

# EXPONENTIAL BOUNDS IMPLYING CONSTRUCTION OF COMPRESSED SENSING MATRICES, ERROR-CORRECTING CODES AND NEIGHORLY POLYTOPES BY RANDOM SAMPLING

DAVID L. DONOHO AND JARED TANNER

ABSTRACT. In [12] the authors proved an asymptotic *sampling theorem for sparse signals*, showing that  $n$  random measurements permit to reconstruct an  $N$ -vector having  $k$  nonzeros provided

$$n > 2 \cdot k \cdot \log(N/n)(1 + o(1));$$

reconstruction uses  $\ell_1$  minimization. They also proved an *asymptotic rate theorem*, showing existence of real error-correcting codes for messages of length  $N$  which can correct all possible  $k$ -element error patterns using just  $n$  generalized checksum bits, where

$$n > 2e \cdot k \log(N/n)(1 + o(1));$$

decoding uses  $\ell_1$  minimization. Both results require an asymptotic framework, with  $N$  growing large. For applications, on the other hand, we are concerned with specific triples  $k, n, N$ .

We exhibit triples  $(k, n, N)$  for which Compressed Sensing Matrices and Real Error-Correcting Codes surely exist and can be obtained with high probability by random sampling. These derive from exponential bounds on the probability of drawing ‘bad’ matrices. The bounds give conditions effective at finite- $N$ , and converging to the known sharp asymptotic conditions for large  $N$ . Compared to other finite- $N$  bounds known to us, they are much stronger, and much more explicit.

Our bounds derive from asymptotics in [12] counting the expected number of  $k$ -dimensional faces of the randomly projected simplex  $T^{N-1}$  and cross-polytope  $C^N$ . We develop here finite- $N$  bounds on the expected discrepancy between the number of  $k$ -faces of the projected polytope  $AQ$  and its generator  $Q$ , for  $Q = T^{N-1}$  and  $C^N$ .

Our bounds also imply existence of interesting geometric objects. Thus, we exhibit triples  $(k, n, N)$  for which polytopes with  $2N$  vertices can be centrally  $k$ -neighborly.

**Keywords.**  $\ell_1$ -minimization, compressed sensing, error-correction by linear programming, neighborly polytopes, random matrices

---

*Date:* July 2008.

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

DLT acknowledges partial support from NSF DMS 05-05303 and a Rothschild Visiting Professorship at the University of Cambridge.

JT acknowledges support from the Alfred P. Sloan Foundation and thanks John E. and Marva M. Warnock for their generous support in the form of an endowed chair.

$N$	Positivity-Constrained				Unconstrained			
	$n$	$k$	$\epsilon$	$\theta$	$n$	$k$	$\epsilon$	$\theta$
$10^4$	3,529	1,253	$10^{-3}$	1/5	4,299	1,208	$10^{-3}$	1/5
$10^6$	30,510	4,472	$10^{-3}$	1/10	31,323	3860	$10^{-3}$	1/10
$10^6$	35,766	5,487	$10^{-10}$	1/10	36,819	4,722	$10^{-10}$	1/10
$10^9$	1,355,580	113,004	$10^{-10}$	1/50	1,365,079	102,646	$10^{-10}$	1/50

TABLE 1. For the specified  $(k, n, N)$  and  $\epsilon$ , the probability of successfully recovering a  $k$ -sparse vector  $x_0 \in \mathbb{R}^N$  from  $n$  samples exceeds  $1 - \epsilon$ .  $\theta$  is a parameter of our method measuring proximity at  $(k, n, N)$  to asymptotic large  $N$ , relationships  $(n \cdot \rho(n/N; Q), n, N)$ ; see Section 2.

## 1. INTRODUCTION

This paper constructs solutions to three apparently different existence problems in geometry, coding theory, and sampling theory. In each problem, for a given triple of integers  $(k, n, N)$  where  $0 < k < n < N$ , we would like to know whether a solution exists for the given  $(k, n, N)$ ; and if so, we would like to construct it.

**1.1. Existence of Compressed Sensing Matrices.** Ordinarily, to reconstruct a vector in  $\mathbb{R}^N$  requires  $N$  measurements. The phrase ‘Compressed Sensing’ refers to the possibility of making fewer than  $N$  measurements, by taking nonstandard measurements and using nonlinear reconstruction to exploit side information about the sparsity of the vector [7, 3].

Say that a vector is  $k$ -sparse if it has only  $k$  nonzeros in some fixed, known basis – without loss of generality, the standard basis. Consider some fixed but unknown  $k$ -sparse vector  $x_0$  in  $\mathbb{R}^N$ . Consider making  $n$  measurements  $b_1, \dots, b_n$ , each one a random linear combination  $b_i = \sum_j a_{ij} x_0(j)$  of entries in  $x_0$ , where the  $a_{ij}$  are iid standard Normal random variables. Equivalently, we may write  $b = Ax_0$ . Now attempt to reconstruct the unknown  $x_0$  by solving

$$(1.1) \quad \min \|x\|_{\ell^1} \quad \text{subject to} \quad Ax = b.$$

Here  $n < N$ , so the linear system is underdetermined. Nevertheless, it turns out that if  $k$  is sufficiently small, this will work with high probability.

An important variation concerns a *nonnegative*  $k$ -sparse vector  $x_0$  in  $\mathbb{R}_+^N$  – again fixed but unknown. We again consider making  $n$  measurements  $b_1, \dots, b_n$ , each one a random linear combination  $b_i = \sum_j a_{ij} x_0(j)$  of entries in  $x_0$ . We attempt to recover  $x_0$  by solving the linear program

$$(1.2) \quad \min 1'x \quad \text{subject to} \quad Ax = b, \quad x \geq 0.$$

In either situation, we are interested in knowing for which  $k$ ,  $n$  and  $N$  the approach can be successful. Since  $A$  is a random matrix, there is also a parameter  $\epsilon$  controlling the probability that reconstruction is successful. As an example of the information our method will generate, see Table 1. It gives some examples of triples for which reconstruction can be successful with high probability.

**1.2. Encoding Matrices for Real Error-Correcting Codes.** Consider a stylized problem of data transmission with immunity to occasional transmission errors. Suppose that we transmit a vector of  $N$  numbers but that  $k$  of these numbers will

$N$	$m = N - n$	$n$	$k$ : Almost All	$k$ : All
5000	4500	500	23	21
5000	4000	1000	117	56
$10^4$	9000	1000	86	45
$10^4$	8000	2000	304	117
$10^4$	5000	5000	1505	441
$10^5$	$9 \times 10^4$	$10^4$	1543	484
$10^5$	$8 \times 10^4$	$2 \times 10^4$	4246	1202

TABLE 2. Messages of  $m$  real numbers are encoded as blocks of length  $N > m$  using the described random encoder-decoder pair. *All*: with probability at least 50%, we draw an encoder-decoder pair that can correct *all* patterns of  $k$  or fewer errors. *Almost All*: We draw an encoder decoder pair that can correct *the vast majority* of patterns of  $k$  or fewer errors; expected error rate one in  $10^{15}$ .

be received with error. A standard strategy encodes  $m = N - n < N$  message numbers as a redundant block of  $N$  numbers, intending that coding redundancy will help in suppressing the transmission errors. (Note: we work over the field  $\mathbf{R}$  of real numbers, not over some discrete alphabet; the relevance to digital settings will be discussed below).

Let  $B$  be an  $m \times N$  matrix with real-valued entries. Given a vector  $\alpha \in \mathbf{R}^m$  to be transmitted, encode it as  $\beta = B^T \alpha \in \mathbf{R}^N$  and transmit. Assume the receiver measures  $\mu = \beta + z$  where  $\mu \in \mathbf{R}^N$  and  $z$  represents transmission errors. It is assumed that  $z$  contains nonzeros in only  $k$  entries – most numbers in  $\mu$  are received without error. The receiver in some way decodes the  $N$  numbers, hoping to produce the  $m$  entries in  $\alpha$ .

Consider a simple encoding decoding scheme based on linear programming [5, 2, 17]. Construct a generalized checksum matrix  $A$  (satisfying  $AB^T = 0$ ) that is  $n \times N$ , with  $m + n = N$ . Given the received data  $\mu$ , form the generalized checksum  $y = A\mu$ . Solve (1.1). The optimization result  $x_1$  is an estimate of  $z$ . Reconstruct by subtracting this estimate of the receiver error out of the received message, and projecting down from  $\mathbf{R}^N$  to  $\mathbf{R}^m$ :  $\alpha_1 = B(\mu - x_1)$ . As (1.1) is a standard convex optimization problem, this can be considered computationally tractable. How many errors can this scheme tolerate? This of course depends on  $A$  and  $B$ .

Letting  $A$  and  $B$  be respectively the first  $n$  rows and the last  $N - n$  rows of a random  $N \times N$  orthogonal matrix, we get a random encoder-decoder pair. For which  $k, n, N$  triples is this likely to be able to correct  $k$  errors? Qualitative results were given in [4, 2]; the present authors gave asymptotic results valid for large  $N$  in [12]. This paper gives results applicable at finite- $N$ ; see Table 2.

**1.3. Existence of Neighborly Polytopes.** A polytope  $P$  is the convex hull of a finite point set of  $N$  points in  $R^n$ . Suppose that  $P$  has  $N$  vertices. It is called *1-neighborly* if, for every pair of vertices, the line segment joining the pair does not meet the interior of  $P$ . More colorfully:  $P$  has no diagonals. Despite the counterintuitive nature of this situation –based on low-dimensional experience, we would expect there to be many such diagonals, but here there are none– one can sometimes find more impressive phenomena. The polytope is called *k-neighborly*

	Neighborly		Central Neighborly	
$N$	$n$	$k$	$n$	$k$
200	150	24	150	14
1000	500	61	500	40
$10^4$	1,000	56	1,000	45

TABLE 3. For the specified  $(k, n, N)$  there exist  $k$ -neighborly (resp. centrally  $k$ -neighborly) polytopes. They can be obtained with positive probability by sampling  $N$  points from the standard Normal in  $\mathbb{R}^n$  (resp. followed by symmetrization).

if the convex hull of every  $k + 1$  vertices does not meet the interior of  $P$ . Then, for some special polytopes  $P$  in high dimensions, every triple, quad, and quintuple span a face of  $P$ . In some sense, all the points and all their  $k$ -th order interactions are ‘on the boundary’ of such  $P$ .

Consider random vectors  $a_i$ ,  $i = 1, \dots, N$  drawn iid standard normal in  $\mathbf{R}^n$ . Consider the random polytope  $P = \text{conv}(\{a_i\})$  formed by taking the convex hull of the points of  $P$ . For what  $k$ ,  $n$ , and  $N$  can  $P$  be expected to be  $k$ -neighborly polytope?

A closely related notion is central neighborliness. Consider centrosymmetric polytopes, formed from  $N$  points  $a_i$  in  $R^n$  by symmetrization  $\{a_1, -a_1, a_2, -a_2, \dots, a_N, -a_N\}$  and then taking the convex hull. Such a centrally-symmetric polytope is called centrally  $k$ -neighborly if the convex hull of every  $k + 1$  vertices not including an antipodal pair does not meet the interior of  $P$ .

