

## COUNTING FACES OF RANDOMLY PROJECTED POLYTOPES WHEN THE PROJECTION RADICALLY LOWERS DIMENSION

DAVID L. DONOHO AND JARED TANNER

### 1. INTRODUCTION

**1.1. Three surprises of high dimensions.** This paper develops asymptotic methods to count faces of random high-dimensional polytopes, a seemingly dry and unpromising pursuit. Yet our conclusions have surprising implications – in statistics, probability, information theory, and signal processing – with potential impacts in practical subjects like medical imaging and digital communications. Before involving the reader in our lengthy analysis of high-dimensional face counting, we describe three implications of our results.

**1.1.1. Convex hulls of Gaussian point clouds.** Consider a random point cloud of  $n$  points  $x_i$ ,  $i = 1, \dots, n$ , sampled independently and identically from a Gaussian distribution in  $\mathbf{R}^d$  with nonsingular covariance. This is a standard model of multivariate data; its properties are increasingly important in a wide range of applications. At the same time, it is an attractive and in some sense timeless object for theoretical study.

Properties of the convex hull of the random point cloud  $\mathcal{X} = \{x_i\}$  have attracted interest for several decades, increasingly so in recent years; there is a now-voluminous literature on the subject. The results could be significant for understanding outlier detection or classification problems in machine learning.

A classical asymptotic result, [23], holds that if the dimension  $d$  stays fixed, while the number of points  $n \rightarrow \infty$ , the convex hull has  $\sim c_d \log^{(d-1)/2}(n)$  vertices, and the remaining points of  $\mathcal{X}$  are all of course in the interior of the convex hull.

The modern trend in statistics and probability is to consider the case where both the number of dimensions  $d$  and the sample size  $n$  are large [20, 22]. In that case, the intuition fostered by the classical fixed-dimension asymptotic is wildly inaccurate. Rather than the relatively few extreme points that we saw in the fixed-dimension asymptotic, there are now many extreme points, many edges, etc. – in fact, the maximal number conceivable. More precisely, let  $k_d^* = k_d^*(\mathcal{X})$  denote the largest number  $k$  such that

- each point  $x_j$  is a vertex of  $\text{conv}(\mathcal{X})$ ;

---

Received by the editors July 5, 2006.

2000 *Mathematics Subject Classification.* Primary 52A22, 52B05, 52B11, 52B12, 62E20, 68P30, 68P25, 68W20, 68W40, 94B20 94B35, 94B65, 94B70.

The first author acknowledges partial support from NSF DMS 05-05303, 01-40698 (FRG), and NIH.

The second author acknowledges support from NSF fellowship DMS 04-03041 and thanks John E. and Marva M. Warnock for their generous support in the form of an endowed chair.

©2008 American Mathematical Society  
Reverts to public domain 28 years from publication

- each line segment  $[x_j, x_i]$ ,  $j \neq i$ , is an edge of  $\text{conv}(\mathcal{X})$ ;
- ...
- every  $k + 1$  distinct points of  $\mathcal{X}$  span a  $k$ -face of  $\text{conv}(\mathcal{X})$ .

Then, not only is  $k_d^*$  defined and positive, it is rather large. Section 7.1 below gives a corollary of our main results saying roughly that, for  $\epsilon > 0$ , with overwhelming probability for large  $d < n$ ,

$$(1.1) \quad k_d^* > \frac{d}{2e \log(n/d)} (1 - \epsilon).$$

Not only are no points of  $\mathcal{X}$  ‘inside’  $\text{conv}(\mathcal{X})$ , it is also true that no edge between any pair of points crosses the interior of  $\text{conv}(\mathcal{X})$ , etc. This is about as far from low-dimensional intuition as it is possible to get!

1.1.2. *Signal recovery from random projections.* Suppose we are interested in a vector  $x_0 \in \mathbf{R}^N$  which, although unknown to us, is known to be  $k$ -sparse – i.e. we know that it has at most  $k$  nonzeros when represented in the standard basis. We are allowed to ask some number  $n$  of ‘questions’ about  $x_0$ , each question coming in the form of a projection  $y_i = \langle a_i, x_0 \rangle$  on a vector  $a_i \in \mathbf{R}^N$ . How big should  $n$  be so that we may recover  $x_0$ , i.e. “how many questions suffice to recover a  $k$ -sparse vector”?

Obviously,  $N$  suffice (simply ask for the  $N$  coordinates in the standard unit vector basis), but in cases where  $x_0$  is very sparse,  $k \ll N$ , many fewer questions will do. Indeed,  $n = 2k + 1$  suffice; simply take the  $a_i$  as independent random vectors with iid Gaussian entries. (The matrix  $A$  having  $a_i$  for rows will then have its columns in general position, which implies that there cannot be two  $k$ -sparse vectors  $x_0$  and  $x_1$  both answering the questions in the same way [9].) Although such a random set of questions determines  $x_0$  uniquely, the task of actually recovering  $x_0$  from such information is daunting; in general, one must enumerate the  $k$ -subsets of columns of  $A$  looking for a subset which can be combined linearly to generate  $y$ .

A more useful question: how many questions are needed in order to permit *computationally tractable* recovery of  $x_0$ ? We will give precise and simply stated results for reconstruction using standard linear programming.

Generate  $n$  questions ‘at random’ by simply taking for  $A$  an  $n \times N$  matrix with iid Gaussian  $N(0, 1/n)$  entries. Obtain a vector of  $n$  measurements  $y = Ax_0$  where  $x_0$  has  $k$  nonzeros. Consider the convex optimization problem

$$(P_1) \quad \min \|x\|_1 \text{ subject to } y = Ax.$$

If  $n$  is large enough relative to  $k$ , then the solution  $x_1$  to  $(P_1)$  is very likely to be *exactly*  $x_0$ . Section 7.3 below gives a corollary of this paper’s main results showing that, for  $N$  much larger than  $k$ , and both large, this exact equality happens as soon as

$$(1.2) \quad n \geq 2k \cdot \log(N/n)(1 + o_p(1)).$$

(Here  $o_p(1)$  denotes a sequence of random variables  $(V_n)$  having  $P\{|V_n| > \epsilon\} = o(1)$  as  $n \rightarrow \infty$  for each  $\epsilon > 0$ .)

Thus if we sample not  $2k + 1$  projections but instead roughly  $2k \log(N/n)$ , we can efficiently reconstruct the  $k$ -sparse vector, and this can be far fewer than the  $N$  samples superficially required.

1.1.3. *How many gross errors can we efficiently correct?* Consider a stylized problem of transmitting  $m$  ‘pieces’ of information – i.e.  $m$  numbers – with immunity to occasional transmission errors. A standard strategy is to encode the data to be transmitted as a block of  $N > m$  numbers and to decode the received block. Let  $B$  be an  $m \times N$  matrix. Given a vector  $u \in \mathbf{R}^m$  to be transmitted, encode it as  $v = B^T u \in \mathbf{R}^N$  and transmit. The receiver measures  $w = v + z$  where  $w \in \mathbf{R}^N$  and  $z$  represents transmission errors. The receiver in some way decodes the  $N$  numbers, hoping to produce the  $m$  original entries in  $u$ .

The nonzeros in  $z$  represent transmission errors; call the number of nonzeros  $k$ . How many errors can such a scheme tolerate? In principle, if  $B$  is well-chosen and  $N - m = 2k + 1$ , it is possible to correct  $k$  errors. To do so, the receiver executes a combinatorial search through all possible locations of the  $k$  nonzeros among the  $N$  received values, to see which error pattern is consistent with the underlying model. Unfortunately, such a brute-force scheme is impractical for all but the smallest  $N$ . More to the point is the question of how many errors a practical decoding scheme can tolerate.

A simple decoding scheme based on  $(P_1)$  can be used if the encoding matrix  $B$  is generated in a specific way. Let  $U$  be a random orthogonal matrix, uniformly distributed on  $O(N)$ , and partition it as  $U = \begin{pmatrix} A \\ B \end{pmatrix}$  where the encoding matrix  $B$  is  $m \times N$  and the generalized checksum matrix  $A$  is  $n \times N$ , with  $m + n = N$ . Given the received data  $w$ , form the generalized checksum  $y = Aw$ . Solve the instance of  $(P_1)$  given by  $(y, A)$ , obtaining  $x_1$ . The generalized checksum is used to estimate the error pattern (as  $BB^T$  is the identity and  $AB^T = 0$ ), and the optimization result  $x_1$  is our estimate of  $z$ . Reconstruct by subtracting this estimate of the error out of the received message and projecting down from  $\mathbf{R}^N$  to  $\mathbf{R}^m$ :  $u_1 = B(w - x_1)$ .

As  $(P_1)$  is a standard convex optimization problem, this can be considered computationally tractable. How many errors can this scheme tolerate?

To answer this quantitatively, let us call  $R = n/N$  the *rate* of the code, and consider the regime of high-rate coding, where  $R$  is nearly one. In this regime we do not want to expand the block length by very much in our encoding, but we still want to gain some immunity to errors.

The results just stated in Section 1.1.2 together with the discussion in Section 7.4 below together imply the following. Consider a sequence of problems  $(n, N_n)$  with  $R_n = n/N_n \rightarrow 1$  sufficiently slowly. Suppose the error vector  $z$  contains  $k$  nonzeros and is stochastically independent of  $(A, B)$ , so the sites and signs of the nonzeros are random and independent of  $A$ . There is *perfect recovery*  $u_1 = u$  provided  $k \leq k_n^W$ , where  $k_n^W$  is a random variable dependent on  $(A, B)$  and obeying

$$(1.3) \quad k_n^W = n / (2 \log(1 / (1 - R_n))) (1 + o_p(1)), \quad n \rightarrow \infty.$$

In short, if we use very long blocks and stipulate a very small loss in transmission rate  $R_n = 1 - \epsilon_n$ , with  $\epsilon_n$  small, we can use linear programming to correct about  $n / 2 \log(\epsilon_n)$  errors.

Results to be stated below – see Section 7.2 – imply an even more impressive result. Again, consider a sequence of problems  $(n, N_n)$  with  $R_n = n/N_n \rightarrow 1$  sufficiently slowly. Suppose the error vector  $z$  contains  $k$  nonzeros *at arbitrary sites* and with *arbitrary nonzeros*. There is *perfect recovery*  $u_1 = u$  provided  $k \leq k_n^S$ ,

where  $k_n^S$  is a random variable dependent on  $(A, B)$  and obeying

$$(1.4) \quad k_n^S \geq n / (2e \log(\sqrt{\pi} / (1 - R_n))(1 + o_p(1))), \quad n \rightarrow \infty.$$

In short, if we use very long blocks and stipulate a very small loss in transmission rate  $R_n = 1 - \epsilon_n$ , with  $\epsilon_n$  small, we can use linear programming to correct all possible patterns of about  $n/2e \log(\epsilon_n)$  errors.

Note that the sites and values of the errors can here be arbitrary; they can be chosen by a malicious opponent who knows  $v$ ,  $B$ ,  $A$ , and  $u$ ! The noise can thus be arbitrarily more energetic than the signal, it can be carefully chosen, and still it is completely suppressed. In contrast to (1.3), which requires errors to be in random positions, (1.4) allows them to occur in bursts or in any other malicious patterns.

**1.2. Random projections of convex polytopes.** The surprises (1.1), (1.2), (1.3), (1.4) are facets of a phenomenon which makes appearances throughout the mathematical sciences, in the fields of statistics, probability, information theory, and signal processing. Underlying all of these surprises are sharp thresholds in the behavior of face counts of random high-dimensional polytopes. We now develop the terminology and framework clarifying what we mean by such thresholds, only later explaining how they imply (1.1)–(1.4).

Let  $T = T^{N-1}$  denote the standard simplex  $\{x : \sum_i x_i = 1, x_i \geq 0\}$  and let  $C = C^N$  denote the standard cross-polytope in  $\mathbf{R}^N$ , i.e. the collection of vectors  $\{x : \|x\|_1 \leq 1\}$  (also called the  $\ell_1^N$ -ball). Here and in what follows, let  $Q$  be either  $T^{N-1}$  or  $C^N$ .

Let  $A$  be an  $n \times N$  random matrix with Gaussian iid entries. The image  $AQ$  is a convex subset of  $\mathbf{R}^n$ , in fact, a convex polytope. We are interested in the case  $n < N$ , so that multiplication by  $A$  lowers the dimension.

It makes sense to count the number of  $k$ -dimensional faces of  $Q$  and  $AQ$ ,  $0 \leq k \leq n$ . In general  $AQ$  will have fewer faces than  $Q$ . More precisely, if we enumerate the  $k$ -faces  $F$  of  $Q$ , each  $AF$  will either be a face of  $AQ$  or will belong to the interior of  $AQ$ . More picturesquely, some of the faces of  $Q$  ‘survive projection’, while some of the faces ‘do not survive’.

**1.2.1. Typical faces of random polytopes.** The  $k$ -dimensional faces of  $Q$  make a finite set,  $\mathcal{F}_k(Q)$  (say); by placing uniform measure on this set, we may speak of *typical faces* as follows.

**Definition 1.0.** Consider a sequence of problem sizes  $(n, N_n)$ . Suppose that, for a given projector  $A$ , a property  $\mathcal{P} = \mathcal{P}(F; A)$  of the projected face  $AF$  holds, at a fraction  $\pi_{k,n} = \pi_{k,n}(A)$  of  $k$ -faces  $F \in \mathcal{F}_k(Q)$ . Suppose that the random variable  $\pi_{k,n} \rightarrow_p 1$  (goes to one in probability) as  $n \rightarrow \infty$ . Then we say that (asymptotically) the *typical  $k$ -face  $F \in \mathcal{F}_k(Q)$  has property  $\mathcal{P}$* .

We now consider the fate of the typical  $k$ -face of  $A$  under the projection  $Q \mapsto AQ$ . In the following statements, fix  $\epsilon > 0$ .

- Let  $F$  be a typical  $k$ -face of  $T^{N-1}$ . Is  $AF$  a face of  $AT^{N-1}$ ? The answer is *yes* provided  $N$  and  $k$  are both large and  $n > 2k \log(N/n)(1 + \epsilon)$  and *no* provided  $n < 2k \log(N/n)(1 - \epsilon)$ .
- Let  $F$  be a typical  $k$ -face of  $C^N$ . Is  $AF$  a face of  $AC^N$ ? The answer is *yes* provided  $N$  and  $k$  are both large and  $n > 2k \log(N/n)(1 + \epsilon)$  and *no* provided  $n < 2k \log(N/n)(1 - \epsilon)$ .

In short, there are well-defined thresholds at which *typical*  $k$ -faces of the simplex and the cross-polytope begin to get ‘swallowed up’ under random lowering of dimension.

1.2.2. *All faces of random polytopes.* We now consider the fate of the whole collection of  $k$ -faces simultaneously.

- For every  $k$ -face  $F$  of  $T^{N-1}$ , is  $AF$  also a  $k$ -face of  $AT^{N-1}$ ? The answer is overwhelmingly likely to be *yes* provided  $N$  and  $k$  are both large and  $n > 2ek \log(N/(n \cdot 2\sqrt{\pi}))(1 + \epsilon)$  and *no* provided  $n < 2ek \log(N/(n \cdot 2\sqrt{\pi}))(1 - \epsilon)$ .
- For every  $k$ -face  $F$  of  $C^N$ , is  $AF$  also a  $k$ -face of  $AC^N$ ? The answer is overwhelmingly likely to be *yes* provided  $N$  and  $k$  are both large and  $n > 2ek \log(N/(n \cdot \sqrt{\pi}))(1 + \epsilon)$  and *no* provided  $n < 2ek \log(N/(n \cdot \sqrt{\pi}))(1 - \epsilon)$ .

Below certain specific bounds on the face dimension  $k$ , no faces are lost in projection.

1.3. **Background: Proportional growth setting.** Our promised applications, such as (1.1) and (1.2), were stated merely with  $n$  and  $N$  (respectively  $k$  and  $N$ ) both large. However, the backbone of our analysis (and the bulk of prior scholarly work) concerns a setting in which  $(k, n, N)$  are large but also comparable in size. We consider this case first and later extend our results to a more general setting.

**Definition 1.1.** A sequence of triples  $((k_n, n, N_n) : n = n_0, n_0 + 1, \dots)$  will be said to **grow proportionally** if there are  $\delta \in (0, 1)$  and  $\rho \in (0, 1)$  so that

$$(1.5) \quad k_n/n \rightarrow \rho, \quad n/N_n \rightarrow \delta, \quad n \rightarrow \infty.$$

We omit subscripts  $n$  on  $k$  and  $N$  unless they are absolutely necessary.

There are several significant prior results concerning thresholds for face counts in the proportional-growth setting. Let  $f_k(Q)$  denote the number of  $k$ -faces of  $Q$ .

1.3.1. *Weak thresholds.* Consider first the question whether the *typical* face survives projection.

- *Simplex.* There is a function  $\rho_W^+ : [0, 1] \mapsto [0, 1]$  with the following property. In the proportional growth setting with  $\rho < \rho_W^+(\delta)$ , we have

$$Ef_k(AT^{N-1}) = f_k(T^{N-1})(1 - o(1)), \quad 0 \leq k < \rho n, \quad n \rightarrow \infty;$$

while if  $\rho > \rho_W^+(\delta)$ , we have that for some  $\epsilon > 0$  and some sequence  $(k_n)$  with  $k_n < \rho n$ ,

$$Ef_k(AT^{N-1}) < f_k(T^{N-1})(1 - \epsilon), \quad n \rightarrow \infty.$$

Informally, the fraction of faces lost:

$$(f_k(T^{N-1}) - Ef_k(AT^{N-1}))/f_k(T^{N-1})$$

is either negligible or nonnegligible depending on which side of  $\rho_W^+(\delta)$  the fraction  $k/n$  sits. In words, for  $k_n$  somewhat below the threshold  $n \cdot \rho_W^+(\delta)$  the typical  $k_n$ -face of the simplex survives projection into  $n$  dimensions; but for  $k_n$  somewhat above the threshold this is no longer true.

- *Cross-polytope.* There is a function  $\rho_W^\pm : [0, 1] \mapsto [0, 1]$  with the following property. In the proportional growth setting with  $\rho < \rho_W^\pm(\delta)$ , we have

$$Ef_k(AC^N) = f_k(C^N)(1 - o(1)), \quad 0 \leq k < \rho n, \quad n \rightarrow \infty;$$

while if  $\rho > \rho_W^\pm(\delta)$ , we have for some  $\epsilon > 0$  and some sequence  $(k_n)$  with  $k_n < \rho n$ ,

$$Ef_k(AC^N) < f_k(C^N)(1 - \epsilon), \quad n \rightarrow \infty.$$

Again, for  $k_n$  somewhat below the threshold  $n \cdot \rho_W^\pm(\delta)$  the typical  $k_n$ -face of the cross-polytope survives projection into  $n$  dimensions; but for some  $k_n$  at or above the threshold this is no longer true.

In view of these results, the square  $0 \leq \delta, \rho \leq 1$  may be decorated with a *phase diagram*. The two  $\rho_W$ -functions mark *phase transitions*; there are two phases for the property “the projected polytope has *approximately* as many faces as the original”. Below the transitions, the property holds asymptotically for large  $n$ , while above the transitions, the property fails asymptotically for large  $n$ . Both transitions are depicted in Figure 1.1, which displays a *phase diagram* in the  $(\delta, \rho)$ -plane. In the region below these curves, typical faces are not lost; in the region above those curves, typical faces are lost. To interpret these curves, note that if  $\delta = 1/2$ , we are lowering dimension by 50%, and if  $n$  is large, then the typical  $k$ -face of the simplex survives, for  $k/n \leq .5581$ , while the typical  $k$ -face of the cross-polytope survives, for  $k/n \leq .3848$ .

Vershik and Sporyshev [34] pioneered study of the proportional growth setting and proved the existence of what we call here the weak threshold for the simplex case. The weak threshold function  $\rho_W^+$  was introduced using our notation and carefully studied by the authors in [12], where numerical methods were developed for its calculation and display. The weak threshold for the cross-polytope  $\rho_W^\pm$  was introduced in [10], calculated, and displayed.

1.3.2. *Strong thresholds.* We now ask when the difference between  $f_k(AQ)$  and  $f_k(Q)$  is small in absolute, not relative, terms.

- *Simplex.* There is a function  $\rho_S^+ : [0, 1] \mapsto [0, 1]$  with the following property. In the proportional growth setting with  $\rho < \rho_S^+(\delta)$ , we have

$$Ef_k(AT^{N-1}) = f_k(T^{N-1}) - o(1), \quad 0 \leq k < \rho n, \quad n \rightarrow \infty.$$

Thus, for  $k$  below  $n \cdot \rho_S^+(\delta)$  there are on average as many  $k$ -faces of the projected simplex as the original simplex. On the other hand, if  $\rho > \rho_S^+(\delta)$ , then there is a sequence  $(k_n)$  with  $k_n < n\rho$  along which

$$f_k(T^{N-1}) - Ef_k(AT^{N-1}) \rightarrow \infty.$$

- *Cross-polytope.* There is a function  $\rho_S^\pm : [0, 1] \mapsto [0, 1]$  with the following property. In the proportional growth setting with  $\rho < \rho_S^\pm(\delta)$ ,

$$Ef_k(AC^N) = f_k(C^N) - o(1), \quad 0 \leq k < \rho n, \quad n \rightarrow \infty.$$

Thus, for  $k$  somewhat below  $n \cdot \rho_S^\pm(\delta)$  there are on average just as many  $k$ -faces of the projected cross-polytope as the standard cross-polytope. On the other hand, if  $\rho > \rho_S^\pm(\delta)$ , then there is  $k_n < n\rho$  with

$$f_k(C^N) - Ef_k(AC^N) \rightarrow \infty.$$

The function  $\rho_S^+$  was introduced and carefully studied by the authors in [12], and numerical methods were developed for its calculation and display. The threshold function  $\rho_S^\pm$  was introduced in [10], calculated, and displayed.

These strong thresholds have another interpretation. Consider the event “all low-dimensional faces survive projection”, i.e.

$$\Omega(k, n, N) = \{f_\ell(AT^{N-1}) = f_\ell(T^{N-1}), \ell = 0, \dots, k\}.$$

Simple arguments as in [10, 12] show that if  $\rho < \rho_S^+(\delta)$ , the probability

$$P(\Omega(k_n, n, N_n)) \rightarrow 1, \quad n \rightarrow \infty.$$

Hence, below the strong phase transition, all low-dimensional faces survive projection. Parallel arguments can be made in the cross-polytope case. Thus in the region where  $k/n$  is below the corresponding  $\rho_S$  function not only are very few  $k$  faces lost on average, but actually, there is overwhelming probability that no faces are lost.

These  $\rho$ -functions are depicted in Figure 1.1. The strong thresholds  $\rho_S^+$  and  $\rho_S^\pm$  fall below the corresponding weak thresholds  $\rho_W^+$ ,  $\rho_W^\pm$ ; indeed a property holding for every  $k$ -face is less likely to hold than one holding for the typical  $k$ -face. To interpret these curves, note that if  $\delta = 1/2$  so we are lowering dimension by 50%, then every  $k$ -face of the simplex survives, for  $k \leq .1335$ , while every  $k$ -face of the cross-polytope survives, for  $k \leq .0894$ .

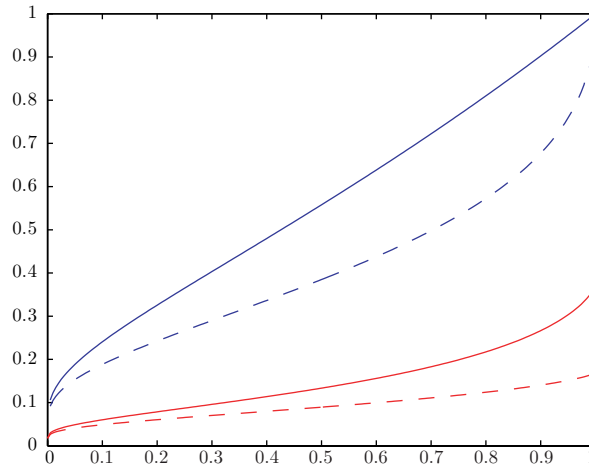


FIGURE 1.1. Thresholds for  $\delta \in (0,1)$  from top to bottom:  $\rho_W^+$  (blue – solid),  $\rho_W^\pm$  (blue – dashed),  $\rho_S^+$  (red – solid), and  $\rho_S^\pm$  (red – dashed).

**1.4. Main results: Proportional growth setting.** For applications, the range where  $\delta$  is small is very interesting; it corresponds to:

- studying convex hulls of Gaussian point clouds where there are many points relative to the number of dimensions – Section 1.1.1;
- recovering a sparse signal from very few samples – Section 1.1.2;
- protecting against errors in digital transmission while sacrificing very little in the transmission rate – Section 1.1.3.

Previous work by the authors [12, 10] considered the asymptotic behavior of the several  $\rho(\delta)$  functions just defined and showed that  $\rho(\delta) \geq c_\epsilon \log(1/\delta)^{-1-\epsilon}$  for

each  $\epsilon > 0$ . Work by others [6, 29, 25] can be seen to imply that actually  $\rho(\delta) \geq c \log(1/\delta)^{-1}$ . In this paper we determine the precise constants in the asymptotic behavior as  $\delta \rightarrow 0$ . These precise constants are important in applications; they can be used to plan how many samples to take in a digital imaging system or how much transmission rate sacrifice to make for a given error resistance.

