Neighborliness of Randomly-Projected Simplices in High Dimensions

David L. Donoho and Jared Tanner

March 2005

Abstract

Let A be a d by n matrix, $d < n$. Let $T = T^{n-1}$ be the standard regular simplex in $\mathbf{R}^n$. We count the faces of the projected simplex AT in the case where the projection is random, the dimension d is large and n and d are comparable: $d \sim \delta n$, $\delta \in (0,1)$. The projector A is chosen uniformly at random from the Grassmann manifold of d-dimensional orthoprojectors of $\mathbf{R}^n$. We derive $\rho_N(\delta) > 0$ with the property that, for any $\rho < \rho_N(\delta)$, with overwhelming probability for large d, the number of k-dimensional faces of $P = AT$ is exactly the same as for T, for $0 \le k \le \rho d$. This implies that P is $\lfloor \rho d \rfloor$-neighborly, and its skeleton $Skel_{\lfloor \rho d \rfloor}(P)$ is combinatorially equivalent to $Skel_{\lfloor \rho d \rfloor}(T)$. We display graphs of ρ_N.

We also study a weaker notion of neighborliness it asks if the k-faces are all simplicial and if the numbers of k-dimensional faces $f_k(P) \ge f_k(T)(1-\epsilon)$. This was already considered by Vershik and Sporyshev, who obtained qualitative results about the existence of a threshold $\rho_{VS}(\delta) > 0$ at which phase transition occurs in k/d. We compute and display ρ_{VS} and compare to ρ_N.

Our results imply that the convex hull of n Gaussian samples in R^d, with n large and proportional to d, 'looks like a simplex' in the following sense. In a typical realization of such a high-dimensional Gaussian point cloud $d \sim \delta n$, all points are on the boundary of the convex hull, and all pairwise line segments, triangles, quadrangles, ..., $\lfloor \rho d \rfloor$-angles are on the boundary, for $\rho < \rho_N(d/n)$.

Our results also quantify a precise phase transition in the ability of linear programming to find the sparsest nonnegative solution to typical systems of underdetermined linear equations; when there is a solution with fewer than $\rho_{VS}(d/n)d$ nonzeros, linear programming will find that solution.

Key Words and Phrases: Neighborly Polytopes. Convex Hull of Gaussian Sample. Underdetermined Systems of Linear Equations. Uniformly-distributed Random Projections.

Acknowledgements. DLD had partial support from NSF DMS 00-77261, and 01-40698 (FRG), and from the Clay Mathematics Institute and an ONR-MURI; he thanks MSRI and its 'neighborly' hospitality in the Winter 2005, while this was prepared. JT was supported by NSF fellowship DMS 04-03041.

1 Introduction

Let $T = T^{n-1}$ be the standard simplex in $\mathbf{R}^n$ and let A be a uniformly-distributed random projection from $\mathbf{R}^n$ to $\mathbf{R}^d$. Some time ago, Goodman and Pollack proposed to study the properties of n points in $\mathbf{R}^d$ obtained as the vertices of $P = AT$; this was called by Schneider the Goodman-Pollack model of a random pointset. Independently, Vershik advocated a 'Grassmann Approach' to high-dimensional convex geometry and began to study the same object P, motivated by average-case analysis of the simplex method of linear programming.

Key insights into the properties of P were obtained by Affentranger and Schneider [1] and Vershik and Sporyshev [13]. Both developed methods to count the number of faces of the randomly-projected simplices $P = AT$. Affentranger and Schneider considered the case where d is fixed and n is large and showed the number of points on the convex hull of P grew logarithmically in n. Vershik and Sporyshev considered the situation where the dimension d was proportional to the number of points n and found that the low-dimensional face numbers of P behaved roughly like those of the simplex.

1.1 New Applications

In the years since [1, 13] first appeared, new reasons have emerged to study this problem:

- *Properties of Gaussian 'Point Clouds'.* Work of Baryshnikov and Vitale [2] has shown that the Goodman-Pollack model is for certain purposes equivalent to the classical model of drawing n samples from a multivariate Gaussian distribution in $\mathbf{R}^d$. Thus, results in this model tell us about the properties of multivariate Gaussian point clouds, in particular, the properties of their convex hull. High-dimensional Gaussian point clouds provide models of modern high-dimensional datasets. Much development of statistical models assumes these clouds behave as low dimensional clouds; as we will see this is wildly inaccurate.

- *Sparse Solution of Linear Systems.* In a companion paper [8], the authors considered the problem of finding the *sparsest* nonnegative solution to an underdetermined system of equations $y = Ax$, $x \geq 0$, A a $d \times n$ matrix. They connected this with the problem of k-neighborliness of the polytope $P_0 = \text{conv}(AT \cup \{0\})$; for more on neighborliness, see below. They showed that, if P_0 is k-neighborly, then for every problem instance (y, A) where $y = Ax_0$ with x_0 having at most k nonzeros, the sparsest solution can be obtained by linear programming.

Inspired by these two more recent developments, we study randomly-projected simplices anew.

1.2 Neighborliness

The polytope P is called *k-neighborly* if every subset of k vertices forms a $k-1$-face [10, Chapter 7]. A k-neighborly polytope 'acts like' a simplex, at least from the viewpoint of its low-dimensional faces. More formally, a k-neighborly polytope with n vertices has several properties of interest:

- It has the same number of ℓ-dimensional faces as the simplex T^{n-1}, $\ell = 0, \dots, k-1$.

- The ℓ-dimensional faces are all simplicial, for $0 \leq \ell < k$.

- The $(k-1)$-dimensional skeleton is combinatorially equivalent to the $(k-1)$-skeleton of the simplex T^{n-1}.