Consider random vectors  $a_i$ ,  $i = 1, \dots, N$  drawn iid standard normal in  $\mathbf{R}^n$ . For what values of  $k$  is there a positive probability that the resulting convex hull is centrally  $k$ -neighborly?

For specific examples, see Table (3). For the  $(k, n, N)$  given, one can obtain  $k$ -neighborly (or centrally  $k$ -neighborly) polytopes by sampling  $N$  points from the standard Normal in  $\mathbb{R}^n$ .

**1.4. What Else Can be Said?** The three existence problems we have just discussed are closely related. A single method is used in subtly different ways to produce the various tables given above. The method can of course produce many other such tables; we make software available to do so.

Our method also provides rigorous memorable formulas that can take the place of tables. For example, we will see that random coder-decoder pairs with block-length  $N$  and  $n$  generalized checksums are able to correct all patterns of  $k$  errors, if  $(k, n, N)$  obey<sup>1</sup>

$$(1.3) \quad n > 2e \cdot k \cdot \log(N/n) + \text{remainder};$$

here the remainder is easily characterized. We will also see that in the problem of compressed sensing, we can exactly reconstruct a  $k$ -sparse object in  $\mathbf{R}^N$  from  $n$  generalized samples provided

$$(1.4) \quad n > 2 \cdot k \log(N/n)(1 + \text{remainder}),$$

where again the remainder is explicitly characterized.

<sup>1</sup>All logarithms throughout this article are “natural base- $e$ ” logarithms.

Such statements offer simple, appealing conclusions. They might be helpful to engineers designing a compressed sensing system; indeed ‘ $n > 2k \log(N/n)$ ’ might be called the ‘sampling theorem for sparse vectors’; Section 4 contains further discussions of this interpretation.

**1.5. The Paper.** Of the three applications presented so far – compressed sensing, error-correcting codes, and neighborly polytopes – we view the polytope setting as the fundamental one; the one in which all the questions can be most precisely stated and their answers most easily obtained.

Most of our paper focuses on counting the number of  $k$ -dimensional faces of certain random polytopes with  $N$  vertices in  $\mathbb{R}^n$ . It turns out that the above applications all flow from making calculations of the expected number of faces of random polytopes.

In the remainder of the introduction, we describe existing knowledge about the counting the faces of random polytopes in an asymptotic, large  $N$  setting, and about the connection to the applications mentioned above.

In Section 2 we state our main results, which give finite- $N$  bounds on the expected number of faces of random polytopes; Section 3 develops the applications discussed earlier and explains how the tables presented above were calculated. Section 4 develops a precise, finite- $N$  sampling theorem. Sections 5-7 provide proofs of our main results.

**1.6. Counting Faces of Polytopes.** Let  $T^{N-1}$  denote the standard unit simplex in  $\mathbb{R}^N$ , and  $C^N$  the standard cross-polytope in  $\mathbb{R}^N$ :

$$T^{N-1} := \left\{ x \in \mathbf{R}^N \mid \sum_{i=1}^n x_i = 1, \quad x_i \geq 0 \right\},$$

$$C^N := \left\{ x \in \mathbf{R}^N \mid \sum_{i=1}^n |x_i| \leq 1, \right\}.$$

Let  $A$  be an  $n \times N$  matrix with  $n < N$ ; the image polytopes  $AT^{N-1}$  and  $AC^N$  lie in  $\mathbb{R}^n$ . Quite general polytopes can be constructed in this way; for  $n$  fixed and  $N$  varying, any polytope can arise as some  $AT^{N-1}$  for appropriate  $N$  and  $A$ , and any centrally-symmetric polytope as some  $AC^N$ .

Earlier papers by the authors [8, 10, 12] studied random matrices  $A$  and observed very different behavior when  $n$  and  $N$  are both large—behavior with several surprising implications. For  $n \times N$  matrices  $A$  with iid standard Normal entries they considered the number of  $k$ -faces of the image polytope  $AQ$ ,  $f_k(AQ)$ . They considered sequences of triples  $(k, n, N)$  indexed by  $n$  such that, for fixed fractions  $\delta, \rho \in (0, 1)$ ,

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

the so-called proportional-dimensional asymptotic, and considered the limiting behavior as  $n \rightarrow \infty$ . They identified curves  $\rho_W(\cdot; Q)$  marking an abrupt change in behavior of the expected fraction of faces

$$(1.6) \quad \lim_{n \rightarrow \infty} \frac{\mathcal{E} f_k(AQ)}{f_k(Q)} \quad \begin{cases} = 1 & \rho < \rho_W(\delta; Q), \\ < 1 & \rho > \rho_W(\delta; Q). \end{cases}$$

Here and below, we follow the convention from [10] that a display like (1.6) containing  $Q$  stands for two displays, one in which  $Q = T^{N-1}$  and one where  $Q = C^N$ .

Note that the functions  $\rho(\delta; Q)$  depend on  $N$  only through the limit  $n/N \rightarrow \delta$ . Here the limit is taken along sequences  $(k_n, n, N_n)$  as  $n \rightarrow \infty$ . Evidently, below the curve there is *weak* agreement between  $f_k(AQ)$  and  $f_k(Q)$ , while above the curve differences are noticeable. They also identified curves  $\rho_S(\cdot; Q)$  such that

$$(1.7) \quad \lim_{n \rightarrow \infty} \text{Prob}\{f_k(AQ) = f_k(Q)\} = \begin{cases} 1 & \rho < \rho_S(\delta; Q), \\ 0 & \rho > \rho_S(\delta; Q); \end{cases}$$

this curve marks *strong* agreement between  $f_k(AQ)$  and  $f_k(Q)$ . Informally, although for general  $A$ , the face count  $f_k(AQ)$  is nearly arbitrary, for random  $A$ , each  $f_k(AQ)$  is exactly or approximately  $f_k(Q)$ , for all low dimensions  $k$  running from 0 up to a threshold approximately of the form  $\rho(\delta; Q) \cdot n$ . These four threshold curves are depicted in Figure 1.1.

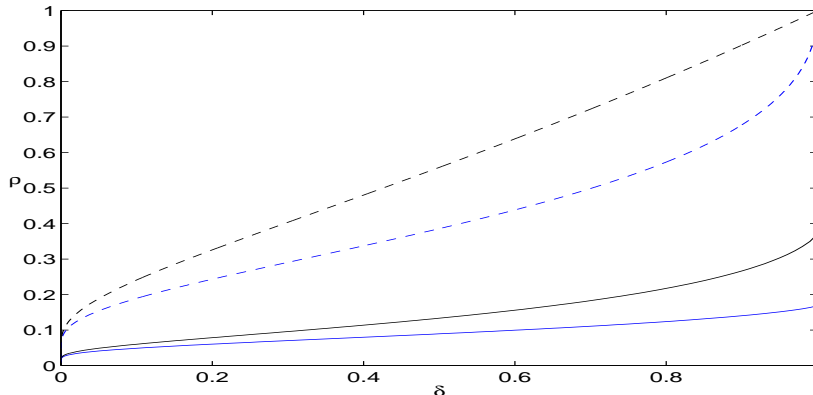


FIGURE 1.1. The phase transition thresholds, from bottom to top,  $\rho_S(\delta; C)$  solid-blue,  $\rho_S(\delta; T)$  solid-black,  $\rho_W(\delta; C)$  dash-blue and  $\rho_W(\delta; T)$  dash-black.

These phenomena imply various existence results connected to Sections 1.1-1.3.

- The relation  $f_k(AC^N) = f_k(C^N)$  implies that the polytope  $P = AC^N$  is centrally  $k$ -neighborly. It follows from (1.7) that for many combinations  $(k, n, N)$  where  $k/n < \rho_S(n/N; C)$ , the randomly-projected polytope will be  $k$ -centrally neighborly, providing a vast supply of such polytopes.
- The relation  $f_k(AC^N) > .99f_k(C^N)$  implies that of all systems  $b = Ax$  having a solution with at most  $k$  non-zeros, for more than 99% of the possible sign patterns in  $x$ , the minimal  $\ell_1$ -norm solution is also the sparsest solution. It follows from (1.6) that for many combinations  $(k, n, N)$  where  $k/n < \rho_W(n/N; C)$ , the random matrix  $A$  will have this property, providing a vast supply of matrices useful for “compressed sensing”, see Section 3.1.
- The relation  $f_k(AC^N) = f_k(C^N)$  implies that  $A$  is a perfect checksum matrix: a linear programming decoder can use it to decode all patterns of  $k$  or fewer errors. Hence for many combinations  $(k, n, N)$  where  $k/n < \rho_S(n/N; C)$ , such checksum matrices are prevalent. The relation  $f_k(AC^N) > (1 - 10^{-12})f_k(C^N)$  implies that the checksum matrix  $A$  is nearly perfect: able to decode almost all patterns of  $k$  or fewer errors – failing once in  $10^{12}$  such patterns. Hence many combinations  $(k, n, N)$  where

$k/n < \rho_W(n/N; Q)$ , offer prevalence of nearly perfect checksum matrices with error rates better than one part in  $10^{12}$ .

Such *asymptotic* results however, are not finite- $N$  results: they don't say what holds for *specific* combinations of  $(k, n, N)$ . The Tables presented in Sections 1.1 - 1.3 require stronger tools.

This paper repackages the inequalities and asymptotic statements developed in our earlier papers [6, 11, 12] and develops exponential bounds on certain probabilities and expectations which are shown to be effective already at moderate values of  $(k, n, N)$ . Moreover, these bounds involve the asymptotic thresholds  $\rho(\cdot)$  in an explicit way, even in this finite- $n$  case. Thus, for a fraction  $\theta \in (0, 1)$ , triples where  $k/n = (1 - \theta) \cdot \rho(n/N)$  are subject to inequalities depending explicitly on the sizes of  $\theta$  and  $n/N$ . In consequence there is a clearly defined family of curves indexed by problem size  $N$ , converging to the asymptotic threshold curves  $\rho(\cdot)$  with increasing  $n$ , and implying lower bounds on the probability of sampling "good" matrices/polytopes for triples along the curve, at a given  $N$ .

These curves are significantly higher than those available by any other finite- $n$  bounds known to us, and they converge to the right asymptotic limit at the right rate. As we show, they allow practical answers to problems of reasonable size that might be of interest in signal processing or other applied fields.

A different asymptotic studied by the authors in [12] involved sequences of triples  $(k_n, n, N_n)$  where  $N_n$  is much larger than  $n$  (but not exponentially larger) so that the projection lowers dimension drastically. For a random  $n \times N_n$  matrix  $A$  with entries are drawn iid from the standard Gaussian, it was found that eventually  $f_k(AQ) = f_k(Q)$  provided

$$(1.8) \quad n > 2e \cdot k \cdot \log(N/n)(1 + o(1)),$$

with an appropriate  $o(1)$  term.<sup>2</sup> Moreover, it was found that  $f_k(AQ) = f_k(Q)(1 + o(1))$  provided

$$(1.9) \quad n > 2 \cdot k \log(N/n)(1 + o(1)),$$

with appropriate  $o(1)$  terms. Such *asymptotic* results have greater force if known to hold for *specific* combinations of  $(k, n, N)$ . In Section 4, we develop inequalities valid for finite  $(k, n, N)$  having the same leading terms but replacing the  $o(1)$  terms by simple and explicit remainders.