**Theorem 1.2** (Weak threshold – simplex).

$$(1.6) \quad \rho_W^+(\delta) \sim |2 \log(\delta)|^{-1}, \quad \delta \rightarrow 0.$$

**Theorem 1.3** (Strong threshold – simplex).

$$(1.7) \quad \rho_S^+(\delta) \sim |2e \log(\delta 2\sqrt{\pi})|^{-1}, \quad \delta \rightarrow 0.$$

Comparing these results:

- Note the leading factor  $e$  in  $\rho_S$ . The highest dimension  $k$  where the *vast majority* of  $k$ -faces survive projection is asymptotically  $e$  times higher than the dimension where we can guarantee that *every*  $k$ -face survives.
- An additional difference is the  $2\sqrt{\pi}$  factor in the argument of the logarithm.

**Theorem 1.4** (Weak threshold – cross-polytope).

$$(1.8) \quad \rho_W^\pm(\delta) \sim |2 \log(\delta)|^{-1}, \quad \delta \rightarrow 0.$$

**Theorem 1.5** (Strong threshold – cross-polytope).

$$(1.9) \quad \rho_S^\pm(\delta) \sim |2e \log(\delta\sqrt{\pi})|^{-1}, \quad \delta \rightarrow 0.$$

Comparing the cross-polytope results to those for the simplex:

- Remarkably, to first order, the thresholds are the same for the simplex and cross-polytope. This is surprising since at moderate values of  $\delta$  the two functions are quite different; see Figure 1.1.
- The bounds on strong thresholds agree, except for factors of 2 in the argument of the logarithm.

The weak-threshold asymptotic behavior (1.2) and (1.8) closely matches  $\rho_W^+$  and  $\rho_W^\pm$  for modest values of  $\delta$  – see Figure 1.2. The strong-threshold asymptotic behavior, on the other hand, slowly approaches  $\rho_S^+$  and  $\rho_S^\pm$  from above – see Figure 1.3.

**1.5. Beyond proportional growth.** Having considered the Vershik-Sporyshev proportional growth scenario, we now generalize to the case where  $N$  can be dramatically larger than  $n$ . This is important for applications where we want to sample very few projections of a high-dimensional object. The paper [8] exhibits stylized imaging problems where an  $N$ -pixel image can be reconstructed by asking  $n = O(N^a)$  questions,  $a < 1$ . This of course lies outside the reach of proportional growth as it is dramatically smaller than  $N$ , underscoring the potential interest of the ‘how many questions’ problem of Section 1.1.2 where the number of questions  $n \ll N$ .

We would naively hope that the same threshold functions  $\rho(\cdot)$  “work” even outside the proportional growth setting. That is, in a setting where  $n/N_n \rightarrow 0$ , we would hope to get the ‘right answer’ for the behavior of face counts by simply ‘plugging in’ a *varying*  $\delta = \delta_n = n/N_n \rightarrow 0$  into the appropriate  $\rho$ -function. Happily, such naive hopes go unpunished.



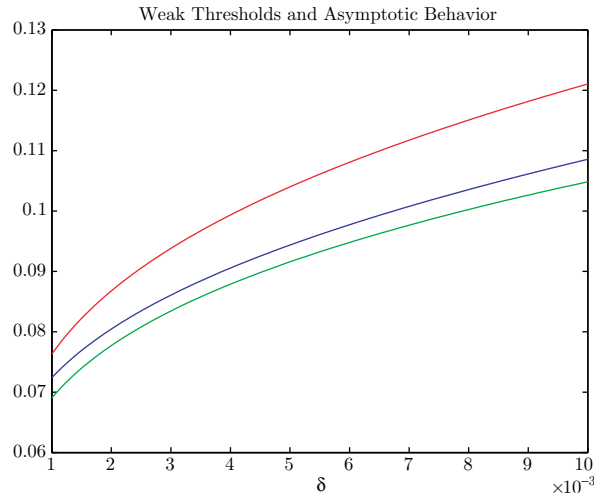


FIGURE 1.2. Weak thresholds  $\rho_W^+$  (red),  $\rho_W^\pm$  (green), and their asymptotic behavior,  $|2\log(\delta)|^{-1}$  (blue), from Theorems 1.2 and 1.4,  $\delta \in [10^{-3}, 10^{-2}]$ .

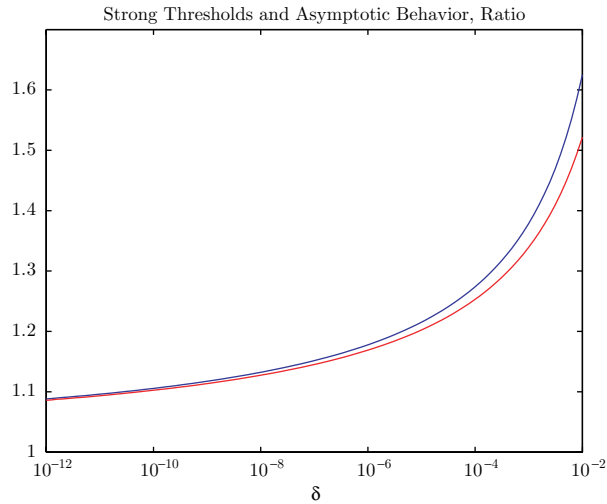


FIGURE 1.3. Ratio of the strong thresholds and their asymptotic behaviors,  $\rho_S^+(\delta)$  in blue and  $\rho_S^\pm(\delta)$  in red. The asymptotic formulae approach slowly from above by the factors shown.

We say that  $N$  grows *subexponentially* relative to  $n$  if

$$(1.10) \quad N_n/n \rightarrow \infty, \quad \frac{\log(N_n)}{n} \rightarrow 0, \quad n \rightarrow \infty.$$

**Theorem 1.6.** Consider a sequence of problem sizes  $(n, N_n)$  where  $N_n$  grows subexponentially relative to  $n$ . Let  $\rho$  be one of the four functions  $\rho_W^+$ ,  $\rho_W^\pm$ ,  $\rho_S^+$ ,  $\rho_S^\pm$ . Fix  $\epsilon > 0$  and consider a sequence  $(k_n)$  obeying  $k_n/n < \rho(n/N_n)(1-\epsilon)$  for  $n > n_0$ . Then

the same statement that was made for that  $\rho$  in the proportional growth scenario holds in this nonproportional growth scenario.

Thus, for example,  $k_n < (1 - \epsilon)\rho_W^+(n/N_n) \cdot n$  for  $n = n_0, n_0 + 1, \dots$  implies

$$Ef_k(AT^{N-1}) = f_k(T^{N-1})(1 - o(1)), \quad n \rightarrow \infty;$$

similarly,  $k_n < (1 - \epsilon)\rho_S^\pm(n/N_n) \cdot n$  for  $n = n_0, n_0 + 1, \dots$  implies that with overwhelming probability for large  $n$ ,

$$f_\ell(AC^N) = f_\ell(C^N), \quad \ell = 0, \dots, k - 1.$$

In short, the limit relations of Theorems 1.2–1.5 are useful both in proportional and nonproportional growth settings.

**1.6. Contents.** Our paper proves Theorems 1.2–1.6. The development is organized as a branching tree, with initial sections mapping out the main concepts, propositions and lemmas and later sections dealing with detailed estimates and proofs.

Section 2 introduces the underlying machinery of face counting and an analytic approach to studying asymptotic behavior. Our starting point is a beautiful and essential formula for the expected number of faces of randomly projected polytopes, due to Affentranger and Schneider and to Vershik and Sporyshev; it involves three factors, representing contributions from combinatorial aspects, from external angles and from internal angles. We focus on the exponential growth and/or decay of the factors by defining associated exponents  $\Psi(\delta, \rho)$ ; we work as if these factors behave exactly as  $\exp\{N \cdot \Psi\}$ . Each  $\rho(\cdot)$  function is defined as the smallest root  $0 = \Psi(\delta, \rho(\delta))$  of an associated exponent function  $\Psi$ , viewed as a function of  $\rho$  with  $\delta$  fixed.

Section 3 gives the proofs for the lower bound half of Theorems 1.2–1.5. The proofs are simple consequences of the asymptotic behavior of the net exponents as a function of  $\delta$  and  $\rho$  in the regime where  $\delta \rightarrow 0$ .

Section 4 develops the basic asymptotic analysis of the net exponents. The exponents in question explicitly involve tail probabilities of the Gaussian distribution; our asymptotic analysis exploits detailed estimates for the Mills ratio of the standard normal density.

Section 5 turns to the proof of Theorem 1.6, going outside the proportional growth setting. Here we have to make careful estimates of the errors incurred by treating the pieces in the Affentranger-Schneider-Vershik-Sporyshev formula as if they grow exactly like  $\exp\{N \cdot \Psi\}$ . We refine our analysis associated with the Mills ratio, getting remainder estimates assuming  $N_n$  is subexponential in  $n$ .

Sections 2–5 are preoccupied largely with proving only half of Theorems 1.2–1.5; namely the bounds  $\rho(\delta) \geq c_1/\log(c_2/\delta)$ . Section 6 gives the arguments establishing inequalities in the other direction, in the process completing the proofs of Theorems 1.2–1.5.

Section 7 shows how our face-counting results generate the applications mentioned in Section 1.1. It also presents empirical results showing that our asymptotic results work at moderate sample sizes and translates our asymptotic results into finite-sample bounds. It also considers extensions of this work and compares our results with other recent work.

2. DEFINITIONS OF  $\rho_S^+$ ,  $\rho_W^+$ ,  $\rho_S^\pm$ ,  $\rho_W^\pm$ 

The various  $\rho$  quantities referred to in Theorems 1.2–1.5 have so far been discussed behaviorally, by their role in locating or bounding phase transitions in face counts. In this section, we review an analytic definition for these quantities given in [10, 12]. The definition unfortunately requires a considerable amount of machinery associated with convex integral geometry. Equipped with such machinery, the claims made by Theorems 1.2–1.5 can be translated into sharply defined questions about the leading-order asymptotics of certain exponents. Sections 3 and 6 answers those questions.

**2.1. Expected face counts of projected polytopes.** Let  $Q$  be a polytope in  $R^N$  and  $A : R^N \mapsto R^n$  a random ortho-projection, uniformly distributed on the Grassmann manifold of all such projectors. Affentranger and Schneider [1] developed a useful identity for the expected number of faces of  $AQ$  [1]:

$$(2.1) \quad Ef_k(AQ) = f_k(Q) - 2 \sum_{\ell}' \sum_{F \in \mathcal{F}_k(Q)} \sum_{G \in \mathcal{F}_\ell(Q)} \beta(F, G) \alpha(G, Q);$$

here  $\mathcal{F}_k(Q)$  denotes the set of  $k$ -faces of  $Q$ , each  $F$  is a subspace of  $G$ , and  $\sum'$  denotes the sum over  $\ell = n+1, n+3, \dots$  where  $\ell < N$ . We are intensely interested in the discrepancy between the expected number of faces of the projected polytope  $AQ$  and the necessarily larger number of faces of the original polytope  $Q$ , i.e. in knowing *on average, how many faces are lost in the projection from  $\mathbf{R}^N$  to  $\mathbf{R}^n$* . The discrepancy in question is

$$(2.2) \quad \Delta(k, n, N; Q) := f_k(Q) - Ef_k(AQ) = 2 \sum_{\ell}' \sum_{F \in \mathcal{F}_k(Q)} \sum_{G \in \mathcal{F}_\ell(Q)} \beta(F, G) \alpha(G, Q).$$

Here the sum covers the external angles between the original polytope  $Q$  and its subfaces  $G$ ,  $\alpha(G, Q)$ , multiplied by the sum of all internal angles between each particular subface  $G$  and its faces  $F$ ,  $\beta(F, G)$ . For definitions of these angles see e.g. Grünbaum [19, Chapter 14], Matousek [26], or Affentranger and Schneider [1].

**2.2. Analytic definition of  $\rho_S^+$ ,  $\rho_S^\pm$ .** In the remainder of the paper we are always interested in just two choices of  $Q$ : the simplex,  $Q = T^{N-1}$ , and the cross-polytope,  $Q = C^N$ . Various quantities associated with the simplex case will be labeled with superscript  $+$  (as the interior of the standard simplex consists of positive vectors) and objects associated with the cross-polytope case will be labelled with superscript  $\pm$  (as the standard cross-polytope contains vectors with entries of both signs). We frequently use  $\star$  as a superscript in a statement which concerns either case, implying two different statements, with obvious substitutions.

In the introduction, the functions  $\rho_S^+$  and  $\rho_S^\pm$  were partially characterized by the claim that, for  $(k_n, n, N_n)$  growing proportionally and limit ratios  $(k_n/n, n/N_n) \rightarrow (\rho, \delta)$  with  $\rho < \rho_S^\star(\delta)$ , then

$$(2.3) \quad \Delta(k_n, n, N_n; Q) \rightarrow 0, \quad n \rightarrow \infty.$$

[Note: To make sure the reader follows our convention for  $\star$ , the previous sentence is actually two sentences, one for the symbol binding  $(\star, Q) = (+, T^{N-1})$  and one for the symbol binding  $(\star, Q) = (\pm, C^N)$ .] It was also stated that if  $\rho > \rho_S^\star$ , then for some sequence  $(k_n)$  obeying  $k_n < \rho n$  and some  $\epsilon > 0$ ,

$$(2.4) \quad \liminf_{n \rightarrow \infty} \Delta(k_n, n, N_n; Q) \geq \epsilon > 0.$$

The papers [10, 12] actually defined  $\rho_S^*(\delta)$  with the following stronger property: if  $\rho < \rho_S^*(\delta)$ , then, in the proportional growth setting (1.5) for some  $\epsilon > 0$  and  $n > n_0(\epsilon, \rho)$ , we have

$$(2.5) \quad \Delta(k_n, n, N_n; Q) \leq N_n \exp(-N_n \epsilon).$$

Those papers implied/stated without proof that if  $\rho > \rho_S^*$ , then for some sequence  $k_n < \rho n$ , some  $\epsilon > 0$ , and  $n_0$  we have

$$(2.6) \quad \Delta(k_n, n, N_n; Q) \geq \exp(N\epsilon), \quad n > n_0.$$

While conceptually both (2.3) and (2.4) are equally important parts of the picture, in practice (2.3) is the more useful/surprising. Hence in Sections 3–5 of this paper we focus on supporting assertions like (2.3) and (2.5) showing that the discrepancy is small, rather than assertions like (2.4) and (2.6) showing that the discrepancy is large. Section 6 will return to (2.4) and (2.6).

The analysis supporting the bound (2.5) for the unit simplex and cross-polytope went by first rewriting (2.2) as a sum of contributions due to faces of different dimensions:

$$\Delta(k_n, n, N_n; Q) = \sum_{\ell}' D_{\ell, n}^*$$

where

$$D_{\ell, n}^* := 2 \sum_{F \in \mathcal{F}_k(Q)} \sum_{G \in \mathcal{F}_\ell(Q)} \beta(F, G) \alpha(G, Q).$$

The papers [12, 10] defined functions  $\Psi_{net}^*(\nu, \gamma)$  for  $\star \in \{+, \pm\}$  associated with our two choices for  $Q$  (restated in Sections 2.4 and 2.5 and illustrated in Figure 2.1 for  $\Psi_{net}^+$ ); these can be used to bound  $D_{\ell, n}^*$  as follows.

Put  $\nu_{\ell, n} = \ell/N_n$  and  $\gamma_{\ell, n} = k_n/\ell$ , and note that  $\nu_{\ell, n} \in [\delta, 1]$  and  $\gamma_{\ell, n} \in [0, \rho]$  over the relevant range  $\ell = n+1, n+2, \dots$  where  $\ell < N_n$ . In the proportional growth setting [12, 10] showed that, for each  $\epsilon > 0$ , there is  $n_0(\epsilon; \delta, \rho)$  so that

$$(2.7) \quad N_n^{-1} \log(D_{\ell, n}^*) \leq \Psi_{net}^*(\nu_{\ell, n}, \gamma_{\ell, n}) + 3\epsilon, \quad \ell = n+1, n+3, \dots, \quad n \geq n_0,$$

Since our focus is the condition (2.5), we of course are interested in conditions guaranteeing that the right side is negative, uniformly over the *admissible domain* of  $(\nu, \gamma)$  pairs obeying  $\nu \geq \delta$ ,  $\gamma \leq \rho$ .

**Definition 2.1.** The **maximal operator**  $M[\cdot]$  associated to the family of rectangles where  $\nu \in [\delta, 1]$ ,  $\gamma \in [0, \rho]$  takes a function  $\psi(\nu, \gamma)$  and delivers the **maximal function**  $M[\psi](\delta, \rho)$  defined by

$$M[\psi](\delta, \rho) = \sup\{\psi(\nu, \gamma) : \nu \in [\delta, 1], \gamma \in [0, \rho]\}.$$

Applying this operator to each  $\Psi_{net}^*$  yields two maximal functions,  $M[\Psi_{net}^+]$  and  $M[\Psi_{net}^\pm]$ , to be studied extensively below. Finally we can give an analytic definition for the key quantities in Theorems 1.3 and 1.5:

**Definition 2.2.** For  $\star \in \{+, \pm\}$ , define the strong phase transition  $\rho_S^*(\delta)$  as the ‘first’ zero of  $M[\Psi_{net}^*]$ :

$$\rho_S^*(\delta) = \inf\{\rho : M[\Psi_{net}^*](\delta, \rho) = 0, \rho \in [0, 1]\}.$$

Definition 2.2 is depicted in Figure 2.1.

Several properties of the  $\rho_S^*$  are known from [10, 12]. These functions are strictly positive on  $[0, 1]$ , strictly increasing, with limit 0 as  $\delta$  tends to 0 and limits  $\approx .3679$  and  $.1685$  ( $\star = +, \pm$ , respectively) as  $\delta$  tends to 1.

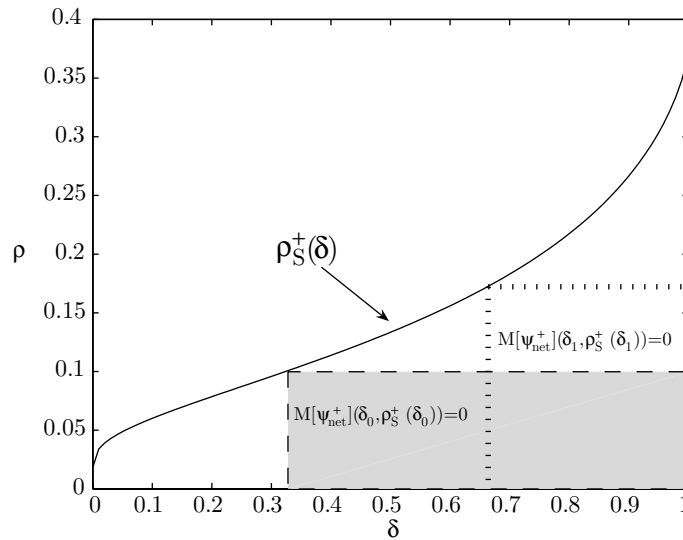


FIGURE 2.1. Throughout the shaded region  $\Psi_{net}^+ \leq 0$ . Therefore the maximal function is  $\leq 0$  at the upper left corner  $(\delta_0, \rho_S^+(\delta_0))$ ; in fact,  $M[\Psi_{net}^+](\delta_0, \rho) < 0$  for  $\rho < \rho_S^+(\delta_0)$ , and  $M[\Psi_{net}^+](\delta_0, \rho_S^+(\delta_0)) = 0$ ; this is the “first” zero of  $M[\ ]$ . The family of such first zeros of  $M[\Psi_{net}^+](\delta, \rho)$  defines  $\rho_S^+(\delta)$ .

The functions  $M[\Psi_{net}^*]$  are continuous. It follows that for  $\rho < \rho_S^*(\delta)$ ,

$$M[\Psi_{net}^*](\delta, \rho) < 0.$$

Setting  $\epsilon = |M[\Psi_{net}^*](\delta, \rho)|/4$ ,

$$(2.8) \quad N^{-1} \log(D_{\ell,n}^*) \leq -\epsilon, \quad \ell = n + 1, n + 3, \dots$$

The result (2.5) follows.

**2.3. Analytic definition of  $\rho_W^+$ ,  $\rho_W^\pm$ .** The papers [12] and [10] also defined phase transitions  $\rho_W^+$  and  $\rho_W^\pm$ . Conceptually, these quantities are defined by the notion that, for  $(k_n, n, N_n)$  growing proportionally with limit ratios  $(k_n/n, n/N_n) \rightarrow (\rho, \delta)$ , then if  $\rho < \rho_W^*(\delta)$ , the relative discrepancy is negligible, i.e.

$$(2.9) \quad \Delta(k, n, N; Q)/f_k(Q) = o(1), \quad k = 0, \dots, \lfloor \rho n \rfloor, \quad n \rightarrow \infty,$$

while for  $\rho > \rho_W^*(\delta)$  the relative discrepancy can be substantial; for some sequence  $(k_n)$  obeying  $k_n < \rho n$  and some  $\epsilon > 0$  and  $n_0$ ,

$$(2.10) \quad \Delta(k, n, N; Q)/f_k(Q) \geq \epsilon > 0, \quad n > n_0.$$

Again while conceptually both (2.9) and (2.10) are equally important, practically speaking, the former is more useful/significant than the latter, which mainly serves to show that we cannot substantially improve on (2.9). We will focus on (2.9) in Sections 3–5 and then return to discussion of (2.10) in Section 6.

Define

$$(2.11) \quad \Psi_{face}^+(\nu, \gamma) = H(\nu\gamma),$$

with  $H(\cdot)$  the type- $e$  Shannon entropy (2.17), so that under proportional growth

$$N_n^{-1} \log f_{k_n}(T^{N_n-1}) \rightarrow \Psi_{face}^+(\rho, \delta), \quad n \rightarrow \infty.$$

Also, put

$$(2.12) \quad \Psi_{face}^\pm(\nu, \gamma) = H(\nu\gamma) + \nu\gamma \log_e(2),$$

so that under proportional growth

$$N_n^{-1} \log f_{k_n}(C^{N_n}) \rightarrow \Psi_{face}^\pm(\rho, \delta), \quad n \rightarrow \infty.$$

**Definition 2.3.** For  $\star \in \{+, \pm\}$ , define  $\rho_{W^\star}^\star(\delta)$  as the ‘first’ zero of  $M[\Psi_{net}^\star - \Psi_{face}^\star]$ :

$$\rho_{W^\star}^\star(\delta) = \inf\{\rho : M[\Psi_{net}^\star - \Psi_{face}^\star](\delta, \rho) = 0, \rho \in [0, 1]\}.$$

If  $\rho < \rho_{W^\star}^\star(\delta)$ , in the proportional growth setting, then for some  $\epsilon > 0$  we have

$$(2.13) \quad \Delta(k, n, N; Q)/f_k(Q) \leq N \exp(-N\epsilon), \quad n > n_0.$$

This establishes (2.9). The  $\rho_w^\star(\delta)$  were calculated and displayed in [10, 12].

**2.4. Simplex exponent  $\Psi_{net}^+$ .** We now give more details about the exponent  $\Psi_{net}^+$  for the simplex  $T^{N-1}$ . We begin with observations by Affentranger and Schneider [1] and Vershik and Sporyshev [34], that

- there are  $\binom{N}{k+1}$   $k$ -faces of  $T^{N-1}$ ;
- for  $\ell > k$ , there are  $\binom{N-k-1}{\ell-k}$   $\ell$ -faces of  $T^{N-1}$  containing a given  $k$ -face of  $T^{N-1}$ ;
- the faces of  $T^{N-1}$  are all simplices, and the internal angle  $\beta(F, G) = \beta(T^k, T^\ell)$ , where  $T^d$  denotes the standard  $d$ -simplex.

Thus, for  $\ell = n+1, n+3, \dots$  we can write

$$\begin{aligned} D_{\ell, n}^+ &= 2 \binom{N}{k+1} \binom{N-k-1}{\ell-k} \cdot \beta(T^k, T^\ell) \cdot \alpha(T^\ell, T^{N-1}) \\ &= C_{\ell, n}^+ \cdot \beta(T^k, T^\ell) \cdot \alpha(T^\ell, T^{N-1}), \end{aligned}$$

with  $C_{\ell, n}^+$  denoting the combinatorial prefactor.

Each of the factors in this product has either exponential growth or decay. We will soon define associated exponents  $\Psi_{com}^+$ ,  $\Psi_{int}^+$ , and  $\Psi_{ext}^+$  so that, for any  $\epsilon > 0$  and  $n > n_0(\delta, \rho)$ ,

$$(2.14) \quad N^{-1} \log(C_{\ell, n}^+) \leq \Psi_{com}^+(\nu_{\ell, n}, \gamma_{\ell, n}) + \epsilon,$$

$$(2.15) \quad N^{-1} \log(\beta(T^k, T^\ell)) \leq -\Psi_{int}^+(\nu_{\ell, n}, \gamma_{\ell, n}) + \epsilon,$$

and

$$(2.16) \quad N^{-1} \log(\alpha(T^\ell, T^{N-1})) \leq -\Psi_{ext}^+(\nu_{\ell, n}) + \epsilon,$$

uniformly in  $\ell = n+1, n+3, \dots$  where  $\ell < N$ .