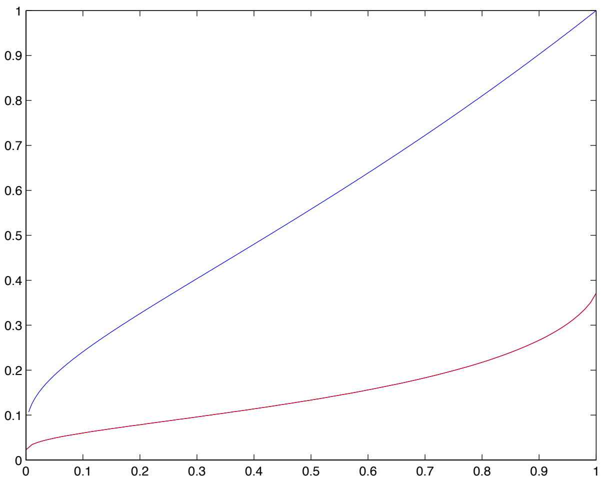


Figure 1: Lower curve: lower bound $\rho_N(\delta)$ on the neighborliness threshold, computed by methods of this paper. Upper curve: Vershik-Sporyshev weak neighborliness threshold ρ_{VS}. Matlab software available from the authors.

Such properties can seem counterintuitive. Comparing $T^{n-1} \subset \mathbf{R}^n$ with $P = AT^{n-1} \subset \mathbf{R}^d$, we note that P is a lower-dimensional projection of T^{n-1} and, it would seem, might 'lose faces' as compared to T^{n-1} because of the projection. For example, it might seem likely that, under projection, some edges of T^{n-1} might fall 'inside' the convex hull $\mathrm{conv}(AT^{n-1})$; yet if P is 2-neighborly, this does not happen. Surprisingly, in high dimensions, the counterintuitive event of 2-neighborliness is quite typical. Even much more extreme things occur – we can have k-neighborliness with k proportional to d.

1.3 Asymptotic Analysis

We adopt the Vershik-Sporyshev asymptotic setting and consider the case where d is proportional to n and both are large. However, to better align with applications, and with our own companion work [6, 7, 8], we use different notation than Vershik and Sporyshev in [13]. In a later section we will harmonize results. We assume $d = d_n = \lfloor \delta n \rfloor$ and consider n large.

Our primary concern is the *neighborliness phase transition*. It turns out that, with overwhelming probability for large n, the polytope $P = AT^{n-1}$ typically has n vertices and is k-neighborly for $k \approx \rho_N(d/n) \cdot d$. The function ρ_N will be characterized and computed below; see Figure 1. For example, that Figure shows that, if $n = 2d$ and n is large, k-neighborliness holds for $k \leq .133d$.

To state a formal result, for a polytope Q, let $f_\ell(Q)$ denote the number of ℓ-dimensional faces.

Theorem 1 Main Result. *Let $\rho < \rho_N(\delta)$ and let $A = A_{d,n}$ be a uniformly-distributed random projection from $\mathbf{R}^n$ to $\mathbf{R}^d$, with $d \geq \delta n$. Then*

$$Prob\{f_\ell(AT^{n-1}) = f_\ell(T^{n-1}), \quad \ell = 0, \ldots, \lfloor \rho d \rfloor\} \to 1, \qquad \text{as } n \to \infty. \tag{1.1}$$

In particular, this agreement of face numbers means that P is k neighborly for $k = \rho_N(\delta)d(1 + o_P(1))$.

We may distinguish this result from the pioneering work of Vershik and Sporyshev [13], who were interested in the question of whether, for k in a fixed proportion to n, the face numbers $f_k(AT^{n-1}) = f_k(T^{n-1})(1+o_P(1))$ or not. They also proved a threshold phenomenon for k in the vicinity of (say) $\rho_{VS} d$, for some implicitly characterized $\rho_{VS} = \rho_{VS}(d/n)$. While Vershik and Sporyshev referred to 'the neighborliness problem' in the title of their article, the notion they studied was not neighborliness in the sense of [10] and classical convex polytopes but instead what we might call *weak* neighborliness. Such weak neighborliness asks whether, for a given random polytope $P = AT^{n-1}$, there are n vertices and whether the overwhelming majority of ℓ-membered subsets of those vertices span $(\ell - 1)$-faces of P, for $\ell \leq k$.

For comparison to Theorem 1, note that the question of *approximate* equality of face numbers $f_k(AT^{n-1}) = f_k(T^{n-1})(1 + o_P(1))$ is weaker than the *exact* equality studied here in Theorem 1; it changes at a different threshold in k/d. Vershik-Sporyshev's result can be stated as follows.

Theorem 2 Vershik-Sporyshev. *There is a function $\rho_{VS}(\delta)$, characterised below, with the following property. Let $d = d(n) \sim \delta n$ and let $A = A_{d,n}$ be a uniform random projection from $\mathbf{R}^n$ to $\mathbf{R}^d$. Then for a sequence $k = k(n)$ with $k/d \sim \rho$, $\rho < \rho_{VS}(\delta)$, we have*

$$f_k(AT^{n-1}) = f_k(T^{n-1})(1 + o_P(1)). \tag{1.2}$$

We emphasize that our notation differs from Vershik and Sporyshev, who studied instead the inverse function $\delta_{VS}(\rho)$ (say). Figure 1 displays the weak-neighborliness phase transition function ρ_{VS} for comparison with the neighborliness phase transition ρ_N.

The Vershik-Sporyshev result is sharp in the sense that for sequences with $k/d \sim \rho > \rho_{VS}$, we do not have the approximate equality (1.2). In this paper we will show how a proof of Theorem 2 can be made similar to the proof of Theorem 1.

