no. 3, 553–620.

9. R. Bott and L. W. Tu, Differential forms in algebraic topology, Graduate Texts in Mathematics, vol. 82, Springer, New York–Berlin, 1982. MR 658304.

10. S. Boucheron, G. Lugosi, and P. Massart, Concentration inequalities: An asymptotic theory of independence, Oxford Univ. Press, 2013.

11. P. Bürgisser, Average Euler characteristic of random real algebraic varieties, C. R. Math. Acad. Sci. Paris 345 (2007), no. 9, 507–512. MR 2375112.

12. P. Burgisser and A. Lerario, Probabilistic Schubert calculus, Crelle’s J. (2018). https://doi.org/10.1515/crelle-2018-0009

13. E. Calabi, Linear systems of real quadratic forms, Proc. Amer. Math. Soc. 15 (1964), 844–846. MR 0166203.

14. M. De Berg, M. Van Kreveld, M. Overmars, and O. C. Schwarzkopf, Visibility graphs, Computational geometry, Springer, Berlin, 2000, pp. 307–317.

15. A. Edelman and E. Kostlan, How many zeros of a random polynomial are real?, Bull. Amer. Math. Soc. (N.S.) 32 (1995), no. 1, 1–37. MR 1290398.

16. P. Erdös and A. Rényi, On random graphs, i, Publ. Math. Debrecen 6 (1959), 290–297.

17. Y. V. Fyodorov, A. Lerario, and E. Lundberg, On the number of connected components of random algebraic hypersurfaces, J. Geom. Phys. 95 (2015), 1–20. MR 3357820.

18. D. Gayet and J.-Y. Welschinger, Expected topology of random real algebraic submanifolds, J. Inst. Math. Jussieu 14 (2015), no. 4, 673–702.

19. E. N. Gilbert, Random graphs, Ann. Math. Stat. 30 (1959), no. 4, 1141–1144.

20. E. N. Gilbert, Random plane networks, J. Soc. Ind. Appl. Math. 9 (1961), no. 4, 533–543.

21. B. Iversen, Cohomology of sheaves, Universitext, Springer, Berlin, 1986. MR 842190 (87m:14013).

22. E. Kostlan, On the distribution of roots of random polynomials, From Topology to Computation: Proceedings of the Smalefest, Springer, New York, N.Y., 1993, pp. 419–431. MR 1246137.

23. A. Lerario, Convex pencils of real quadratic forms, Discrete Comput. Geom. 48 (2012), no. 4, 1025–1047.

24. A. Lerario, Random matrices and the average topology of the intersection of two quadrics, Proc. Amer. Math. Soc. 143 (2015), no. 8, 3239–3251. MR 3348768.

25. A. Lerario, Complexity of intersections of real quadrics and topology of symmetric determinantal varieties, J. Eur. Math. Soc. (JEMS) 18 (2016), no. 2, 353–379. MR 3459954.

26. A. Lerario and E. Lundberg, Gap probabilities and Betti numbers of a random intersection of quadrics, Discrete Comput. Geom. 55 (2016), no. 2, 462–496.

27. S. N. Majumdar, C. Nadal, A. Scardicchio, and P. Vivo, How many eigenvalues of a gaussian random matrix are positive?, Phys. Rev. E 83 (2011), no. 4, 041105.

28. R. Schneider, Convex bodies: the Brunn - Minkowski theory, no. 151, Cambridge Univ. Press, Cambridge, 2014.

29. M. Shub and S. Smale, Complexity of Bezout’s theorem II - volumes and probabilities, Computational algebraic geometry, Springer, Berlin, 1993, pp. 267–285.

30. T. Tao, Topics in random matrix theory, vol. 132, Amer. Math. Soc., Providence, R.I., 2012.