The exponents were introduced in [12], which showed (2.14)–(2.16); we repeat the definitions, although the reader should not expect much insight at this point. The definitions are restated in Section 4: equations (4.1), (4.5), (4.11), and (4.16), where further details emerge. The combinatorial exponent involves the base- $e$  Shannon entropy:

$$(2.17) \quad H(p) = p \log(1/p) + (1-p) \log(1/(1-p)).$$

Thus,

$$(2.18) \quad \Psi_{com}^+(\nu, \gamma) := H(\nu) + \nu H(\gamma).$$

The internal exponent is

$$(2.19) \quad \Psi_{int}^+(\nu, \gamma) := \nu(1 - \gamma) \left[ \log(y_\gamma/\gamma) + \frac{1}{2} \log(2\pi) + \frac{\gamma - 1}{2\gamma} y_\gamma^2 \right].$$

Here  $y_\gamma$  is defined implicitly by

$$(2.20) \quad \frac{1 - \gamma}{\gamma} y_\gamma = s_\gamma \quad \text{with} \quad R(s_\gamma) = 1 - \gamma,$$

where

$$(2.21) \quad R(s) := s e^{s^2/2} \int_s^\infty e^{-y^2/2} dy.$$

The function  $R(s)$  is closely related to a fundamental tool for studying tail probabilities of the standard Normal distribution – the so-called *Mills' ratio* of the normal distribution, [24, Sec. 5.37] about which more will be said in later sections. The fact that  $R(s) \rightarrow 1$  as  $s \rightarrow \infty$  signifies that the tail probability under the normal distribution is asymptotic to  $s^{-1}$  times the normal density. Details of this approximation will be crucial for our work here. Finally, the external exponent is

$$(2.22) \quad \Psi_{ext}^+(\nu) := \nu x_\nu^2 - (1 - \nu) \log Q(x_\nu),$$

with  $x_\nu$  the solution of

$$(2.23) \quad \frac{2xQ(x)}{q(x)} + 1 - \nu^{-1} = 0;$$

here  $q(x) := \pi^{-1/2} e^{-x^2}$  and  $Q(x) = \int_{-\infty}^x q(y) dy$ . Note that  $Q(\cdot)$  is the normal distribution with mean zero and standard deviation  $1/\sqrt{2}$ , and so  $x_\nu$  is again associated with the relationship between tail probabilities and density. The definition of the external angle seems at first very similar to the definition of the internal angle; however, note that  $Q(x) \rightarrow 1$  as  $x \rightarrow \infty$ , while  $q(x) \rightarrow 0$  rapidly. This difference is reflected in the behavior of the  $x_\nu$  as a function of  $\nu$  which is very different than the behavior of  $y_\gamma$  as a function of  $\gamma$ .

These  $\Psi$ -functions are all smooth functions of their arguments. For details on these exponents, see either the original source [12], where graphical displays are provided, or Section 4 below.

It follows from (2.14)–(2.16) that for  $\ell = n + 1, n + 3, \dots$ ,

$$N^{-1} \log(D_{\ell,n}^+) \leq \Psi_{com}^+(\nu_{\ell,n}, \gamma_{\ell,n}) - \Psi_{int}^+(\nu_{\ell,n}, \gamma_{\ell,n}) - \Psi_{ext}^+(\nu_{\ell,n}, \gamma_{\ell,n}) + 3\epsilon.$$

Defining now

$$\Psi_{net}^+(\nu, \gamma) := \Psi_{com}^+(\nu, \gamma) - \Psi_{int}^+(\nu, \gamma) - \Psi_{ext}^+(\nu)$$

provides us the desired property (2.7) referred to earlier, in the simplex case. Graphs were presented in [12] showing the exponent's behavior for  $\gamma = .5555$  over the range  $\nu \in (.5555, 1]$ . Software is available to make similar graphs for other parameter choices.

2.5.  $\Psi_{net}^\pm$ , **cross-polytope exponent.** Böröczky and Henk [3] previously studied the expected number of faces for the randomly projected cross-polytope  $C^N$ , and although the analysis is quite different, we utilize a number of their observations.

- There are  $2^{k+1} \binom{N}{k+1}$   $k$ -faces of  $C^N$ .
- For  $\ell > k$ , there are  $2^{\ell-k} \binom{N-k-1}{\ell-k}$   $\ell$ -faces of  $C^N$  containing a given  $k$ -face of  $C^N$ .
- The faces of  $C^N$  are all simplices, and the internal angle  $\beta(F, G) = \beta(T^k, T^\ell)$ .
- The external angle  $\alpha(G^\ell, C^N)$  is the same for all  $\ell$ -faces of  $C^N$ , the closed form expression of which was originally given in [3]. A version written in our notation was developed in [10] and is spelled out below in (5.11).

Thus, for  $\ell = n+1, n+3, \dots$  we can write

$$\begin{aligned} D_{\ell,n}^\pm &= 2 \cdot 2^\ell \cdot \binom{N}{k+1} \binom{N-k-1}{\ell-k} \beta(T^k, T^\ell) \alpha(F^\ell, C^N) \\ &= C_{\ell,n}^\pm \cdot \beta(T^k, T^\ell) \cdot \alpha(F^\ell, C^N), \end{aligned}$$

with  $C_{\ell,n}^\pm$  the combinatorial prefactor.

The factors in this product again have either exponential growth or decay. We will soon define associated exponents  $\Psi_{com}^\pm$ ,  $\Psi_{int}^\pm$ , and  $\Psi_{ext}^\pm$  so that, for any  $\epsilon > 0$  and  $n > n_0(\epsilon)$ ,

$$(2.24) \quad N^{-1} \log(C_{\ell,n}^\pm) \leq \Psi_{com}^\pm(\nu_{\ell,n}, \gamma_{\ell,n}) + \epsilon,$$

$$(2.25) \quad N^{-1} \log(\beta(T^k, T^\ell)) \leq -\Psi_{int}^\pm(\nu_{\ell,n}, \gamma_{\ell,n}) + \epsilon,$$

and

$$(2.26) \quad N^{-1} \log(\alpha(F^\ell, C^N)) \leq -\Psi_{ext}^\pm(\nu_{\ell,n}) + \epsilon,$$

uniformly in  $\ell = n+1, n+3, \dots$  where  $\ell < N$ . It follows that for  $n > n_0$ ,

$$N^{-1} \log(D_{\ell,n}^\pm) \leq \Psi_{com}^\pm(\nu_{\ell,n}, \gamma_{\ell,n}) - \Psi_{int}^\pm(\nu_{\ell,n}, \gamma_{\ell,n}) - \Psi_{ext}^\pm(\nu_{\ell,n}, \gamma_{\ell,n}) + 3\epsilon.$$

The exponents were introduced in [10], which showed (2.24)–(2.26); we rehearse the definitions, admitting they yield little insight at this point. The definitions are restated in Section 4; where further information can be obtained. The combinatorial exponent again involves the base- $e$  Shannon entropy  $H$ :

$$(2.27) \quad \Psi_{com}^\pm(\nu, \gamma) := H(\nu) + \nu H(\gamma) + \nu \log_e(2);$$

thus  $\Psi_{com}^\pm = \Psi_{com}^+ + \nu \log_e(2)$ . The internal exponent is actually the same as in the simplex case:  $\Psi_{int}^+ = \Psi_{int}^\pm$ . Finally, the external exponent is

$$(2.28) \quad \Psi_{ext}^\pm(\nu) := \nu y_\nu^2 - (1 - \nu) \log G(y_\nu),$$

with  $y_\nu$  the solution to

$$(2.29) \quad \frac{2yG(y)}{g(y)} + 1 - \nu^{-1} = 0,$$

and  $g(y) := 2\pi^{-1/2}e^{-y^2}$ ,  $G(y) = \text{erf}(y) = \int_0^y g(w)dw$ . Here  $G(\cdot)$  is the error function, also called the Half-Normal distribution  $HN(0, \frac{1}{2})$ . Again the  $\Psi$ 's are smooth functions of their arguments.

Defining now

$$\Psi_{net}^\pm(\nu, \gamma) := \Psi_{com}^\pm(\nu, \gamma) - \Psi_{int}^\pm(\nu, \gamma) - \Psi_{ext}^\pm(\nu)$$



provides us, in the cross-polytope case, the property (2.7) referred to earlier. In [10] it was shown that this is a well-defined and in fact nicely behaved quantity as a function of  $\gamma$  for each fixed  $\nu$ . Graphs in [10] portray its behavior over the range  $\nu \in (.5555, 1]$  for  $\delta = .5555$ ; software is available to compute similar graphs as other values for  $\nu$ .

### 3. ASYMPTOTICS OF $\Psi_{net}^*$ AS $\delta \rightarrow 0$

We now turn to the asymptotics at the heart of Theorems 1.2–1.5. As indicated earlier, in Sections 3–5 we focus on establishing lower bounds on  $\rho$ -functions, practically the most ‘important’ or ‘surprising’ part of our results.

We introduce a parametrized family of simple comparison functions  $r^*(\delta)$  of the form  $|\tau \log(c \cdot \delta)|^{-1}$  and control the behavior of  $\rho^*(\delta)$  by studying the maximal functions along the trajectories  $(\delta, r^*(\delta))$  as  $\delta \rightarrow 0$ . The central point will be that for  $\tau > \tau_0$ , each associated maximal functions is asymptotically negative along the trajectory  $(\delta, r^*(\delta))$ . This forces  $\rho^*(\delta) > r^*(\delta)$ .

We will also glean insights useful for establishing upper bounds on  $\rho$ -functions. It will emerge that fixing  $\tau > \tau_0$  defines a trajectory along which the net exponents are asymptotically positive and that fixing  $\tau = \tau_0$  defines a trajectory such that the difference between net and face exponents is vanishing; it will be explained in Section 6 how this implies the upper bound half of Theorems 1.2–1.5.

It is convenient to develop the results in a permuted order.

**3.1. Theorem 1.3.** Fix  $\tau > 2e$  and define  $r_S^+(\delta) := r_S^+(\delta; \tau) := |\tau \log(\delta 2\sqrt{\pi})|^{-1}$ . In what follows,  $\tau$  is always held fixed throughout an argument, while  $\delta$  is sent towards 0.

We intend to show that there is  $\delta_S^+ = \delta_S^+(\tau) > 0$  so that

$$(3.1) \quad M[\Psi_{net}^+](\delta, r_S^+(\delta)) < 0, \quad 0 < \delta < \delta_S^+.$$

This establishes the lower-bound half of Theorem 1.3, i.e. that

$$\rho_S^+(\delta) \geq |2e \log(\delta 2\sqrt{\pi})|^{-1} \cdot (1 + o(1)).$$

The other half of Theorem 1.3 can be inferred from the fact that if we instead have  $\tau < 2e$ , there is  $\delta_0 = \delta_0(\tau) > 0$  with

$$\Psi_{net}^+(\delta, r_S^+(\delta)) > 0, \quad 0 < \delta < \delta_0.$$

See further discussion in Section 6 below.

We start the proof of (3.1) by observing that the maximal operator  $M[\ ]$  becomes ‘transparent’ in the limit  $\delta \rightarrow 0$  if we stay along the trajectory  $(\delta, r_S^+(\delta; \tau))$ . Corollary 4.1 below shows that, if  $\tau > 2e$ , for some  $\delta_1 = \delta_1(\tau) > 0$ ,

$$(3.2) \quad \Psi_{net}^+(\delta, r_S^+(\delta)) = M[\Psi_{net}^+](\delta, r_S^+(\delta)), \quad 0 < \delta < \delta_1.$$

The following limiting behavior of the individual exponents as  $\delta \rightarrow 0$  and/or  $\rho \rightarrow 0$  will be derived in Section 4; see (4.3), (4.8), (4.15):

$$(3.3) \quad \Psi_{com}^+(\delta, \rho) = \delta [\log(1/\delta) + 1 + \mathcal{O}(\delta \vee \rho \log \rho)], \quad \max(\rho, \delta) \rightarrow 0,$$

$$(3.4) \quad \Psi_{int}^+(\delta, \rho) = -\frac{1}{2}\delta [\log \rho + \log(e/2\pi) + \mathcal{O}(\rho \log \rho)], \quad \rho \rightarrow 0,$$

$$(3.5) \quad \Psi_{ext}^+(\delta) = \delta \left[ \log z_\delta^+ - \frac{1}{2} \log \log z_\delta^+ + 1 + \mathcal{O} \left( \frac{\log \log z_\delta^+}{\log z_\delta^+} \right) \right], \quad \delta \rightarrow 0;$$

here  $z_\delta^+ := (\delta 2\sqrt{\pi})^{-1}$ ,  $\mathcal{O}(x)$  denotes a term bounded by  $Const \cdot |x|$  for all sufficiently small  $|x|$ , and  $x \vee y$  is the maximum of  $x$  and  $y$ .

From  $\Psi_{net}^+ = \Psi_{com}^+ - \Psi_{int}^+ - \Psi_{ext}^+$  we have, with  $\rho = r_S^+(\delta)$ ,

$$\begin{aligned} M[\Psi_{net}^+](\delta, \rho) &= \delta \frac{1}{2} \left[ \log \rho + \log \log z_\delta^+ + \log(2e) + \mathcal{O} \left( \delta \vee \rho \log \rho \vee \frac{\log \log z_\delta^+}{\log z_\delta^+} \right) \right], \\ (3.6) \quad &= \delta \frac{1}{2} \left[ \log \left( \frac{2e}{\tau} \right) + \mathcal{O} \left( \frac{\log \log z_\delta^+}{\log z_\delta^+} \right) \right], \quad \delta \rightarrow 0. \end{aligned}$$

The  $\mathcal{O}(\cdot)$  term tends to zero with  $\delta$ . Now  $\tau > 2e$  so  $\log(2e/\tau) < 0$ ; for some  $\delta_2(\tau) > 0$  the bracketed term stays negative on  $0 < \delta < \delta_2(\tau)$ . Formula (3.1) follows with  $\delta_S^+(\tau) = \min(\delta_1(\tau), \delta_2(\tau))$ .  $\square$

**3.2. Theorem 1.2.** With (3.6) in hand, it is now convenient to prove the lower bound in Theorem 1.2.

Fix  $\tau > 2$  and define  $r_W^+(\delta) := r_W^+(\delta; \tau) := [\tau \log(1/\delta)]^{-1}$ . We will show that there is  $\delta_W^+ = \delta_W^+(\tau) > 0$  so that

$$(3.7) \quad M[\Psi_{net}^+ - \Psi_{face}^+](\delta, r_W^+(\delta)) < 0, \quad 0 < \delta < \delta_W^+.$$

Below, Corollary 4.3 shows that the maximal function becomes ‘transparent’ – namely that, fixing  $\tau > 2$ , there is  $\delta_1 = \delta_1(\tau) > 0$  so that

$$(3.8) \quad M[\Psi_{net}^+ - \Psi_{face}^+](\delta, r_W^+(\delta)) = (\Psi_{net}^+ - \Psi_{face}^+)(\delta, r_W^+(\delta)), \quad \text{for } \delta < \delta_1.$$

Recall (3.3)–(3.5) and (2.11). From (4.4) as the asymptotics for (2.11) we obtain the following display, in which  $\rho = r_W^+(\delta)$ :

$$\begin{aligned} M[\Psi_{net}^+ - \Psi_{face}^+](\delta, \rho) &= \delta \frac{1}{2} \left[ \log \rho - 2\rho \log(1/\delta) + \log \log z_\delta^+ + \log(2e) \right. \\ &\quad \left. + \mathcal{O} \left( \delta \vee \rho \log \rho \vee \frac{\log \log z_\delta^+}{\log z_\delta^+} \right) \right], \\ (3.9) \quad &= \delta \frac{1}{2} \left[ \log \left( \frac{2e}{\tau} \right) - \frac{2}{\tau} + \mathcal{O} \left( \frac{\log \log 1/\delta}{\log 1/\delta} \right) \right], \quad \delta \rightarrow 0. \end{aligned}$$

Since  $\log(1+x) < x$  for  $x \in (-1, \infty)$ , by setting  $1+x = 2/\tau$ , we see that  $\tau > 2$  implies  $\log(2e/\tau) - 2/\tau < 0$ . Hence there is  $\delta_2(\tau) > 0$  so that the term in brackets is negative for  $\delta$  sufficiently small. Define now  $\delta_W^+(\tau) = \min(\delta_1(\tau), \delta_2(\tau))$ , establishing (3.7).

Looking further ahead to proving the upper bound half of the theorem, we record the following remark. Fix  $\tau = 2$ . Then as  $\delta \rightarrow 0$ ,

$$(3.10) \quad (\Psi_{net}^+ - \Psi_{face}^+)(\delta, r_W^+(\delta)) \rightarrow 0.$$

The implications will emerge in Section 6.  $\square$

**3.3. Theorem 1.5.** The proof of this lower bound is structurally analogous to the proof of the lower bound in Theorem 1.3.

Fix  $\tau > 2e$ , and define  $r_S^\pm(\delta) := r_S^\pm(\delta; \tau) := |\tau \log(\delta\sqrt{\pi})|^{-1}$ . As in the proof of Theorem 1.3, we will show there is  $\delta_S^\pm = \delta_S^\pm(\tau) > 0$  so that

$$(3.11) \quad M[\Psi_{net}^\pm](\delta, r_S^\pm(\delta)) < 0, \quad 0 < \delta < \delta_S^\pm.$$

This establishes half of Theorem 1.5. Again, the other half can be inferred from the fact that for  $\tau < 2e$  there is  $\delta_0 = \delta_0(\tau) > 0$  with

$$M[\Psi_{net}^\pm](\delta, r_S^\pm(\delta)) > 0, \quad 0 < \delta < \delta_0.$$

Section 6 will give the details.

Corollary 4.2 below shows that the maximal operator  $M[\ ]$  becomes transparent in the limit  $\delta \rightarrow 0$ ; for some  $\delta_1 = \delta_1(\tau) > 0$ ,

$$(3.12) \quad \Psi_{net}^\pm(\delta, r_S^\pm(\delta)) = M[\Psi_{net}^\pm](\delta, r_S^\pm(\delta)), \quad 0 < \delta < \delta_1.$$

The following limiting behavior of the individual exponents as  $\delta \rightarrow 0$  and/or  $\rho \rightarrow 0$  will be derived in Section 4; see (4.3), (4.8), and (4.18):

$$(3.13) \quad \Psi_{com}^\pm(\delta, \rho) = \delta [\log(1/\delta) + 1 + \log_e(2) + \mathcal{O}(\delta \vee \rho \log \rho)], \quad \max(\rho, \delta) \rightarrow 0,$$

$$(3.14) \quad \Psi_{int}^\pm(\delta, \rho) = -\frac{1}{2}\delta [\log \rho + \log(e/2\pi) + \mathcal{O}(\rho \log \rho)], \quad \rho \rightarrow 0,$$

$$(3.15) \quad \Psi_{ext}^\pm(\delta) = \delta \left[ \log z_\delta^\pm - \frac{1}{2} \log \log z_\delta^\pm + 1 + \mathcal{O} \left( \frac{\log \log z_\delta^\pm}{\log z_\delta^\pm} \right) \right], \quad \delta \rightarrow 0,$$

where  $z_\delta^\pm := (\delta \sqrt{\pi})^{-1}$ .

Combining asymptotics using  $\Psi_{net}^\pm = \Psi_{com}^\pm - \Psi_{int}^\pm - \Psi_{ext}^\pm$  yields, with  $\rho = r_S^\pm(\delta)$ ,

$$(3.16) \quad \begin{aligned} M[\Psi_{net}^\pm](\delta, \rho) &= \delta \frac{1}{2} \left[ \log \rho + \log \log z_\delta^\pm + \log(2e) + \mathcal{O} \left( \delta \vee \rho \log \rho \vee \frac{\log \log z_\delta^\pm}{\log z_\delta^\pm} \right) \right], \\ &= \delta \frac{1}{2} \left[ \log \left( \frac{2e}{\tau} \right) + \mathcal{O} \left( \frac{\log \log z_\delta^\pm}{\log z_\delta^\pm} \right) \right], \quad \delta \rightarrow 0. \end{aligned}$$

As  $\log(2e/\tau) < 0$ , there is  $\delta_2(\tau) > 0$  so the term in brackets is negative for  $0 < \delta < \delta_2(\tau)$ . Setting  $\delta_S^\pm(\tau) = \min(\delta_1(\tau), \delta_2(\tau))$ , (3.11) follows.  $\square$

**3.4. Theorem 1.4.** Structurally, the argument for this lower bound resembles that in the proof of Theorem 1.2, in the same way as the proof of the lower bound in Theorem 1.5 resembles that in Theorem 1.3.

Fix  $\tau > 2$  and define  $r_W^\pm(\delta) := r_W^\pm(\delta; \tau) := [\tau \log(1/\delta)]^{-1}$ . Note that  $r_W^+(\delta) = r_W^+(\delta)$ , unlike the strong threshold comparison functions  $r_S^+(\delta)$  and  $r_S^-(\delta)$ , which are not equal. We will show that for  $\delta_W^\pm = \delta_W^\pm(\tau) > 0$ ,

$$(3.17) \quad M[\Psi_{net}^\pm - \Psi_{face}^\pm](\delta, r_W^\pm(\delta)) < 0, \quad 0 < \delta < \delta_W^\pm.$$

Corollary 4.4 shows that the maximal function machinery again simplifies for small  $\delta$ . Thus for  $\tau > 2$  and for  $\delta_1 = \delta_1(\tau) > 0$ ,

$$(3.18) \quad M[\Psi_{net}^\pm - \Psi_{face}^\pm](\delta, r_W^\pm(\delta)) = (\Psi_{net}^\pm - \Psi_{face}^\pm)(\delta, r_W^\pm(\delta)), \quad 0 < \delta < \delta_1.$$

Recall (3.16) and (2.12) with asymptotic behavior following from (4.4). We have the following display, in which  $\rho = r_W^*(\delta)$ :

$$(3.19) \quad \begin{aligned} M[\Psi_{net}^\pm - \Psi_{face}^\pm](\delta, \rho) &= \delta \frac{1}{2} \left[ \log \rho - 2\rho \log(1/\delta) + \log \log z_\delta^+ + \log(2e) \right. \\ &\quad \left. + \mathcal{O} \left( \delta \vee \rho \log \rho \vee \frac{\log \log z_\delta^\pm}{\log z_\delta^\pm} \right) \right], \\ &= \delta \frac{1}{2} \left[ \log \left( \frac{2e}{\tau} \right) - \frac{2}{\tau} + o(1) \right], \quad \delta \rightarrow 0. \end{aligned}$$

As in the proof of the lower bound for Theorem 1.2, for each  $\tau > 2$  there is  $\delta_2(\tau) > 0$  so that the term in brackets is negative for all  $\delta \in (0, \delta_2)$ . Setting  $\delta_W^\pm(\tau) = \min(\delta_1(\tau), \delta_2(\tau))$ , (3.17) follows.  $\square$

#### 4. ANALYSIS OF THE EXPONENTS

We now verify earlier claims about the asymptotic behavior of the exponents.

**4.1. Combinatorial exponents,  $\Psi_{com}^*$ .** The combinatorial exponents for the simplex and cross-polytope were defined in (2.18) and (2.27), respectively; they obey

$$(4.1) \quad \Psi_{com}^+(\nu, \gamma) = \Psi_{com}^\pm(\nu, \gamma) - \nu \log_e(2) := H(\nu) + \nu H(\gamma),$$

where again  $H(p) = p \log(1/p) + (1-p) \log(1/(1-p))$  is the (base- $e$ ) Shannon entropy. Both identities in (4.1) derive from the limit

$$(4.2) \quad n^{-1} \log \binom{n}{\lfloor pn \rfloor} \rightarrow H(p), \quad n \rightarrow \infty, \quad p \in [0, 1],$$

which of course is fundamental in asymptotic analysis and has proven useful in earlier research concerning polytopes [34, 9, 10]. The asymptotic behavior

$$(4.3) \quad \Psi_{com}^+(\nu, \gamma) = \nu [\log(1/\nu) + 1 + \mathcal{O}(\nu \vee \gamma \log \gamma)], \quad \nu \vee \gamma \rightarrow 0,$$

follows directly from that of the Shannon entropy,

$$(4.4) \quad H(p) = p \log(1/p) + p + \mathcal{O}(p^2), \quad p \rightarrow 0.$$

**4.2. Internal exponents,  $\Psi_{int}^*$ .** The internal-angle exponent is the same for both  $\star = +$  and  $\star = \pm$ ; it was defined in (2.19) by

$$(4.5) \quad \Psi_{int}^*(\nu, \gamma) := \nu(1-\gamma) \left[ \log(y_\gamma/\gamma) + \frac{1}{2} \log(2\pi) + \frac{\gamma-1}{2\gamma} y_\gamma^2 \right],$$

where  $y_\gamma$  was defined implicitly by

$$(4.6) \quad \frac{1-\gamma}{\gamma} y_\gamma = s_\gamma, \quad \text{and } s_\gamma \text{ solves } R(s_\gamma) = 1-\gamma;$$

here  $R(s)$  – defined at (2.21) – is closely associated to a famous quantity in probability theory, the Mills' ratio of the standard Normal distribution [24, Sec 5.38]. The asymptotic properties of  $s_\gamma$  as  $\gamma \rightarrow 0$  (and hence also of  $y_\gamma$ ) were studied in [10] using properties of Laplace's asymptotic series for  $R$ . In the Appendix, we refine that approach, obtaining the following error bounds.