1.4 Numerical results

Our work contributes the first study of the neighborliness phase transition and the first numerical information about the Vershik-Sporyshev weak-neighborliness phase transition. Our Matlab software for computing these curves is available from the authors. In particular, Figure 1 depicts substantial numerical differences in the critical proportion ρ_{VS} and the lower bounds ρ_N. The most striking property of ρ_{VS} is that it crosses the line $\rho = 1/2$ near $\delta = .425$ and increases to 1 as $\delta \to 1$. This has implications for sparse solution of linear equations with n equations and $2n$ unknowns; see [8]. For comparison, we compute that

$$.371 \approx \lim_{\delta \to 1} \rho_N(\delta). \tag{1.3}$$

1.5 Solid Simplices

There are two natural variations on the notion of simplex to which the above results also apply. The first, T_0^n, is the convex hull of $\{0\}$ and T^{n-1}. This is a 'solid' n-simplex in $\mathbf{R}^n$, but not a regular simplex, since the vertex at 0 is closer to the other vertices than they are to each other. The second, T_1^n, is the convex hull of the vector $-\alpha 1$ with T^{n-1}, where α solves $(1 + \alpha)^2 + (n - 1)\alpha^2 = 2$. This is also a 'solid' n-simplex in $\mathbf{R}^n$, this time a regular one, with $n + 1$ vertices all spaced $\sqrt{2}$ apart. For applications where random projections of one or both of these alternate simplices could be of interest, we make the following remark.

Theorem 3 *Theorems 1 and 2 hold for AT_1^n, with the same functions ρ_N and ρ_{VS} and the comparable conclusions. Theorems 1 and 2 hold for AT_0^n, with the same functions ρ_N and ρ_{VS} and the comparable conclusions, provided 'neighborliness' is replaced by 'outward neighborliness'.*

'Outward neighborliness' is a slight variation of the concept of 'neighborliness', see the paper [8]. We give the (simple) proof of Theorem 3 in the Appendix.

1.6 Applications

We briefly indicate how these new results give information about the applications sketched in Section 1.1.

1.6.1 Gaussian Point Clouds.

Suppose we sample $X_1, X_2, \ldots, X_n$ i.i.d. according to a multivariate Gaussian distribution on $\mathbf{R}^d$ with nonsingular covariance. By Baryshnikov-Vitale [2], any affine-invariant property of the point configuration will have the same probability distribution under this model as it would under the model where A is a uniform random projection and X_i is the i-th column of A. We conclude the following.

Corollary 1.1 *Let $\delta \in (0,1)$ be fixed and let $d = d_n = \lfloor \delta n \rfloor$. Let $\rho < \rho_N(\delta)$. Let $X_1, X_2 \ldots, X_n$ be i.i.d. samples from a Gaussian distribution on $\mathbf{R}^d$ with nonsingular covariance. Consider the convex hull P of $(X_i)_{i=1}^n$. Then with overwhelming probability for large n,*

- *every X_i is a vertex of the convex hull P;*
- *every pair X_i, X_j generates an edge of the convex hull;*
- *...*
- *every $k = \lfloor \rho d \rfloor$ points generate a $(k-1)$-face of P.*

In short, not only are the points on the convex hull, but all reasonable-sized subsets span faces of the convex hull.

This is wildly different than the behavior that would be expected by traditional low-dimensional thinking. If we consider the case of d fixed and n tending to infinity, Affentranger and Schneider showed that there are a constant times $\log(n)^{(d-1)/2}$ points on the convex hull; in contrast, in the high-dimensional asymptotic considered here, all n points are on the convex hull. Even more exotically, Theorem 3 implies that a result just like Corollary 1.1 is true for the point set of $n+1$ points with X_i $i = 1, \ldots, n$ random as before, this time with zero mean, and the additional point $X_0 = 0$. Even though 0 is the most likely value for a standard Gaussian vector, it is a very highly exposed point in high dimensions!

1.6.2 Sparse Solution by Linear Programming

Finding the sparsest nonnegative solution to $y = Ax$ is an NP-hard problem in general when $d < n$. Surprisingly, many matrices have a sparsity threshold: for all instances y such that $y = Ax$ has a sufficiently sparse nonnegative solution, there is a unique nonnegative solution, which can be found by linear programming. Interestingly, the neighborliness phase transitions ρ_N and ρ_{VS} describe the threshold behavior of typical matrices A. This connection is discussed at length in [8]. Consider the standard linear program:

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

Corollary 1.2 *Fix $\epsilon, \delta > 0$. Let $d = \lfloor \delta n \rfloor$, and let A be a d times n matrix whose columns are independent and identically distributed according a multivariate normal distribution with nonsingular covariance. Let $k = \lfloor (\rho_N(\delta) - \epsilon) d \rfloor$. With overwhelming probability for large n, A has the property that, for every nonnegative vector x_0 containing at most k nonzeros, the corresponding $y = Ax_0$ generates an instance of the minimization problem (LP) which has x_0 for its unique solution.*

In words, for a typical A, for all problem instances permitting sufficiently sparse solutions, the linear programming problem (LP) computes the sparsest solution. Here sufficiently sparse is determined by $\rho_N(d/n)$.

The weak neighborliness threshold has implications in terms of 'most' underdetermined systems. Consider the collection $S_+(n, d, k)$ of all systems of linear equations with n unknowns, d equations, permitting a solution by $\leq k$ nonzeros. As explained in [8], one can place a measure on S_+ in which different matrices with the same row space are identified and different vectors y are identified if their sparsest decompositions have the same support. The result is a compact space, on which a natural uniform measure exists: the uniform measure on d-subspaces of $\mathbf{R}^n$ times the uniform measure on k-subsets of n objects.

