Neighborliness of randomly projected simplices in high dimensions

David L. Donoho* and Jared Tanner

Department of Statistics, Sequoia Hall, Serra Mall 370, Stanford University, Stanford, CA 94305-4065

Contributed by David L. Donoho, March 30, 2005

Let A be a $d \times n$ matrix and $T = T^{n-1}$ be the standard simplex in R^n. Suppose that d and n are both large and comparable: $d \approx \delta n$, $\delta \in (0, 1)$. We count the faces of the projected simplex AT when the projector A is chosen uniformly at random from the Grassmann manifold of d-dimensional orthoprojectors of 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 $0 \leq k \leq \rho d$. 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 also study a weaker notion of neighborliness where the numbers of k-dimensional faces $f_k(P) \geq f_k(T)(1 - \varepsilon)$. Vershik and Sporyshev previously showed existence of a threshold $\rho_{VS}(\delta) > 0$ at which phase transition occurs in k/d. We compute and display ρ_{VS} and compare with ρ_N. Corollaries are as follows. (1) The convex hull of n Gaussian samples in R^d, with n large and proportional to d, has the same k-skeleton as the $(n - 1)$ simplex, for $k < \rho_N(d/n)d(1 + o_P(1))$. (2) There is a "phase transition" in the ability of linear programming to find the sparsest nonnegative solution to systems of underdetermined linear equations. For most systems having a solution with fewer than $\rho_{VS}(d/n)d(1 + o(1))$ nonzeros, linear programming will find that solution.

neighborly polytopes | convex hull of Gaussian sample | underdetermined systems of linear equations | uniformly distributed random projections | phase transitions

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 model 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 (2). 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 if 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 refs. 1 and 2 first appeared, new connections arose, motivating a fresh study of this problem.

- The first connection involves properties of Gaussian "point clouds." Work of Baryshnikov and Vitale (3) 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 data sets. Much development of statistical models assumes these clouds behave as low-dimensional clouds; as we will see, low-dimensional intuition is wildly inaccurate.
- The second connection involves sparse solution of linear systems. In a companion paper (4), we 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. We connected this with the problem of k-neighborliness of the polytope $P_0 = \mathrm{conv}(AT \cup \{0\})$; for more on neighborliness, see below. We 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 (ref. 5, Chap. 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 as follows:

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

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 with T^{n-1} because of the projection. For example, it might seem plausible 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, the plausible 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 companion work (4, 6, 7), we use different notation than Vershik and Sporyshev in ref. 2. 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 Fig. 1). For example, Fig. 1 shows that if $n = 2d$ and n is large, k-neighborliness holds for $k \leq 0.133d$.

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

*To whom correspondence should be addressed. E-mail: donoho@stat.stanford.edu.

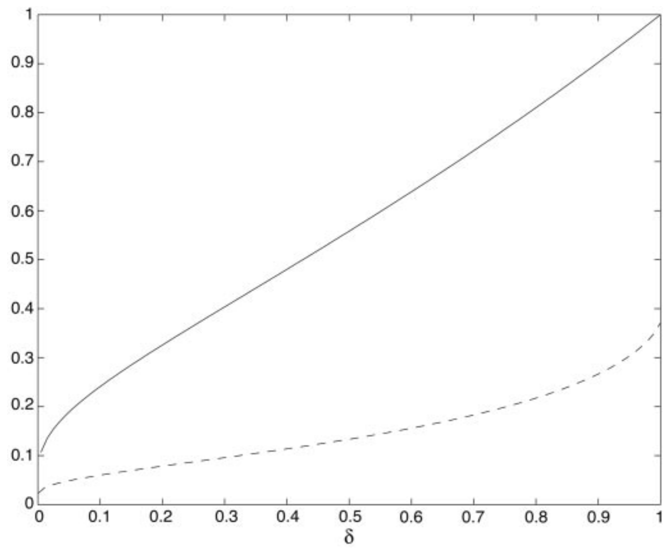

Fig. 1. ρ_{vs} and ρ_N. The lower curve (dashed) shows the lower bound $\rho_N(\delta)$ on the neighborliness threshold, computed by methods described in this work. The upper curve (solid) shows Vershik–Sporyshev weak neighborliness threshold ρ_{VS}. MATLAB software is available from the authors.