**Lemma 4.1.**

$$(4.7) \quad y_\gamma = \frac{\gamma^{1/2}}{1-\gamma} + r_2(\gamma), \quad \text{with } |r_2(\gamma)| \leq 4\gamma^{3/2} \quad \text{for } \gamma \leq 1/30.$$

The behavior (3.4) of the internal exponent as  $\gamma \rightarrow 0$  follows from this lemma directly. Indeed, substitute the behavior of  $y_\gamma$  given by Lemma 4.1, and rearrange

terms:

$$\begin{aligned}
 \Psi_{int}^*(\nu, \gamma) &= \nu(1 - \gamma) \left[ \log(y_\gamma/\gamma) + \frac{1}{2} \log(2\pi) + \frac{\gamma - 1}{2\gamma} y_\gamma^2 \right] \\
 &= \nu(1 - \gamma) \frac{1}{2} \left[ -\log \gamma + \log(2\pi) - \frac{1}{1 - \gamma} \right. \\
 &\quad \left. - 2 \left( \log(1 - \gamma) + \gamma^{-1/2} r_2(\gamma) + \frac{1 - \gamma}{2\gamma} r_2^2(\gamma) + \log \left( 1 + \frac{(1 - \gamma)r_2(\gamma)}{\gamma^{1/2}} \right) \right) \right] \\
 &= \nu(1 - \gamma) \frac{1}{2} [-\log \gamma + \log(2\pi/e) + \mathcal{O}(\gamma)], \quad \gamma \rightarrow 0 \quad (\text{by (4.7)}) \\
 (4.8) \quad &= -\frac{1}{2} \nu [\log \gamma + \log(e/2\pi) + \mathcal{O}(\gamma \log \gamma)], \quad \gamma \rightarrow 0;
 \end{aligned}$$

this is (3.4). □

**4.3. External exponents,  $\Psi_{ext}^*$ .** Each external exponent  $\Psi_{ext}^*$  is defined implicitly through a relation qualitatively resembling

$$(4.9) \quad f(x(z), z) = 0, \quad \text{where } f(x, z) = xe^{x^2} - z,$$

that is to say, we will soon be interested in quantities resembling the solution  $x(z)$ . We briefly sketch an analysis technique for such quantities.

Our approach approximates the asymptotic behavior of  $x(z)$  for  $z$  large by

$$x_2(z) = \sqrt{\log z - \frac{1}{2} \log \log z};$$

the approximation error obeys

$$|x(z) - x_2(z)| \leq \frac{\log \log z}{2 \log z} \quad \text{as } z \rightarrow \infty.$$

The subscript 2 signals that  $x_2(z)$  is the second in a *sequence* of approximations. The sequence starts from a very crude approximation,  $x_1(z)$ , and then improves with each stage. The initial approximation,  $x_1 := \sqrt{\log z}$ , is obtained by treating the factor  $x$  in (4.9) as if it were constant, so that instead of solving (4.9), we simply solve

$$e^{x_1^2} = z.$$

This approximation, substituted into equation (4.9), yields an error

$$(4.10) \quad f(x_1, z) = z((\log z)^{1/2} - 1).$$

The next approximation,  $x_2$ , comes from attempting to cancel the  $(\log z)^{1/2}$  factor in the above error. This is done by solving

$$e^{x_2^2} = z(\log z)^{-1/2},$$

which indeed yields  $x_2(z)$ . This sequence continues on to increasingly accurate approximations, but we stop here because the second term is sufficiently accurate for our purposes.

4.3.1. *Simplex case*  $\Psi_{ext}^+$ . Recall the definition given in (2.22):

$$(4.11) \quad \Psi_{ext}^+(\nu) := \nu x_\nu^2 - (1 - \nu) \log Q(x_\nu),$$

where  $x_\nu$  solves

$$(4.12) \quad \frac{2xQ(x)}{q(x)} + 1 - \nu^{-1} = 0;$$

here  $q(x) := \pi^{-1/2}e^{-x^2}$  and  $Q(x) = \int_{-\infty}^x q(y)dy$  is related to the error function by  $Q(x) = 2(1+\operatorname{erf}(x))$ . Since there is no closed form solution to  $Q(x) = c$  as a function of  $c$ , to analyze the implicitly defined  $x_\nu$ , we develop an asymptotic approximation using the technique just sketched. Define

$$(4.13) \quad z^+ = z^+(\nu) := (\nu 2\sqrt{\pi})^{-1}, \quad \tilde{x}_\nu := \left[ \log z^+ - \frac{1}{2} \log \log z^+ \right]^{1/2}.$$

In the Appendix, we prove the approximation result:

**Lemma 4.2.** *There is  $\nu_0 > 0$  so that*

$$\begin{aligned} x_\nu &= \tilde{x}_\nu + r_3(\nu), \quad \tilde{x}_\nu := \left[ \log z^+ - \frac{1}{2} \log \log z^+ \right]^{1/2}, \\ |r_3(\nu)| &\leq \frac{1}{2} \left( \log z^+ - \frac{1}{2} \log \log z^+ \right)^{-1/2} \frac{\log \log z^+}{\log z^+}, \quad 0 < \nu < \nu_0. \end{aligned}$$

We now plug this approximation into (4.11) and derive the asymptotic behavior. As the cumulative distribution of normal  $Q(\cdot)$  famously has no known closed form expression, we approximate  $Q(x)$  for large  $x$  using the asymptotic series [24, Sec. 5.38],

$$Q(x) = 1 - \frac{e^{-x^2}}{2\sqrt{\pi}x} \sum_{r=0}^{\infty} \frac{(r-1/2)!}{(-x^2)^r}.$$

Keeping the first two terms and applying bounds from [24, eq (5.109)], we have

$$Q(x) = 1 - \frac{1}{2\sqrt{\pi}x} e^{-x^2} + \mathcal{O}(x^{-3}e^{-x^2}), \quad x \rightarrow \infty.$$

Recalling (4.12), we now substitute the approximation to  $x_\nu$  from Lemma 4.2; note that  $x_\nu^2 = \tilde{x}_\nu^2 + r_4(\nu)$  with  $|r_4(\nu)| \leq 2(\log \log z^+)/\log z^+$  and  $z^+$  as in (4.13). Hence,

$$\begin{aligned} Q(x_\nu) &= Q(\tilde{x}_\nu + r_3(\nu)) \\ &= 1 - \nu \left[ 1 - \frac{\log \log z^+}{2 \log z^+} \right]^{-1/2} \cdot [1 + r_3(\nu)/\tilde{x}_\nu]^{-1} e^{-r_4(\nu)} + \mathcal{O}(\nu/\log z^+), \quad \nu \rightarrow 0, \\ &= 1 - \nu + \mathcal{O}\left(\nu \frac{\log \log z^+}{\log z^+}\right), \quad \nu \rightarrow 0, \end{aligned}$$

from which follows

$$(4.14) \quad \log Q(x_\nu) = \nu \left[ 1 + \mathcal{O}\left(\frac{\log \log z^+}{\log z^+}\right) \right], \quad \nu \rightarrow 0.$$

We obtain, finally,

$$(4.15) \quad \Psi_{ext}^+(\nu) = \nu \left[ \log z^+ - \frac{1}{2} \log \log z^+ + 1 + \mathcal{O}\left(\frac{\log \log z^+}{\log z^+}\right) \right], \quad \nu \rightarrow 0.$$

This is (3.5). □

4.3.2. *Cross-polytope case:*  $\Psi_{ext}^\pm$ . The definition given in (2.28) was

$$(4.16) \quad \Psi_{ext}^\pm(\nu) := \nu y_\nu^2 - (1 - \nu) \log G(y_\nu),$$

with  $y_\nu$  the solution of

$$(4.17) \quad \frac{2yG(y)}{g(y)} + 1 - \nu^{-1} = 0,$$

where we recall from before  $g(y) = 2\pi^{-1/2}e^{-y^2}$  on  $y \geq 0$ , and  $G(y) = \operatorname{erf}(y) = \int_0^y g(w)dw$  is the error function. The procedure just used in Section 4.3.1 also works here. We merely state results, omitting proofs.

Let  $z^\pm = z^\pm(\nu) := (\nu\sqrt{\pi})^{-1}$ , and set  $\tilde{y}_\nu = [\log z^\pm - \frac{1}{2} \log \log z^\pm]^{1/2}$ .

**Lemma 4.3.** *There is  $\nu_0 > 0$  so that*

$$y_\nu = \tilde{y}_\nu + r_5(\nu), \quad |r_5(\nu)| \leq \frac{1}{2} \tilde{x}_\nu^{-1} \frac{\log \log z^\pm}{\log z^\pm}, \quad 0 < \nu < \nu_0.$$

This approximation is motivated by the asymptotic series of  $2yG(y)/g(y)$ , giving

$$y_\nu e^{y_\nu^2} - \pi^{-1/2} \nu^{-1} = \mathcal{O}(y_\nu^{-2}).$$

The series is identical to the series motivating  $x_\nu$  in  $\Psi_{ext}^+$  but now  $z^\pm(\nu) = z^\pm := \pi^{-1/2} \nu^{-1}$ . The precise bound on the remainder,  $r_5(\nu)$ , can be recovered by following the same steps as in the proof of Lemma 4.2, replacing  $J(x, \nu)$  in that proof by  $2ye^{-y^2} \int_0^y e^{-w^2} dw + 1 - \nu$ . □

The asymptotic behavior of the external exponent,

$$(4.18) \quad \Psi_{ext}^\pm(\nu) = \nu \left[ \log z^\pm - \frac{1}{2} \log \log z^\pm + 1 + \mathcal{O}\left(\frac{\log \log z^\pm}{\log z^\pm}\right) \right], \quad \nu \rightarrow 0,$$

follows by substituting  $\tilde{y}_\nu$ , as justified by Lemma 4.3.

**4.4. Maximal function for  $\Psi_{net}^*$ .** We now support our earlier claim (3.2) that  $M[\Psi_{net}^*] = \Psi_{net}^*$  along the trajectory  $(\delta, r_S^+(\delta))$  for  $\delta$  small enough.

**Corollary 4.1.** *Fix  $\tau > 2e$ , and recall the definition  $r_S^+(\delta) := r_S^+(\delta; \tau) := |\tau \log(\delta 2\sqrt{\pi})|^{-1}$ . There is  $\delta_1(\tau) > 0$  so that*

$$\Psi_{net}^+(\delta, r_S^+(\delta)) = M[\Psi_{net}^*](\delta, r_S^+(\delta)), \quad 0 < \delta < \delta_1(\tau).$$

This follows from two lemmas, proved in the Appendix, which clarify how  $\Psi_{net}$  changes with  $\nu$  in the regime of interest.

**Lemma 4.4.** *Fix  $\tau > 2e$ . There is  $\delta_1 = \delta_1(\tau) > 0$  so that for  $0 < \delta < \delta_1$  and  $0 < \gamma \leq r_S^+(\delta)$ ,  $\Psi_{net}^+(\nu, \gamma)$  is a decreasing function of  $\nu$  for  $\nu \in [\delta, 1)$ .*

**Lemma 4.5.** *For  $0 < \gamma < \gamma_0$ ,  $\Psi_{net}^+(\nu, \gamma)$  is an increasing function of  $\gamma$ .*

Similar results hold for the cross-polytope [note the slight difference in definition between  $r_S^+(\delta)$  and  $r_S^\pm(\delta)$ ].

**Corollary 4.2.** *Pick  $\tau > 2e$  and again set  $r_S^\pm(\delta) := r_S^\pm(\delta; \tau) := |\tau \log(\delta\sqrt{\pi})|^{-1}$ . For  $\delta < \delta_0(\tau)$ ,  $\Psi_{net}^\pm(\nu, \gamma)$  obtains its maximum value over  $\nu \in [\delta, 1)$  and  $\gamma \leq r_S^\pm(\delta)$  at  $(\nu, \gamma) = (\delta, r_S^\pm(\delta))$ :*

$$\Psi_{net}^\pm(\delta, r_S^\pm(\delta)) = M[\Psi_{net}^\pm](\delta, r_S^\pm(\delta)).$$

We omit the proof, whose arguments parallel those for Lemmas 4.4 and 4.5.

4.5. **Maximal function for  $\Psi_{net}^* - \Psi_{face}^*$ .** We now consider the maximal function associated with the weak exponent, establishing the earlier claim (3.8).

**Corollary 4.3.** *Fix  $\tau > 2$ . There is  $\delta_1 = \delta_1(\tau) > 0$  so that*

$$(\Psi_{net}^+ - \Psi_{face}^+)(\delta, r_W^+(\delta)) = M[\Psi_{net}^+ - \Psi_{face}^+](\delta, r_W^+(\delta)), \quad 0 < \delta < \delta_1.$$

This follows immediately from the next lemmas, which are proven in the Appendix.

**Lemma 4.6.** *Fix  $\tau > 2$ . For  $0 < \delta < \delta_1(\tau)$ ,  $0 < \gamma \leq r_W^+(\delta)$ ,  $(\Psi_{net}^+ - \Psi_{face}^+)(\nu, \gamma)$  is a decreasing function of  $\nu$  over  $\nu \in [\delta, 1]$ .*

**Lemma 4.7.** *Fix  $\tau > 2$ . For  $\delta \in (0, \delta_1(\tau))$ ,  $\rho \in (0, r_W^+(\delta))$  and  $\nu \in [\delta, 1]$ ,  $(\Psi_{net}^+ - \Psi_{face}^+)(\nu, \gamma)$  is an increasing function of  $\gamma$ ,  $0 \leq \gamma \leq r_W^+(\delta)$ .*

Similar results for the cross-polytope are obtained by following the same arguments line-by-line with appropriate substitutions. One obtains the following, though we omit the argument.

**Corollary 4.4.** *Fix  $\tau > 2$ . There is  $\delta_1 = \delta_1(\tau) > 0$  so that*

$$(\Psi_{net}^\pm - \Psi_{face}^\pm)(\delta, r_W^\pm(\delta)) = M[\Psi_{net}^\pm - \Psi_{face}^\pm](\delta, r_W^\pm(\delta)), \quad 0 < \delta < \delta_1.$$

## 5. BEYOND PROPORTIONAL GROWTH

Theorem 1.6 can be reformulated as follows.

**Theorem 5.1.** *Let  $N_n$  grow subexponentially with  $n$ .*

- *Strong exponents.* Fix  $\tau > 2e$  and consider a sequence  $(k_n)$  with  $k_n \leq n \cdot r_S^*(n/N_n; \tau)$ . There is a sequence  $(\epsilon_n)$  with  $N_n \epsilon_n \rightarrow \infty$  and

$$(5.1) \quad N_n^{-1} \log(D_{\ell,n}^*) \leq -\epsilon_n, \quad \ell = n+1, n+3, \dots$$

- *Weak exponents.* Fix  $\tau > 2$  and consider a sequence  $(k_n)$  with  $k_n \leq n \cdot r_W^*(n/N_n; \tau)$ . There is a sequence  $(\epsilon_n)$  with  $N_n \epsilon_n \rightarrow \infty$  and

$$(5.2) \quad N_n^{-1} (\log(D_{\ell,n}^*) - \log f_k(Q)) \leq -\epsilon_n, \quad \ell = n+1, n+3, \dots$$

To venture outside the proportional growth setting requires to strengthen all previous arguments. First, we have to show not just that each maximal function is negative before its first zero, but that it is sufficiently negative in a quantitative sense. Fortunately, the hard work has already been done; summarizing the implications of (3.6), (3.9), (3.16), and (3.19), we have

**Lemma 5.2.** *Let  $\delta_n = n/N_n$ .*

- *Strong exponents.* Fix  $\tau > 2e$ . There are  $\zeta_S^*(\tau) > 0$  so that for  $n > n_0$

$$(5.3) \quad M[\Psi_{net}^*](\delta_n, r_S^*(\delta_n)) < -\zeta_S^*(\tau) \delta_n.$$

- *Weak exponents.* Fix  $\tau > 2$ . There are  $\zeta_W^*(\tau) > 0$  so that for  $n > n_0$

$$M[\Psi_{net}^* - \Psi_{face}^*](\delta_n, r_W^*(\delta_n)) < -\zeta_W^*(\tau) \delta_n.$$

We must also strengthen the previously discussed inequalities (2.14), (2.15), (2.16), (2.24), (2.25), and (2.26), giving precise information about the remainders. We start with the combinatorial exponent.



**Lemma 5.3.**

$$(5.4) \quad N^{-1} \log C_{\ell,n}^* \leq \Psi_{com}^*(\nu_{\ell,n}, \gamma_{\ell,n}) + O(N^{-1} \log(N)),$$

where the  $O(\cdot)$  term is uniform in  $\ell = n + 1, n + 3, \dots$ .

The proof is given in Section 5.1. We next consider the external angles.

**Lemma 5.4.**

$$(5.5) \quad N^{-1} \log \alpha(T^\ell, T^{N-1}) \leq -\Psi_{ext}^+(\nu_{\ell,n}) + O(N^{-1} \log(N)),$$

where the  $O(\cdot)$  is uniform in  $\ell = n + 1, n + 3, \dots$ . Similarly,

$$(5.6) \quad N^{-1} \log \alpha(F^\ell, C^N) \leq -\Psi_{ext}^\pm(\nu_{\ell,n}) + O(N^{-1} \log N),$$

where the  $O(\cdot)$  is uniform in  $\ell = n + 1, n + 3, \dots$ .

For the proof see Section 5.2. We finally consider the internal angles.

**Lemma 5.5.** Fix  $\tau > 2e$ .

$$(5.7) \quad N^{-1} \log \beta(T^k, T^\ell) \leq -\Psi_{int}^*(\nu_{\ell,n}, \gamma_{\ell,n}) + o(1)\Psi_{net}^*(\nu_{\ell,n}, \gamma_{\ell,n}) + O(N^{-1} \log N),$$

where the  $o(\cdot)$  is uniform in  $\ell = n + 1, n + 3, \dots$  and in  $k = 1, \dots, \lfloor n \cdot r_S^*(\delta_n) \rfloor$ . Fix  $\tau > 2$ .

$$(5.8) \quad \begin{aligned} N^{-1} \log \beta(T^k, T^\ell) &\leq -\Psi_{int}^*(\nu_{\ell,n}, \gamma_{\ell,n}) \\ &+ o(1)(\Psi_{net}^*(\nu_{\ell,n}, \gamma_{\ell,n}) - \Psi_{face}^*(\nu_{\ell,n}, \gamma_{\ell,n})) + O(N^{-1} \log N), \end{aligned}$$

where the  $o(\cdot)$  is uniform in  $\ell = n + 1, n + 3, \dots$  and in  $k = 1, \dots, \lfloor n \cdot r_W^*(\delta_n) \rfloor$ .

We also need analogous results for the number of faces of  $T^{N-1}$  and  $C^N$ .

**Lemma 5.6.**

$$-N^{-1} \log f_k(T^{N-1}) \leq \Psi_{face}^+(\nu, \gamma) + O(N^{-1} \log(N)),$$

where the  $O(\cdot)$  is uniform in  $k = 1, 2, \dots, n$ . Similarly,

$$-N^{-1} \log f_k(C^N) \leq \Psi_{face}^\pm(\nu, \gamma) + O(N^{-1} \log(N)),$$

where the  $O(\cdot)$  is uniform in  $k = 1, 2, \dots, n$ .

For the proof see Section 5.4.

These lemmas easily combine to finish the argument for Theorem 5.1. Under the subexponential growth assumption  $\log(N_n) = o(n)$ , the remainder terms

$$O(N_n^{-1} \log(N_n)) = o(n/N_n) = o(\delta_n).$$

Hence the remainders are much smaller than the bounds on  $M[\ ]$  terms associated with (5.2), (5.3). Consider the case of the strong exponent for the cross-polytope. Uniformly in  $\ell = n + 1, n + 3, \dots$ ,

$$\begin{aligned} N^{-1} \log(D_{\ell,n}^\pm) &\leq M[\Psi_{net}^\pm](\delta_n, r_S^\pm(\delta_n)) \cdot (1 + o(1)) + O(N_n^{-1} \log(N_n)) \\ &\leq -\zeta_S^\pm(\tau) \cdot \delta_n \cdot (1 + o(1)) + O(N_n^{-1} \log(N_n)) \\ &< -(\zeta_S^\pm(\tau)/2) \cdot \delta_n, \quad n > n_0. \end{aligned}$$

Hence (5.1) follows, with  $\star = \pm$  and  $\epsilon_n = (\zeta_S^\pm(\tau)/2)\delta_n$ . The rest of Theorem 5.1 follows similarly.  $\square$

It remains to prove Lemmas 5.3–5.6. This we do in the coming subsections.

**5.1. Combinatorial exponents.** Stirling's formula provides error bounds for the combinatorial exponents.

**Lemma 5.7** (Stirling's inequality, [7]).

$$(2\pi n)^{1/2} \left(\frac{n}{e}\right)^n \leq n! \leq \frac{5}{4}(2\pi n)^{1/2} \left(\frac{n}{e}\right)^n \quad \text{for } n \geq 1.$$

To verify (5.4), recall the combinatorial factors

$$C_{\ell,n}^+ = 2 \binom{N}{\ell} \binom{\ell}{k+1} \quad \text{and} \quad C_{\ell,n}^\pm = 2^{\ell+1} \binom{N}{\ell+1} \binom{\ell+1}{k+1}$$

and that  $\nu_{\ell,n} = \ell/N$  and  $\gamma_{\ell,n} = k/\ell$ .

Using Lemma 5.7, we arrive at

$$(5.9) \quad C_{\ell,n}^* \leq \frac{5}{8\pi} \cdot N e^{N\Psi_{com}^+(\nu,\gamma)},$$

establishing (5.4).

## 5.2. External angle.

5.2.1. *Simplex case*  $\Psi_{ext}^+$ . It is enough to show that uniformly over  $\ell = n+1, n+3, \dots$ ,

$$(5.10) \quad N^{-1} \log \alpha(T^\ell, T^{N-1}) \leq -\Psi_{ext}^+(\nu_{\ell,n}) + N^{-1} \log(N);$$

of course, the remainder term is  $O(\log(N)/N)$ . The simplex part of Lemma 5.5 follows.

The external angle for the simplex is given by

$$(5.11) \quad \alpha(T^\ell, T^{N-1}) = \sqrt{\frac{\ell+1}{\pi}} \int_0^\infty e^{-(\ell+1)x^2} \left( \frac{1}{\sqrt{\pi}} \int_{-\infty}^x e^{-y^2} dy \right)^{N-\ell-1} dx.$$

As before,  $Q(x) := \pi^{-1/2} \int_{-\infty}^x e^{-y^2} dy$ . Recall that  $\nu_{\ell,n} = \ell/N$  and rewrite the simplex external angle as

$$(5.12) \quad \alpha(T^\ell, T^{N-1}) = \sqrt{\frac{\ell+1}{\pi}} \int_0^\infty \exp(-N[\nu_{\ell,n}x^2 + (\nu_{\ell,n} - 1) \log Q(x)]) \frac{e^{-x^2}}{Q(x)} dx.$$

The factor  $N$  in the integral might suggest the use of Laplace's method as in [10]. A simpler, direct approach is possible. The following is obvious but very useful.