Corollary 1.3 *Fix $\delta > 0$, and set $\rho < \rho_{VS}(\delta)$. For large n, in the overwhelming majority of systems in $S_+(n, \delta n, (\rho\delta)n)$, (LP) delivers the sparsest solution.*

We read off of Figure 1 that $\rho_{VS}(1/2) > .55$. Thus, for large n, in most n by $2n$ systems permitting a sparse solution with 55% as many nonzeros as equations, that is the solution delivered by (LP). This phenomenon is studied further in [8] and material cited there.

In both such results about solutions of linear equations, Theorem 3's applicability to the solid simplices AT_0^n is crucial.

1.7 Contents

In this paper we develop a viewpoint that allows to prove Theorems 1 and 2 in the same way, and that is essentially parallel to proofs of face-counting results in [7]. While necessarily our proofs have much to do with Vershik and Sporyshev's proof of Theorem 2, the viewpoint we adopt has the benefit of solving a range of problems, not only in this setting.

Section 2 proves Theorem 1, while Section 3 defined certain exponents used in the proof. Section 4 explains how the proof may be adapted to obtain Theorem 2. Section 5 sketches the proof of Theorem 3.

2 Random Projections of Simplices

We now outline the proof of Theorem 1. Key lemmas and inequalities will be justified in a later section.

2.1 Angle Sums

As remarked in the introduction, our proof proceeds by refining a line of research in convex integral geometry. Affentranger and Schneider [1] (see also Vershik and Sporyshev [13]) studied the properties of random projections $P = AT$ where T is an $n-1$-simplex and P is its d-dimensional orthogonal projection. [1] derived the formula

$$Ef_k(P) = f_k(T) - 2 \sum_{s \geq 0} \sum_{F \in \mathcal{F}_k(Q)} \sum_{G \in \mathcal{F}_{d+1+2s}(Q)} \beta(F, G)\gamma(G, T);$$

where E denotes the expectation over realizations of the random orthogonal projection, and the sum is over pairs (F, G) where F is a face of G. In this display, $\beta(F, G)$ is the internal angle at face F of G and $\gamma(G, T)$ is the external angle of T at face G; for definitions and derivations of these terms see eg. Grünbaum, Chapter 14, as well as [9, 11, 12]. Write

$$Ef_k(P) = f_k(T) - \Delta(k, d, n) \tag{2.1}$$

with

$$\Delta(k, d, n) = 2 \sum_{s \geq 0} \sum_{F \in \mathcal{F}_k(T)} \sum_{G \in \mathcal{F}_{d+1+2s}(T)} \beta(F, G)\gamma(G, T). \tag{2.2}$$

2.2 Exact Equality from Expectation

We view (2.1) as showing that on average $f_k(P)$ is about the same as $f_k(T)$, except for a nonnegative 'discrepancy' Δ. We will show that under the stated conditions on k,d, and n, for some $\epsilon > 0$

$$\Delta(k, d, n) \leq n \exp(-n\epsilon). \tag{2.3}$$

Now as $f_k(P) \leq f_k(T)$,

$$Prob\{f_k(P) \neq f_k(T)\} \leq E(f_k(T) - f_k(P)) = \Delta(k, d, n).$$

Hence (2.3) implies that with overwhelming probability we get equality of $f_k(P)$ with $f_k(T)$, as claimed in the theorem. To extend this into the needed simultaneous result - that $f_\ell(P) = f_\ell(T)$, $\ell = 0, \ldots, k-1$ – one defines events $E_k = \{f_k(P) \neq f_k(T)\}$ and notes that by Boole's inequality

$$Prob(\cup_0^{k-1} E_\ell) \leq \sum_0^{k-1} Prob(E_k) \leq \sum_{\ell=0}^{k-1} \Delta(\ell, d, n).$$

The exponential decay of $\Delta(k, d, n)$ will guarantee that the sum converges to 0 whenever the $k-1$-th term does. Hence by establishing (2.3) we get

$$Prob\{f_\ell(P) = f_\ell(T), \quad \ell = 0, \ldots, k-1\} \to 1$$

as is to be proved.

To establish (2.3), we rewrite (2.2) as

$$\Delta(k, d, n) = \sum_{s \geq 0} D_s$$

where, for $\ell = d + 1 + 2s$, $s = 0, 1, 2, \ldots$

$$D_s = 2 \cdot \sum_{F \in \mathcal{F}_k(T)} \sum_{G \in \mathcal{F}_{d+1+2s}(T)} \beta(F, G)\gamma(G, T).$$

We will show that, for $\rho < \rho_N$ (still to be defined) and for sufficiently small $\epsilon > 0$, then for $n > n_0(\epsilon; \rho, \delta)$

$$n^{-1} \log(D_s) \leq -\epsilon, \qquad s = 0, 1, 2, \ldots.$$

This implies (2.3) and hence our main result follows.

2.3 Decay and Growth Exponents

Following Affentranger and Schneider [1] and Vershik and Sporyshev [13], observe that:

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

Thus we can write

$$\begin{aligned} D_s &= 2 \cdot \binom{n}{k+1}\binom{n-k-1}{\ell-k}\beta(T^k,T^\ell)\gamma(T^\ell,T^{n-1}) \\ &= C_s\beta(T^k,T^\ell)\gamma(T^\ell,T^{n-1}), \end{aligned} \tag{2.4}$$

say, with C_s the combinatorial prefactor.

