Discrete & Computational

Geometry

© 2005 Springer Science+Business Media, Inc.

High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension*

David L. Donoho

Statistics Department, Stanford University,
Stanford, CA 94305, USA
donoho@stat.stanford.edu

Abstract. Let A be a d by n matrix, $d < n$. Let C be the regular cross polytope (octahedron) in $\mathbf{R}^n$. It has recently been shown that properties of the centrosymmetric polytope $P = AC$ are of interest for finding sparse solutions to the underdetermined system of equations $y = Ax$ [9]. In particular, it is valuable to know that P is centrally k-neighborly.

We study the face numbers of randomly projected cross polytopes in the proportional-dimensional case where $d \sim \delta n$, where 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 $P = AC$ is the same as for C, for $0 \leq k \leq \rho d$. This implies that P is centrally $\lfloor \rho d \rfloor$-neighborly, and its skeleton $Skel_{\lfloor \rho d \rfloor}(P)$ is combinatorially equivalent to $Skel_{\lfloor \rho d \rfloor}(C)$. We display graphs of ρ_N.

Two weaker notions of neighborliness are also important for understanding sparse solutions of linear equations: weak neighborliness and sectional neighborliness [9]; we study both. *Weak* (k, ε)-*neighborliness* asks if the k-faces are all simplicial and if the number of k-dimensional faces $f_k(P) \geq f_k(C)(1 - \varepsilon)$. We characterize and compute the critical proportion $\rho_W(\delta) > 0$ such that weak (k, ε) neighborliness holds at k significantly smaller than $\rho_W \cdot d$ and fails for k significantly larger than $\rho_W \cdot d$. *Sectional* (k, ε)-*neighborliness* asks whether all, except for a small fraction ε, of the k-dimensional intrinsic sections of P are k-dimensional cross polytopes. (Intrinsic sections intersect P with k-dimensional subspaces spanned by vertices of P.) We characterize and compute a proportion $\rho_S(\delta) > 0$ guaranteeing this property for $k/d \sim \rho < \rho_S(\delta)$. We display graphs of ρ_S and ρ_W.

1. Introduction

1.1. *Neighborliness and Central Neighborliness*

In the classical theory of convex polytopes, the notion of neighborliness offers a beautiful glimpse of the surprises of high dimensions. k-Neighborliness asks if every k vertices of a polytope span a $(k-1)$-face. In low dimensions this is difficult for beginners to arrange—outside the trivial case of the simplex—because it seems that some candidate edges easily get "swallowed up" crossing "inside" the polytope. It can be surprising to students that in higher dimensions, $d > 3$, this can be managed easily, by simply taking $n > d$ points $x_i = M(t_i)$ along the moment curve $M(t) = (t, t^2, \ldots, t^d)$ [12], [14]. The convex hull of these points is a polytope with n vertices which is $\lfloor d/2 \rfloor$-neighborly, for each $n > d$; for $n > d + 1$. this is the maximum possible value. See, e.g., Chapter 7 of [14] for more.

For centrosymmetric polytopes, a modified notion of neighborliness is needed; one asks if every k vertices *not including an antipodal pair* span a $(k-1)$-face. This property is called *central k-neighborliness*. The "not including" proviso detracts a bit from the beauty of the notion; and perhaps also from the interest in studying it. There is no known general construction of centrally k-neighborly for large n and d, and the achievable upper bound is smaller: $k \leq \lfloor (d+1)/3 \rfloor$, according to McMullen and Shephard [17]. For n not much larger than d, Schneider [21] showed the existence of centrally symmetric polytopes which are centrally k-neighborly for $k \approx 0.2309d$; however, Schneider's polytopes have only $2d(1 + o(1))$ vertices. Burton [4] showed that for fixed d and large enough n, even 2-central-neighborliness is impossible. Until recently, not much else was known.

1.2. *Central Neighborliness and Optimization*

In a companion paper [9], the author shows that central neighborliness of centrally symmetric polytopes is important for understanding solvability of certain combinatorial optimization problems by convex relaxation. Specifically, suppose A is a d by n matrix with $d < n$ and we are interested in finding the solution to the underdetermined system $y = Ax$ having *fewest nonzeros*. Although this problem is NP-hard in general, the sparsest solution can be often found by solving the *convex* optimization problem $\min \|x\|_1$ subject to $y = Ax$. The conditions on A and y guaranteeing success are: first, that a solution with at most k nonzeros exist; and, secondly, that the convex polytope $P = AC$ be centrally k-neighborly. Here C denotes the cross polytope (ℓ^1 ball) in $\mathbf{R}^n$.