**Lemma 5.8.** *Let  $\psi : [0, \infty) \mapsto \mathbf{R}$  achieve its global minimum at  $x^*$  and let  $\varphi : [0, \infty) \mapsto [0, \infty)$  be integrable. Then*

$$(5.13) \quad \int_0^\infty \exp(-N\psi(x))\varphi(x)dx \leq \exp(-N\psi(x^*)) \int_0^\infty \varphi(x)dx.$$

Recall that  $x_\nu$  is the minimizer of  $[\nu x^2 + (\nu - 1) \log Q(x)]$  and

$$\Psi_{ext}^+(\nu) := \nu x_\nu^2 + (\nu - 1) \log Q(x_\nu).$$

Apply Lemma 5.8 to the integral (5.12); set  $\psi(x) = [\nu x^2 + (\nu - 1) \log Q(x)]$  and  $\varphi(x) = e^{-x^2}/Q(x)$ . Because  $\psi(x^*) = \Psi_{ext}^+(\nu)$  and  $\int \varphi = (3\pi)/8$ , (5.13) yields

$$(5.14) \quad \alpha(T^\ell, T^{N-1}) \leq e^{-N\Psi_{ext}^+(\nu)} \cdot \sqrt{\frac{\ell+1}{\pi}} \cdot \frac{3\pi}{8} \leq \sqrt{N+1} \cdot e^{-N\Psi_{ext}^+(\nu)}.$$

Inequality (5.10) follows.  $\square$

5.2.2. *Cross-polytope case  $\Psi_{ext}^\pm$ .* Our goal is to prove (5.6). We introduce a perturbed version of  $\nu_{\ell,n}$ :

$$\widehat{\nu}_{\ell,n} = \nu_{\ell,n} + \frac{1}{2N}.$$

Note that  $\widehat{\nu}_{\ell,n} \in [0, 1)$  as is the unperturbed  $\nu_{\ell,n}$ . This is used in our first step, where we find that it appears naturally in the bound

$$(5.15) \quad N^{-1} \log \alpha(F^\ell, C^N) \leq -\Psi_{ext}^\pm(\widehat{\nu}_{\ell,n}) + N^{-1} \log(N), \quad N > 3.$$

Note that the remainder is  $O(N^{-1} \log(N))$  uniformly over  $\ell \in [n+1, N-1]$ , as our goal requires. Indeed (5.6) is an inequality like (5.15) but with  $\nu_{\ell,n}$  rather than  $\widehat{\nu}_{\ell,n}$ . In our second step, we verify that the perturbation of the argument is unimportant:

$$(5.16) \quad \Psi_{ext}^\pm(\widehat{\nu}_{\ell,n}) = \Psi_{ext}^\pm(\nu_{\ell,n}) + O(N^{-1} \log(N)),$$

uniformly over  $\ell = n+1, n+3, \dots$  where  $\ell < N$ . The cross-polytope half of Lemma 5.5 then follows. It remains to show (5.15)–(5.16).

The external angle for the cross-polytope is given by

$$(5.17) \quad \alpha(F^\ell, C^N) = \sqrt{\frac{\ell+1}{\pi}} \int_0^\infty e^{-(\ell+1)y^2} \left( \frac{2}{\sqrt{\pi}} \int_0^y e^{-w^2} dw \right)^{N-\ell-1} dy.$$

Following the same approach as for the simplex, recall  $G(y) := 2\pi^{-1/2} \int_0^y e^{-w^2} dw$  and rewrite the cross-polytope external angle as

$$(5.18) \quad \alpha(F^\ell, C^N) = \sqrt{\frac{\ell+1}{\pi}} \int_0^\infty \exp(-N[\widehat{\nu}_{\ell,n}y^2 + (\widehat{\nu}_{\ell,n} - 1) \log G(y)]) \left( \frac{e^{-y^2}}{G(y)} \right)^{1/2} dy.$$

Let  $\psi(y; \nu) = \nu y^2 + (\nu - 1) \log G(y)$ , and let  $\widehat{y}_\nu$  be the minimizer of  $\psi(\cdot; \nu)$ . Set

$$\Psi_{ext}^\pm(\widehat{\nu}) := \nu \widehat{y}_\nu^2 + (\nu - 1) \log G(\widehat{y}_\nu).$$

Apply Lemma 5.8 to the integral (5.18); set  $\psi = \psi(\cdot; \widehat{\nu}_{\ell,n})$  and  $\varphi(y) = \left( \frac{e^{-y^2}}{G(y)} \right)^{1/2}$ . The factor  $\exp(-N\psi(x^*)) = \exp(-N\Psi_{ext}^\pm(\widehat{\nu}_{\ell,n}))$ , while  $\int \varphi \leq 2.175$ ; we obtain

$$(5.19) \quad \alpha(F^\ell, C^N) \leq \frac{5}{4} \sqrt{\ell+1} \exp(-N\Psi_{ext}^\pm(\widehat{\nu}_{\ell,n})).$$

Hence

$$N^{-1} \log \alpha(F^\ell, C^N) \leq -\Psi_{ext}^\pm(\widehat{\nu}_{\ell,n}) + N^{-1} \log \left( \frac{5}{4} \sqrt{\ell+1} \right),$$

from which (5.15) follows.

We earlier studied the asymptotic behavior of  $\Psi_{ext}^\pm(\nu)$ ; see (4.18). The effect of the perturbation  $1/2N$  in  $\widehat{\nu}_{\ell,n}$  can be bounded simply. Put  $z_n^\pm = z^\pm(\widehat{\nu}_{\ell,n})$ . Then

$$\begin{aligned} \Psi_{ext}^\pm(\widehat{\nu}_{\ell,n}) &= (\nu_{\ell,n} + 1/2N) \left[ \log z_n^\pm - \frac{1}{2} \log \log z_n^\pm + 1 + \mathcal{O} \left( \frac{\log \log z_n^\pm}{\log z_n^\pm} \right) \right], \quad n \rightarrow \infty, \\ &= \nu_{\ell,n} \left[ \log z_n^\pm - \frac{1}{2} \log \log z_n^\pm + 1 \right] + \mathcal{O} \left( \log z_n^\pm \sqrt{\frac{\log \log z_n^\pm}{\log z_n^\pm}} \right) / N, \quad n \rightarrow \infty. \end{aligned}$$

Our goal (5.16) follows. Combined with (5.15), we obtain (5.6).  $\square$

**5.3. Internal angle.** We aim to demonstrate (5.7). We again introduce perturbed variables:

$$\tilde{\nu} = \tilde{\nu}_{\ell,n} = \frac{\ell+2}{N}, \quad \tilde{\gamma} = \tilde{\gamma}_{\ell,n} := \frac{k+1}{\ell+2}.$$

Our plan is to first show that for  $n > n_0$

$$(5.20) \quad N^{-1} \log \beta(T^k, T^\ell) \leq -\Psi_{int}(\tilde{\nu}, \tilde{\gamma}) + N^{-1} \log \left[ \frac{2}{\pi} (N+3)^{5/2} \right].$$

The remainder here is  $O(N^{-1} \log(N))$ . We then show that the perturbation of variables has a negligible impact:

$$(5.21) \quad \Psi_{int}(\tilde{\nu}, \tilde{\gamma}) - \Psi_{int}(\nu, \gamma) = o(\Psi_{net}^*)$$

uniformly in  $0 \leq k \leq n \cdot r_S^*(\delta_n)$ . Our goal (5.7) follows. It remains to prove (5.20), (5.21).

An expression for the internal angle was developed in [10]:

$$(5.22) \quad \beta(T^k, T^\ell) = \left( \pi \frac{\ell+2}{k+1} \right)^{1/2} 2^{k-\ell} g_{T+W_m}(0);$$

here  $g_{T+W_m}(0)$  denotes the probability density of a certain random variable expressible as a sum of  $m+1$  independent random variables; here  $m = \ell - k + 1$ . The paper [10] used large deviations analysis to bound this term using a certain nonnegative convex rate function  $\Lambda^* : [0, \infty) \mapsto [0, \infty)$ ; the bound was

$$g_{T+W_m}(0) \leq \frac{2}{\sqrt{\pi}} \frac{m^2}{2\theta} \int_0^{\sqrt{2/\pi}} y \exp\left(-m \left[ \left( \frac{m}{2\theta} \right) y^2 + \Lambda^*(y) \right]\right) dy + \frac{2}{\sqrt{\pi}} \exp\left(-\frac{m^2}{\pi\theta}\right) \\ =: I_m + II_m,$$

say, with  $m = \ell - k + 1$  and  $\theta = k + 1$ . The second term was argued to be negligible in the proportional growth setting by soft analysis; later below we will check that it is still negligible in the current nonproportional growth setting.

Focusing on the supposedly dominant term  $I_m$ , substitute in the values for  $m$  and  $\theta$  and recall that  $\tilde{\gamma} = (k+1)/(\ell+2)$ :

$$I_m = \frac{1}{\sqrt{\pi}} \frac{(\ell-k+1)^2}{(k+1)} \int_0^{\sqrt{2/\pi}} y \exp\left(-(\ell-k+1) \left[ \left( \frac{1-\tilde{\gamma}}{\tilde{\gamma}} \right) \frac{y^2}{2} + \Lambda^*(y) \right]\right) dy.$$

The integral here can be rewritten as

$$J_m := \int_0^{\sqrt{2/\pi}} y \exp(-N\tilde{\nu}(1-\tilde{\gamma})\xi_{\tilde{\gamma}}(y)) dy,$$

where, consistent with earlier definitions,

$$\xi_{\tilde{\gamma}}(y) = \left[ \left( \frac{1-\tilde{\gamma}}{\tilde{\gamma}} \right) \frac{y^2}{2} + \Lambda^*(y) \right].$$

Note that  $y_{\tilde{\gamma}}$  is the minimum of  $\xi_{\tilde{\gamma}}(y)$ . Again apply Lemma 5.8 to bound  $J_m$ ; setting  $\psi = \tilde{\nu}(1-\tilde{\gamma})\xi_{\tilde{\gamma}}(y)$  and  $\varphi = y \mathbf{1}_{[0, \sqrt{2/\pi}]}$ , (5.13) gives

$$J_m \leq \exp(-N\tilde{\nu}(1-\tilde{\gamma})\xi_{\tilde{\gamma}}(y_{\tilde{\gamma}}))/\pi.$$

Note that

$$\Psi_{int}(\tilde{\nu}, \tilde{\gamma}) := \tilde{\nu}(1-\tilde{\gamma})[\xi_{\tilde{\gamma}}(y_{\tilde{\gamma}}) + \log 2],$$

and so

$$2^{k-\ell-1} \exp(-N\tilde{\nu}(1-\tilde{\gamma})\xi_{\tilde{\gamma}}(y_{\tilde{\gamma}})) = \exp(-N\Psi_{int}(\tilde{\nu}, \tilde{\gamma})).$$

Noting the presence of a factor  $2^{k-\ell}$  in (5.22) and noting that  $\ell + 1 - k \leq N$ , we obtain

$$\begin{aligned} \beta(T^k, T^\ell) &\leq 2 \left(\frac{\ell+2}{k+1}\right)^{1/2} \frac{(\ell-k+1)^2}{(k+1)} \exp(-N\Psi_{int})/\pi + \sqrt{\pi(N+2)} \cdot 2^{k-\ell} II_m \\ (5.23) \quad &\leq 2(N+2)^{5/2} \cdot \exp(-N\Psi_{int}(\tilde{\nu}, \tilde{\gamma})) + \sqrt{\pi(N+2)} \cdot 2^{k-\ell} II_m. \end{aligned}$$

This essentially verifies (5.20).

However, it remains to verify that  $II_m \ll I_m$ . Put  $\mu = \sqrt{2/\pi}$  and recall from [10] that  $\mu = E(T + W_m)$ . We focus on  $y = \mu = \sqrt{2/\pi}$  and use the fact that the large deviations rate function always vanishes at the underlying mean, i.e.  $\Lambda^*(\mu) = 0$  essentially by definition. Then

$$-m \left[ \left(\frac{m}{2\theta}\right) \mu^2 + \Lambda^*(\mu) \right] = \frac{-m^2}{\pi\theta}.$$

It follows that

$$II_m = \exp(-N \cdot \tilde{\nu}(1-\tilde{\gamma})\xi_{\tilde{\gamma}}(\mu)) \cdot \sqrt{\frac{2}{\pi}}.$$

But by definition of  $y_{\tilde{\gamma}}$  as the minimizer of  $\xi_{\tilde{\gamma}}$  and the asymptotic  $y_{\tilde{\gamma}} \rightarrow 0$ ,

$$\xi_{\tilde{\gamma}}(\mu) > \xi_{\tilde{\gamma}}(y_{\tilde{\gamma}});$$

in fact  $\xi_{\tilde{\gamma}}(\mu) \sim \tilde{\gamma}^{-1}\mu^2 \gg \log(\tilde{\gamma}^{-1}) \sim \xi_{\tilde{\gamma}}(y_{\tilde{\gamma}})$  as  $\tilde{\gamma} \leq r_W^*(\delta_n)(1+o(1)) \rightarrow 0$ . Hence  $II_m$  is exponentially smaller than  $J_m$ , and (5.20) is fully proven.

As for (5.21), recall that

$$(5.24) \quad \Psi_{int}(\tilde{\nu}, \tilde{\gamma}) = -\frac{1}{2}\tilde{\nu} \left[ \log \gamma + \log(e(1+k^{-1})/2\pi) + o(1) \right],$$

while, if  $\rho = r_S^*(\delta)$ ,

$$\Psi_{net}^*(\nu, \eta\rho) = \nu \frac{1}{2} \left[ \log \left( \frac{2e}{\tau} \right) + \log(\eta) + o(1) \right].$$

Look now in the vicinity of  $k = \gamma n$ , where  $\gamma = \eta \cdot r_S^*(\delta_n)$ .

$$\frac{|\Psi_{int}^*(\tilde{\nu}, \tilde{\gamma}) - \Psi_{int}^*(\nu, \gamma)|}{|\Psi_{net}^*(\nu, \gamma)|} \leq \frac{\min(1, 1/\eta \cdot \frac{1}{n \cdot r_S^*(\delta_n)}) + o(1)}{|\frac{2e}{\tau} + \log(\eta) + o(1)|} = o(1).$$

Here all the  $o(1)$ 's are uniform in  $0 \leq \eta \leq 1$ .

The argument for (5.8) is similar to that of (5.7) detailed above, replacing  $\Psi_{net}^*$  by  $\Psi_{net}^* - \Psi_{face}^*$ .  $\square$

**5.4. Face counts of  $T^{N-1}$  and  $C^N$ .** The number of  $k$ -faces for the simplex and cross-polytope are

$$f_k(T^{N-1}) = \binom{N}{k+1} \quad \text{and} \quad f_k(C^N) = 2^{k+1} \binom{N}{k+1}.$$

Invoking Lemma 5.7 and recalling that  $\nu_{\ell,n} = \ell/N$  and  $\gamma_{\ell,n} = k/\ell$ , we arrive at

$$(5.25) \quad f_k(T^{N-1}) \geq \frac{8}{25} \sqrt{\frac{2}{\pi}} N^{-1} e^{N\Psi_{face}^+(\nu, \gamma)}$$

and

$$(5.26) \quad f_k(C^N) \geq \frac{16}{25} \sqrt{\frac{2}{\pi}} N^{-1} e^{N\Psi_{\text{face}}^\pm(\nu, \gamma)},$$

establishing Lemma 5.6.

## 6. UPPER BOUNDS ON PHASE TRANSITIONS

Until this point, we have focused on establishing lower bounds on the several  $\rho$ -functions introduced in Section 1. Our work so far has given the lower-bound “half” of Theorems 1.2–1.5; we now give upper bounds on the  $\rho$ -functions and complete the proof of Theorems 1.2–1.5.

We remark, parenthetically, that the “half” already proven is the more surprising/interesting part of the result, in view of applications. However, the remaining part settles any question about whether the lower bounds have slack, i.e. whether they actually agree with the precise phase transitions.

For establishing *lower* bounds on the  $\rho$ 's, we have been applying *upper* bounds on the combinatorial factor and on the internal and external angles. Now that we want *upper* bounds on the  $\rho$ 's, we will turn to *lower* bounds on the combinatorial factor and the angles.

The required lower bounds will be developed in later subsections of this section; effectively we will be using standard ideas such as Stirling's inequality, Laplace's method and the Saddlepoint method.

Before turning to those lower bounds, we give the arguments completing the proofs of Theorems 1.2–1.5.

**6.1. Upper bounds on strong phase transition.** The key to tying down the strong phase transitions  $\rho_S^*$  is to use the fact that  $\Psi_{\text{net}}^*(\delta, \cdot)$  makes a sign change at  $\rho_S^*$ . Indeed, by definition,  $\Psi_{\text{net}}^*(\delta, \rho)$  has a zero at  $\rho = \rho^*(\delta)$ ; but actually it is strictly increasing in the vicinity of this zero. For sufficiently small  $\epsilon > 0$ , we can find  $\rho = \rho_\epsilon > \rho_S^*(\delta)$  so that

$$\Psi_{\text{net}}^*(\delta, \rho) > 2\epsilon.$$

Set now  $k_n = \lfloor \rho_\epsilon n \rfloor$ ; for all sufficiently large  $n$ ,

$$(6.1) \quad \Psi_{\text{net}}^* \left( \frac{n+2}{N}, \frac{k+1}{n+2} \right) > \epsilon.$$

We now invoke lemmas placing lower bounds on the combinatorial, internal and external angle factors.

**Lemma 6.1.** *There is an absolute constant  $c_1 > 0$  so that*

$$C_{\ell, n}^* \geq c_1 \cdot \frac{\ell^{1/2}}{N^{3/2}} \cdot \exp \left( N \Psi_{\text{com}}^* \left( \frac{\ell+1}{N}, \frac{k+1}{\ell+1} \right) \right).$$

The next lemma is more than we really need at this stage; the extra generality will be useful in discussion of the weak phase transition in the next subsection.

**Lemma 6.2.** *In the proportional growth setting, we have constants  $c_2, c_3$ , and  $c_4$  depending at most on  $\delta$ , so that, for  $\ell = n+1, n+3, \dots, \ell \leq n + \sqrt{N}$ ,  $k = k_n =$*

$\lfloor \rho_\epsilon n \rfloor$ , and  $n > n_0$ ,

$$(6.2) \quad \beta(T^k, T^\ell) \geq c_2 \cdot \exp\left(-N\Psi_{int}^*\left(\frac{\ell+1}{N}, \frac{k+1}{\ell+1}\right)\right),$$

$$(6.3) \quad \alpha(T^\ell, T^{N-1}) \geq c_3 \cdot \exp\left(-N\Psi_{ext}^+\left(\frac{\ell+1}{N}, \frac{k+1}{\ell+1}\right)\right),$$

$$(6.4) \quad \alpha(F^\ell, C^N) \geq c_4 \cdot \exp\left(-N\Psi_{ext}^\pm\left(\frac{\ell+1}{N}, \frac{k+1}{\ell+1}\right)\right).$$

Combining the last two lemmas, we get, specializing to the case  $\ell = n + 1$ ,

$$\begin{aligned} D_{n+1,n} &\geq c_5 \ell^{1/2} N^{-3/2} \exp\left(N\Psi_{net}^*\left(\frac{n+2}{N}, \frac{k+1}{n+2}\right)\right) \\ &\geq c_5 \ell^{1/2} N^{-3/2} \exp(N\epsilon) \rightarrow \infty, \quad N_n \rightarrow \infty. \end{aligned}$$

As  $f_k(Q) - Ef_k(AQ) > D_{n+1,n}^*$ , we conclude that  $\Delta(k_n, n, N_n) \rightarrow \infty$  as  $n \rightarrow \infty$ ; this completes the upper bound for the strong phase transition  $\rho_S^*$ .  $\square$

**6.2. Upper bounds on the weak phase transition.** We now aim to show that, in the proportional growth setting with  $n/N_n \rightarrow \delta > 0$  and  $k = \lfloor \rho_W^*(\delta)n \rfloor$ ,

$$(6.5) \quad (f_k(Q) - Ef_k(AQ))/f_k(Q) > \epsilon > 0, \quad n > n_0.$$

In words, ‘above  $\rho_W^*(\delta)$  a nonvanishing fraction of faces gets lost under projection’.

In fact we will show that for all large enough  $n$  and for all  $\ell$  in the range  $n + 1, n + 3, \dots, \ell \leq n + \sqrt{N}$ ,

$$(6.6) \quad D_{\ell,n}^*/f_k(Q) \geq cn^{-1/2}, \quad \ell = n + 1, n + 3, \dots, \ell \leq n + \sqrt{N}.$$

Since this inequality holds for at least  $\frac{1}{2}\sqrt{N}$  terms from the sum  $\Sigma'$ , we have

$$\Delta(k_n, n, N_n) = \Sigma' D_{\ell,n}^* \geq \epsilon f_k(Q),$$

for  $\epsilon = c/2$ , which implies (6.5).

The different structure of our argument in the weak transition case can be traced to the fact that  $\Psi_{net}^* - \Psi_{face}^*$  does not change sign at  $\rho = \rho_W^*$ . Instead, it achieves its global maximum 0. This means that

$$\frac{\partial}{\partial \nu} (\Psi_{net}^* - \Psi_{face}^*)(\nu, \delta\rho/\nu) = 0$$

from which it follows that, for  $\nu \in [\delta, \delta + 1/\sqrt{N}]$  and some  $c > 0$ ,

$$(\Psi_{net}^* - \Psi_{face}^*)(\nu, \delta\rho/\nu) \geq c/N.$$

The combinatorial identity

$$\binom{n}{k+1} \binom{n-k-1}{\ell-k} = \binom{n}{\ell} \binom{\ell}{k+1}$$

implies

$$D_{\ell,n}^+/f_k(T^{N-1}) = 2 \cdot \binom{N-k-1}{\ell-k} \cdot \beta(T^k, T^\ell) \alpha(T^\ell, T^{N-1})$$

and

$$D_{\ell,n}^\pm/f_k(C^N) = 2 \cdot \binom{N-k-1}{\ell-k} \cdot \beta(T^k, T^\ell) \alpha(T^\ell, C^N).$$

We need the following combinatorial result; it follows from Stirling’s inequalities (Lemma 5.7) and we omit the proof.

**Lemma 6.3.**

$$\binom{N-k-1}{\ell-k} \geq \frac{1}{3}(N-k-1)^{-1/2} \exp\left((N-k-1)H\left(\frac{\ell-k}{N-k-1}\right)\right).$$

We combine this with Lemma 6.2 and get that, under proportional growth,

$$D_{\ell,n}^*/f_k(Q) \geq cN^{-1/2} \cdot \exp\left(N(\Psi_{net}^* - \Psi_{face}^*)\left(\frac{\ell+1}{N}, \frac{k+1}{\ell+1}\right)\right).$$

This implies (6.6) and (6.5) follows.

**6.3. Analysis of the external angle. Simplex case.** We recall the exact formula

$$(6.7) \quad \alpha(T^\ell, T^{N-1}) = \sqrt{\frac{\ell+1}{\pi}} \int_0^\infty \exp(-N\psi_\nu(x)) dx,$$

where  $\psi_\nu(x) = \nu x^2 - (1-\nu) \log Q(x)$  and  $\nu = \frac{\ell+1}{N}$ . Note that throughout Section 6.3, we use the convention  $\nu = \frac{\ell+1}{N}$  consistent with Lemma 6.2. In the next subsection, we use Laplace's method to obtain lower bounds on general integrals of this type. That lemma requires estimates which are in turn supplied by Lemma 6.4 below. Against the textbook situation in the study of such integrals, here the exponent  $\psi_\nu$  should not be viewed as constant in  $N$ ; it depends on the variable  $\nu = \frac{\ell+1}{N}$  where  $\ell$  is a variable with effective range "in the vicinity of  $n$ ". Lemma 6.4 gives lower bounds on Laplace integrals with uniform multiplicative remainders; this yields that for a fixed subinterval  $0 < \nu_0 < \nu_1 < 1$  and for each  $\epsilon > 0$  there is  $N_0(\epsilon)$  so that for all  $\nu = (\ell+1)/N$  in  $(\nu_0, \nu_1)$ ,

$$\sqrt{\frac{\ell+1}{\pi}} \int_0^\infty \exp(-N\psi_\nu(x)) dx \geq \sqrt{\frac{2\pi}{N\psi_\nu''(x_\nu)}} \exp(-N\Psi_\nu(x_\nu)) \cdot (1-\epsilon).$$

We conclude from (6.7) and (6.8) that for each  $\epsilon > 0$  and all sufficiently large  $N > N_0(\nu, \epsilon)$ ,

$$\alpha(T^\ell, T^{N-1}) \geq \sqrt{(1-\nu)(1+2x_\nu^2)} \cdot \exp(-N\Psi_{ext}^+(x_\nu))(1-\epsilon),$$

where the threshold  $N_0$  may be taken locally uniform in  $\nu \in [0, 1)$ . Here, again,  $x_\nu$  is the minimizer of  $\psi_\nu(x)$ . It follows that there is a constant  $c > 0$  so that for all sufficiently large  $N$  and all  $\nu \in I_N$ ,  $I_N = [\delta, \delta + 1/\sqrt{N}]$ ,

$$\alpha(T^\ell, T^{N-1}) \geq c \cdot \exp(-N\Psi_{ext}^+(x_\nu)).$$

Equation (6.4) follows.

**Lemma 6.4.** *Let  $\psi_\nu(x) = \nu x^2 - (1-\nu) \log Q(x)$ . Then  $\psi_\nu(x)$  is  $C^4(0, \infty)$ ,*

$$(6.8) \quad \begin{aligned} \psi_\nu''(x_\nu) &= \frac{2\nu}{1-\nu}(1+2x_\nu^2), \\ \psi_\nu'''(x_\nu) &= -(1-\nu) \left[ \frac{4\nu}{1-\nu}(2x_\nu^3 - x_\nu) + \frac{24\nu}{1-\nu}x_\nu^2 - \frac{16\nu}{1-\nu}x_\nu^3 \right]. \end{aligned}$$

For  $\epsilon > 0$ , set

$$C(\nu, \epsilon) = \sup_{|x-x_\nu| < \epsilon} \frac{|\psi_\nu'''(x)|}{\psi_\nu''(x_\nu)}.$$

Then for small  $\epsilon > 0$ ,  $C(\nu, \epsilon) < \infty$ , and as  $\nu \rightarrow 0$ ,  $C(\nu, \epsilon) \sim 2x_\nu$ .