## 2. MAIN RESULTS

**2.1. Absolute Agreement of Face Numbers.** In [8, 10] the authors established bounds on the absolute *face deficit*  $f_k(Q) - \mathcal{E}f_k(AQ)$  for  $A$  with entries drawn iid Gaussian and for  $Q = C^N$  and  $T^{N-1}$ . They defined functions  $\Psi_{net}(\delta, \rho; Q)$  affording the inequalities

$$(2.1) \quad \pi_l(N; Q)e^{N\Psi_{net}(n/N, k/n; Q)} < f_k(Q) - \mathcal{E}f_k(AQ) < \pi_u(N; Q)e^{N\Psi_{net}(n/N, k/n; Q)},$$

where  $\pi_l(N; Q)$  and  $\pi_u(N; Q)$  are known polynomial functions in  $N$ .

In display (2.1), the exponents  $\Psi_{net}$  are overwhelmingly more important for large  $N$  than the polynomial terms. For any triple  $(k, n, N)$  set  $\rho = k/n$  and  $\delta = n/N$ . This pair of coordinates  $(\delta, \rho)$  drives all of the asymptotic large  $N$  behaviour of  $\mathcal{E}f_k(AQ)$ . In the portion of the  $(\delta, \rho)$ -plane where  $\Psi_{net}(\delta, \rho; Q) < 0$ ,

<sup>2</sup> $o(1)$  indicates a term tending to zero as  $n \rightarrow \infty$ .

the face deficit tends to zero with increasing  $N$ , and in the portion of the  $(\delta, \rho)$ -plane where  $\Psi_{net}(\delta, \rho; Q) > 0$ , the face deficit grows with  $N$ . The level curve  $\{\Psi_{net}(\delta, \rho; Q) = 0\}$  is the boundary between these two “phases”. As it turns out, for each fixed  $\delta \in (0, 1)$ ,  $\Psi_{net}(\delta, \rho; Q)$  is monotone increasing in  $\rho$ , and has a unique zero crossing; call this  $\rho_S(\delta; Q)$ . The two curves  $\rho_S(\delta; T)$  and  $\rho_S(\delta; C)$  are displayed in Figure 1.1; they are both smooth and monotone increasing. “Below”  $\rho_S(\delta; Q)$ , the face deficit  $f_k(Q) - \mathcal{E}f_k(AQ)$  tends to zero exponentially fast in  $N$ ; “above”  $\rho_S(\delta; Q)$ , the face deficit grows exponentially fast. Of course, the notion of “exponentially fast” is qualitative. We now supply an effective quantitative notion, valid for specific triples  $(k, n, N)$ .

**Theorem 2.1** (Exponential Bounds for Face Deficit). *Let  $\Omega_S(\delta; Q)$  denote the function defined in (5.5) below and displayed in Figure 2.1, panel (a). Let  $c_S(T) := \frac{15}{32\pi^{3/2}}$  and  $c_S(C) := \frac{25}{16\pi^2}$ . Let  $A$  be a random matrix with iid standard Normal entries, and  $f_k(Q)$  denote the number of  $k$ -dimensional faces of polytope  $Q$ . Consider a triple  $(k, n, N)$  obeying  $k/n \leq (1 - \theta)\rho_S(\delta; Q)$ , where  $\theta \in (0, 1)$  and  $\delta = n/N$ . Then*

$$(2.2) \quad f_k(Q) - \mathcal{E}f_k(AQ) < c_S(Q) \cdot N^2(N + 2)^3 \exp(-n\theta\Omega_S(\delta; Q)).$$

In this inequality, the driving factor is  $n\theta$ , with  $\theta$  the fractional distance from the asymptotic phase transition  $\rho_S(\delta; Q)$ .

The exponent function  $\Omega_S(\delta; Q)$  can be seen in Figure 2.1 Panel (a) to be strictly positive, in fact always at least  $1/2$ ; see subsection 2.3.

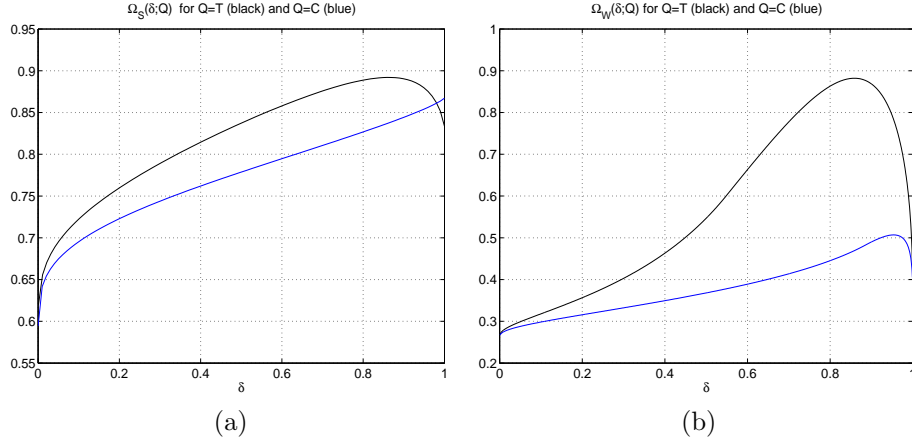


FIGURE 2.1. Panel (a)  $\Omega_S(\delta; Q)$ ; panel (b)  $\Omega_W(\delta; Q)$  for  $Q = T$  (black) and  $Q = C$  (blue).

**2.2. Relative Agreement of Face Numbers.** Also in [8, 10, 12] the authors established bounds on the *fractional face deficit*  $(f_k(Q) - \mathcal{E}f_k(AQ))/f_k(Q)$  for  $Q = T^{N-1}$  and  $C^N$ . They defined functions  $\Psi_{face}(\delta, \rho; Q)$  giving the inequalities

$$(2.3) \quad \begin{aligned} \pi_l(N; Q) \max_{\nu \in (0, k/n]} e^{N(\Psi_{net} - \Psi_{face})(n/N, \nu; Q)} &< \frac{f_k(Q) - \mathcal{E}f_k(AQ)}{f_k(Q)} \\ &< \pi_u(N; Q) \max_{\nu \in (0, k/n]} e^{N(\Psi_{net} - \Psi_{face})(n/N, \nu; Q)}, \end{aligned}$$



where  $\pi_l(N; Q)$  and  $\pi_u(N; Q)$  are known polynomials (different from those in (2.1)).

Again, the exponents are overwhelmingly more important for large  $N$  than the polynomial terms. Because  $f_k(AQ)$  must be less than  $f_k(Q)$ , it follows that  $(\Psi_{net} - \Psi_{face})(\delta, \rho; Q) \leq 0$  everywhere in the  $(\delta, \rho)$  parameter space for each  $Q = T$  and  $Q = C$ , and that each function  $\Psi_{net} - \Psi_{face}$  reaches its maximum only along the corresponding curve  $\rho_W(\delta; Q)$ , also displayed in Figure 1.1. Controlling the *fractional* face deficit is a weaker notion than controlling the absolute face deficit; we call  $\rho_W(\delta; Q)$  the *weak phase transition*. Consider parameters  $(\delta, \rho)$  falling “below” the curve  $\rho_W(\delta; Q)$  and a sequence of triples  $(k, n, N)$ , the fractional face deficit decays to zero exponentially fast. We now quantify this effect.

**Theorem 2.2** (Exponential Bounds on Fractional Face Deficit). *Let  $\Omega_W(\delta; Q)$  be the function defined in (6.2) below and displayed in Figure 2.1, panel (b). Let  $c_W(T) = \frac{375\sqrt{2}}{512\pi}$  and  $c_W(C) = \frac{625\sqrt{2}}{512\pi^{3/2}}$ . Let  $A$  be a random matrix with iid standard Normal entries, and  $f_k(Q)$  denote the number of  $k$ -dimensional faces of polytope  $Q$ . Consider a triple  $(k, n, N)$  where  $k/n \leq (1 - \theta)\rho_W(n/N; Q)$  for  $\theta \in (0, 1)$ . Set  $\delta = n/N$ . Then*

$$(2.4) \quad \frac{f_k(Q) - \mathcal{E}f_k(AQ)}{f_k(Q)} < c_W(Q) \cdot N^3(N+2)^3 \exp(-n\theta^2\Omega_W(\delta; Q)).$$

In this inequality, the driving factor is  $n\theta^2$ ; again,  $\theta$  is the fractional distance below phase transition.

The exponent function  $\Omega_W(\delta; Q)$  can be seen in Figure 2.1 Panel (b) to be strictly positive, in fact always at least  $1/4$ ; see subsection 2.3.

**2.3. Convenient Simplifications.** Theorems 2.1 and 2.2 give the strongest known finite- $N$  bounds on the expected discrepancy between the number of  $k$ -faces of the polytopes in question and the corresponding randomly-projected polytope. However, because they involve rate exponents  $\Omega_S(\delta; Q)$  and  $\Omega_W(\delta; Q)$  they are not completely explicit and transparent. One can develop simplified results which are slightly weaker, but more memorable. These follow from numerical evaluation of the rate exponents which are consequently framed as ‘Findings’ rather than as theorems or lemmas.

We first develop a simplified form of (2.2), in which  $\Omega_S(\delta; Q)$  does not appear.

*Finding 2.1* (Lower bound of  $\Omega_S(\delta; Q)$ ).  $\Omega_S(\delta; Q) \geq 1/2$  for all  $\delta \in [0, 1]$ .

*Demonstration of Finding 2.1.* Lemma 5.2 shows that the lower bound  $1/2$  is approached from above as  $\delta \rightarrow 0$ . In the limit as  $\delta \rightarrow 1$ , all factors comprising  $\Omega_S(\delta; Q)$  approach finite values, and those values, numerically evaluated, imply the corresponding limit of  $\Omega_S(\delta; Q)$  is greater than  $1/2$ , see Figure 2.1 Panel (a). For moderate values of  $\delta$ ,  $\Omega_S(\delta; Q)$  is again observed in Figure 2.1 Panel (a) to be greater than  $1/2$ .  $\square$

*Corollary 2.1* (Exponential Bounds Uniform in  $\theta$ ). Under the same conditions and notation as in Theorem 2.1, assuming the validity of Finding 2.1. Then

$$(2.5) \quad f_k(Q) - \mathcal{E}f_k(AQ) < c_S(Q) \cdot N^2(N+2)^3 \exp(-n\theta/2).$$

This finding shows that, everywhere on or below the curves  $(1 - \theta)\rho_S^*(\delta)$ , the exponential term in (2.2) is bounded by  $\exp(-n\theta/2)$ ; this follows from Theorem 2.1 and Finding 2.1.

Clearly, the face deficit bounds (2.2) and (2.5) become effective at moderately large  $N$ . As a simple example, we have

*Corollary 2.2.* Let  $n_S(\epsilon; N_0, \theta_0, Q)$  denote the solution to the equation

$$\log \epsilon = \log c_S(Q) + 5 \log(N_0 + 2) - n\theta_0/2.$$

Assume the validity of Finding 2.1. Then for every triple  $(k, n, N_0)$  where  $k/n \leq (1 - \theta_0)\rho_S(n/N; Q)$ ,

$$n \geq n_S(\epsilon; N_0, \theta_0, Q) \quad \text{implies} \quad f_k(Q) - \mathcal{E}f_k(AQ) \leq \epsilon.$$

This simple result directly makes the point that our bounds become effective for many triples  $(k, n, N)$  where  $n$  and  $N$  are not very large. It is not best possible - Theorem 2.1 yields stronger bounds than Corollaries 2.1 and 2.2. More precise information and examples will be given in Section 3 below.