We now estimate $n^{-1}\log(D_s)$, decomposing it into a sum of terms involving logarithms of the combinatorial prefactor, the internal angle and the external angle. Formally, we will define exponents Ψ_{com}, Ψ_{int} and Ψ_{ext} so that for $\epsilon > 0$, and $n > n_0(\epsilon, \delta, \rho)$

$$n^{-1}\log(C_s) \leq \Psi_{com}(\ell/n; \rho, \delta) + \epsilon, \quad s = 0, 1, 2, \ldots,$$

and

$$n^{-1}\log(\beta(T^k, T^l)) \leq -\Psi_{int}(\ell/n; k/n) + \epsilon, \tag{2.5}$$

uniformly in $\ell \geq \delta n$, $k \geq \rho n$, $(\ell - k) \geq (\delta - \rho)n$.

$$n^{-1}\log(\gamma(T^l, T^{n-1})) \leq -\Psi_{ext}(\ell/n) + \epsilon, \tag{2.6}$$

uniformly in $\ell \geq \delta n$. It follows that for any fixed choice of ρ, δ, for $\epsilon > 0$, and for $n \geq n_0(\rho, \delta, \epsilon)$ we have the inequality

$$n^{-1}\log(D_s) \leq \Psi_{com}(\nu; \rho, \delta) - \Psi_{int}(\nu; \rho\delta) - \Psi_{ext}(\nu) + 3\epsilon, \tag{2.7}$$

valid uniformly in s. Exactly the same approach (with different details) has been used in [7], and the approach is related to [13].

To see where the exponents come from, we consider the simpest case, Ψ_{com}. Define the Shannon entropy:

$$H(p) = p\log(1/p) + (1-p)\log(1/(1-p));$$

noting that here the logarithm base is e, rather than the customary base 2. As did Vershik and Sporyshev [13] (and also [5, 7]), we note that

$$n^{-1}\log\binom{n}{\lfloor pn \rfloor} \to H(p), \quad p \in [0,1], \quad n \to \infty \tag{2.8}$$

so this provides a convenient summary for combinatorial terms. Defining $\nu = \ell/n \geq \delta$, we have

$$n^{-1}\log(C_s) = H(\rho\delta) + H(\frac{\nu - \rho\delta}{1 - \rho\delta})(1 - \rho\delta) + R_1 \tag{2.9}$$

with remainder $R_1 = R_1(s,k,d,n)$. Define then the *growth* exponent

$$\Psi_{com}(\nu;\rho,\delta) \equiv H(\rho\delta) + H(\frac{\nu - \rho\delta}{1-\rho\delta})(1-\rho\delta),$$

describing the exponential growth of the combinatorial factors. It is banal to apply (2.8) and see that the remainder R_1 in (2.9) is $o(1)$ uniformly in the range $k - \ell > (\delta - \rho)n$, $n > n_0$.

The definitions for the exponent functions (2.5)-(2.6) are significantly more involved, and are postponed to the following section. There it will be seen that these are continuous functions.

Define now the *net exponent* $\Psi_{net}(\nu;\rho,\delta) = \Psi_{com}(\nu;\rho,\delta) - \Psi_{int}(\nu;\rho\delta) - \Psi_{ext}(\nu)$. We can define at last the mysterious ρ_N as the threshold where the net exponent changes sign. It can be seen that the components of Ψ_{net} are all continuous over sets $\{\rho \in [\rho_0, 1], \delta \in [\delta_0, 1], \nu \in [\delta, 1]\}$, and so Ψ_{net} has the same continuity properties.

Definition 1 *Let $\delta \in (0,1]$. The critical proportion $\rho_N(\delta)$ is the supremum of $\rho \in [0,1]$ obeying*

$$\Psi_{net}(\nu;\rho,\delta) < 0, \qquad \nu \in [\delta, 1).$$

Continuity of Ψ_{net} shows that if $\rho < \rho_N$ then, for some $\epsilon > 0$,

$$\Psi_{net}(\nu;\rho,\delta) < -4\epsilon, \qquad \nu \in [\delta, 1).$$

Combine this with (2.7). Then for all $s = 0, 2, \ldots, (n-d)/2$ and all $n > n_0(\delta,\rho,\epsilon)$

$$n^{-1}\log(D_s) \le -\epsilon.$$

This implies (2.3) and our main result follows.

3 Properties of Exponents

We now define the exponents Ψ_{int} and Ψ_{ext} and discuss properties of ρ_N.

3.1 Exponent for External Angle

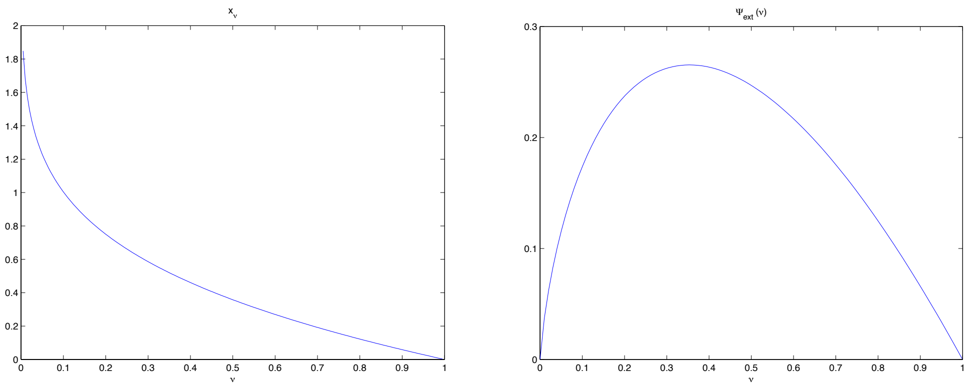
Let Q denote the cumulative distribution function of a normal $N(0,1/2)$ random variable, i.e. $X \sim N(0,1/2)$, and $Q(x) = Prob\{X \le x\}$. It has density $q(x) = \exp(-x^2)/\sqrt{\pi}$. Writing this out,

$$Q(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{x} e^{-y^2} dy. \tag{3.1}$$

For $\nu \in (0,1]$, define x_ν as the solution of

$$\frac{2xQ(x)}{q(x)} = \frac{1-\nu}{\nu}; \tag{3.2}$$

noting that possible values of x_ν are non-negative. Since xQ is a smooth strictly increasing function ~ 0 as $x \to 0$ and $\sim x$ as $x \to \infty$, and $q(x)$ is strictly decreasing, the function $2xQ(x)/q(x)$ is one-one on the positive axis, and x_ν is well-defined, and a smooth, decreasing function of ν. See Figure 2 for a depiction.