*Cross-polytope case.* We recall the exact formula

$$\alpha(F^\ell, C^N) = \sqrt{\frac{\ell+1}{\pi}} \int_0^\infty \exp(-N\psi_\nu(x)) dx,$$

where  $\psi_\nu(x) = \nu x^2 - (1-\nu) \log G(x)$  and  $\nu = (\ell+1)/N$ . We apply Lemma 6.4 bounding Laplace integrals with multiplicative remainder to conclude

$$\sqrt{\frac{\ell+1}{\pi}} \int_0^\infty \exp(-N\psi_\nu(x)) dx \geq \sqrt{\frac{2\pi}{N\psi''_\nu(x_\nu)}} \exp(-N\psi_\nu(x_\nu)) \cdot (1+o(1)).$$

Here the  $o(1)$ -term is locally uniform over  $\nu \in [0, 1)$ . We conclude that for each subinterval  $(\nu_0, \nu_1)$  with  $0 < \nu_0 < \nu_1 < 1$  and for  $\epsilon > 0$  we have for  $N > N_0(\nu, \epsilon)$ ,

$$\alpha(F^\ell, C^N) \geq (1 + \frac{4\nu}{1-\nu} x_\nu^2)^{-1/2} \cdot \exp(-N\Psi_{ext}^\pm(\nu))(1-\epsilon).$$

Here, again,  $x_\nu$  is the minimizer of  $\psi_\nu(x)$ . It follows that there is a constant  $c > 0$  so that for all sufficiently large  $N$  and all  $\nu \in I_N$ ,  $I_N = [\delta, \delta + 1/\sqrt{N}]$ ,

$$\alpha(F^\ell, C^N) \geq c \cdot \exp(-N\Psi_{ext}^\pm(\nu)).$$

Inequality (6.3) follows.

**Lemma 6.5.** *Let  $\psi_\nu(x) = \nu x^2 - (1-\nu) \log G(x)$ . Then  $\psi_\nu(x)$  is  $C^4(0, \infty)$ ,*

$$\begin{aligned} \psi''_\nu(x_\nu) &= 2\nu \cdot (1 + x_\nu^2 \frac{4\nu}{1-\nu}), \\ \psi'''_\nu(x_\nu) &= (1-\nu) \left[ \frac{4\nu}{1-\nu} (2 - 4x_\nu^2) + 6x_\nu z^2 + 2x_\nu z^3 \right], \end{aligned}$$

where  $z_\nu = \frac{2\nu x_\nu}{1-\nu}$ . For  $\epsilon > 0$ , set

$$C(\nu, \epsilon) = \sup_{|x-x_\nu|<\epsilon} \frac{|\psi'''_\nu(x)|}{\psi''_\nu(x_\nu)}.$$

Then for small  $\epsilon > 0$ ,  $C(\nu, \epsilon) < \infty$ , and as  $\nu \rightarrow 0$ ,  $C(\nu, \epsilon) \sim 4x_\nu^3$ .

**6.4. Uniform Laplace method.** We use a uniform variant of Laplace's method, suitable for bounding a collection of integrals uniformly. The approach is similar to [10].

**Lemma 6.6.** *Let  $I = [-\epsilon, \epsilon]$  and suppose that  $f$  attains its minimum on  $I$  at 0. Let*

$$C = \sup_I \frac{|f'''(x)|}{f''(0)}.$$

Then

$$\begin{aligned} \int_I \exp(-Nf(x)) dx &\geq \sqrt{\frac{2\pi}{Nf''(0)}} \cdot \exp(-Nf(0)) \cdot R(\epsilon, N), \\ R(\epsilon, N) &= (1 - \sqrt{\frac{2}{\pi}} \exp(-N\epsilon^2 f''(0))) \cdot \exp(-Nf''(0)C\epsilon^3/16). \end{aligned}$$

The derivation of the lemma is similar to that of Lemma 8.4 in [10] (although with all inequalities reversed).

**Lemma 6.7.** *Consider the collection of integrals*

$$J(N, \lambda) = \int_0^\infty \exp(-Nf_\lambda(x)) dx$$

and suppose either that  $\lambda = \lambda_0$  independent of  $N$  or that  $\lambda = \lambda_N \rightarrow \lambda_0$  as  $N \rightarrow \infty$ . Suppose that  $f_\lambda$  has a unique minimizer  $x_\lambda$  interior to  $(0, \infty)$  and suppose that  $f_\lambda$  is  $C^4(0, \infty)$ . Let

$$C(\lambda, \epsilon) = \sup_{|x-x_\lambda|<\epsilon} \frac{|f_\lambda'''(x)|}{f_\lambda''(x_\lambda)}.$$

Suppose that

$$Nf_{\lambda_N}(x_{\lambda_N}) \rightarrow \infty$$

and

$$\frac{C(\lambda_N, \epsilon_N)}{\sqrt{N\psi''(x_{\lambda_N})}} \rightarrow 0.$$

Then

$$J(N, \lambda_N) \geq \sqrt{\frac{2\pi}{Nf_\lambda''(x_\lambda)}} \cdot \exp(-Nf_\lambda(x_\lambda))(1 + o(1)).$$

To prove Lemma 6.7, simply translate coordinates so that  $x_\lambda = 0$ , pick  $\epsilon_N = N^{-2/5}$  and set  $I = [-\epsilon, \epsilon]$ ; then apply Lemma 6.6.

**6.5. Analysis of the internal angle.** Our earlier analysis of the internal angle employed an upper bound derived in [10] from large-deviations theory. We now develop a lower bound using complex analysis techniques; our analysis is related to the approach of Vershik and Sporyshev [34].

Let  $X \sim HN(0, 1)$  be a real half-normal random variable, i.e.  $X = |Z|$  where  $Z$  is standard normal. The moment generating function  $M(t) = Ee^{tX}$  can be continued to the complex plane. We have the explicit formula  $M(t) = e^{t^2/2} \cdot 2\Phi(t)$ , where  $\Phi$  denotes the standard  $N(0, 1)$  cumulative distribution function. Operations with Taylor series show that for  $\omega$  real,  $\Phi(i\omega)$  has real part  $1/2$  along the imaginary axis and so the cumulant generating function  $\log(2\Phi(z))$  can be consistently defined in a neighborhood of both the real and imaginary axes. Define

$$\psi_\gamma(z) = z^2/2 + (1 - \gamma) \log(2\Phi(z)).$$

We begin by justifying our interest in the complex domain:

**Lemma 6.8.** *For  $\gamma = \frac{k+1}{\ell+2}$ ,*

$$\beta(T^k, T^\ell) = \sqrt{\ell+3} \cdot 2^{-\ell-k+1} \cdot \frac{1}{\sqrt{2\pi}} \int_{-i\infty}^{i\infty} e^{(\ell+2)\psi_\gamma(z)} dz.$$

Contour integration was previously used in the analysis of the internal angle by Vershik and Sporyshev, without making the connection to the cumulant generating function. The contour integral and the form of the integrand suggests to use the method of steepest descents [2]. An analysis of  $\psi_\gamma(z)$  is easily performed computationally. One learns that there is a path  $C_\gamma$  along which  $\psi_\gamma(z)$  is purely real and which is asymptotic, for large  $|z|$ , to the imaginary axis; see Figure 6.1. This path crosses the real axis at a point  $z_\gamma$ . Because  $\psi_\gamma$  is real for real  $z$ ,  $z_\gamma$  is necessarily a

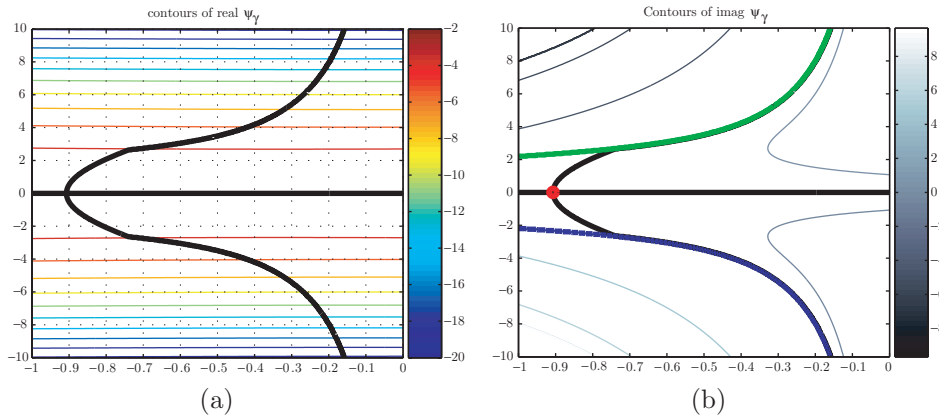


FIGURE 6.1. Level curves for the exponent  $\psi_\gamma(z)$  with  $\gamma = 3/8$ : real (a) and imaginary (b) components. The path  $C_\gamma$  along which  $\psi_\gamma(z)$  is purely real is overlaid as the wider black line. Panel (b) additionally overlays the level curves with the imaginary part of  $\psi_\gamma(z)$  being equal to  $-\pi$  (green) and  $\pi$  (blue). The path  $C_\gamma$  lies between these hyperbolae and the imaginary axis, allowing the deformation in (6.9) without necessitating branch cuts. The *saddlepoint*,  $z_{3/8} \approx -0.907 + i0$ , is indicated by the red circle in panel (b).

saddlepoint of  $\psi_\gamma$ . Within the region bounded by the imaginary axis and  $C_\gamma$ ,  $\psi_\gamma$  is analytic, and so we have the identity

$$(6.9) \quad \int_{-i\infty}^{i\infty} e^{(\ell+2)\psi_\gamma(z)} dz = \int_{C_\gamma} e^{(\ell+2)\psi_\gamma(z)} dz,$$

provided the orientation of the path  $C_\gamma$  is chosen properly. Parametrizing by arc-length, the contour integral can be rewritten purely in terms of real variables:

$$\int_{-\infty}^{\infty} e^{(\ell+2)\tilde{\psi}_\gamma(t)} dt$$

where  $\tilde{\psi}_\gamma(t) = \psi_\gamma(z(t))$ ; this of course is in the form of a Laplace integral. Taking into account that

$$\tilde{\psi}_\gamma(0) = \psi_\gamma(z_\gamma), \quad \tilde{\psi}_\gamma''(0) = \psi_\gamma''(z_\gamma)$$

and that  $\tilde{\psi}_\gamma(t)$  is  $C^4(-\infty, \infty)$ , we immediately have

**Lemma 6.9.** *Let  $\gamma$  be fixed in  $(0, 1)$ . Then  $\psi_\gamma(z)$  has a saddlepoint  $z_\gamma$  on the negative real axis and*

$$\int_{-i\infty}^{i\infty} e^{(\ell+2)\psi_\gamma(z)} dz = \sqrt{\frac{2\pi}{(\ell+2)\psi_\gamma''(z_\gamma)}} \cdot \exp\{(\ell+2)\psi_\gamma(z_\gamma)\} \cdot (1 + o(1)), \quad \ell \rightarrow \infty.$$

Actually, however, we are interested in the case where  $\gamma$  is changing slightly with  $n$ , i.e.  $\gamma = \gamma_n = \frac{k_n+1}{\ell_n+2}$ , and we need a stronger result. We note that the third and fourth derivatives of  $\psi_\gamma(t)$  near  $t = 0$  are bounded locally uniformly in  $\gamma$ . We conclude

**Lemma 6.10.** *Fix  $c > 0$ . Let  $\gamma_n = \frac{k_n+1}{\ell_n+2}$ . In the proportional growth setting, we have*

$$\int_{-i\infty}^{i\infty} e^{(\ell_n+2)\psi_{\gamma_n}(z)} dz = \sqrt{\frac{2\pi}{(\ell+2)\psi''_{\gamma_n}(z_{\gamma_n})}} \exp\{(\ell_n+2)\psi_{\gamma_n}(z_{\gamma_n})\} \cdot (1+o(1)), \quad n \rightarrow \infty,$$

with the term  $o(1)$  uniform in  $n \leq \ell_n \leq n + c\sqrt{n}$ .

To complete the evaluation of the asymptotics of the internal angle, we need

**Lemma 6.11.**

$$\psi''_\gamma(z_\gamma) = 1 - z_\gamma^2 \cdot \frac{\gamma}{1-\gamma}.$$

Let  $\xi_\gamma(y)$  denote the function introduced earlier in connection with the internal angle. Then

$$\psi_\gamma(z_\gamma) = -(1-\gamma) \cdot \xi_\gamma(y_\gamma).$$

We conclude that

$$(6.10) \quad \beta(T^k, T^\ell) \geq c_2 \cdot \exp\left(-N\Psi_{int}^*\left(\frac{\ell+1}{N}, \frac{k+1}{\ell+1}\right)\right).$$

The result (6.2) follows.

## 6.6. Proof of Lemmas 6.8 and 6.11.

6.6.1. *Proof of Lemma 6.8.* Böröczky and Henk gave the formula

$$(6.11) \quad \beta(T^k, T^\ell) = \theta^{(m-1)/2} \cdot \sqrt{(m-1)\alpha+1} \cdot \pi^{-m/2} \cdot \alpha^{-1/2} \cdot J(m, \theta),$$

where

$$(6.12) \quad \theta = k+1, \quad \alpha = 1/(k+2), \quad m = \ell - k + 1,$$

and

$$J(m, \theta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-\lambda^2} \left( \int_0^{\infty} \exp(-\theta v^2 + 2iv\lambda) dv \right)^m d\lambda.$$

Note that

$$\int_0^{\infty} \exp(-\theta v^2 + 2iv\lambda) dv = \frac{\sqrt{\pi}}{2\sqrt{\theta}} \cdot E e^{i\sqrt{\frac{2}{\theta}}\lambda X},$$

where  $X$  is standard half normal,  $X = |Z|$ ,  $Z \sim N(0, 1)$ . Using now the cumulant generating function of the half-normal,

$$\Lambda(z) = \log E e^{zX},$$

we write

$$J(m, \theta) = \frac{\pi^{m/2-1/2}}{2^m \theta^{m/2}} \cdot \int_{-\infty}^{\infty} e^{(i\lambda)^2} e^{m\Lambda(i\sqrt{\frac{2}{\theta}}\lambda)} d\lambda.$$

Now change variables  $\omega = \sqrt{\frac{2}{\theta}}\lambda$ , and write

$$J(m, \theta) = \frac{\pi^{m/2-1/2}}{2^m \theta^{m/2}} \cdot \sqrt{\frac{\theta}{2}} \int_{-\infty}^{\infty} e^{\theta(i\omega)^2/2} e^{m\Lambda(i\omega)} d\omega.$$

Recalling (6.11)–(6.12) and noting that

$$\theta = k + 1, \quad ((m + 1)\alpha + 1)/\alpha = \ell + 3,$$

we have

$$\beta(T^k, T^\ell) = \sqrt{\ell + 3} \cdot 2^{-m} \cdot \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-(k+1)\omega^2/2 + (\ell-k+1)\Lambda(i\omega)} d\omega.$$

The cumulant generating function of the half-normal obeys  $\Lambda(s) = e^{s^2/2} \cdot 2\Phi(s)$ . Setting  $\gamma = \frac{k+1}{\ell+2}$ , the exponent can be rewritten as  $(\ell + 2)\psi_\gamma(z)$ .  $\square$

6.6.2. *Proof of Lemma 6.11.* Note that

$$\Lambda'(z) = \phi(z)/\Phi(z),$$

where  $\psi$  is the standard normal density and  $\Phi$  is the standard normal cumulative distribution function. Hence from  $\psi'_\gamma(z) = z + (1 - \gamma)\phi(z)/\Phi(z)$  and  $\psi_\gamma(z_\gamma) = 0$  we have

$$(6.13) \quad -z_\gamma/(1 - \gamma) = \phi(z_\gamma)/\Phi(z_\gamma).$$

We also have

$$\psi''_\gamma(z) = 1 + (1 - \gamma)\left[\frac{\phi'}{\Phi} - \frac{\phi^2}{\Phi^2}\right]$$

and  $\phi' = (-z)\phi$ . Hence

$$\psi''_\gamma(z_\gamma) = 1 - z_\gamma^2 \cdot \frac{\gamma}{1 - \gamma}.$$

This proves half the lemma.

For the other half of the lemma, we need to establish a connection between the values of  $\psi_\gamma(z_\gamma) = z_\gamma^2/2 + (1 - \gamma)\Lambda(z_\gamma)$  and  $\xi_\gamma(y_\gamma)$ , where

$$\xi_\gamma(y) = \frac{1 - \gamma}{\gamma}y^2/2 + \Lambda^*(y).$$

Here  $\Lambda^*(x) = \max_s sx - \Lambda(s)$  is the classical Fenchel-Legendre transform of the cumulant generating function on the real axis. It is worth reviewing Sections 6.4 and 6.5 of [10]. The definition of  $\Lambda^*$  sets up a one-one relationship between variables  $(y, s)$ , where  $y = y(s)$  and  $s = s(y)$ , so that

$$\Lambda^*(y) = s(y)y - \Lambda(s).$$

Hence

$$\xi_\gamma(y_\gamma) = s(y_\gamma)y_\gamma - \Lambda(s_\gamma) + \frac{1 - \gamma}{\gamma}y_\gamma^2/2.$$

Formula (6.12) in [10] reads

$$\frac{1 - \gamma}{\gamma}y_\gamma = -s_\gamma;$$

this implies

$$\begin{aligned} \xi_\gamma(y_\gamma) &= -\frac{\gamma}{1 - \gamma}s_\gamma^2/2 - \Lambda(s_\gamma) \\ &= -\frac{\gamma}{1 - \gamma}s_\gamma^2/2 - s_\gamma^2/2 - \log(2\Phi(s_\gamma)) \\ &= -\frac{1}{1 - \gamma}s_\gamma^2/2 - \log(2\Phi(s_\gamma)). \end{aligned}$$

We note – parenthetically – that the variable  $s$  is in this subsection the argument to a cumulant generating function, and elsewhere in the paper, the same symbol

denotes the negative of this same quantity. Moreover the dual relationship between  $s, y$  variables is expressed through  $\Lambda'(s_\gamma) = y_\gamma$ . We compute that  $\psi_\gamma(s_\gamma + i0) = 0$ , i.e.  $z_\gamma = s_\gamma + i0$ . In words, the saddlepoint value  $z_\gamma$  is identical to the dual variable  $s_\gamma$ . Finally we have

$$-\Psi_{int}(\nu, \gamma) = -(\xi_\gamma(y_\gamma) + \log_e(2)) \cdot \nu \cdot (1 - \gamma) = (\psi_\gamma(z_\gamma) + \log_e(2)(1 - \gamma)) \cdot \nu.$$

Compare also Section 6.5 of [10]. □

## 7. DISCUSSION

In this section, we first show how the applications (1.1)–(1.4) follow from Theorems 1.2–1.6. We next consider the performance of these rules at finite  $n$ . Finally we discuss extensions, open questions, and relations to other work.

**7.1. Convex hulls of Gaussian point clouds. Proof of (1.1).** In the 1950's, David Gale [17] introduced an important extremal property of polytopes; the following is now classical:

**Definition 7.1** ([19, Chapter 7]). A convex polytope  $P$  is called  $k$ -neighborly if every subset of  $k + 1$  vertices spans a  $k$ -face of  $P$ .

By mere face counting, we can determine whether a polytope is  $k$ -neighborly. In this section, put for short  $T = T^{N-1}$ .

**Lemma 7.2** ([19, Chapter 7]). *Let  $P = AT$ . Suppose that*

- $P$  has  $N$  vertices,
- $P$  has  $\binom{N}{k+1}$   $k$ -faces.

*Then  $P$  is  $k$ -neighborly.*

Combining Theorems 1.3 and 1.6 we have

**Corollary 7.1.** *Let  $(k_n, n, N_n)$  be a sequence of triples with  $n$  tending to  $\infty$  and  $N_n$  growing subexponentially with  $n$ . Fix  $\epsilon > 0$  and suppose that*

$$k_n < (1 - \epsilon) \cdot \frac{n}{2e \cdot \log(N_n/(n \cdot 2\sqrt{\pi}))}, \quad n > n_0.$$

*Let  $A$  be a random  $n \times N$  matrix with iid  $N(0, 1/n)$  entries. Define the event*

$$\Omega(k, n, N) = \{P = AT \text{ is } k\text{-neighborly}\}.$$

*Then*

$$P(\Omega(k_n, n, N_n)) \rightarrow 1, \quad \text{as } n \rightarrow \infty.$$

In words, with overwhelming probability for large  $n$ ,  $P = AT$  is at least  $k_n$ -neighborly.

This is simply (1.1) in another language. To see why, note that, for each  $k > 1$ , a  $k$ -neighborly polytope is also  $k - 1$ -neighborly. If  $a_1, \dots, a_N$  are vertices of  $\mathcal{A} = \text{conv}(a_1, \dots, a_N)$ , then  $k$ -neighborliness of  $\mathcal{A}$  is equivalent to the following  $k$  simultaneous properties:

- every pair  $(a_i, a_j)$  spans an edge of  $\mathcal{A}$ ,
- ...
- every  $k + 1$ -tuple  $(a_{i_1}, \dots, a_{i_k})$  spans a  $k$ -face of  $\mathcal{A}$ .

This is precisely the condition mentioned in Section 1.1.1 with the substitutions  $x_i \leftrightarrow a_i$ ,  $\mathcal{A} \leftrightarrow \mathcal{X}$ ,  $n \leftrightarrow d$ , and  $N \leftrightarrow n$ .

To conclude, we note that  $P = AT$  has  $N$  vertices with probability 1, those vertices are simply the columns of  $A$ , and so  $P = \text{conv}(a_1, \dots, a_N)$ . Invoking now the above corollary, we obtain the conclusion (1.1).  $\square$

**7.2. Correcting all patterns of  $k$  or fewer errors. Proof of (1.4).** A convex polytope is centrosymmetric if it has  $2N$  vertices made of  $N$  antipodal pairs. Neighborliness *per se* does not apply to centrosymmetric polytopes; instead one needs the following notion (see e.g. [19, Chapter 8]).

**Definition 7.3.** A centrosymmetric convex polytope  $P$  with vertices  $\pm a_1, \dots, \pm a_N$  is called **centrally  $k$ -neighborly** if every subset of  $k + 1$  vertices *not including an antipodal pair* spans a  $k$ -face of  $P$ .

By face counting, we can determine whether a polytope is centrally  $k$ -neighborly. In this section, put for short  $C = C^N$ .

**Lemma 7.4** ([10, Lemma 1]). *Let  $P = AC$ . Suppose that*

- $P$  has  $2N$  vertices and
- $P$  has  $2^{k+1} \cdot \binom{N}{k+1}$   $k$ -faces.

*Then  $P$  is centrally  $k$ -neighborly.*

Combining Theorems 1.5 and 1.6, we have

**Corollary 7.2.** *Let  $(k_n, n, N_n)$  be a sequence of triples with  $n$  tending to  $\infty$  and  $N_n$  growing subexponentially with  $n$ . Fix  $\epsilon > 0$  and suppose that*

$$k_n < (1 - \epsilon) \cdot \frac{n}{2e \cdot \log(N_n/(n \cdot \sqrt{\pi}))}, \quad n > n_0.$$

*Let  $A$  be a random  $n \times N$  matrix with iid  $N(0, 1/n)$  entries. Then  $P = AC$  is a random centrosymmetric polytope. Define the event*

$$\Omega(k, n, N) = \{P = AC \text{ is centrally } k\text{-neighborly}\}.$$

*Then*

$$P(\Omega(k_n, n, N_n)) \rightarrow 1, \quad \text{as } n \rightarrow \infty.$$

In words, with overwhelming probability for large  $n$ ,  $P = AC$  is at least  $k_n$ -centrally neighborly.

We now relate central  $k$ -neighborliness to (1.4). Recall the optimization problem

$$(P_1) \quad \min_x \|x\|_1 \text{ subject to } y = Ax.$$

Call the solution  $x_1$ ; it obviously depends on  $y$  and  $A$ .

**Theorem 7.5** ([11]). *The following statements about an  $n \times N$  matrix  $A$  are equivalent.*

- *The polytope  $AC$  has  $2N$  vertices and is centrally  $k$ -neighborly.*
- *For every problem instance  $y = Ax_0$  where  $x_0 \in \mathbf{R}^N$  has at most  $k$  nonzeros, the solution  $x_1$  to the corresponding instance of  $(P_1)$  is unique and is equal to  $x_0$ .*

To apply this, recall the setting of Section 1.1.3. The encoding matrix  $B$  mentioned there was obtained as follows: a random orthogonal matrix  $U$  is generated, and  $B$  makes up  $N - n$  rows of this matrix. The checksum matrix  $A$  makes up the other  $n$  rows of  $U$ .