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*

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

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 (2), 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 $\rho_{VS}d$, for some implicitly characterized $\rho_{VS} = \rho_{VS}(d/n)$. Although 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 ref. 5 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 with *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 $\rho_{VS}(\delta)$, characterized below, with the following property. Let $d = d(n) \approx \delta n$ 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*

$$f_k(AT^{n-1}) = f_k(T^{n-1})(1 + o_P(1)). \qquad \textbf{[1.2]}$$

We emphasize that our notation differs from Vershik and Sporyshev, who studied instead the inverse function $\delta_{VS}(\rho)$. Fig. 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 work, we will show how a proof of *Theorem 2* can be made similar to the proof of *Theorem 1*.

1.4. Numerical Result. Our work contributes to the study of the neighborliness phase transition and to the numerical information about the Vershik–Sporyshev weak-neighborliness phase transition. Our MATLAB software for computing these curves is available from D.L.D. or J.T. on request. In particular, Fig. 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 $\delta = 0.425$ and increases to 1 as $\delta \rightarrow 1$. This property has implications for sparse solution of linear equations with n equations and $2n$ unknowns (see ref. 4). For comparison, we compute that

$$0.371 \approx \lim_{\delta \rightarrow 1} \rho_N(\delta). \qquad \textbf{[1.3]}$$

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}. It is a "solid" n-simplex in $\mathbf{R}^n$ but not a regular simplex, because 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 $(1 + \alpha)^2 + (n - 1)\alpha^2 = 2$. It 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 ref. 4). To save space we give the (simple) proof of *Theorem 3* in the technical report (ref. 8, 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 (3), 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 ith 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 behavior 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 (1) 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.

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 ref. 4. Consider the standard linear program

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

Corollary 1.2. *Fix ε, $\delta > 0$. Let $d = \lfloor \delta n \rfloor$, and let A be a $d \times n$ matrix whose columns are independent and identically distributed according to a multivariate normal distribution with nonsingular covariance. Let $k = \lfloor (\rho_N(\delta) - \varepsilon)d \rfloor$. 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 other 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 we explain in our companion paper (4), 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 Fig. 1 that $\rho_{VS}(1/2) > 0.55$. Thus, for large n, in most $n \times 2n$ systems permitting a sparse solution with 55% as many nonzeros as equations, that is the solution delivered by (LP). This phenomenon is studied further by us in ref. 4 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 work, we develop a viewpoint that allows us to prove *Theorems 1* and *2* in the same way, and that is essentially parallel to proofs of face-counting results in ref. 7. Although 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*. Theorem 3 is proven in ref. 8.

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 (2)] studied the properties of random projections $P = AT$ where T is an $(n - 1)$-simplex and P is its d-dimensional orthogonal projection. Ref. 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, e.g., Grünbaum, Chap. 14 (5) as well as refs. 9–11. Write

$$Ef_k(P) = f_k(T) - \Delta(k, d, n) \qquad \textbf{[2.1]}$$

with

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

2.2. Exact Equality from Expectation. We view Eq. **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 $\varepsilon > 0$

$$\Delta(k, d, n) \leq n \exp(-n\varepsilon). \qquad \textbf{[2.3]}$$

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

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

Hence, Eq. **2.3** implies that with overwhelming probability, we get equality of $f_k(P)$ with $f_k(T)$, as claimed in the theorem. For 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

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

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 Eq. **2.3** we get

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

as is to be proved.

To establish Eq. **2.3**, we rewrite Eq. **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 $\varepsilon > 0$, then for $n > n_0(\varepsilon; \rho, \delta)$

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

Eq. **2.3** follows, as well as our main result.

2.3. Decay and Growth Exponents. Following Affentranger and Schneider (1) and Vershik and Sporyshev (2), observe the following:

- 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

$$D_s = 2\cdot\binom{n}{k+1}\binom{n-k-1}{\ell-k}\beta(T^k, T^\ell)\gamma(T^\ell, T^{n-1})$$
$$= C_s\beta(T^k, T^\ell)\gamma(T^\ell, T^{n-1}),$$

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 $\varepsilon > 0$, and $n > n_0(\varepsilon, \delta, \rho)$

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

and

$$n^{-1}\log(\beta(T^k, T^l)) \le -\Psi_{\text{int}}(\ell/n;\ k/n) + \varepsilon, \qquad \textbf{[2.4]}$$

uniformly in $\ell \ge \delta n$, $k \ge \rho n$, $(\ell - k) \ge (\delta - \rho)n$;

$$n^{-1}\log(\gamma(T^l, T^{n-1})) \le -\Psi_{\text{ext}}(\ell/n) + \varepsilon, \qquad \textbf{[2.5]}$$

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

$$n^{-1}\log(D_s) \le \Psi_{\text{com}}(\nu;\ \rho,\ \delta) - \Psi_{\text{int}}(\nu;\ \rho\delta) - \Psi_{\text{ext}}(\nu) + 3\varepsilon, \qquad \textbf{[2.6]}$$

valid uniformly in s. Exactly the same approach (with different details) has been used in ref. 7, and the approach is related to ref. 2.