We now develop a simplified form of (2.4), in which  $\Omega_W(\delta; Q)$  does not appear.

*Finding 2.2* (Lower bound of  $\Omega_W(\delta; Q)$ ).  $\Omega_W(\delta; Q) \geq 1/4$  for all  $\delta \in [0, 1]$ .

*Demonstration of Finding 2.2.* Lemma 6.2 shows that the lower bound  $1/4$  is approached from above as  $\delta \rightarrow 0$ . In the limit as  $\delta \rightarrow 1$ , all factors comprising  $\Omega_W(\delta; Q)$  approach finite values, and those values, numerically evaluated, imply the corresponding limit of  $\Omega_W(\delta; Q)$  is greater than  $1/4$ , see Figure 2.1 Panel (b). For moderate values of  $\delta$ ,  $\Omega_W(\delta; Q)$  is again observed in Figure 2.1 Panel (b) to be greater than  $1/4$ .  $\square$

*Corollary 2.3* (Exponential Bounds uniform in  $\theta$ ). Under the same conditions and notation of Theorem 2.2, assume also the validity of Finding 2.2. Then

$$(2.6) \quad \frac{f_k(Q) - \mathcal{E}f_k(AQ)}{f_k(Q)} < c_W(Q) \cdot N^3(N+2)^3 \exp(-n\theta^2/4).$$

In Corollary 2.3, everywhere on or below the curves  $(1 - \theta)\rho_W(\delta; Q)$ , the exponential term in (2.4) is bounded by  $\exp(-n\theta^2/4)$ ; this follows from Theorem 2.2 and Finding 2.2. While the dependence on  $\theta^2$  is markedly different than in the earlier case, we still see that these fractional face deficit bounds become effective at moderately large  $N$ .

*Corollary 2.4.* Let  $n_W(\epsilon; N_0, \theta_0, Q)$  denote the value of  $n$  solving the equation

$$\log \epsilon = \log c_W(Q) + 6 \log(N_0 + 2) - n\theta_0^2/4.$$

Assume the validity of Finding 2.2. Then for every triple  $(k, n, N_0)$  where  $k/n \leq (1 - \theta_0)\rho_W(n/N_0; Q)$ ,

$$n \geq n_W(\epsilon; N_0, \theta_0, Q) \quad \text{implies} \quad \frac{f_k(Q) - \mathcal{E}f_k(AQ)}{f_k(Q)} \leq \epsilon.$$

### 3. COROLLARIES AND APPLICATIONS

The explicit bounds in Theorems 2.1 and 2.2 allow us to make non-asymptotic statements concerning specific triples  $(k, n, N)$ , producing the tables in introduction. We now review the meaning of those tables; it is convenient to do so in reversed order. We also provide figures summarizing a large range of other triples. We also describe the computational approach, the software which implements it, our philosophy towards reproducibility, and make comparisons to previous work.

**3.1. Existence of Compressed Sensing Matrices.** Table 1 arose from considering the following questions.

**Q1:** *Compressed Sensing Matrices.* A random matrix  $A$  will be called *acceptable for Compressed Sensing* at failure fraction  $\epsilon > 0$  if, among all  $k$ -sparse problem instances  $x_0$ , sensing  $b = Ax_0$  then solving (1.1) perfectly recovers  $x_0$  in at least a fraction  $(1 - \epsilon)$  of such instances. Can we set a threshold  $n \geq n_{CS}(k, N, \epsilon; C)$  implying that  $n \times N$  iid Gaussian matrices are acceptable for CS?

**Q2:** *Compressed Sensing with positivity.* A random matrix  $A$  will be called *acceptable for Compressed Sensing of positive objects* at failure probability  $\epsilon > 0$  if, among all *nonnegative*  $k$ -sparse problem instances  $x_0$ , sensing  $b = Ax_0$  then solving (1.2) perfectly recovers  $x_0$  in at least a fraction  $(1 - \epsilon)$  of such instances. Can we set a threshold  $n \geq n_{CS}(k, N, \epsilon; T)$  implying that  $n \times N$  iid Gaussian matrices are acceptable for CS of positive objects?

These questions can be answered for  $\epsilon = 0.01$  by establishing the bounds [6, 11]

$$\frac{\mathcal{E}f_k(AT^{N-1})}{f_k(T^{N-1})} \geq 0.99, \quad \text{for (Q2); or } \frac{\mathcal{E}f_k(AC^N)}{f_k(C^N)} \geq 0.99, \quad \text{for (Q1);}$$

for the given  $(k, n, N)$  of interest. Such inequalities can be developed using our bounds on the fractional face deficit. (Equivalence of the recovery of sparse vectors from linear measurements by solving (1.1)-(1.2) and the face counts of the convex hull of  $A$  and its associated cross-polytope were proven by the authors in [6, 11].)

Figures 3.1(c-d) provide graphical answers to Q1-Q2. They show the domain in the phase diagram in which, on average, 99% of faces survive the prescribed dimension reduction. Let  $Bound(Q, k, n, N)$  denote the right-hand side of (2.4) for the given values of  $k, n$ , and  $N$  and the specific binding  $Q$ . The Figures display the level sets  $Bound(Q, k, n, N) = 10^{-2}$  for  $N = 200, 1000$ , and  $5000$ .

*Corollary 3.1.* For a given  $N \in \{200, 1000, 5000\}$ , consider values of  $k$  and  $n$  such that  $(n/N, k/n)$  lies strictly beneath the curve for that  $N$  depicted in Figures 3.1(c). Fix a given face  $F$  of  $T^{N-1}$  independently of  $A$ . There is at least a 99% chance that  $AF$  is a face of  $AT^{N-1}$ . Again, we refer to  $A$  generated by random sampling from the standard Gaussian distribution.

Consider values of  $k$  and  $n$  such that  $(n/N, k/n)$  lies strictly beneath the curve for that  $N$  depicted in Figures 3.1(d). Fix a given face  $F$  of  $C^N$  independently of  $A$ . There is at least a 99% chance that  $AF$  is a face of  $AC^N$ . Here probability refers to random sampling from the standard Gaussian distribution on  $\mathbf{R}^{n \times N}$ .

For specific numerical values, see Table 1.

**3.2. Encoding Matrices for Error-Correcting Codes.** Consider again the encoding/decoding scheme described in Section 1.2. Let  $U$  be a random orthogonal matrix, drawn uniformly from the distribution on orthogonal  $N \times N$  matrices,  $O(N)$ , 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  $\mu$ , form the generalized checksum  $y = A\mu$ . Solve (1.1). Reconstruct by subtracting this estimate of the receiver error out of the received message, and projecting down from  $\mathbf{R}^N$  to  $\mathbf{R}^m$ :  $\alpha_1 = B(\mu - x_1)$ .

**Q3: Perfect ECC Matrices.** Call the pair  $(A, B)$  a *perfect*  $k$ -error correcting code ( $k$ -ECC) if the above encoder-decoder correctly decodes *every* received message corrupted by an error vector with  $k$  or fewer nonzeros. For particular values of  $k$ ,  $N$  and  $n$ , is the probability at least 50% that random sampling of  $U = (A, B)$  from the uniform distribution on  $O(N)$  yields a perfect ECC?

**Q4: Almost-Perfect ECC Matrices.** Call an *error pattern* a configuration of  $k$  nonzeros with specified signs. Say that the pair  $(A, B)$  is *almost perfect*, with failure rate  $\epsilon > 0$ , if the above encoder-decoder pair can correct all but a fraction  $\epsilon$  of  $k$ -error patterns. For a given failure rate  $\epsilon = 10^{-15}$ , at what specific values of  $k$ ,  $N$  and  $n$  do we expect an almost-perfect  $k$ -ECC code by random sampling  $U$ ?

These questions can be answered by establishing the bounds [6]

$$f_k(C^N) - \mathcal{E}f_k(AC^N) < 1/2, \text{ for (Q3) or } \frac{f_k(C^N) - \mathcal{E}f_k(AC^N)}{f_k(C^N)} < 10^{-15}, \text{ for (Q4)}$$

Exemplar values of  $(k, n, N)$ , with  $m = N - n$ , were given in Table 2.

*Corollary 3.2.* For a given  $N \in \{200, 1000, 5000\}$ , consider values of  $k$  and  $n$  such that  $(n/N, k/n)$  lies strictly beneath the curve corresponding to that  $N$  depicted in Figure 3.1(b). There exist perfect ECC matrix pairs for that  $(k, n, N)$ . They can be obtained with positive probability by a random draw from the uniform distribution on  $O(N)$ .

Consider values of  $k$  and  $n$  such that  $(n/N, k/N)$  lies strictly beneath the curve corresponding to that  $N$  depicted in Figure 3.1(d). There exist almost-perfect ECC matrix pairs for that  $(k, n, N)$ , at error rate  $\epsilon = 10^{-2}$ . They can be obtained with positive probability by a random draw from the uniform distribution on  $O(N)$ .  $\square$

Note that the ECC matrix pairs referred to in Corollary 3.2 have real-valued rather than integer-valued entries. Empirically, various ensembles of random matrices with rational entries behave similarly; compare [13]. Theorems 2.1 and 2.2 supply triples  $(k, n, N)$  that satisfy specified bounds for  $A$  with entries drawn Gaussian iid, and we can empirically test if these bounds are also satisfied for other random matrix ensembles. For instance, consider checksum matrices  $A$  with entries drawn the Bernoulli ensemble: iid uniform on  $\{-1, 1\}$ :

- Our answer to Q3 implies prevalence of perfect ECC pairs  $(A, B)$  with real-valued matrices at  $(k = 7, n = 85, N = 100)$ , but not at  $k = 8$  for the same  $n, N$ . An integer-valued checksum matrix  $A$ , 85 by 100, was generated with entries iid from the uniform distribution on  $\{-1, 1\}$ . It was consistently able to correct errors corrupting 7 entries of the encoded vector  $\beta$ ; when challenged with 100,000 instances of uniformly distributed patterns of errors, this  $A$  allowed to correct every set of errors – consistent with the prediction of Theorem 2.1.
- Our answer to Q4 furnishes triples  $(k, n, N)$  for which almost perfect ECC pairs  $(A, B)$  with  $\epsilon = .01$  are prevalent. We selected triples  $(20, 150, 200)$ ,  $(20, 200, 400)$ , and  $(20, 250, 400)$  at the edge of this regime, and for each triple generated a random integer-values matrix  $A$  with entries iid uniform  $\{-1, 1\}$ ; in each case the  $A$  so obtained was challenged 1000 times with an encoded message corrupted by a  $k$ -error pattern. In each case, the decoder

was able to correct each set of errors - consistent with the predictions of Theorem 2.2.