Given received data  $w \in \mathbf{R}^N$ , form the generalized checksum  $y = Aw \in \mathbf{R}^n$ . Then solve the instance of  $(P_1)$  defined by  $(y, A)$ . Define the reconstruction  $u_1 = B(w - x_1)$ . Inequality (1.4) now follows from the above and the following:

*Claim.* If  $AC$  is centrally  $k$ -neighborly and if the error vector  $z$  has at most  $k$  nonzeros, one has perfect error-correction:

$$u = u_1.$$

*Proof.* The received message  $w = B^T u + z$  where the error vector  $z$  has, by hypothesis, nonzeros in at most  $k$  positions. Since  $AB^T = 0$ ,  $y = Az$ . Invoking Theorem 7.5, we have  $x_1 = z$ . Hence  $B(w - x_1) = B(w - z) = BB^T u + z - z = u$ .  $\square$

**7.3. How many projections? Proof of (1.2).** We first transform the “how many questions” problem into face counting.

**Definition 7.6.** The random  $n \times N$  matrix  $A$  will be called *orthant-symmetric* if, for every signed permutation  $\Pi$  and for every measurable  $\Omega \subset \mathbf{R}^{n \times N}$ ,

$$P\{A \in \Omega\} = P\{A\Pi \in \Omega\}.$$

**Theorem 7.7** ([11]). *Let  $A$  be an orthant symmetric random  $n \times N$  matrix. Let  $x_0$  be a fixed vector with  $k$  nonzeros. Form a random problem instance  $(y, A)$  of  $(P_1)$ , where  $y = Ax_0$ . Let  $x_1$  denote the solution of this instance of  $(P_1)$ .*

$$P\{x_1 = x_0\} \geq \frac{Ef_{k-1}(AC)}{f_{k-1}(C)}.$$

Theorems 1.4 and 1.6 imply the following precise version of (1.2).

**Corollary 7.3.** *Let  $(k_n, n, N_n)$  be a sequence of triples with  $n$  tending to  $\infty$  and  $N_n$  growing subexponentially with  $n$ . Fix  $\epsilon > 0$  and suppose that*

$$(7.1) \quad k_n < (1 - \epsilon) \cdot \frac{n}{2 \cdot \log(N_n/n)}, \quad n > n_0.$$

Then

$$\frac{Ef_{k-1}(AC)}{f_{k-1}(C)} \rightarrow 1, \quad n \rightarrow \infty.$$

In words, for  $(k, n, N)$  obeying the asymptotics (7.1), an overwhelming fraction of the  $k - 1$  faces  $F$  of  $C$  induce  $k - 1$  faces  $AF$  of  $AC$ .

$P = AT$  is at least  $k_n$ -neighborly.

**7.4. Correcting random patterns of  $k$  errors or fewer. Proof of (1.3).** Let  $\|z\|_0$  count the number of nonzeros in  $z$ .

**Definition 7.8.** The random vector  $z$  is a *symmetric  $k$ -sparse random vector* if

- $P\{z \in \Omega\} = P\{-z \in \Omega\}$  for all measurable sets  $\Omega$  and
- $P\{\|z\|_0 \leq k\} = 1$ .



Suppose that the received message  $w = B^T u + z$  where  $u$  is arbitrary and  $z$  is a symmetric  $k$ -sparse random vector stochastically independent of  $A, B$ . Define  $y = A\mu$  and consider the resulting instance of  $(P_1)$ . Then, conditional on each fixed realization of  $z$ , put  $x_0 := z$  and apply Theorem 7.7 to get that

$$E\{f_{k-1}(AC)|z\} \geq (1 - \epsilon)f_{k-1}(C)$$

implies

$$P\{x_1 = z|z\} \geq 1 - \epsilon.$$

By independence of  $z$  and  $A$ ,

$$E\{f_{k-1}(AC)|z\} = Ef_{k-1}(AC).$$

Apply now Corollary 7.3 to infer (1.3) □

**7.5. Empirical results.** The phenomena uncovered by Theorems 1.2 and 1.4 can be observed empirically. For a given  $(\delta, \rho)$  pair, pick a large  $N$ , generate a random  $A$  of dimensions  $n = \lfloor \delta N \rfloor$  by  $N$ , and check whether for  $k = \lfloor \rho \cdot n \rfloor$ , a randomly chosen  $k$ -face  $F$  of  $Q$  yields a projected simplex  $AF$  that is also a face of  $AQ$ ; here  $Q = T^{N-1}$  or  $Q = C^N$ . This can be verified by linear programming.

Let  $F$  be a  $k-1$ -face of  $Q = T^{N-1}$  or  $C^N$ . Then the elements of  $F$  have nonzeros in only  $k$  coordinates. If  $Q = T^{N-1}$ , the nonzeros are nonnegative; if  $Q = C^N$ , the nonzeros have a definite sign pattern particular to the interior of  $F$ . In the following result, let  $\chi_F$  denote the barycenter of the face  $F$ , and let  $(\mathcal{P})$  denote problem  $(P_1)$  if  $Q = C^N$  or problem  $(LP)$  if  $Q = T^{N-1}$ , where

$$(LP) \quad \min 1'x \text{ subject to } y = Ax, x \geq 0.$$

**Theorem 7.9** ([11, 13]). *Let  $Q = T^{N-1}$  or  $C^N$ . The following statements about a face  $F$  of  $Q$  are equivalent.*

- $AF$  is a face of  $AQ$ .
- Let  $y_F = A\chi_F$ . Then  $\chi_F$  is the unique solution of the instance of  $(\mathcal{P})$  defined by  $(y_F, A)$ .

Thus, to check Theorems 1.2 and 1.4, one checks that for a randomly-generated vector  $x_0$  with  $k$  nonzeros, the corresponding vector  $y = Ax_0$  generates an instance of either  $(LP)$  or  $(P_1)$  uniquely solved by  $x_0$ ;  $(LP)$  corresponds to  $x_0 \geq 0$  and  $Q = T^{N-1}$  whereas  $(P_1)$  corresponds to  $x_0$  with entries of either sign and  $Q = C^N$ . If such uniqueness holds, we call that experiment a success. Theorems 1.2 and 1.4 imply that for  $k$  below a given threshold, success is very likely while above that threshold, success is very unlikely.

We conducted 44,000 such experiments with the common value  $N = 10,000$ , exploring the  $(\delta, \rho)$  domain as follows. We considered  $n = 10, 15, 20, \dots, 100$ ; for each value of  $n$ , eleven values of the sparsity,  $k$ , were chosen near the asymptotic thresholds,  $n \cdot |2 \log(\delta)|^{-1}$ . At each combination of  $k$  and  $n$ , two hundred random problem instances were generated.

Figures 7.1(a)–(c) summarize our results. A region of the  $(\delta, \rho)$  plane is decorated with a shaded attribute depicting the fraction of successful experiments. Figure 7.1(a) shows the simplex case, along with the threshold  $\rho_W^+(\delta)$  and its asymptotic approximant,  $|2 \log(\delta)|^{-1}$ . Figure 7.1(b) shows the cross-polytope case, with the threshold  $\rho_W^\pm(\delta)$  and the approximant,  $|2 \log(\delta)|^{-1}$ . To better highlight the (subtle) difference between the simplex and cross-polytope cases, Figure 7.1(c) shows the

fraction of cases where the simplex experiments were successful and the cross-polytope experiments were not.

Figures 7.1(a) and (b) display a remarkable match between the thresholds  $\rho_W^+(\delta)$ ,  $\rho_W^\pm(\delta)$  and their asymptotic approximations,  $|2 \log(\delta)|^{-1}$ . Both curves track the observed empirical phase transition. This empirical transition is of course not a true discontinuity, because we are working with finite problem size  $N = 10,000$ ; instead it is a relatively abrupt change. Still, some relatively sharp distinctions can be made; there is a definite region where the simplex experiment is typically successful but the cross-polytope experiment is not – see Figure 7.1(c).

For  $\delta$  near  $1/100$  the empirical transitions at  $N = 10,000$  show a clear agreement with the theoretical thresholds  $\rho_W(\delta)$  and the  $|2 \log(\delta)|^{-1}$  asymptotic approximant. Fixing the region  $\delta \in [1/1000, 1/100]$  explored in Figure 7.1 and increasing  $N$  offers better resolution in  $k/n$ ; the sharper empirical transition is again in agreement with the theoretical thresholds.

**7.6. Relation to other work.** We discussed face-counting related work in the body of the text as the opportunity arose. We now mention several categories of related literature.

**7.6.1. How neighborly can a polytope be?** Theorems 1.3, 1.5, and 1.6 imply the following. *For  $N \gg n$ , both large, and  $N$  subexponential in  $n$ , there exist polytopes  $P$  which are*

- *$k$ -neighborly with*

$$k \sim \frac{n}{2e \log(N/(n \cdot \sqrt{\pi}))};$$

*indeed, simply take  $P = AT^{N-1}$  where  $A$  has Gaussian i.i.d. entries;*

- *centrally  $k$ -neighborly with*

$$k \sim \frac{n}{2e \log(N/(n \cdot 2\sqrt{\pi}))};$$

*indeed, simply take  $P = AC^N$  where  $A$  has Gaussian i.i.d. entries.*

Recently, the problem of showing the existence of high-dimensional neighborly polytopes has attracted a resurgence of interest. After fundamental work in the 1950's–1970's starting with D. Gale [17, 18] and extending through P. McMullen and G.C. Shephard [27] and R. Schneider [31], the subject was very quiet. Now, as Schneider wrote one of us, “the subject has come to life again”. Our own work [12, 13, 10, 11] carefully studied the questions of neighborliness and central neighborliness of projections of random polytopes in the proportional growth setting. Our attempt was to characterize the exact location of the asymptotic phase transitions associated with strong and weak neighborliness. Linial and Novik [25] gave exponential bounds on the probability that  $AC^N$  is centrally neighborly; note that Rudelson and Vershynin's work [29] came earlier and implies similar bounds by duality. Both [25, 29] use a geometric functional analysis approach which gave inequalities akin to

$$(7.2) \quad P\{f_k(AC) \neq f_k(C)\} \leq \psi_1 \exp(-\psi_2 n), \quad n > n_0,$$

valid for  $k < cn/\log(N/n)$  with unspecified constants.

Other authors apparently sought simply to obtain a formula showing a threshold of the form  $k \leq cn/\log(N/n)$ , which is indeed the qualitatively correct formula.

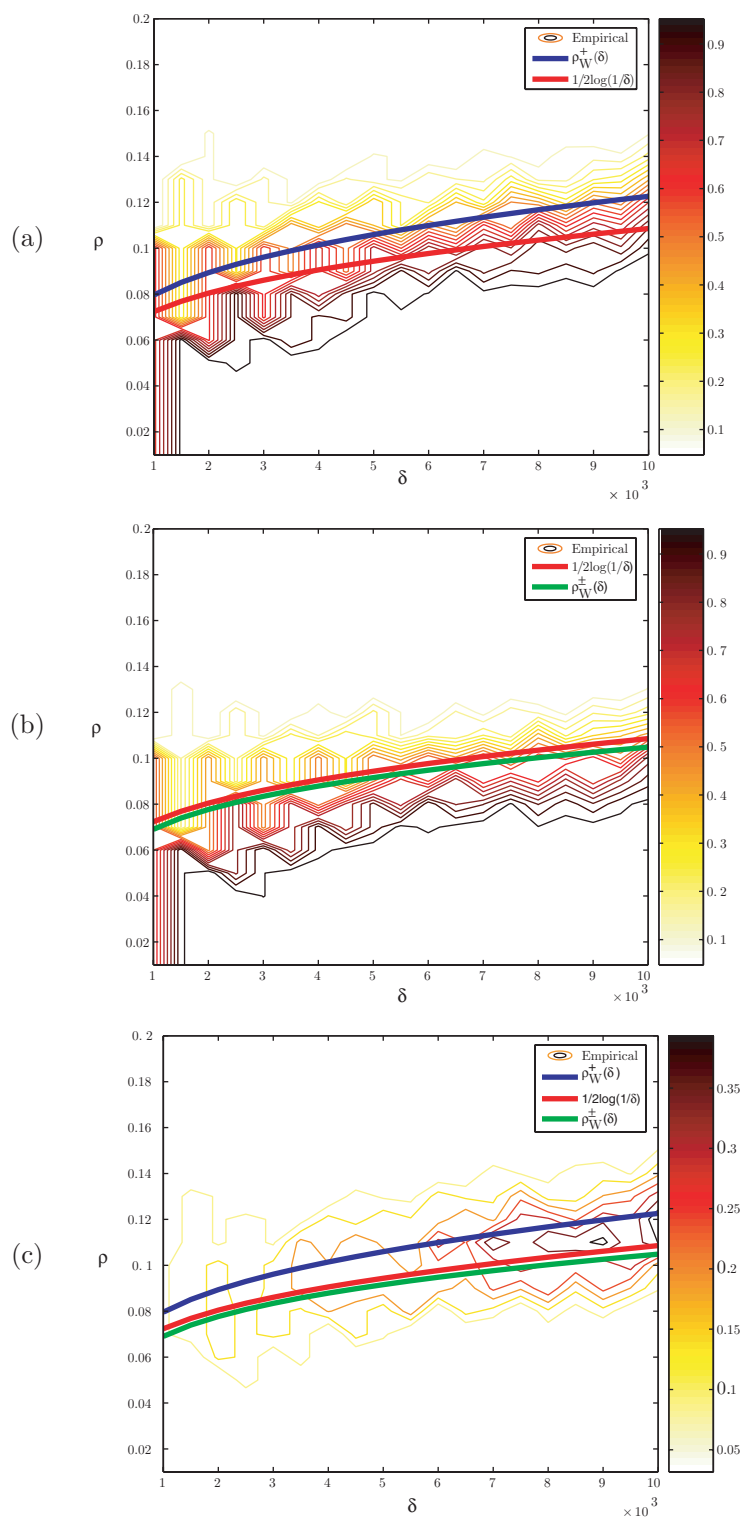


FIGURE 7.1. Panel (a): success fraction, simplex; panel (b): success fraction, cross-polytope; panel (c): fraction successful for simplex but not for cross-polytope.  $N = 10,000$ .

Our work serves to advance a viewpoint with three concepts: (a) phase diagram; (b) existence of sharp phase transitions; (c) precise location of phase transitions. At the same time the inequalities developed here will be shown elsewhere to give bounds on finite, i.e.  $N$ , probabilities that are much stronger than bounds currently available from the best known bounds of the form currently available from threshold relations obtained by other methods,  $k \leq cn/\log(N/n)$ . However, if the reader just wants a simple argument showing that no transition can occur for  $k \leq cn/\log(N/n)$ , then we recommend the admirably short proof of Rudelson and Vershynin, [30].

*7.6.2. How many projections are needed to recover a  $k$ -sparse object?* One reason that study of neighborliness “has come to life again” is the surprising implications for speeding up key processes in medical imaging and proteomics. The general idea, often labeled *compressed sensing* [8] is that images, spectra, and other real-world objects are highly compressible and that this compressibility makes it possible to reconstruct such objects accurately from relatively few carefully chosen generalized samples. In effect Section 1.1.2 has described an abstract model of compressed sensing.

In the application scenario,  $x_0$  represents the coefficients of an image to be acquired and the rows of  $A$  represent a random set of linear combinations (measurements) which will be used to reconstruct  $x_0$ . In effect, we are saying that if  $x_0$  has  $N$  pixels but only  $k \ll N$  nonzeros in (say) a wavelet basis and if  $k$  and  $N$  are large, then we only need  $n$  measurements, where

$$n \geq 2k \log(N/n)(1 + o(1)).$$

In contrast,  $N$  is the ‘standard’ number of samples; the point is that for objects which are  $k$ -sparse with  $k$  small, we can easily have  $n \ll N$  if  $x_0$  is highly sparse. (In fact real objects will not exhibit such strict sparsity ( $k$  zeros and  $N - k$  nonzeros) but because the  $(P_1)$  has an  $\ell_1$  stability property [8], we can pretend that this is so without distorting the problem.)

The interested reader may pursue the papers of Candès and collaborators [4, 6], other theoretical work [29, 21, 32] and much recent applied work [33, 15].

The quantitative approach developed here is precise about how much data would be needed. Most of the cited theoretical work is qualitative, often leaving the constants unspecified. An important point: in Section 1.1.2 and in (1.2) we are studying the equivalent of *weak central neighborliness*. We argued in Section 7.5 that this is the empirically relevant notion; we repeat here that ordinary (strong) central neighborliness is simply not empirically observable. Nevertheless, most authors have effectively studied implications of ordinary (strong) central neighborliness. That notion is hard to analyze and appears to indicate a far more pessimistic view of what is possible than what one actually observes in practice.

*7.6.3. Fast decoding of error-correcting codes.* In general, decoding of linear error-correcting codes is NP-hard [16]. However, fast decoding of specific error-correcting codes has been an object of great practical and theoretical attention over the last 10 years, with great advances in turbo codes and in LDPC codes (Gallager codes). We proposed in Section 1.1.3 above a simple scheme for fast decoding of random linear codes over  $\mathbf{R}$  using  $(P_1)$ . The scheme we proposed is equivalent to one proposed by Candès and Tao [5] and studied further by Rudelson and Vershynin [29]. Using

the notation of our Section 1.1.3, their decoder solves the  $\ell_1$ -minimization problem

$$\min_u \|w - B^T u\|_1,$$

yielding the reconstruction  $u_1$ , say. The equivalence of such minimization with the one proposed in Section 1.1.3 is shown in [11, Section 8].

It is of course crucial to know how many errors such a scheme can correct. The theoretical literature (Candès and Tao/Rudelson and Vershynin) gives qualitative results, saying that one can correct at least  $cn/\log(N/n)$  errors, with  $c$  left unspecified, or else specified as a constant which seems much smaller than what would be expected based on a comparison of those papers' results with the results obtained here.

The problem solved in those visionary papers is to show existence of integer-valued matrix pairs  $A, B$  allowing block coding of messages of length  $m$  as blocks of length  $N$ , such that all patterns of at most  $k$  errors can be corrected; here  $n = N - m$ . Our results here change the problem so that  $A$  and  $B$  are generated by partitioning a uniformly distributed random projection matrix (n.b. not with integer-valued entries); with this change, we get a precise asymptotic formula  $k = n\rho_S^\pm(n/N)(1 + o_P(1))$ . For the case  $N \gg n$  we have proven the formula  $\rho_S^\pm(n/N) \sim 1/2e \log(N/(n\sqrt{\pi}))$ .

If we change the problem again slightly so that the goal is to correct *nearly all* rather than *all* error patterns, then for the case of long block codes, we get a precise asymptotic formula  $k = n\rho_W^\pm(n/N)(1 + o_P(1))$ . For the case  $N \gg n$  we have proven the formula  $\rho_W^\pm(n/N) \sim 1/2 \log(N/n)$ .

Conceivably, such results for the “changed problems” we just mentioned may be better than for the original problem; i.e. the situation for general random matrices may be more optimistic than for matrices with integer entries. However, our empirical results with Rademacher random matrices indicate that our formula  $n\rho_W^\pm(n/N)$  accurately describes the integer-valued case as well, i.e. accurately describes the number of errors which can typically be corrected by such random matrices with integer-valued entries.

APPENDIX. PROOFS OF KEY LEMMAS

**A.1. Proof of Lemma 4.1.** We develop (4.7) in two stages. Initially, we derive the asymptotic behavior of  $s_\gamma$  as  $\gamma \rightarrow 0$ ; we then substitute that into equation (2.20). To motivate our approximation of  $s_\gamma$ , we use an asymptotic series for  $R(s)$  appropriate for the regime of  $s$  large,

$$R(s) := se^{s^2/2} \int_s^\infty e^{-y^2/2} dy = 1 - \frac{1}{s^2} + \frac{1 \cdot 3}{s^4} - \frac{1 \cdot 3 \cdot 5}{s^6} + \frac{1 \cdot 3 \cdot 5 \cdot 7}{s^8} + \dots$$

This is derived as follows. The ratio  $R(s) = s \cdot Mills(s)$  for  $s > 0$ , where  $Mills(s)$  is the usual Mills ratio for the normal distribution. The corresponding asymptotic series for the Mills ratio is developed in [24, Secs. 5.37,5.38]; H. Ruben [28] credits this series to Laplace.

In [24, Eq. (5.106)] it is shown that the error in truncating the series for  $Mills(\cdot)$  at the  $s$ -th term is at most as large as the  $s$ -th term itself. Here  $R(\cdot)$  inherits this property.

It is now convenient to define  $L(s, \gamma) := R(s) - 1 + \gamma$  and note that  $s_\gamma$  is defined by  $L(s_\gamma, \gamma) = 0$ .

Keeping the first two terms in the series expansion for  $R(s)$  and applying the bounds from [24, Eq. (5.106)] yields  $L(s, \gamma) = \gamma - s^{-2} + 3s^{-4} + \mathcal{O}(s^{-6})$ , which suggests the approximation,

$$(A.1) \quad s_\gamma \approx \tilde{s}_\gamma := \gamma^{-1/2} - \frac{3}{2}\gamma^{1/2}.$$

To quantify the error in this approximation, invoke the mean value theorem; given a smooth function  $F(x)$ , there is always a point  $w \in [\min(x, y), \max(x, y)]$  satisfying

$$(A.2) \quad F(y) = F(x) + (y - x) \frac{d}{dy} F(y)|_{y=w}.$$

Hence we can bound  $|y - x|$  if we have suitable bounds on  $|F(y) - F(x)|$  and  $\frac{d}{dy} F(y)|_{y=w}$ . Apply this principle to  $F(s) = L(s, \gamma)$  about  $s_\gamma$ , getting

$$(A.3) \quad |\tilde{s}_\gamma - s_\gamma| \leq \left| L(\tilde{s}_\gamma, \gamma) \right/ \left. \frac{\partial}{\partial s} L(s, \gamma) \right|_{s=s_{mid}} \right|,$$

for some point  $s_{mid} \in [\min(s_\gamma, \tilde{s}_\gamma), \max(s_\gamma, \tilde{s}_\gamma)]$ .

The following bounds follow from [24, Eq. (5.106)]:

$$(A.4) \quad 1 - s^{-2} + \frac{5}{2}s^{-4} < R(s) < 1 - s^{-2} + 3s^{-4} \quad \text{for } s > \sqrt{30},$$

yielding in turn

$$(A.5) \quad |L(\tilde{s}_\gamma, \gamma)| < \frac{1}{2}\gamma^2, \quad \text{for } \gamma < 1/30.$$

To bound the denominator, note that

$$\frac{\partial}{\partial s} L(s, \gamma) = [s + s^{-1}] \cdot R(s) - s,$$

which is a positive decreasing function of  $s$ ; this attains its lower bound on the interval  $s \in [\min(s_\gamma, \tilde{s}_\gamma), \max(s_\gamma, \tilde{s}_\gamma)]$  at one of the endpoints  $\{s_\gamma, \tilde{s}_\gamma\}$ . At  $\tilde{s}_\gamma$  we again make use of the lower bound on the Mills ratio in equation (A.4),

$$(A.6) \quad \begin{aligned} \frac{\partial}{\partial s} L(s, \gamma)|_{s=\tilde{s}_\gamma} &= [\tilde{s}_\gamma + \tilde{s}_\gamma^{-1}]R(\tilde{s}_\gamma) - \tilde{s}_\gamma \\ &> [\tilde{s}_\gamma + \tilde{s}_\gamma^{-1}](1 - \tilde{s}_\gamma^{-2} + \frac{5}{2}\tilde{s}_\gamma^{-4}) - \tilde{s}_\gamma \\ &= \frac{3}{2}\tilde{s}_\gamma^{-3} + \frac{5}{2}\tilde{s}_\gamma^{-5} > \frac{3}{2}\gamma^{3/2}. \end{aligned}$$

For the lower bound at  $s_\gamma$  we assume  $|\tilde{s}_\gamma - s_\gamma| \leq \frac{1}{2}\gamma^{1/2}$  (which we will verify momentarily), which gives the upper bound  $s_\gamma \leq \gamma^{-1/2} - \gamma^{1/2}$ . From this we have the lower bound,

$$\begin{aligned} \frac{\partial}{\partial s} L(s, \gamma)|_{s=s_\gamma} &= [s_\gamma + s_\gamma^{-1}] \cdot (1 - \gamma) - s_\gamma = (1 - \gamma)s_\gamma^{-1} - \gamma s_\gamma \\ &\geq (1 - \gamma) \cdot \frac{1}{\gamma^{-1/2} - \gamma^{1/2}} - \gamma(\gamma^{-1/2} - \gamma^{1/2}) = \gamma^{3/2}. \end{aligned}$$

Using these bounds in display (A.3), we have

$$(A.7) \quad |s_\gamma - \tilde{s}_\gamma| \leq \frac{1}{2}\gamma^{1/2}, \quad \text{for } \gamma \leq 1/30,$$

which justifies the earlier claim that  $|s_\gamma - \tilde{s}_\gamma| \leq \frac{1}{2}\gamma^{1/2}$ . For the following calculations the following estimate suffices:

$$(A.8) \quad s_\gamma = \gamma^{-1/2} + r_1(\gamma), \quad \text{with } |r_1(\gamma)| \leq 2\gamma^{1/2}, \quad \text{for } \gamma \leq 1/30.$$

Combined with (2.20), this gives (4.7) and hence Lemma 4.1. □

**A.2. Proof of Lemma 4.2.** We first motivate our approximation for  $x_\nu$ , which solves  $2xQ(x)/q(x) = 1 - \nu^{-1}$ . The truncated asymptotic series

$$\frac{2xQ(x)}{q(x)} = 2\pi^{1/2}xe^{x^2} - 1 + \mathcal{O}(x^{-2}), \quad x \rightarrow \infty,$$

suggests approximating  $x_\nu$  as the solution to

$$(A.9) \quad xe^{x^2} - \frac{1}{2}\pi^{-1/2}\nu^{-1} = 0;$$

this is exactly of the form (4.9) with  $z = z^+ := (2\nu\sqrt{\pi})^{-1}$ . Our approach for approximate solution of (4.9), carried out to two stages, yields the approximant  $x_\nu$ , obeying

$$(A.10) \quad \tilde{x}_\nu^2 := \log z^+ - \frac{1}{2} \log \log z^+.$$

Our claim that  $\tilde{x}_\nu$  accurately approximates  $x_\nu$  as  $\nu \rightarrow 0$ , as stated in Lemma 4.2, will be supported by arguments similar to those used in proving Lemma 4.1.