To see where the exponents come from, we consider the simplest 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 (2) and also the authors of refs. 7 and 12, we note that

$$n^{-1}\log\binom{n}{\lfloor pn \rfloor} \to H(p), \quad p \in [0, 1],\ n \to \infty, \qquad \textbf{[2.7]}$$

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

$$n^{-1}\log(C_s) = H(\rho\delta) + H\left(\frac{\nu - \rho\delta}{1 - \rho\delta}\right)(1 - \rho\delta) + R_1, \qquad \textbf{[2.8]}$$

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

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

describing the exponential growth of the combinatorial factors. It is banal to apply Eq. **2.7** and see that the remainder R_1 in Eq. **2.8** is $o(1)$ uniformly in the range $k - \ell > (\delta - \rho)n$, $n > n_0$.

The definitions for the exponent functions Eqs. **2.4** and **2.5** 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 $\Psi_{\text{net}}(\nu; \rho, \delta) = \Psi_{\text{com}}(\nu; \rho, \delta) - \Psi_{\text{int}}(\nu; \rho\delta) - \Psi_{\text{ext}}(\nu)$. 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

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

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

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

Recall now Eq. **2.6**. Then for all $s = 0, 2, \ldots, (n-d)/2$ and all $n > n_0(\delta, \rho, \varepsilon)$

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

Eq. **2.3** follows, and so also our main result.

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 $Q(x) = \text{Prob}\{X \le x\}$. It has density $q(x) = \exp(-x^2)/\sqrt{\pi}$. Writing things out explicitly,

$$Q(x) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{x} e^{-y^2}\,dy. \qquad \textbf{[3.1]}$$

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

$$\frac{2xQ(x)}{q(x)} = \frac{1-\nu}{\nu}, \qquad \textbf{[3.2]}$$

noting that possible values of x_ν are nonnegative. 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 Fig. 2 for a depiction.

3.2. Exponent for Internal Angle. Let Y be a standard half-normal random variable $HN(0, 1)$, with 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 (13)]

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

Λ^* is smooth and convex on $(0, \infty)$, strictly positive except at $\mu = EY = \sqrt{2/\pi}$. More details are provided in ref. 7.

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 $\gamma = \rho\delta/\nu \le \rho$,