(a) (b)

Figure 2: Panel (a): The minimizer x_ν of ψ_ν, as a function of ν; Panel (b): The exponent Ψ_{ext}, a function of ν.

3.2 Exponent for Internal Angle

Let Y be a standard half-normal random variable $HN(0,1)$; this has cumulant generating function $\Lambda(s) = \log(E\exp(sY))$. Very convenient for us is the exact formula

$$\Lambda(s) = s^2/2 + \log(2\Phi(s)),$$

where Φ is the usual cumulative distribution function of a standard Normal $N(0,1)$. The cumulant generating function Λ has a rate function (Fenchel-Legendre dual [4])

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

This is smooth and convex on $(0,\infty)$, strictly positive except at $\mu = EY = \sqrt{2/\pi}$. More details are provided in [7]. See Figure 3.

For $\gamma \in (0,1)$ let

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

The function $\xi_\gamma(y)$ is strictly convex and positive on $(0,\infty)$ and has a minimum at a unique y_γ in the interval $(0, \sqrt{2/\pi})$. We define, for $\gamma = \frac{\rho\delta}{\nu} \le \rho$,

$$\Psi_{int}(\nu;\rho\delta) = \xi_\gamma(y_\gamma)(\nu - \rho\delta) + \log(2)(\nu - \rho\delta).$$

This is depicted in Figure 4. For fixed ρ, δ, Ψ_{int} is continuous in $\nu \ge \delta$. Most importantly, [7, Section 6] gives the asymptotic formula

$$\xi_\gamma(y_\gamma) \sim \frac{1}{2} \cdot \log(\frac{1-\gamma}{\gamma}), \quad \gamma \to 0. \tag{3.3}$$

3.3 Combining the Exponents

We now consider the combined behavior of Ψ_{com}, Ψ_{int} and Ψ_{ext}. We think of these as functions of ν with ρ, δ as parameters. The combinatorial exponent Ψ_{com} involves a scaled, shifted version of the Shannon entropy, which is a symmetric, roughly parabolic shaped function. This is the

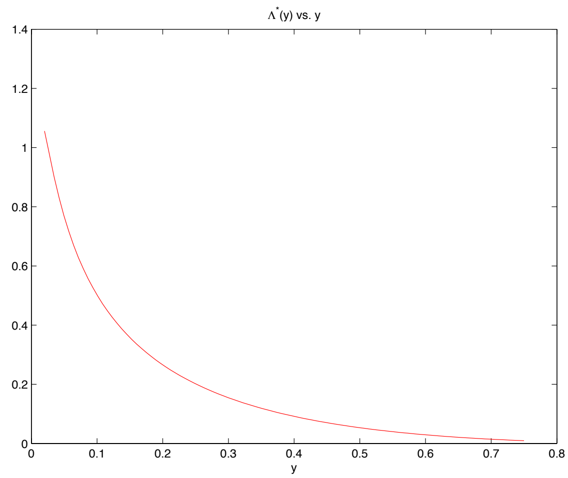


Figure 3: $\Lambda^*(y)$, rate function for Half-normal distribution; only the 'left-half' $0 < y < \mu$ is depicted. The function diverges at 0.

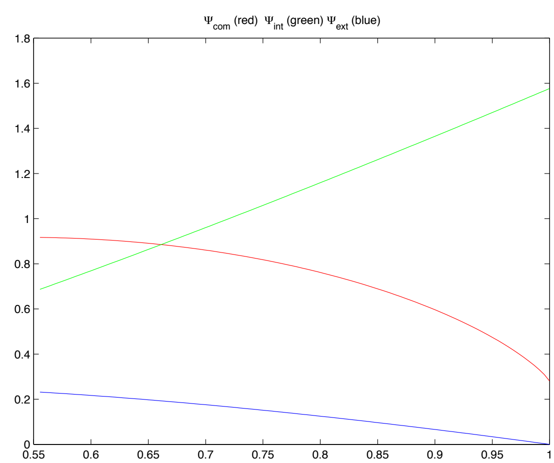


Figure 4: The exponents $\Psi_{com}(\nu; \rho, \delta)$ (red) and $\Psi_{int}(\nu; \rho\delta)$ (green), for $\rho = .145$, $\delta = .5555$. For comparison, Ψ_{ext} is displayed in blue.

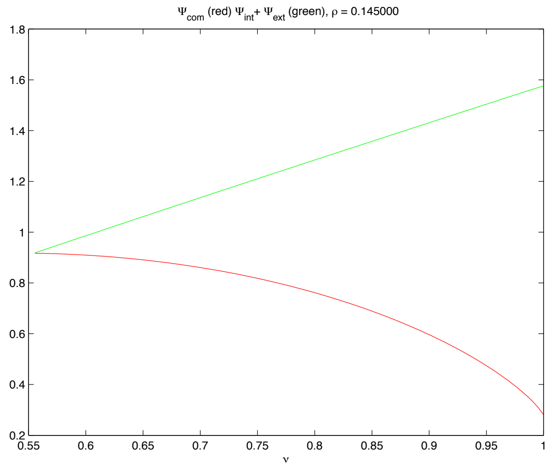


Figure 5: The exponents $\Psi_{com}(\nu;\rho,\delta)$ and $\Psi_{int}(\nu;\rho\delta)+\Psi_{ext}(\nu)$, for $\rho=.145$, $\delta=.5555$. The graph of Ψ_{com} (red) falls below that of $\Psi_{int}+\Psi_{ext}$ (green) and so $\Psi_{net}<0$.

exponent of a growing function which must be outweighed by the sum $\Psi_{ext}+\Psi_{int}$. It is depicted in Figure 4.