The relation to optimization brings new significance into the study of neighborliness in the centrosymmetric case. As [9] shows, we can interpret recent results in the study of sparse solutions by ℓ^1 optimization as constructions of centrosymmetric polytopes which are centrally k-neighborly for reasonably large k. For example, a result of the author [8] relying on Banach space geometry implies that for large d and n, d proportional to n, if we randomly take points $x_1, \ldots, x_n$ from the uniform distribution on the unit sphere in $\mathbf{R}^d$, then the centrosymmetric polytope generated by taking the convex hull of these points and their antipodes is overwhelmingly likely to be k-neighborly, for $k < \rho d$. Here ρ is a positive constant depending on n/d; until now little was known about the possible values for ρ.

Clearly, we would like to know more about the possible/prevalent ranges of neighborliness.

1.3. *Analysis in High Dimensions*

In this paper we adopt the high-dimensional viewpoint, and construct polytopes by projecting from n dimensions down to d dimensions, n large, d proportional to n. The resulting families of high-dimensional centrosymmetric polytopes are *proportionally neighborly*, in the sense that for some $\rho > 0$ and large d, typical realizations are centrally $\lfloor \rho d \rfloor$-neighborly. Our approach gives quantitative information about the size of ρ achievable. We present numerical evidence that $k \geq 0.089d$ when $n = 2d$ and n is large.

Our analysis considers the ensemble of polytopes $P = AC$ where A is a random projection from $\mathbf{R}^d$ to $\mathbf{R}^n$ and $C = C^n$ is the standard cross polytope. We study a function $\rho_N \colon (0, 1] \mapsto [0, 1]$, depicted in Fig. 1.1 and defined in detail later. The definition is unfortunately implicit, involving large deviation properties of certain random variables (see Sections 3.4 and 4.4), but sufficiently concrete that accurate numerical calculations are possible. ρ_N provides a lower bound on the degree of central neighborliness of the random polytope P.

Corollary 1.1. *Let A be a uniform random projection from $\mathbf{R}^n$ to $\mathbf{R}^d$ with $d = \lfloor \delta n \rfloor$. Fix $\varepsilon > 0$. With overwhelming probability for large n, $P = AC$ is centrally k-neighborly with $k/d \geq \rho_N(\delta) - \varepsilon$.*
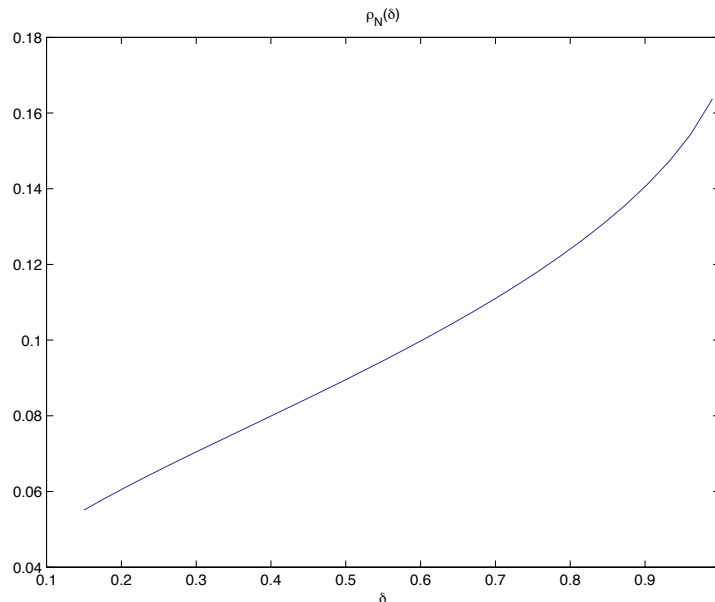

Fig. 1.1. The lower bound $\rho_N(\delta)$ on the neighborliness threshold, computed by methods of this paper. MATLAB software available from the author.

Much of the paper is concerned with introducing the components in the definition of ρ_N and proving the validity of its role as a lower bound on neighborliness.