**3.3. Neighborly Polytope Existence Questions.** Table 3 arose from these theoretical questions:

**Q5:** *Random Neighborly Polytopes.* At a particular choice of  $N$  and  $n$ , for what values of  $k$  is there a positive chance that a standard Gaussian point cloud of  $N$  points in  $\mathbf{R}^n$  has a  $k$ -neighborly convex hull?

**Q6:** *Random Centrally Neighborly Polytopes.* For a particular choice of  $n$  and for a given value of  $N$ , consider random vectors  $a_i$ ,  $i = 1, \dots, N$  iid standard normal in  $\mathbf{R}^n$ . Consider the symmetrized point cloud with  $N$  points  $\{a_1, -a_1, a_2, -a_2, \dots, a_N, -a_N\}$ . For what values of  $k$  is there a positive probability that the resulting convex hull is centrally  $k$ -neighborly?

These questions can be answered by establishing the bounds [6, 11]

$$f_k(T^{N-1}) - \mathcal{E}f_k(AT^{N-1}) < 1, \text{ for (Q5) or } f_k(C^N) - \mathcal{E}f_k(AC^N) < 1, \text{ for (Q6)}$$

The existence of  $k$ -neighborly polytopes follows from these bounds, as the expected number of lost faces is less than one, implying that there must exist projected polytopes  $AQ$  which have exactly as many  $k-1$  dimensional faces as  $Q$ . Tightening the discrepancy even further than one, say to  $1 - \epsilon$ , implies that such polytopes not only exist, but can even be found with positive probability (at least)  $\epsilon$  when drawn at random from the distribution of  $AQ$ . In polytope theory random projection is the most powerful known way to generate highly centrally neighborly polytopes having many vertices. (Other techniques exist for few vertices).

Figures 3.1(a-b) are relevant to Q5 and Q6. Let  $Bound(Q, k, n, N)$  denote the right-hand side of display (2.2), for a given choice of  $k$ ,  $n$ ,  $N$ , and binding  $Q = T$  and  $Q = C$ . The figures portray the unit level set  $Bound(Q, k, n, N) = 1$  for  $N = 200, 1000$ , and  $5000$ .

*Corollary 3.3.* For a given  $N \in \{200, 1000, 5000\}$ , consider values of  $k$  and  $n$  such that  $(n/N, k/n)$  lies strictly beneath the curve corresponding to that  $N$  in Figure 3.1(a). There exist  $n \times N$  matrices  $A$  such that  $AT^{N-1}$  is  $k$ -neighborly. One can be obtained with positive probability by iid random sampling from the standard Gaussian distribution on  $\mathbf{R}^{n \times N}$ .

Consider values of  $k$  and  $n$  such that  $(n/N, k/n)$  lies strictly beneath the curve corresponding to that  $N$  depicted in Figure 3.1(b). There exist  $n \times N$  matrices  $A$  so that  $AC^N$  is centrally  $k$ -neighborly. One can be obtained with positive probability by iid random sampling from the standard Gaussian distribution on  $\mathbf{R}^{n \times N}$ . □

Specific examples were given in Table 3. For the  $(k, n, N)$  given, one can obtain  $k$ -neighborly (resp. centrally  $k$ -neighborly) polytopes by sampling  $N$  points from the standard Normal in  $\mathbb{R}^n$  (resp. then symmetrizing).

**3.4. About the Calculations.** The graphical and numerical results presented in this section were obtained by plugging specified triples  $(k, n, N)$  into the expressions in Theorems 2.1 and 2.2. We have developed computational tools to evaluate  $\Omega_S(\delta; Q)$ ,  $\Omega_W(\delta; Q)$ ,  $\rho_S(\delta; Q)$ , and  $\rho_W(\delta; Q)$ . These allowed us to obtain specific numerical results.

Less precise results can be obtained by hand using Corollaries 2.1 and 2.3.

Although some calculations assumed specific thresholds  $\epsilon$  on fractional face deficits, the results are relatively robust against changes in those thresholds. Due to the exponentiality of the bounds in Theorems 2.1 and 2.2, there would be barely perceptible changes in Figure 3.1 if the specified levels used in calculating those figures were changed moderately. Thus if we changed from 99% success rate to 50% success rate in panels (c-d), or from existence ( $|Bound(k, n, N)| = 1$ ) to prevalence ( $|Bound(k, n, N)| = 1/2$ ) or even highly prevalent ( $|Bound(k, n, N)| = 1/1000$ ) in panels (a-b), the figures would not change substantially.

It should also be noted from Figure 3.1 that even for small  $N$ , say 200, when  $\delta = n/N$  is relatively large there is already a large region below the level curves. However, for  $N$  and  $n/N$  simultaneously small, the bounds in Theorems 2.1 and 2.2 become weak or useless. For instance, the  $N = 200$  contour in Figure 3.1(a) reaches zero at about  $n/N = 0.17$ , corresponding to  $n = 34$ . In such instances, as a method of last resort, more accurate results can be obtained for particular triples by resorting to the full  $(\delta, \rho)$  dependent exponents  $\Psi_{net}(\delta, \rho; Q)$  in (2.1) and  $\Psi_{face}(\delta, \rho; Q)$  in (2.3); however, significant gains are only obtained when  $N$  and/or  $n/N$  are small.

**3.5. Computational Epistemology.** Our existence claims can be verified empirically: one can take a specific triple we identify and, by random sampling and subsequent testing, attempt to verify/falsify our claim.

To facilitate such empirical work we follow the principle of reproducible computational science [9]. Software is available [1] that can reproduce our numerical results or can empirically test specific triples.

The tabulated values we publish here offer information about a small selection of triples  $(k, n, N)$ . We make it possible to study other triples. Matlab software available for download [1] allows users to set parameters for use in a variety of situations. The ability to accurately evaluate the functions  $\Omega_S(\delta; Q)$  and  $\Omega_W(\delta; Q)$  is the central contribution of this article. Less accurate but simplified bounds are provided: Corollaries 2.1-2.4 and 4.2-4.3. Making the software for our calculations publicly available also allows further investigations into how Findings 2.1 and 2.2 were obtained.

Reproducibility of computational results helps to allow errors to be easily identified and corrected; to ensure that results are widely applied; and to provide precise numerical benchmarks which future research can aim to outperform. In particular, our article and software provide a collection of known good finite- $N$  triples that others may now try to expand.

**3.6. Comparison to Other Approaches.** Rudelson and Vershynin (RV) raised the issue of developing finite-sample bounds in a presentation at Princeton in 2006. Their recent publication [18] developed bounds that can be interpreted as enabling finite- $N$  statements concerning  $C^N$ . Table 4 lists values derived from their approach allowing for direct comparison. Comparing our tabulated values with those of Rudelson and Vershynin, it seems that the approach developed here is generally much stronger. For similar values of  $n$  and  $N$ , their bounds have implications only for significantly smaller values of  $k$ ; for similar values of  $k$  they require a substantially larger  $n$  than ours. It should be noted that Rudelson and Vershynin were not aiming for sharp numerical bounds. Their motivation was economy of argument; they gave short proofs using existing techniques from geometric functional analysis.

$N$	Here		RV [18]	
	$n$	$k$	$n$	$k$
$10^3$	500	40	652	1
$10^4$	1,000	45	1,089	3
$10^5$	1,000	27	4,377	27
$10^7$	100,000	2,993	100,090	895

TABLE 4. Comparison of existence results implied by the approach of this paper with those implied by work of Rudelson and Vershynin.

In comparison, our approach is asymptotically completely precise, at the expense of much more involved analysis.

Other results considering neighborliness of polytopes can in some instances also offer finite dimensional examples. For instance, in an investigation primarily focusing on the rough asymptotics of “How neighborly can a centrally symmetric polytope be?”, Linial and Novik [15] also established that there exist  $\frac{n}{400}$ -neighborly centrally symmetric polytopes in  $\mathbb{R}^n$  with  $4n$  vertices ( $N = 2n$ ). Focusing on the same parameter ratio,  $n/N = 1/2$ , our bounds in Theorem 2.1 show the existence of 0-neighborly polytopes for  $n \geq 37$ ; moreover, for  $n \geq 600$  there exist  $\frac{n}{12}$ -neighborly centrally symmetric polytopes in  $\mathbb{R}^n$  with  $4n$  vertices.

In principle, finite-sample implications can also be drawn from Restricted Isometry Principle arguments, [4]; however, there does not appear to have been a concerted effort to obtain effective quantitative bounds for finite dimensions, and our calculations show that they are generally weaker in finite samples than the approach of Rudelson and Vershynin, [18].

#### 4. THE SAMPLING THEOREM FOR SPARSE SIGNALS

As the introduction pointed out, displays (1.3) and (1.4) provide an appealing pair of easily-memorable asymptotic ‘laws’ about the behavior of face lattices of random projections. They also have applications: (1.3) implies the prevalence of  $k$ -neighborly polytopes and perfect checksum matrices; (1.4) implies the prevalence of compressed sensing matrices and almost-perfect checksum matrices.

Thus for example we can design a practical system able to reconstruct almost all vectors of length  $N$ , whose entries have at most  $k$  nonzeros; we only need to draw a Gaussian random matrix with  $N$  columns and roughly  $n > 2k \log(N/k)$  rows, obtain the vector of  $n$  measurements  $b = Ax$ , and apply minimum  $\ell_1$  reconstruction. This amounts to  $2k \log(N/k)$  generalized samples – dramatically fewer than  $N$  samples when  $N$  is much larger than  $k$ . This formula is simple and memorable. Similarly, in linear programming decoding with long blocklengths, we can successfully correct every pattern of  $k$  errors if we devote roughly  $n$  out of the  $N$  entries in a block to generalized checksums, where  $n \geq 2e \cdot k \log(N/n)$ .

Unfortunately, the word ‘roughly’ appears in the previous paragraph, meaning that in both statements there are some loose items that still need to be made precise. It would be preferable to have precise statements effective at specific finite  $n$ , but maintaining essentially the same form.