Figure 5 shows both Ψ_{com} and $\Psi_{ext}+\Psi_{int}$ with $\delta=.5555$ and $\rho=.145$. The desired condition $\Psi_{net}<0$ is the same as $\Psi_{com}<\Psi_{ext}+\Psi_{int}$, and this is distinctly obeyed except near $\nu=\delta$, where the two curves are close. We have $\rho_N(\delta)\approx .145$.

3.4 Justifying the Exponents

It remains to justify (2.5)-(2.6).

We sketch the argument for (2.6). The key point is the closed-form expression for $\gamma(T^\ell,T^{n-1})$:

$$\gamma(T^\ell,T^{n-1})=\sqrt{\frac{\ell+1}{\pi}}\int_0^\infty e^{-(\ell+1)x^2}\left(\frac{1}{\sqrt{\pi}}\int_{-\infty}^x e^{-y^2}dy\right)^{n-\ell-1}dx;$$

see [1]. We recognize the inner integral as involving Q from (3.1). Set $\nu_{\ell,n}=(\ell+1)/n$. The integral formula can be rewritten as

$$\sqrt{\frac{n\nu_{\ell,n}}{\pi}}\int_0^\infty \exp\{-n\nu_{\ell,n}x^2+n(1-\nu_{\ell,n})\log Q(x)\}dx. \qquad (3.4)$$

The appearance of n in the exponent suggests to use Laplace's method; we define, for ν fixed,

$$f_{\nu,n}(y)=\exp\{-n\psi_\nu(y)\}\cdot\sqrt{\frac{n\nu}{\pi}}$$

with

$$\psi_\nu(y)\equiv \nu y^2-(1-\nu)\log Q(y).$$

We note that ψ_ν is smooth and in the obvious way can develop expressions for its second and third derivatives. Applying Laplace's method to ψ_ν in the usual way, but taking care about regularity conditions and remainders, gives a result with uniformity in ν. Arguing in a fashion paralleling Section 5 of [7], one obtains:

Lemma 3.1 *For $\nu \in (0,1)$ let x_ν denote the minimizer of ψ_ν. Then*

$$\int_0^\infty f_{\nu,n}(x)dx \le \exp(-n\psi_\nu(x_\nu))(1 + R_n(\nu)),$$

where, for $\delta, \eta > 0$,

$$\sup_{\nu \in [\delta, 1-\eta]} R_n(\nu) = o(1) \text{ as } n \to \infty.$$

The minimizer x_ν mentioned in this lemma is the same x_ν defined earlier in (3.2) in terms of the error function. Also, the minimum value identified in this Lemma as driving the exponential rate is the same as our exponent Ψ_{ext}:

$$\Psi_{ext}(\nu) = \psi_\nu(x_\nu). \tag{3.5}$$

Hence (2.6) follows.

The decay estimate (2.5) for the internal angle was derived in [7] and details can be found there. Vershik and Sporyshev [13] used a related but seemingly different approach. The argument starts from a closed-form integral expression for $\beta(T^k, T^\ell)$. By [3], $\beta(T^k, T^\ell) = B(\frac{1}{k+2}, \ell - k + 1)$, where

$$B(\alpha, m) = \theta^{(m-1)/2}\sqrt{(m-1)\alpha + 1}\pi^{-m/2}\alpha^{-1/2}J(m, \theta) \tag{3.6}$$

with $\theta \equiv (1-\alpha)/\alpha$ and

$$J(m, \theta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^\infty (\int_0^\infty e^{-\theta v^2 + 2iv\lambda} dv)^m e^{-\lambda^2} d\lambda. \tag{3.7}$$

It was shown in [7] that Laplace's method applied to this last integral yields exponential bounds on the decay of β of the form (2.5).

3.5 Properties of ρ_N

We mention two key facts about ρ_N Firstly, the concept is nontrivial:

Lemma 3.2

$$\rho_N(\delta) > 0, \qquad \delta \in (0,1). \tag{3.8}$$

Secondly, one can show that, although $\rho_N(\delta) \to 0$ as $\delta \to 0$, it goes to zero slowly.

Lemma 3.3 *For $\eta > 0$,*

$$\rho_N(\delta) \ge \log(1/\delta)^{-(1+\eta)}, \qquad \delta \to 0.$$

These results require only a simple observation. The paper [7] studied uniform random projections AC^n of the cross-polytope C^n, namely the unit ℓ^1 ball in $\mathbf{R}^n$. A function $\rho_N^\pm$ was derived, giving the threshold below which a certain event $E_{n,\rho}$ happens with overwhelming probability for large n. Under the event $E_{n,\rho}$ the images under A of all $\lfloor \rho d \rfloor$-dimensional faces of C appeared as faces of AC. Viewing T^{n-1} as a face of C^n, when $E_{n,\rho}$ holds, it follows that every low-dimensional face of T^{n-1} must therefore appear as a face of AT^{n-1}, meaning that

$$\rho_N(\delta) \ge \rho_N^\pm(\delta), \quad \delta \in (0,1).$$

Lower bounds completely parallel in form to those in Lemmas 3.2 and 3.3 were already proven for $\rho_N^\pm$ in [7]. Hence Lemmas 3.2 and 3.3 follow from those.

4 Weak Neighborliness

We now explain how the above proof can be adapted to handle Vershik-Sporyshev's result – Theorem 2.

Observe that $f_{k-1}(T^{n-1}) = \binom{n}{k}$; this combinatorial factor has exponential growth with n according to an exponent $\Psi_{face}(\rho\delta) \equiv H(\rho\delta)$; thus, if $k = k(n) \sim \rho\delta n$,

$$n^{-1}\log(f_{k-1}(T^{n-1})) \to \Psi_{face}(\rho\delta), \qquad n \to \infty.$$

We again define Ψ_{net} as in the proof of Theorem 1.