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

APPLIED MATHEMATICS

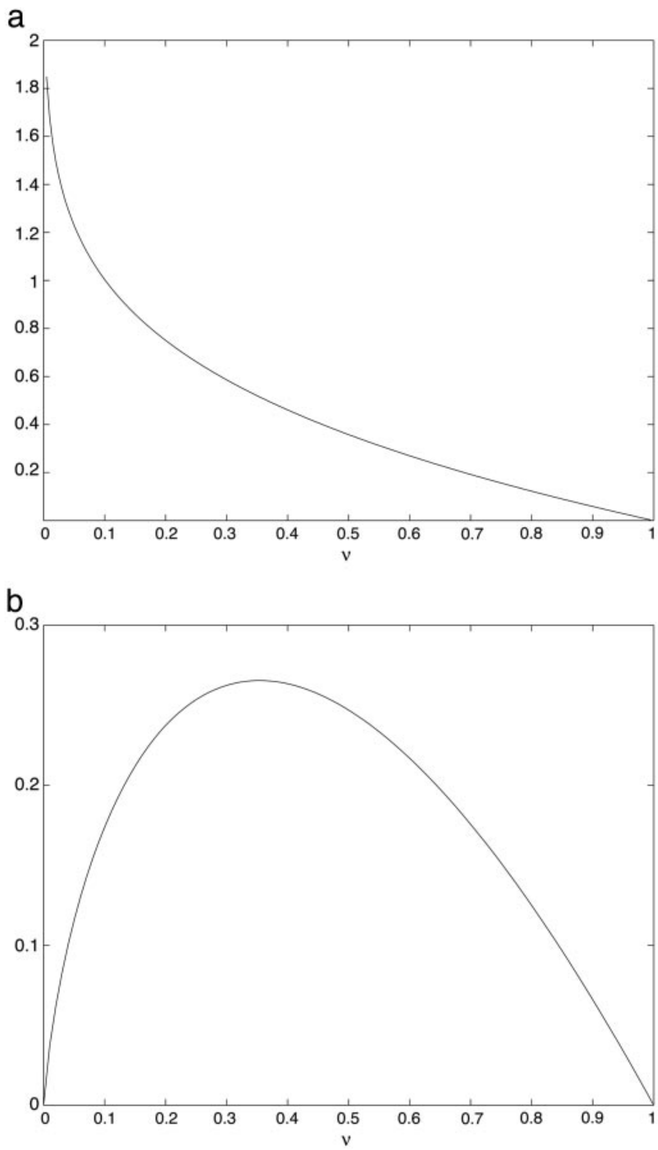

Fig. 2. Key notions associated with external angle. (*a*) The minimizer x_ν of ψ_ν, as a function of ν. (*b*) The exponent $\Psi_{\rm ext}$, a function of ν.

For fixed ρ, δ, $\Psi_{\rm int}$ is continuous in $\nu \geq \delta$. Most importantly, section 6 of ref. 7 gives the asymptotic formula

$$\xi_\gamma(y_\gamma) \approx \frac{1}{2} \cdot \log\left(\frac{1-\gamma}{\gamma}\right), \ \gamma \rightarrow 0. \qquad \textbf{[3.3]}$$

3.3. Combining the Exponents. We now consider the combined behavior of $\Psi_{\rm com}$, $\Psi_{\rm int}$, and $\Psi_{\rm ext}$. We think of these as functions of ν with ρ, δ as parameters. The combinatorial exponent $\Psi_{\rm com}$ involves a scaled, shifted version of the Shannon entropy, which is a symmetric, roughly parabolic shaped function. $\Psi_{\rm com}$ is the exponent of a growing function that must be outweighed by the sum $\Psi_{\rm ext} + \Psi_{\rm int}$.

Fig. 3 shows both $\Psi_{\rm com}$ and $\Psi_{\rm ext} + \Psi_{\rm int}$ with $\delta = 0.5555$ and $\rho = 0.145$. The desired condition $\Psi_{\rm net} < 0$ is the same as $\Psi_{\rm com} < \Psi_{\rm ext} + \Psi_{\rm int}$ and holds with plenty of slack except near $\nu = \delta$, where the two curves are close. We have $\rho_N(\delta) \approx 0.145$.

3.4. Justifying the Exponents. It remains to justify Eqs. **2.4** and **2.5**.

We sketch the argument for Eq. **2.5**. The key point is the closed-form expression for $\gamma(T^\ell, T^{n-1})$

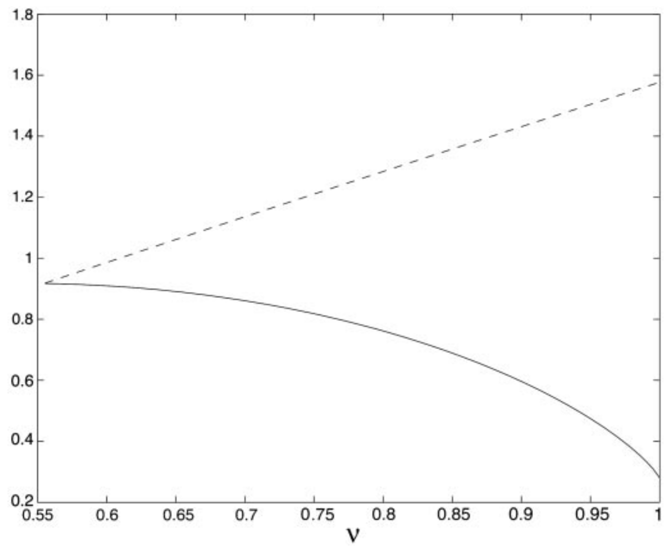

Fig. 3. The exponents $\Psi_{\rm com}(\nu; \rho, \delta)$ and $\Psi_{\rm int}(\nu; \rho\delta) + \Psi_{\rm ext}(\nu)$, for $\rho = 0.145$, $\delta = 0.5555$. The graph of $\Psi_{\rm com}$ (solid line) falls below that of $\Psi_{\rm int} + \Psi_{\rm ext}$ (dashed line), and so $\Psi_{\rm net} < 0$.