1.4. *Face Numbers*

In fact, this article does not much discuss neighborliness per se. Instead, we consider the properties of face numbers of the projected cross polytope, getting the following result:

Theorem 1. *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 $d \geq \delta n$. Let $f_\ell(C)$ and $f_\ell(AC)$ denote the number of ℓ-dimensional faces of C and AC, respectively. Then*

$$Prob\{f_\ell(AC) = f_\ell(C), \ \ell = 0, \ldots, \lfloor \rho d \rfloor\} \to 1, \qquad as \quad n \to \infty. \qquad (1.1)$$

Central k-neighborliness follows from this equality of face numbers; see Section 2 below.

Our proof of Theorem 1 starts from work of Böröczky and Henk [2], who considered face numbers of the randomly projected cross polytope with d fixed and $n \to \infty$. We modify the analysis, letting d and n both go to infinity in a proportional way. The approach of [2] depends on the framework for computing Grassmann angles of a polytope due to Affentranger and Schneider [1] and Vershik and Sporyshev [23]. This uses exact analytical work in integral geometry of convex sets by McMullen [16] (nonlinear angle-sum relations), Grünbaum [13] (Grassmann angles), and Ruben [18] (volumes of spherical simplices).

Our approach is to develop formulas for the internal and external angles of cross polytope faces in the n-proportional-to-d setting, obtaining inequalities of a substantially different form than in the d-fixed setting. We use these inequalities to characterize and compute $\rho_N(\delta)$.

The study of face numbers in the proportional-dimensional case, where $d \sim \delta n$, was pioneered by Vershik and Sporyshev [23] in the "projection of simplex" case $P = AT^n$, with T^n the regular simplex in $\mathbf{R}^n$. Most importantly, Vershik and Sporyshev [23] developed, in addition to the proportional-to-dimension viewpoint, several analytical tools relevant to the proportional-dimensional case, for studying internal and external angles of simplices; these are also used here.

1.5. *Weaker Notions of Neighborliness*

Vershik and Sporyshev [23] were interested in the question of whether, for k in a fixed proportion to n, the face numbers obeyed $f_k(AT^n) = f_k(T^n)(1 + o_P(1))$. (Here we use standard "order in probability" notation o_P; a sequence of random variables Y_n is $o_P(1)$ if it tends to zero in probability, or, equivalently, if for all $\varepsilon > 0$, $Prob\{|Y_n| > \varepsilon\} \to 0$, $n \to \infty$.) Thus, they wanted to know if, for large n, the overwhelming majority of random projections A, yielded AT^n with approximately the same number of k-faces as T^n. The answer exhibited sharp threshold behavior in the vicinity of $\rho_{VS} \cdot d$, for some implicitly characterized $\rho_{VS} = \rho_{VS}(d/n)$.

For comparison with Theorem 1, note that the question of *approximate* equality of face numbers $f_k(AT^n) = f_k(T^n)(1 + o_P(1))$ is weaker than the *exact* equality studied here in Theorem 1; it changes at a different threshold in k/d. The comparable question in our setting is approximate equality of face numbers $f_k(AC) = f_k(C)(1 + o_P(1))$.

This notion may be called *weak (central) neighborliness*; it asks for the *typical k*-tuple of vertices to span a $k - 1$ face, rather than for *every k*-tuple of vertices to span a face.

It turns out that weak central neighborliness exhibits sharp threshold behavior in k/d, in a fashion similar to what Vershik and Sporyshev found in the non-centrosymmetrric case, but at a different threshold. The threshold function ρ_W is again implicitly defined in terms of certain large deviation exponents, but is amenable to high-accuracy numerical calculations. Figure 1.2 displays thresholds computed using the following result.