Definition 2 *Let $\delta \in (0,1]$. The critical proportion $\rho_{VS}(\delta)$ is the supremum of $\rho \in [0,1]$ obeying*

$$\Psi_{net}(\nu; \rho, \delta) < \Psi_{face}(\rho\delta), \qquad \nu \in [\delta, 1). \tag{4.1}$$

Recall Section 2's definition $\Delta(k,d,n) = f_{k-1}(T) - f_{k-1}(AT) \geq 0$. The proof of Theorem 2 is based on observing that (4.1) implies

$$\Delta(k,d,n) = o(f_{k-1}(T^{n-1})). \tag{4.2}$$

We immediately get (1.2). Showing that (4.1) implies (4.2) requires no new ideas; one proceeds as in Section 2 almost line-by-line; we omit the exercise. □

We remark that the criticial proportion ρ_{VS} defined in this way does not immediately resemble the result of Vershik and Sporyshev's result. Section 6 of [7] explains how to translate between the two notational systems.

5 Proof of Theorem 3

We now sketch the arguments supporting Theorem 3.

5.1 Solid Simplex T_1^n

The standard n simplex with $n+1$ vertices, T^n, lives in $\mathbf{R}^{n+1}$. However, in fact it lies in an n-plane orthogonal to the main diagonal. We think of that n-plane as a copy of n-space, which is to say that by rotating and translating $\mathbf{R}^{n+1}$ and dropping the last coordinate, we get isometrically a convex body in $\mathbf{R}^n$; this is in fact T_1^n.

Applying a random projection $B : R^{n+1} \mapsto \mathbf{R}^d$ to T^n gives a result which is identically distributed (up to a translation) with a random projection $A : \mathbf{R}^n \mapsto \mathbf{R}^d$. Indeed, $BT^n = B\binom{U}{0}T_1^n + v$ where U is a fixed $n \times n$ orthogonal matrix and $v \in \mathbf{R}^d$ is a fixed vector. But $\tilde{A} = B\binom{U}{0}$ defines a uniform random projection from $\mathbf{R}^n \mapsto \mathbf{R}^d$. As $\tilde{A}$ and A are identically distributed, hence AT_1^n and $BT^n - v$ are identically distributed. Translations of a pointset do not affect neighborliness properties.

Now in the asymptotic setting $d \sim \delta n$, BT^n obeys Theorem 1 with $\rho_N(d/(n+1))d$ in place of $\rho_N(d/n)d$, and similarly for ρ_{VS} in Theorem 2; all we are really doing is renaming n as $n+1$. And of course the limiting $\delta \sim d/n \sim d/(n+1)$.

5.2 Solid Simplex T_0^n

We think of T^{n-1} as the 'outward' face of T_0^n. AT_0^n is called *outwardly* k-neighborly if every $k-1$ face of AT^{n-1} is also a face of AT_0^n. For more discussion, see [8] where the following result is proved as Lemma A.1.

Lemma 5.1 *Suppose that $0 \notin conv\{a_j\}$. Suppose that there exist $b \neq 0$ so that*

$$Q = conv(\{a_j\}_{j=1}^n \cup \{b\})$$

has $n+1$ vertices, is k-neighborly, and has $0 \in Q$. Then $P = conv(\{0\} \cup \{a_j\}_{j=1}^n)$ has $n+1$ vertices and is outwardly k-neighborly.

We remark that $AT_0^n = \text{conv}(\{0\} \cup \{a_j\})$ while $AT_1^n = \text{conv}(\{-\alpha A1\} \cup \{a_j\})$. Hence AT_1^n is exactly of the form Q given by this lemma, and AT_0^n is of the form P. Hence, k-neighborliness of AT_1^n implies outward k-neighborliness of AT_0^n.

References

[1] Fernando Affentranger and Rolf Schneider. Random projections of regular simplices. *Discrete Comput. Geom.*, 7(3):219–226, 1992.

[2] Yuliy M. Baryshnikov and Richard A. Vitale. Regular simplices and Gaussian samples. *Discrete Comput. Geom.*, 11(2):141–147, 1994.

[3] Károly Böröczky, Jr. and Martin Henk. Random projections of regular polytopes. *Arch. Math. (Basel)*, 73(6):465–473, 1999.

[4] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 1998.

[5] David L. Donoho. For most large systems of underdetermined equations, the minimum ℓ^1-norm solution is the sparsest solution. Technical report, Department of Statistics, Stanford University, 2004.

[6] David L. Donoho. Neighborly polytopes and sparse solutions of underdetermined linear equations. Technical report, Department of Statistics, Stanford University, 2004.

[7] David L. Donoho. High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension. Technical report, Department of Statistics, Stanford University, 2005.

[8] David L. Donoho and Jared Tanner. Sparse nonnegative solutions of underdetermined linear equations by linear programming. Technical report, Department of Statistics, Stanford University, 2005.

[9] Branko Grünbaum. Grassmann angles of convex polytopes. *Acta Math.*, 121:293–302, 1968.

[10] Branko Grünbaum. *Convex polytopes*, volume 221 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2003. Prepared and with a preface by Volker Kaibel, Victor Klee and Günter M. Ziegler.

[11] Peter McMullen. Non-linear angle-sum relations for polyhedral cones and polytopes. *Math. Proc. Cambridge Philos. Soc.*, 78(2):247–261, 1975.

[12] Harold Ruben. On the geometrical moments of skew-regular simplices in hyperspherical space, with some applications in geometry and mathematical statistics. *Acta Math.*, 103:1–23, 1960.

[13] 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(2):181–201, 1992.