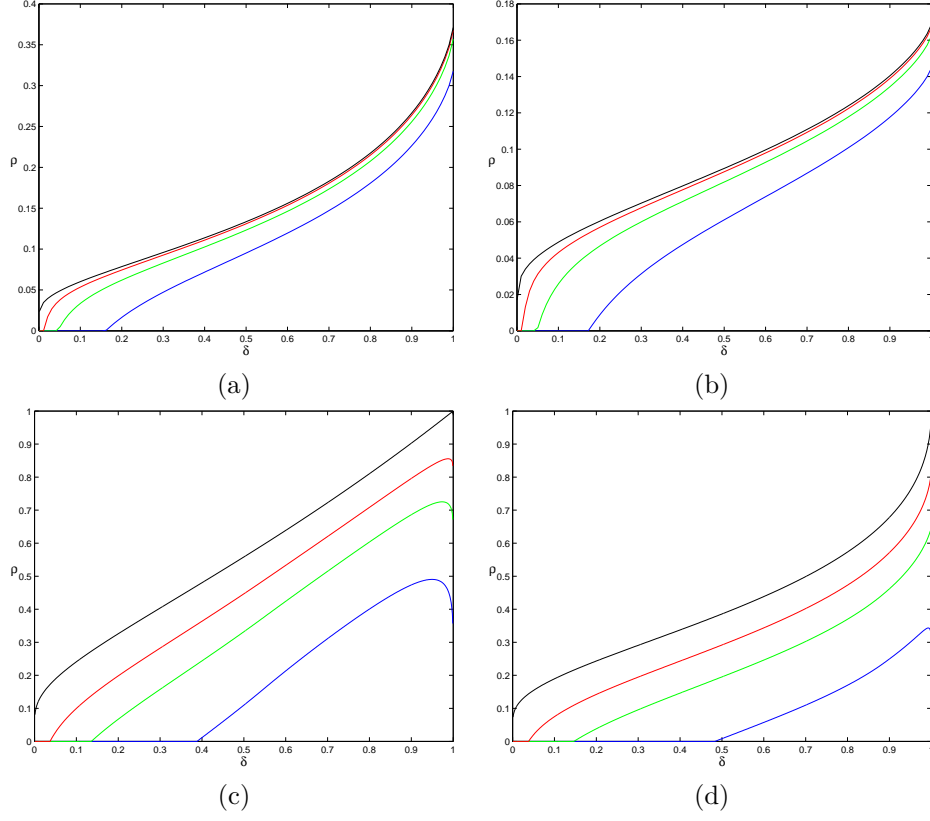


FIGURE 3.1. Panels (a) and (b): Unit level curves ( $|Bound(k, n, N)| = 1$ ) for the upper bounds in Theorem 2.1 for  $N = 200$  (blue), 1000 (green), and 5,000 (red);  $Q = T^{N-1}$  (left) and  $Q = C^N$  (right). The asymptotic,  $N \rightarrow \infty$ , limits  $\rho_S(\delta; Q)$  are also shown, (black). Panels (c) and (d): The  $10^{-2}$  level curves for the lower bounds ( $|Bound(k, n, N)| = 10^{-2}$ ) in Theorem 2.2 again for  $N = 200$  (blue), 1000 (green), and 5,000 (red);  $Q = T^{N-1}$  (left) and  $Q = C^N$  (right). The asymptotic,  $N \rightarrow \infty$ , limits  $\rho_W(\delta; Q)$  are also shown, (black).

We develop such statements, in two simple steps. First, we show that the asymptotic thresholds have effective finite  $n$  implications of the form  $n > k/\rho(n/N) + remainder$  and  $n > k/\rho(n/N)(1 + remainder)$ .

*Corollary 4.1* (Bounds Using Asymptotic Thresholds and Remainders). Set  $\delta = n/N$ . For  $\epsilon > 0$  and for  $Q \in \{T^{N-1}, C^N\}$  set

$$R_S(\epsilon, n, N; Q) := \frac{\log(c_S(Q) \cdot (N + 2)^5 / \epsilon)}{\Omega_S(\delta; Q)}$$

where  $c_S(Q)$  was defined in Theorem 2.1; ( $c_S(Q) < 1/6$ ). Then for

$$(4.1) \quad n > k/\rho_S(n/N; Q) + R_S(\epsilon, n, N; Q),$$



we have

$$\text{Prob}\{f_k(Q) = f_k(AQ)\} \geq 1 - \epsilon.$$

For  $\epsilon > 0$  and for  $Q \in \{T^{N-1}, C^N\}$  set

$$R_W(\epsilon, n, N; Q) := \left[ \frac{\log(c_W(Q) \cdot (N+2)^6/\epsilon)}{n\Omega_W(\delta; Q)} \right]^{1/2}$$

where  $c_W(Q)$  was defined in Theorem 2.2; ( $c_W(Q) < 4$ ). Then for  $R_W(\epsilon, n, N; Q) < 1$ ,

$$(4.2) \quad n > k/\rho_W(n/N; Q) \cdot (1 - R_W(\epsilon, n, N; Q))^{-1},$$

we have

$$\frac{f_k(Q) - \mathcal{E}f_k(AQ)}{f_k(Q)} \leq \epsilon.$$

*Proof.* By Theorem 2.1, if  $n\theta > R_S(\epsilon, n, N; Q)$ , the expected face deficit is bounded by  $\epsilon$  for all  $k \leq n(1-\theta) \cdot \rho_S(n/N; Q)$ . Rewriting the bound as  $n\theta < n - k/\rho_S(\delta; Q)$ , we obtain the inequalities

$$n - k/\rho_S(\delta; Q) > n\theta > R_S(\epsilon, n, N; Q),$$

which implies (4.1).

By Theorem 2.2, if  $n\theta^2 > \frac{\log(c_W(Q)(N+2)^6/\epsilon)}{\Omega_W(\delta; Q)}$ , the expected relative face deficit is bounded by  $\epsilon$  for all  $k \leq n(1-\theta) \cdot \rho_W(n/N; Q)$ . Rewriting this as  $\theta < 1 - k/(n\rho_W(n/N; Q))$ , and employing the condition on  $n\theta^2$  we have

$$1 - \frac{k}{n\rho_W(n/N; Q)} > \theta > R_W(\epsilon, n, N; Q).$$

Dropping the intermediate  $\theta$  and solving for  $n$  we arrive at (4.2).  $\square$

Remarks.

- Notice once again an explicit appearance of the asymptotic thresholds in a finite-sample bound; this reaffirms the significance of the asymptotic thresholds.
- By inspection, the remainders in this corollary are small if  $N$  is not exponentially larger than  $n$ , i.e. if  $\log(N)$  is small relative to  $n$ .
- The reader may be concerned that the upper bounds presented could be excessively pessimistic; in fact, this is not the case. These explicit conditions on  $(k, n, N)$  for the ‘strong agreement’ differ from the asymptotic limit  $n > k/\rho_S(\delta; Q)$  by a logarithmic additive term asymptotically negligible compared to the asymptotic limit; for the ‘weak agreement’ these explicit conditions on  $(k, n, N)$  differ from the asymptotic limit  $n > k/\rho_W(\delta; Q)$  by a multiplicative term tending to 1.

Corollary 4.1 gives the best known, finite dimensional, sampling theorem for  $k$ -sparse vectors from Gaussian measurements. This bound is given in terms of the asymptotic phase transitions  $\rho_S(n/N; Q)$  and  $\rho_W(n/N; Q)$  and the special functions  $\Omega_S(\delta; Q)$  and  $\Omega_W(\delta; Q)$ . The following Corollaries 4.2 and 4.3 are simplifications of Corollary 4.1 which are independent of all of the above special functions. They

follow from Corollary 4.1, the bounds on  $\Omega_S(\delta; Q)$  and  $\Omega_W(\delta; Q)$  in Findings 2.1 and 2.3, and the asymptotic behaviour of the phase transitions [12],

$$\begin{aligned}\rho_S(\delta; Q) &\sim [2e \log(1/\delta)]^{-1} \\ \rho_W(\delta; Q) &\sim [2 \log(1/\delta)]^{-1};\end{aligned}$$

here, by  $f(\delta) \sim g(\delta)$  we mean that the ratio tends to 1 as  $\delta \rightarrow 0$ .

To get bounds in place of limit statements for  $\rho_S(\delta; Q)$  and  $\rho_W(\delta; Q)$ , we add a ‘constant offset’.

*Finding 4.1* (Lower bounds on  $\rho_S(\delta; Q)$  and  $\rho_W(\delta; Q)$ ). We find numerically that

$$\rho_S(\delta; Q) > [2e \log(1/\delta) + d_S(Q)]^{-1}$$

for all  $\delta \geq 10^{-12}$  with  $d_s(T) := 6$ ,  $d_s(C) := 10$ ; and we find numerically that

$$\rho_W(\delta; Q) > [2 + 2 \log(1/\delta)]^{-1}$$

is valid for all  $\delta$ .

*Corollary 4.2* (Sampling Theorem - Strong Agreement). Define  $\tilde{R}_S(\epsilon, n, N; Q) := 2 \cdot \log(c_S(Q) \cdot (N + 2)^5 / \epsilon)$ . Conditional on  $\rho_S(\delta) > [2e \log(1/\delta) + d_S(Q)]^{-1}$  (Finding 4.1), the relation,

$$n > 2ek \cdot [\log(N/n) + d_s(Q)] + \tilde{R}_S(\epsilon, n, N; Q)$$

implies, for  $n/N > 10^{-12}$ ,

$$\text{Prob}\{f_k(Q) = f_k(AQ)\} \geq 1 - \epsilon.$$

*Corollary 4.3* (Sampling Theorem - Weak Agreement). Define  $\tilde{R}_W(\epsilon, n, N; Q) := 2[n^{-1} \log(c_W(Q) \cdot (N + 2)^6 / \epsilon)]^{1/2}$ . Conditional on  $\rho_W(\delta; Q) > [2 + 2 \log(1/\delta)]^{-1}$  (Finding 4.1),  $\tilde{R}_W(\epsilon, n, N; Q) < 1$  and

$$(4.3) \quad n > 2k \cdot [\log(N/n) + 1] \cdot (1 - \tilde{R}_W(\epsilon, n, N; Q))^{-1}$$

imply

$$\frac{f_k(Q) - \mathcal{E}f_k(AQ)}{f_k(Q)} \leq \epsilon.$$

## 5. ANALYSIS OF FACE DEFICIT

Both sides of the expected face deficit bound (2.1) are of interest, but only the upper bound has positive implications for existence questions - our focus in this paper. We would like the upper bound to be small, and so wish to document the extent to which the exponent  $\Psi_{net}(\delta, \rho; Q)$  is negative at  $\rho$  below the zero crossing  $\rho_S(\delta; Q)$ . The partial derivative of  $\Psi_{net}(\delta, \rho; Q)$  with respect to  $\rho$  has a simple multiplicative form,

$$(5.1) \quad \frac{\partial}{\partial \rho} \Psi_{net}(\delta, \rho; Q) = \delta F(\rho);$$

here we define

$$(5.2) \quad F(\rho) := \log\left(\rho^{-2}(1-\rho)y_\rho\sqrt{2\pi}\right) + (1-\rho) \cdot \left(\rho^{-1} - \frac{\dot{y}_\rho}{y_\rho} - \frac{1+\rho}{2\rho^2}y_\rho^2 + (\rho^{-1}-1)\dot{y}_\rho y_\rho\right),$$

where  $y_\rho$  is a quantity considered in our earlier papers [12] and defined implicitly by

$$(5.3) \quad \frac{1-\rho}{\rho}y_\rho = s_\rho, \quad \text{and } s_\rho \text{ solves } R(s_\rho) = 1 - \rho;$$

with  $R(s)$  also considered in our earlier papers, and defined by

$$(5.4) \quad R(s) := se^{s^2/2} \int_s^\infty e^{-y^2/2} dy.$$

$R$  is closely associated to a famous quantity in probability theory, the Mills' ratio of the standard Normal distribution [14, Sec 5.38].

Functions  $y_\rho$  and  $F(\rho)$  are depicted in Figure 5.1, panels (a) and (b) respectively.