Theorem 2. *There is a function $\rho_W(\delta)$, characterized 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 any sequence $k = k_n$ with $k/d \to \rho$, $\rho < \rho_W(\delta)$, we have*

$$f_k(AC) = f_k(C)(1 + o_P(1)). \tag{1.2}$$

Theorem 2 is sharp in the sense that for sequences with $k/d \to \rho > \rho_W$ there is no more the approximate equality (1.2); but we do not prove this here. Thus, we distinguish between ρ_W which is really a threshold and ρ_N which is a lower bound on a threshold.
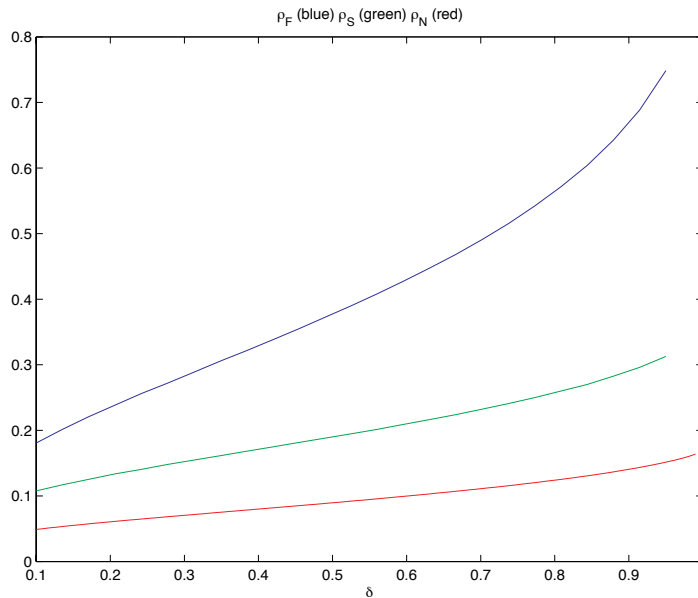

Fig. 1.2. The threshold $\rho_W(\delta)$ for approximate equality of ρd-dimensional face numbers of C and AC (blue), and the lower bound $\rho_S(\delta)$ for sectional neighborliness (green). Plot of ρ_N overlaid in red for comparison.

The companion paper [9] discusses applications of weak neighborliness. This notion of neighborliness is easier to satisfy than orthodox central neighborliness and so $\rho_W > \rho_N$. A notion of *sectional* neighborliness, intermediate between weak neighborliness and neighborliness, is also defined in [9]. In this notion we take any k vertices not including an antipodal pair and section P by the linear subspace spanned by those vertices. If the overwhelming majority of such sections are k-dimensional cross polytopes, we say that P is typically *sectionally k-neighborly*. In Fig. 1.2 we also display a bound on the sectional neighborliness of quotient polytopes, based on the following result.

Theorem 3. *There is a function $\rho_S(\delta)$, characterized below, with the following property. Let $\rho < \rho_S(\delta)$ and let A be a uniform random projection from $\mathbf{R}^n$ to $\mathbf{R}^d$, with $d = d_n \sim \delta n$. Then for $k = k_n \sim \rho d_n$, we have with overwhelming probability for large n that $P = AC$ is typically sectionally k-neighborly.*

All three theorems are proved in more or less the same way; we spend the bulk of this article on the proof of Theorem 1 and in a final section indicate the changes needed to prove Theorems 2 and 3.

Figure 1.2 depicts substantial numerical differences in the critical proportion ρ_W and the lower bounds ρ_N and ρ_S. The most striking differences between ρ_W and the other two proportions are that ρ_W crosses the line $\rho = \frac{1}{2}$ near $\delta = 0.701$ and increases to 1 as $\delta \to 1$. This means that if $n > d + 1$, although the degree of central neighborliness can never exceed $d/2$, weak neighborliness can certainly exceed $d/2$. Although if $n > d + 1$ there must always be collections of $d/2 + 2$ vertices which do not span $d/2 + 1$ faces, for these random polytopes such collections are quite rare. The Appendix proves the following.

Theorem 4.

$$\lim_{\delta \to 1} \rho_W(\delta) = 1. \tag{1.3}$$

For some $\delta_0 \in (0, 1)$ (numerically $\delta_0 \approx 0.701$)

$$\rho_W(\delta) > \tfrac{1}{2}, \qquad \delta_0 < \delta < 1. \tag{1.4}$$

For comparison, one can compute that

$$0.168 \approx \lim_{\delta \to 1} \rho_N(\delta) \tag{1.5}$$

and

$$0.352 \approx \lim_{\delta \to 1} \rho_S(\delta). \tag{1.6}$$

These comparisons can be very interesting from the applications viewpoint, where they can be interpreted as saying that average-case behavior is far more favorable than worst-case behavior. See the discussion in [9].