Let

$$(A.11) \quad J(x, \nu) := 2xe^{x^2} \int_{-\infty}^x e^{-y^2} dy + 1 - \nu^{-1},$$

whose level curve  $J(x_\nu, \nu) = 0$  defines  $x_\nu$ . To bound the error in the approximation,  $\tilde{x}_\nu$ , we again use the mean value approach (A.2), getting

$$(A.12) \quad |\tilde{x}_\nu - x_\nu| \leq \left| J(\tilde{x}_\nu, \nu) \right/ \left. \frac{\partial}{\partial x} J(x, \nu) \right|_{x=x_{mid}},$$

for some point  $x_{mid} \in [\min(x_\nu, \tilde{x}_\nu), \max(x_\nu, \tilde{x}_\nu)]$ . The magnitude of  $J(\tilde{x}_\nu, \nu)$  can be bounded by

$$(A.13) \quad \begin{aligned} |J(\tilde{x}_\nu, \nu)| &= -J(\tilde{x}_\nu, \nu) = -2\pi^{1/2}\tilde{x}_\nu e^{\tilde{x}_\nu^2} + \nu^{-1} + r_6(\nu) \\ &\leq -2\pi^{1/2}\tilde{x}_\nu e^{\tilde{x}_\nu^2} + \nu^{-1}, \quad \nu < 1/10, \\ &= \nu^{-1} \left[ 1 - \left( 1 - \frac{1}{2} \frac{\log \log z^+}{\log z^+} \right)^{1/2} \right] \end{aligned}$$

$$(A.14) \quad \leq \nu^{-1} \frac{3}{8} \frac{\log \log z^+}{\log z^+}$$

where the transition from the first to second line utilizes  $r_6(\nu) := 2\tilde{x}_\nu e^{\tilde{x}_\nu^2} \int_x^\infty e^{-y^2} dy - 1 \leq 0$  for  $\nu < 1/10$ .

Turning to the denominator in (A.12), we observe that on the half-line  $x \geq 0$  the derivative is a positive increasing function,

$$(A.15) \quad \begin{aligned} \frac{\partial}{\partial x} J(x, \nu) &= 2x + 2(1 + 2x^2)e^{x^2} \int_{-\infty}^x e^{-y^2} dy \\ &= \frac{1 + 2x^2}{x} [J(x, \nu) + \nu^{-1}] - x^{-1}; \end{aligned}$$

a lower bound for  $\partial J/\partial x$  over  $[\min(x_\nu, \tilde{x}_\nu), \max(x_\nu, \tilde{x}_\nu)]$  is attained at one of the endpoints  $x_\nu$  or  $\tilde{x}_\nu$ . At  $x_\nu$  a simple lower bound is

$$(A.16) \quad \frac{\partial}{\partial x} J(x, \nu)|_{x=x_\nu} = \frac{1 + 2x_\nu^2}{x_\nu} \nu^{-1} - x_\nu^{-1} \geq 2x_\nu \nu^{-1} \quad \text{for } \nu \leq 1.$$

A similar lower bound holds at  $\tilde{x}_\nu$ ,

$$(A.17) \quad \begin{aligned} \frac{\partial}{\partial x} J(x, \nu)|_{x=\tilde{x}_\nu} &= \frac{1 + 2\tilde{x}_\nu^2}{\tilde{x}_\nu} [J(\tilde{x}_\nu, \nu) + \nu^{-1}] - \tilde{x}_\nu^{-1} \\ &\geq \frac{1 + 2\tilde{x}_\nu^2}{\tilde{x}_\nu} \nu^{-1} \left[ 1 - \frac{1}{2} \frac{\log \log z^+}{\log z^+} \right] - \tilde{x}_\nu^{-1} \quad [\text{by (A.13)}] \\ &= \tilde{x}_\nu \nu^{-1} + \tilde{x}_\nu \nu^{-1} \left[ 1 + (\log z^+)^{-1} - \frac{\log \log z^+}{\log z^+} - \nu \right] \\ &\geq \tilde{x}_\nu \nu^{-1} \quad \text{for } \nu \leq 1/4. \end{aligned}$$

Combining (A.16) and (A.17),

$$(A.18) \quad \frac{\partial}{\partial x} J(x, \nu)|_{x=x_{mid}} \geq \nu^{-1} \min(x_\nu, \tilde{x}_\nu);$$

although crude, this is sufficient for later purposes.

Shortly we will prove there is  $\nu_0 > 0$  such that

$$(A.19) \quad \min(x_\nu, \tilde{x}_\nu) \geq \frac{3}{4} \tilde{x}_\nu, \quad 0 < \nu < \nu_0.$$

Substituting (A.14) and (A.18) into equation (A.12) gives

$$(A.20) \quad |x_\nu - \tilde{x}_\nu| \leq \frac{1}{2} \tilde{x}_\nu^{-1} \frac{\log \log z^+}{\log z^+}.$$

Lemma 4.2 follows by simple substitution of terms.

We now show (A.19). Recall that  $\int_{-\infty}^{\infty} e^{-y^2} dy = \sqrt{\pi}$ . Hence on  $x \geq 0$ ,  $J(x, \nu) \leq \tilde{J}(x, \nu) := 2\sqrt{\pi} x e^{x^2} + 1 - \nu^{-1}$ . As  $J(x, \nu)$  is monotone increasing on  $(0, \infty)$ , it follows that  $\tilde{J}(x', \nu) < 0$  implies  $x_\nu > x'$ .

We now show that if  $0 < a < 1$ , then

$$(A.21) \quad \tilde{J}(a\tilde{x}_\nu, \nu) < 0$$

for  $\nu$  sufficiently small. Setting  $a = 3/4$ , this will imply  $J(\frac{3}{4}\tilde{x}_\nu, \nu) < 0$  for all sufficiently small  $\nu$ , and so, for such  $\nu$ ,  $\min(x_\nu, \tilde{x}_\nu) \geq \frac{3}{4}\tilde{x}_\nu$ ; (A.19) follows.

Proceed thus:

$$\begin{aligned} \tilde{J}(a\tilde{x}_\nu, \nu) &= 2\sqrt{\pi} \cdot \frac{a\sqrt{\log(z^+) - \frac{1}{2}\log \log z^+}}{\log(z^+)^{a^2/2}} \cdot (z^+)^{a^2} + 1 - \nu^{-1} \\ &= 2\sqrt{\pi} a \cdot \log(\nu^{-1})^{(1-a^2)/2} (1 + o(1)) \cdot \nu^{-a^2} + 1 - \nu^{-1} \\ &= o(\nu^{-1}) + 1 - \nu^{-1}, \quad \nu \rightarrow 0. \end{aligned}$$

Inequality (A.21) follows. □



**A.3. Proof of Lemma 4.4.** We will show that

$$(A.22) \quad M\left[\frac{\partial}{\partial \nu} \Psi_{net}^+(\delta, r_S^+(\delta))\right] \leq \frac{1}{2} \log\left(\frac{2e}{\tau}\right) + o(1), \quad \delta \rightarrow 0.$$

Because  $\tau > 2e$ , the leading term on the RHS is a negative constant, showing that for small enough  $\delta$  the function  $\Psi_{net}^+$  is monotone decreasing in  $\nu$  on the admissible domain, implying the assertions of the lemma. Now

$$(A.23) \quad \begin{aligned} \frac{\partial}{\partial \nu} \Psi_{net}^+(\nu, \gamma) &= \frac{\partial}{\partial \nu} (\Psi_{com}^+ - \Psi_{int}^+ - \Psi_{ext}^+) \\ &= \log x_\nu + \frac{1}{2} \log(4\pi) + H(\gamma) - (1 - \gamma) \left[ \log\left(\frac{y_\gamma}{\gamma}\right) + \frac{1}{2} \log(2\pi) + \frac{\gamma - 1}{2\gamma} y_\gamma^2 \right]. \end{aligned}$$

Over the interval  $\nu \in [\delta, 1)$ , the first component,  $\log x_\nu$ , is largest at  $\nu = \delta$ . Applying Lemma 4.2, we have

$$(A.24) \quad \begin{aligned} \log x_\nu \leq \log x_\delta &= \frac{1}{2} \log \log z_\delta^+ + \log \left( \left[ 1 - \frac{1}{2} \frac{\log \log z_\delta^+}{\log z_\delta^+} \right]^{1/2} + r_3(\delta) (\log z_\delta^+)^{-1/2} \right) \\ &< \frac{1}{2} \log \log z_\delta^+ \quad \text{for } \delta < 1/50. \end{aligned}$$

The RHS of (A.23) is an increasing function of  $\gamma$ , maximized at  $\gamma = r_S^+(\delta)$ . Using Lemma 4.1, gives, for all  $\gamma$  small enough,

$$(A.25) \quad \begin{aligned} &H(\gamma) - (1 - \gamma) \left[ \log\left(\frac{y_\gamma}{\gamma}\right) + \frac{1}{2} \log(2\pi) + \frac{\gamma - 1}{2\gamma} y_\gamma^2 \right] \\ &< \frac{1}{2} \log \gamma + \frac{1}{2} \log(e/2\pi) + 6\gamma - \frac{1}{2} \gamma \log \gamma \\ &= \frac{1}{2} \log \left[ \frac{e}{2\pi\tau \log z_\delta^+} \right] + \mathcal{O}\left(\frac{\log \log z_\delta^+}{\log z_\delta^+}\right), \quad \delta \rightarrow 0. \end{aligned}$$

Combining (A.23)–(A.25) yields (A.22).  $\square$

**A.4. Proof of Lemma 4.5.** It is sufficient to show that for some  $\gamma_0 > 0$

$$(A.26) \quad \frac{\partial}{\partial \gamma} \Psi_{net}^*(\nu, \gamma) > \nu/2, \quad \nu \in [\delta, 1), \quad 0 < \gamma < \gamma_0.$$

Now

$$(A.27) \quad \begin{aligned} \frac{\partial}{\partial \gamma} \Psi_{net}^*(\nu, \gamma) &= \nu \left[ \gamma^{-1} - 2 \log \gamma + \log(y_\gamma) + \frac{\gamma^2 - 1}{2\gamma^2} y_\gamma^2 + \log(1 - \gamma) \right. \\ &\quad \left. + \frac{1}{2} \log(2\pi) + 1 + (1 - \gamma)^2 \left( \frac{y_\gamma}{\gamma} - \frac{1}{y_\gamma(1 - \gamma)} \right) \frac{d}{d\gamma} y_\gamma \right]. \end{aligned}$$

Lower bounds for each but the last term follow either directly or from Lemma 4.1. For  $\gamma < 1/10$ ,  $y_\gamma$  satisfies

$$(A.28) \quad y_\gamma \geq \frac{\gamma^{1/2}}{1 - \gamma} - 4\gamma^{3/2} > \gamma^{1/2}(1 - 3\gamma)$$

and

$$y_\gamma \leq \frac{\gamma^{1/2}}{1 - \gamma} + 4\gamma^{3/2} < \gamma^{1/2}(1 + 6\gamma),$$

from which follow both

$$\log y_\gamma > \frac{1}{2} \log \gamma - 4\gamma$$

and

$$\frac{\gamma^2 - 1}{2\gamma^2} y_\gamma^2 > \frac{-1}{2\gamma} (1 + 6\gamma)^2 (1 - \gamma^2) \frac{-1}{2\gamma} (1 + 16\gamma),$$

respectively. The last term in (A.27) requires estimating

$$\frac{d}{d\gamma} y_\gamma = \frac{s_\gamma}{(1 - \gamma)^2} \left[ 1 - \frac{\gamma(\gamma - 1)}{\gamma s_\gamma^2 + \gamma - 1} \right].$$

From (A.28)

$$\gamma(\gamma - 1) > \gamma s_\gamma^2 + \gamma - 1 > 4\gamma^2 - 3\gamma > 4\gamma(\gamma - 1)$$

yielding

$$(A.29) \quad 0 < \frac{d}{d\gamma} y_\gamma < \frac{3}{4} \frac{s_\gamma}{(1 - \gamma)^2} < \frac{3}{4} \gamma^{-1/2} + \gamma^{1/2},$$

for  $\gamma \leq 1/30$ . As the above quantity is positive, a lower bound for the last term in (A.27) is obtained with a lower bound on its multiplicative factor,

$$(1 - \gamma)^2 \left( \frac{y_\gamma}{\gamma} - \frac{1}{y_\gamma(1 - \gamma)} \right) > -8\gamma^{1/2}(1 - \gamma)^2$$

which is obtained from (A.28). With (A.29) we arrive at

$$(1 - \gamma)^2 \left( \frac{y_\gamma}{\gamma} - \frac{1}{y_\gamma(1 - \gamma)} \right) \frac{d}{d\gamma} y_\gamma > -6 - 8\gamma \quad \text{for } \gamma < 1/30.$$

Combining these bounds, we have that

$$(A.30) \quad \frac{\partial}{\partial \gamma} \Psi_{net}^*(\nu, \gamma) > \nu \left[ \frac{1}{2} \gamma^{-1} + \frac{3}{2} \log(1/\gamma) - 13 - 14\gamma \right];$$

for  $\gamma < 1/30$ , the term in brackets exceeds  $1/2$ . Inequality (A.26) follows.  $\square$

**A.5. Proof of Lemma 4.6.** We will show that over the admissible domain,

$$(A.31) \quad \frac{\partial}{\partial \nu} (\Psi_{net}^+ - \Psi_{face}^+) < \left[ \frac{1}{2} \log \left( \frac{2e}{\tau} \right) - \frac{1}{\tau} + o(1) \right], \quad \delta \rightarrow 0.$$

As  $\tau > 2$ , this proves Lemma 4.6. For sufficiently small  $\delta$ , we have the inequality

$$(A.32) \quad \begin{aligned} \frac{\partial}{\partial \nu} (\Psi_{net}^+ - \Psi_{face}^+) &= \frac{\partial}{\partial \nu} (\Psi_{net}^+) + \gamma \log \nu + \gamma \log \gamma - \gamma \log(1 - \nu\gamma) \\ &< \log x_\nu + \gamma \log \nu + \frac{1}{2} \log \gamma + \frac{1}{2} \log(2e) + 8\gamma + \frac{1}{2} \gamma \log \gamma \\ &:= \Omega(\nu, \gamma), \end{aligned}$$

say. We will show that

$$M[\Omega](\delta, r_W^+(\delta)) < \left[ \frac{1}{2} \log \left( \frac{2e}{\tau} \right) - \frac{1}{\tau} \right], \quad 0 < \delta < \delta_0;$$

this implies (A.31). We first note that

$$\begin{aligned}
\frac{\partial}{\partial \gamma} \Omega(\nu, \gamma) &= \log \nu + \frac{17}{2} + \frac{1}{2} \log \gamma + \frac{1}{2} \gamma^{-1} \\
&\geq \log \delta + \frac{17}{2} + \frac{1}{2} \log(\gamma) + \frac{1}{2} \gamma^{-1} \\
\text{(A.33)} \quad &\geq \left[ \frac{\tau}{2} - 1 \right] \log(1/\delta) - \frac{1}{2} \log \log(1/\delta) + \frac{17}{2} - \frac{1}{2} \log(\tau)
\end{aligned}$$

which for any  $\tau > 2$  becomes arbitrarily large as  $\delta$  approaches zero. As a result,  $\Omega(\nu, \gamma)$  obtains its maximum where  $\gamma$  is largest within the admissible domain, i.e. at  $\gamma = r_W^+(\delta)$ . To find the overall maximum, we now examine the  $\nu$  direction along  $\gamma = r_W^+(\delta)$ :

$$\text{(A.34)} \quad \frac{\partial}{\partial \nu} \Omega(\nu, \gamma) = \frac{\frac{\partial}{\partial \nu} x_\nu}{x_\nu} + \frac{\gamma}{\nu} = \nu^{-1} \left[ \frac{1}{\tau \log 1/\delta} - \frac{1}{1 + 2x_\nu^2 - \nu} \right].$$

From Lemma 4.2 it follows that for any  $\tau > 2$ , for  $\delta$  sufficiently small,

$$\tau \log(1/\delta) > 2 \log(1/\delta) > 2x_\nu^2 + 1 - \delta > 2x_\nu^2 + 1 - \nu$$

for  $\nu \in [\delta, 1)$ . As a result (A.34) is negative for  $\delta$  sufficiently small, indicating that the maximum of  $\Omega(\nu, \gamma)$  over the domain of interest is obtained at  $(\delta, r_W^+(\delta))$ . Moreover,

$$\Omega(\nu, \gamma) \leq \Omega(\delta, \rho) < \frac{1}{2} \log \left( \frac{2e}{\tau} \right) - \frac{1}{\tau} + \mathcal{O} \left( \frac{\log \log z^+}{\log z^+} \right), \quad \delta \rightarrow 0,$$

giving (A.31). □

**A.6. Proof of Lemma 4.7.** From (A.30) and

$$\frac{\partial}{\partial \gamma} \Psi_{face}^+(\nu, \gamma) = -\nu [\log \gamma + \log \nu - \log(1 - \nu\gamma)]$$

we have the lower bound,

$$\begin{aligned}
\frac{\partial}{\partial \gamma} (\Psi_{net}^+ - \Psi_{face}^+)(\nu, \gamma) &> \nu \left[ \frac{1}{2} \gamma^{-1} + \log \nu + \frac{1}{2} \log(1/\gamma) - 13 - 14\gamma \right] \\
&> \delta \left[ \left( \frac{\tau}{2} - 1 \right) \log(1/\delta) + \frac{1}{2} \log \log(1/\delta) - 13 \right. \\
&\quad \left. + \frac{1}{2} \log \tau - 14[\tau \log(1/\delta)]^{-1} \right];
\end{aligned}$$

with the last inequality due to  $\gamma \leq r_W^+(\delta)$ . For any  $\tau > 2$  the above bound is positive for  $\delta$  sufficiently small. □

## REFERENCES

1. Fernando Affentranger and Rolf Schneider, *Random projections of regular simplices*, Discrete Comput. Geom. **7** (1992), no. 3, 219–226. MR1149653 (92k:52008)
2. Norman Bleistein and Richard A Handelsman, *Asymptotic expansions of integrals*, Dover, New York, 1986. MR863284 (89d:41049)
3. Károly Böröczky, Jr. and Martin Henk, *Random projections of regular polytopes*, Arch. Math. (Basel) **73** (1999), no. 6, 465–473. MR1725183 (2001b:52004)
4. Emmanuel J. Candès, Justin Romberg, and Terence Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory **52** (2006), no. 2, 489–509. MR2236170 (2007e:94020)

5. Emmanuel J. Candès and Terence Tao, *Decoding via linear programming*, IEEE Trans. Inform. Theory **51** (2005), no. 12, 4203–4215. MR2243152 (2007b:94313)
6. ———, *Near optimal signal recovery from random projections and universal encoding strategies*, IEEE. Trans. Info. Thry. **52** (2006), 5406–5425. MR2300700
7. E. W. Cheney, *Introduction to approximation theory*, Chelsea, New York, 1982. MR1656150 (99f:41001)
8. David L. Donoho, *Compressed sensing*, IEEE. Trans. Info. Thry. **52** (2006), no. 4, 1289–1306. MR2241189 (2007e:94013)
9. ———, *For most large systems of underdetermined equations, the minimum  $\ell^1$ -norm solution is the sparsest solution*, Comm. Pure Appl. Math. **59** (2006), no. 7, 907–934. MR2222440 (2006m:65079)
10. ———, *High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension*, Disc. Comput. Geometry **35** (2006), no. 4, 617–652. MR2225676 (2007h:52012)
11. ———, *Neighborly polytopes and sparse solutions of underdetermined linear equations*, Technical report Stanford University, Department of Statistics, # 2005-04 (2005).
12. David L. Donoho and Jared Tanner, *Neighborliness of randomly-projected simplices in high dimensions*, Proc. Natl. Acad. Sci. USA **102** (2005), no. 27, 9452–9457. MR2168716
13. ———, *Sparse nonnegative solutions of underdetermined linear equations by linear programming*, Proc. Natl. Acad. Sci. USA **102** (2005), no. 27, 9446–9451. MR2168715 (2006d:90089)
14. ———, *Finite dimensional phase transition for the expected number of randomly-projected simplices and cross-polytopes*, preprint (2007).
15. M. F. Duarte, M.B. Wakin, and R.G. Baraniuk, *Fast reconstruction of piecewise smooth signals from random projections*, Proceedings SPARS 05, Rennes, France, 2005.
16. E.R. Berlekamp, R.J. McEliece, and H.C.A. van Tilborg, *On the inherent intractability of certain coding problems*, IEEE Trans. Info. Thry. **24** (1978), 384–386. MR0495180 (58:13912)
17. David Gale, *Neighboring vertices on a convex polyhedron*, Linear inequalities and related system, Annals of Mathematics Studies, no. 38, Princeton University Press, Princeton, N.J., 1956, pp. 255–263. MR0085552
18. ———, *Neighborly and cyclic polytopes*, Proc. Sympos. Pure Math., Vol. VII, Amer. Math. Soc., Providence, R.I., 1963, pp. 225–232. MR0152944 (27:2915)
19. Branko Grünbaum, *Convex polytopes*, second ed., Graduate Texts in Mathematics, vol. 221, Springer-Verlag, New York, 2003, prepared and with a preface by Volker Kaibel, Victor Klee and Günter M. Ziegler. MR1976856
20. P. Hall, J.S. Marron, and A. Neeman, *Geometric representation of high dimensional low sample size data*, J. Roy. Stat. Soc. B **67** (2005), 427–444. MR2155347
21. Jarvis Haupt and Robert A. Nowak, *Signal reconstruction from noisy randomized projections with applications to wireless sensing*, Tech. report, Electrical Engineering, University of Wisconsin, 2005.
22. J. Hoffman-Jorgensen, J.A. Wellner, and M.B. Marcus, *High-dimensional probability III*, Birkhauser, Boston, 2004. MR2033877 (2004j:60005)
23. Irene Hueter, *Limit theorems for the convex hull of random points in higher dimensions*, Trans. Amer. Math. Soc. **351** (1999), no. 11, 4337–4363. MR1670156 (2000a:52008)
24. M. Kendall, A. Stuart, and J.K. Ord, *Kendall’s advanced theory of statistics*, Edward Arnold, London, 1991.
25. Nathan Linial and Isabella Novik, *How neighborly can a centrally symmetric polytope be?*, Disc. Comput. Geometry **36** (2006), no. 2, 273–281. MR2252105 (2007f:52027)
26. Jiri Matousek, *Lectures on discrete geometry*, Graduate Texts in Mathematics, vol. 212, Springer-Verlag, New York, 2002. MR1899299 (2003f:52011)
27. Peter McMullen and Geoffrey C. Shephard, *Diagrams for centrally symmetric polytopes*, Mathematika **15** (1968), 123–138. MR0238180 (38:6456)
28. Harold Ruben, *On the geometrical moments of skew-regular simplices in hyperspherical space, with some applications in geometry and mathematical statistics*, Acta Math. **103** (1960), 1–23. MR0121713 (22:12447)
29. M. Rudelson and R. Vershynin, *Geometric approach to error-correcting codes and reconstruction of signals*, Tech. report, Department of Mathematics, University of California, Davis, 2005. MR2206919 (2006j:94116)

30. ———, *Sparse reconstruction by convex relaxation: Fourier and gaussian measurements*, Proceedings Conference on Information Science and Systems (CISS) 2006, Princeton University, 2006.
31. Rolf Schneider, *Neighbourliness of centrally symmetric polytopes in high dimensions*, *Mathematika* **22** (1975), no. 2, 176–181. MR0405244 (53:9038)
32. J.A. Tropp and Anna Gilbert, *Signal recovery from partial information by orthogonal matching pursuit*, Tech. report, Mathematics Department, University of Michigan, 2005.
33. Y. Tsaig and D.L. Donoho, *Extensions of compressed sensing*, *EURASIP Journal of Applied Signal Processing* **86** (2006), no. 3, 549–571.
34. A. M. Vershik and P. V. Sporyshev, *Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem*, *Selecta Math. Soviet.* **11** (1992), no. 2, 181–201. MR1166627 (93d:60017)

DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY, 390 SERRA MALL, SEQUOIA HALL, STANFORD, CALIFORNIA 94305

*E-mail address:* `donoho@stanford.edu`

DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY 390 SERRA MALL, SEQUOIA HALL, STANFORD, CALIFORNIA 94305

*E-mail address:* `tanner@math.utah.edu`