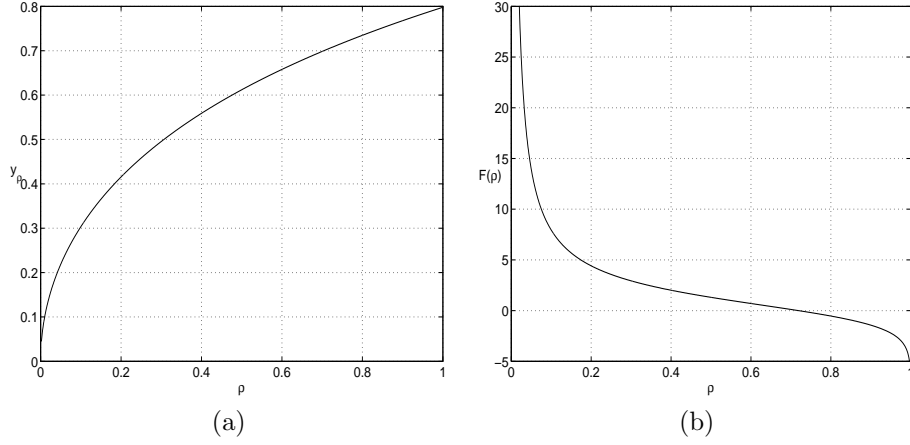


FIGURE 5.1. Panel (a)  $y_\rho$  and, panel (b)  $F(\rho)$ .

**Lemma 5.1.** *Define*

$$(5.5) \quad \Omega_S(\delta; Q) := \rho_S(\delta; Q) \min_{0 < \rho \leq \rho_S(\delta; Q)} F(\rho).$$

Then for  $\theta \in (0, 1)$

$$(5.6) \quad \Psi_{net}(\delta, (1-\theta)\rho_S(\delta; Q); Q) < -\theta\delta\Omega_S(\delta; Q).$$

**Proof.** For fixed  $\delta \in (0, 1)$ , consider the univariate Taylor series of  $G(\rho) \equiv \Psi_{net}(\delta, \rho; Q)$ , expanded at  $\rho = \rho_S(\delta; Q)$ . At some point  $\xi(\delta) \in [(1-\theta)\rho_S(\delta; Q), \rho_S(\delta; Q)]$  we have the equality:

$$(5.7) \quad \begin{aligned} \Psi_{net}(\delta, (1-\theta)\rho_S(\delta; Q); Q) &= \Psi_{net}(\delta, \rho_S(\delta; Q)) \\ &\quad - \theta\rho_S(\delta; Q) \frac{\partial}{\partial \rho} \Psi_{net}(\delta, \rho; Q) \Big|_{\rho=\xi(\delta; Q)} \\ &\leq -\theta\delta\rho_S(\delta; Q) \min_{0 < \rho \leq \rho_S(\delta; Q)} F(\rho) \end{aligned}$$

By definition of  $\rho_S(\delta; Q)$  as the zero level curve of  $\Psi_{net}(\delta; \rho; Q)$  the first term is zero.

Lemma 5.1 implies the form of the exponential term in Theorem 2.1. The polynomial term in Theorem 2.1 follows directly from bounds previously obtained by the authors in [12].

*Finding 5.1* (Monotonicity of  $F$ ).  $\frac{dF}{d\rho} < 0$  for all  $\rho \in [0, 1]$ .

*Demonstration of Finding 5.1.* See Figure 5.1 Panel (b).  $\square$

*Corollary 5.1.* Conditional on the Monotonicity of  $F$  (Finding 5.1),

$$\Omega_S(\delta; Q) = \rho_S(\delta; Q)F(\rho_S(\delta; Q)).$$

We now focus attention on the behavior of  $\Omega_S(\delta; Q)$ .

**Lemma 5.2.**  $\lim_{\delta \rightarrow 0} \Omega_S(\delta; Q)(\delta) \rightarrow 1/2$  from above.

**Proof.** That  $\Omega_S(\delta; Q) \rightarrow 1/2$  from above as  $\delta \rightarrow 0$  follows from an asymptotic expansion of the behavior of  $F(\rho)$  summarized in Lemma 5.3. From the definition of  $\Omega_S(\delta; Q)$  in (5.5) and the asymptotic behaviour of  $F(\rho)$  in (5.8), the asymptotic expansion of  $\Omega_S(\delta; Q)$  as  $\delta \rightarrow 0$  has a leading term  $1/2$ ; the additive remainder is asymptotically negligible, but strictly positive for all sufficiently small  $\delta \rightarrow 0$ .  $\square$

**Lemma 5.3.** For  $\rho \leq 1/100$ ,

$$(5.8) \quad F(\rho) = \frac{1}{2}\rho^{-1} - \frac{3}{2}\log(\rho) + \frac{1}{2}(\log(2\pi) - 3) + r_5(\rho), \quad \text{with } |r_5(\rho)| < 60.4\rho.$$

and

$$(5.9) \quad \frac{d}{d\rho}F(\rho) = \frac{-1}{2}\rho^{-2} - \frac{3}{2}\rho^{-1} + r_6(\rho), \quad \text{with } |r_6(\rho)| < 40.1.$$

Lemma 5.3 follows from Lemma 7.1 and the definition of  $F(\rho)$  in (5.2).

## 6. ANALYSIS OF FRACTIONAL FACE DEFICIT

The phase transitions in fractional face deficit occur at the zero crossings in  $\rho$  of

$$(6.1) \quad \frac{\partial}{\partial \rho}(\Psi_{net} - \Psi_{face})(\delta, \rho; Q) \equiv \delta(F(\rho) + \log(\delta\rho) - \log(1 - \delta\rho) - c(Q)).$$

$$\text{where } c(Q) = \begin{cases} 0 & Q = T \\ \log 2 & Q = C \end{cases}.$$

Level sets of (6.1) for  $Q = T$  are depicted in Figure 6.1 (a); the zero set is decorated with a dashed line and the  $\log 2$  level set is indicated by the solid line. These sets define  $\rho_W(\delta; Q)$  for  $Q = C$  and  $Q = T$  respectively.

**Lemma 6.1.** *Define*

$$(6.2) \quad \Omega_W(\delta; Q) := -\frac{1}{2}(\rho_W(\delta; Q))^2 \max_{0 < \rho \leq \rho_W(\delta; Q)} \left[ \frac{\partial}{\partial \rho}F(\rho) + \rho^{-1} + \frac{\delta}{1 - \delta\rho} \right]$$

Then for  $\theta \in (0, 1)$

$$(\Psi_{net} - \Psi_{face})(\delta, (1 - \theta)\rho_W(\delta; Q); Q) \leq -\delta\theta^2\Omega_W(\delta; Q).$$

**Proof.** Consider the Taylor series of the univariate function  $G(\rho) = (\Psi_{net} - \Psi_{face})(\delta, \rho; Q)$  at  $\rho = \rho_W(\delta; Q)$ . At any point  $\xi(\delta) \in [(1 - \theta)\rho_W(\delta; Q), \rho_W(\delta; Q)]$  we have the equality

$$(6.3) \quad \begin{aligned} (\Psi_{net} - \Psi_{face})(\delta, (1 - \theta)\rho_W(\delta; Q); Q) &= (\Psi_{net} - \Psi_{face})(\delta, \rho_W(\delta; Q); Q) \\ &\quad - \theta\rho_W(\delta; Q) \frac{\partial}{\partial \rho}(\Psi_{net} - \Psi_{face})(\delta, \rho; Q) \Big|_{\rho=\rho_W(\delta; Q)} \\ &\quad + \frac{1}{2}(\theta\rho_W(\delta; Q))^2 \frac{\partial^2}{\partial \rho^2}(\Psi_{net} - \Psi_{face})(\delta, \rho; Q) \Big|_{\rho=\xi(\delta)} \end{aligned}$$

The definition of  $\rho_W(\delta; Q)$  as the zero of (6.1) yields  $\frac{\partial}{\partial \rho}(\Psi_{net} - \Psi_{face})(\delta, \rho_W(\delta; Q); Q) = 0$ . By definition  $(\Psi_{net} - \Psi_{face})(\delta, \rho; Q) \leq 0$  for all  $(\delta, \rho) \in (0, 1)^2$ . Hence the first two out of three lines in the previous display are bounded above by zero. The definition of  $\Omega_W(\delta; Q)$  and  $\frac{\partial^2}{\partial \rho^2}(\Psi_{net} - \Psi_{face})(\delta, \rho; Q)$  bounds the third line by  $-\delta\theta^2\Omega_W(\delta; Q)$ .  $\square$

Lemma 6.1 determines the form of the exponential factor in Theorem 2.2. The polynomial factor in Theorem 2.2 follows directly from bounds previously obtained by the authors in [12].

Level sets of

$$\frac{\partial^2}{\partial \rho^2}(\Psi_{net} - \Psi_{face})(\delta, \rho; Q) = \delta \left( \frac{\partial}{\partial \rho} F(\rho) + \rho^{-1} + \frac{\delta}{1 - \delta\rho} \right)$$

can be studied in Figure 6.1 (b). The level set of the maximum value is decorated in magenta, while  $\rho_W(\delta; Q)$  for  $Q = T$  is depicted by the black curve and for  $Q = C$  by the blue curve.

For a given fixed  $\delta$ , the maximum in (6.2) occurs at  $\rho = \rho_W(\delta; Q)$  provided  $0 < \delta < 0.549$  and  $Q = C$  or provided  $0 < \delta < 0.886$  and  $Q = T$ . For  $\delta$  outside those intervals, the maximum in (6.2) occurs at the maximum level set depicted in Figure 6.1 (b) (dash-dot).

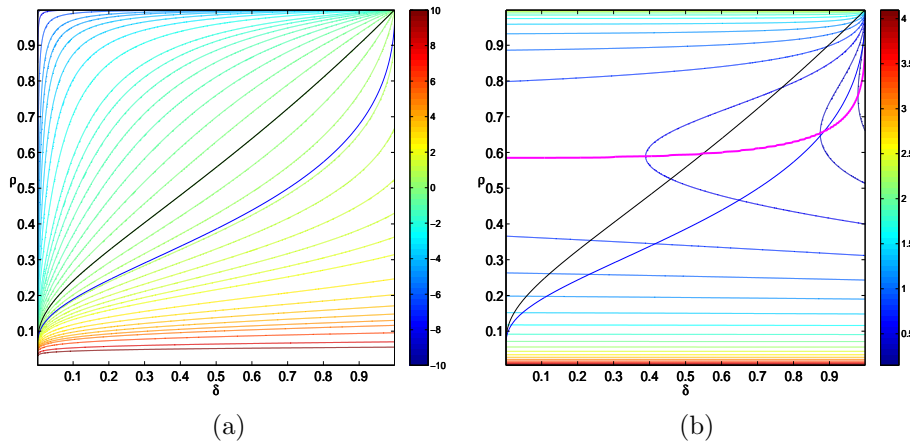


FIGURE 6.1. Panel (a): Contours of  $\frac{\partial}{\partial \rho}(\Psi_{net} - \Psi_{face})(\delta, \rho; T)$ . Panel (b): Level sets of  $-\frac{\partial^2}{\partial \rho^2}(\Psi_{net} - \Psi_{face})(\delta, \rho; T)$  ( $\log_{10}$  scale). The magenta curve in Panel (b) indicates the location  $(\delta, \rho)$  of minimum value attained in each constant- $\delta$  section. Both panels include overlays of  $\rho_W(\delta; C)$  (blue) and  $\rho_W(\delta; T)$  (black).

We now focus attention on the behaviour of  $\Omega_W(\delta; Q)$ .