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

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

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

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 ref. 7, one obtains:

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

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

where, for δ, $\eta > 0$,

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

The minimizer x_ν mentioned in this lemma is the same x_ν defined earlier in Eq. **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},

$$\Psi_{ext}(\nu) = \psi_\nu(x_\nu). \quad [3.5]$$

Hence, Eq. **2.5** follows.

The decay estimate Eq. **2.4** for the internal angle was derived in ref. 7, and details can be found there. Vershik and Sporyshev (2) used a related but seemingly different approach. The argument starts from a closed-form integral expression for $\beta(T^k, T^\ell)$. By ref. 14, $\beta(T^k, T^\ell) = B(1/k+2, \ell - k + 1)$, where

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

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

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

It was shown in ref. 7 that Laplace's method applied to this last integral yields exponential bounds on the decay of β of the form Eq. **2.4**.

3.5. Properties of ρ_N. We mention two key facts about ρ_N. First, the concept is nontrivial.

Lemma 3.2.

$$\rho_N(\delta) > 0, \quad \delta \in (0, 1). \quad [3.8]$$

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

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

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

These results require only a simple observation. Ref. 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

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

Lower bounds completely parallel in form to those in *Lemmas 3.2* and *3.3* were already proven for $\rho_N^\pm$ in ref. 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 $k = k(n) \approx \rho\delta n$,

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

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

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

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 Eq. **4.1** implies

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

We immediately get Eq. **1.2,** showing that Eq. **4.1** implies Eq. **4.2** requires no new ideas; one proceeds as in Section 2 almost line by line, so we omit the exercise. □

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

D.L.D. thanks the Mathematical Sciences Research Institute for its "neighborly" hospitality in Winter 2005 while this work was prepared. D.L.D. was supported in part by National Science Foundation Division of Mathematical Sciences Grant 00-77261 and Focused Research Group Grant 01-40698, by the Clay Mathematics Institute, and by an Office of Naval Research Multidisciplinary University Research Initiative. J.T. was supported by National Science Foundation Division of Mathematical Sciences Fellowship 04-03041.

1. Affentranger, F. & Schneider, R. (1992) *Discrete Comput. Geometry* **7,** 219–226.
2. Vershik, A. M. & Sporyshev, P. V. (1992) *Selecta Math. Sov.* **11,** 181–201.
3. Baryshnikov, Y. M. & Vitale, R. A. (1994) *Discrete Comput. Geometry* **11,** 141–147.
4. Donoho, D. L. & Tanner, J. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 9446–9451.
5. Grünbaum, B. (2003) *Convex Polytopes*, Graduate Texts in Mathematics (Springer, New York), 2nd Ed., Vol. 221.
6. Donoho, D. L. (2005) *Neighborly Polytopes and the Sparse Solution of Underdetermined Systems of Linear Equations*, Technical Report 2005-04 (Department of Statistics, Stanford University, Stanford, CA).
7. Donoho, D. L. (2005) *High-Dimensional Centrosymmetric Polytopes with Neighborliness Proportional to Dimension*, Technical Report 2005-05 (Department of Statistics, Stanford University, Stanford, CA).
8. Donoho, D. L. & Tanner, J. *Neighborliness of Randomly-Projected Simplices in High Dimensions*, Technical Report 2005-06 (Department of Statistics, Stanford University, Stanford, CA).
9. Grünbaum, B. (1968) *Acta Math.* **131,** 293–302.
10. McMullen, P. (1975) *Math. Proc. Cambridge Philos. Soc.* **78,** 247–261.
11. Ruben, H. (1960) *Acta Math.* **103,** 1–23.
12. Donoho, D. L. (2004) *For Most Large Underdetermined Systems of Linear Equations the Minimal ℓ^1-Normal Solution Is Also the Sparsest Solution*, Technical Report 2004-10 (Department of Statistics, Stanford University, Stanford, CA).
13. Dembo, A. & Zeitouni, O. *Large Deviations Techniques and Applications*, Applications of Mathematics (Springer, New York), 2nd Ed., Vol. 38.
14. Böröczky, K., Jr., & Henk, M. (1999) *Arch. Math. (Basel)* **73,** 465–473.