**Lemma 6.2.**  $\lim_{\delta \rightarrow 0} \Omega_W(\delta; Q) \rightarrow 1/4$  from above.

**Proof.** From the definition of  $\Omega_W(\delta; Q)$  in (6.2) and the asymptotic behaviour of  $\frac{\partial F(\rho)}{\partial \rho}$  in (5.9), the asymptotic expansion of  $\Omega_W(\delta; Q)$  as  $\delta \rightarrow 0$  has a leading term  $1/4$ ; the additive remainder is asymptotically negligible, but strictly positive for all sufficiently small  $\delta \rightarrow 0$ .  $\square$

7. APPENDIX: ASYMPTOTIC BEHAVIOUR OF  $y_\rho$ 

The asymptotic properties of  $s_\rho$ , (5.3), as  $\rho \rightarrow 0$  (and hence also of  $y_\rho$ ) were studied in [8] using properties of an asymptotic series due to Laplace. We refine that approach to obtain

**Lemma 7.1.** *For  $\rho \leq 1/100$ ,*

$$(7.1) \quad y_\rho = \rho^{1/2} \left( 1 - \frac{1}{2}\rho + \frac{11}{8}\rho^2 + r_2(\rho) \right), \quad \text{with } |r_2(\rho)| < 48.4\rho^3,$$

$$(7.2) \quad \dot{y}_\rho := \frac{d}{d\rho}y_\rho = \rho^{-1/2} \left( \frac{1}{2} - \frac{3}{4}\rho + r_3(\rho) \right), \quad \text{with } |r_3(\rho)| < 28.62\rho^2,$$

and

$$(7.3) \quad \frac{d^2}{d\rho^2}y_\rho = \frac{-1}{4}\rho^{-3/2} (1 + r_4(\rho)), \quad \text{with } |r_4(\rho)| < 79\rho.$$

These results are developed below in the following order. Expansion (7.1) is obtained in Section 7.1 from properties of Laplace's asymptotic series. Once (7.1) has been established, we invoke the exact expressions for  $\dot{s}_\rho$  and  $\dot{y}_\rho$ :

$$\dot{y}_\rho = \frac{s_\rho(1-\rho) + \rho\dot{s}_\rho}{(1-\rho)^2} \quad \text{with} \quad \dot{s}_\rho = (\rho s_\rho - (1-\rho)s_\rho^{-1})^{-1}.$$

(These follow from the definitions of  $s_\rho$  and  $y_\rho$  in (5.3)). We obtain the bound (7.2) by simply combining (7.1) and bounds for  $(1-\rho)^{-\alpha}$  with these expressions.

The bounds for (7.3), (5.8), and (5.9) are arrived at similarly. We made no effort to carefully control the size of the constants in remainder terms in Lemmas 7.1 and 5.3. We have observed numerically that setting the remainder terms to zero yields approximations which are surprisingly accurate over ranges of  $\rho$  much larger than might be expected for such asymptotic approximations.

**7.1. Proof of (7.1).** We develop (7.1) in two stages. Initially, we develop the asymptotic behavior of  $s_\rho$  as  $\rho \rightarrow 0$ ; then we substitute it into equation (5.3). Our approximation of  $s_\rho$  uses the asymptotic series for  $R(s)$

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

appropriate for the regime of  $s$  large. To obtain this series, note that  $R(s) = s \cdot \text{Mills}(s)$  for  $s > 0$ , where  $\text{Mills}(s)$  is the usual Mills' ratio for the standard normal distribution. The corresponding asymptotic series for Mills' ratio is developed in [14, Secs 5.37, 5.38]; H. Ruben [16] credits this series to Laplace. In [14, Eq. (5.106)] it is shown that the error in truncating the series for  $\text{Mills}(s)$  at the  $s$ -th term is at most as large as the  $s$ -th term itself.  $R(s)$  inherits this property.

Now define  $L(s, \rho) := R(s) - 1 + \rho$  and note that  $s_\rho$  is defined so  $L(s_\rho, \rho) = 0$ . The asymptotic series for  $R$ , combined with the Taylor series of  $L(s, \rho)$  in variable  $s$ , suggest the asymptotic approximation

$$(7.4) \quad s_\rho \approx \tilde{s}_\rho := \rho^{-1/2} \left( 1 - \frac{3}{2}\rho + \frac{15}{8}\rho^2 \right).$$

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

$$(7.5) \quad G(y) = G(x) + (y - x) \frac{d}{dy} G(y)|_{y=\xi}.$$

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

$$(7.6) \quad |\tilde{s}_\rho - s_\rho| \leq \left| L(\tilde{s}_\rho, \rho) \left/ \frac{\partial}{\partial s} L(s, \rho) \Big|_{s=s_{mid}} \right. \right|,$$

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

Adapting bounds from [14, Eq. (5.106)] for Mills' ratio gives the following bound on  $R(\cdot)$ : for  $s > 9.8$ ,

$$(7.7) \quad 1 - s^{-2} + 3s^{-4} - 15s^{-6} + 96s^{-8} < R(s) < 1 - s^{-2} + 3s^{-4} - 15s^{-6} + 105s^{-8}.$$

Inserting  $\tilde{s}_\rho$  into the above upper bound and recalling  $L(s, \rho) = R(s) - 1 + \rho$  yields

$$(7.8) \quad |L(\tilde{s}_\rho, \rho)| < 46.4\rho^4, \quad \text{for } \rho < 1/100.$$

To bound the denominator, note that

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

which is a positive decreasing function of  $s$ ; this attains its lower bound on the interval  $s \in [\min(s_\rho, \tilde{s}_\rho), \max(s_\rho, \tilde{s}_\rho)]$  at one of the endpoints  $\{s_\rho, \tilde{s}_\rho\}$ . At  $\tilde{s}_\rho$  we again make use of a lower bound on Mills' ratio, a simplified variant of equation (7.7) that  $R(s) > 1 - s^{-2} + \frac{5}{2}s^{-4}$  for  $s > 9.8$  is sufficient,

$$(7.9) \quad \begin{aligned} \frac{\partial}{\partial s} L(s, \rho) \Big|_{s=\tilde{s}_\rho} &= [\tilde{s}_\rho + \tilde{s}_\rho^{-1}] R(\tilde{s}_\rho) - \tilde{s}_\rho \\ &> [\tilde{s}_\rho + \tilde{s}_\rho^{-1}] (1 - \tilde{s}_\rho^{-2} + \frac{5}{2}\tilde{s}_\rho^{-4}) - \tilde{s}_\rho \\ &= \frac{3}{2}\tilde{s}_\rho^{-3} + \frac{5}{2}\tilde{s}_\rho^{-5} > \frac{3}{2}\rho^{3/2}. \end{aligned}$$

For the lower bound at  $s_\rho$  we first note that for each  $\rho$ ,  $L(s, \rho)$  is a strictly increasing function of  $s$ , equaling zero at  $s_\rho$ . Employing a simpler variant of (7.7),  $R(s) > 1 + s^{-2} + 2s^{-4}$  for  $s > 3.9$ , we note that  $L(\rho^{-1/2}(1 - \rho), \rho) \geq 4\rho^2$  for  $\rho < 1/100$  which then implies the upper bound  $s_\rho \leq \rho^{-1/2} - \rho^{1/2}$ . From this we have the lower bound,

$$(7.10) \quad \begin{aligned} \frac{\partial}{\partial s} L(s, \rho) \Big|_{s=s_\rho} &= [s_\rho + s_\rho^{-1}] \cdot (1 - \rho) - s_\rho = (1 - \rho)s_\rho^{-1} - \rho s_\rho \\ &\geq (1 - \rho) \cdot \frac{1}{\rho^{-1/2} - \rho^{1/2}} - \rho(\rho^{-1/2} - \rho^{1/2}) = \rho^{3/2}. \end{aligned}$$

Using bounds (7.8) and (7.10) in equation (7.6) we have,

$$(7.11) \quad |s_\rho - \tilde{s}_\rho| \leq 46.4\rho^{5/2}, \quad \text{for } \rho \leq 1/100,$$

and obtain the estimate

$$s_\rho = \rho^{-1/2} \left( 1 - \frac{3}{2}\rho + \frac{15}{8}\rho^2 \right) + r_1(\rho), \quad \text{with } |r_1(\rho)| \leq 46.4\rho^{5/2}, \quad \text{for } \rho \leq 1/100;$$

Combined with (5.3), this yields the claimed (7.1) of Lemma 7.1.  $\square$

## REFERENCES

- [1] *Sparselab*.
- [2] Emmanuel Candès, Mark Rudelson, Terence Tao, and Roman Vershynin, *Error correction via linear programming*, FOCS 2005 (46th Annual Symposium on Foundations in Computer Science (2005)), 295–308.
- [3] 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.
- [4] Emmanuel J. Candès and Terence Tao, *Decoding via linear programming*, IEEE Trans. Inform. Theory **51** (2005), no. 12, 4203–4215.
- [5] ———, *Near optimal signal recovery from random projections and universal encoding strategies*, IEEE. Trans. Info. Thry. **52** (2006), no. 12, 5406–5425.
- [6] David L. Donoho, *Neighborly polytopes and sparse solutions of underdetermined linear equations*, Tech. report, Department of Statistics, Stanford University, 2005.
- [7] ———, *Compressed sensing*, IEEE. Trans. Info. Thry. **52** (2006), no. 4, 1289–1306.
- [8] ———, *High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension*, Disc. Comput. Geometry **35** (2006), no. 4, 617–652.
- [9] David L. Donoho, Arian Maleki, Inam Rahman, Morteza Shahram, and Victoria Strodren, *15 years of reproducible research in computational harmonic analysis*, Computing in Science and Engineering **11** (2009), 8–18.
- [10] 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.
- [11] ———, *Sparse nonnegative solutions of underdetermined linear equations by linear programming*, Proc. Natl. Acad. Sci. USA **102** (2005), no. 27, 9446–9451.
- [12] ———, *Counting faces of randomly projected polytopes when the projection radically lowers dimension*, J. AMS **22** (2009), no. 1, 1–53.
- [13] ———, *Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing*, Phil. Trans. of the Royal Society A **367** (2009), no. 1906, 4273–4293.
- [14] M. Kendall, A. Stuart, and J.K. Ord, *Kendall’s advanced theory of statistics*, Edward Arnold, London, 1991.
- [15] Nathan Linial and Isabella Novik, *How neighborly can a centrally symmetric polytope be?*, Disc. Comput. Geometry **36** (2006), no. 2, 273–281.
- [16] 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. MR MR0121713 (22 #12447)
- [17] M. Rudelson and R. Vershynin, *Geometric approach to error-correcting codes and reconstruction of signals*, Internatinoal Mathematical Research Notices **64** (2005), 4019–4041.
- [18] Mark Rudelson and Roman Vershynin, *On sparse reconstruction from fourier and gaussian measurements*, Communications on Pure and Applied Mathematics **61** (2008), 1025–1045.

DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY

*Current address:* Department of Statistics, Stanford University

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

SCHOOL OF MATHEMATICS, UNIVERSITY OF EDINBURGH

*Current address:* School of Mathematics, University of Edinburgh

*E-mail address:* jared.tanner@ed.ac.uk