1.6. *Recent Work*

When this research was done, it seemed that the study of high-dimensional centrally neighborly polytopes was a lonely endeavor; the history such as is was recounted above at the beginning of the Introduction. While this manuscript was under review, interest seems to have picked up, driven by activity in signal processing and geometric functional analysis. As indicated in [9], certain results in the field of signal processing can be re-interpreted as implicitly establishing the existence of high-dimensional neighborly polytopes; [9] gives a pointer to activity in that field. That activity led to [8], which showed combinatorial isomorphism between the k-skeleta of certain high-dimensional polytopes in $\mathbf{R}^d$ with n vertices and the k-skeleton of the cross polytope $C \subset \mathbf{R}^n$. This of course implies k-neighborliness. The key point was that k could be proportional to d: for some unspecified $\rho^*(\delta)$, i.e. that there is k-neighborliness with $k \sim \rho^*(\delta)d(1 + o_P(1))$ and $d \sim \delta n$.

Paper [8] used techniques from geometry of Banach spaces and eigenvalues of random matrices rather than polytope techniques. It gave weaker, more qualitative information about neighborliness compared with what has been obtained here. Nevertheless, that qualitative result gave a hint that geometric Banach space theory and related tools could be brought to bear on the neighborliness question. In fact, the other recent work relevant to neighborliness all takes the geometric functional analysis viewpoint.

The author's papers [8], [7] implicitly showed that $\rho^*(\delta) \geq c/\log(1/\delta)$, and Candès and Tao [5] explicitly proved results which can be seen to imply $\rho^*(\delta) \geq c/\log(1/\delta)$. After submission of this manuscript, two other manuscripts appeared which cover related territory. Their methods are again based on geometric Banach space techniques and their conclusions are again qualitative. Rudelson and Vershynin [20] develop a viewpoint dual to ours, concerning the number of faces of sections of high-dimensional cubes. Their work can be seen to imply that for the randomly projected cross polytope $P = AC$ as studied in this article, k-neighborliness holds for $k \approx c/\log(1/\delta)(1 + o_P(1))$, for unspecified c. Linial and Novik [15] explicitly study neighborliness of projected cross polytopes and obtain a similar implication, along with a matching upper bound showing no qualitatively better behavior is possible for small δ.

Incidentally, the success of Banach space techniques in giving qualitative insights on this problem is a vindication of the intuitions of Vershik and Sporyshev. They had already suggested in [23] that their results on neighborliness of randomly projected simplices were in some way connected to ongoing developments in Banach space geometry.

Final note: the very recent papers of the author and Tanner [10], [11] apply the techniques of this article to the study of randomly projected simplices and their applications.

2. Neighborliness and Face Numbers

We first justify our claim that face numbers of the quotient polytope alone are enough to determine neighborliness.

We also fix notation concerning convex polytopes; see [14] for more details. In discussing the (closed, convex) polytope P we commonly refer to its vertices $v \in \text{vert}(P)$ and k-dimensional faces $F \in \mathcal{F}_k(P)$. A point $v \in P$ will be called a vertex of P if

there is a linear functional λ_v separating v from $P \setminus \{v\}$, i.e. a value c so that $\lambda_v(v) = c$ and $\lambda_v(x) < c$ for $x \in P$, $x \neq c$. We write conv for the convex hull operation; thus $P = \text{conv}(\text{vert}(P))$. Vertices are just 0-dimensional faces, and a k-dimensional face is a k-dimensional set F for which there exists a separating linear functional λ_F, so that $\lambda_F(x) = c$, $x \in F$, and $\lambda_F(x) < c$, $x \notin F$. Faces are convex polytopes, each one representable as the convex hull of a subset $\text{vert}(F) \subset \text{vert}(P)$; thus if F is a face, $F = \text{conv}(\text{vert}(F))$. A k-dimensional face will be called a *k-simplex* if it has $k + 1$ vertices. We note the elementary

$$\text{vert}(AC) \subset A\,\text{vert}(C),$$

which implies

$$f_0(AC) \leq f_0(C); \tag{2.1}$$

in fact, we have more generally $f_k(AC) \leq f_k(C)$, $0 \leq k < d$; this hints at the fact that projections can only "lose faces," or at least merely preserve them. The following lemma shows that neighborliness of $P = AC$ is equivalent to not "losing faces" under the projection $C \mapsto AC$. The statement and proof given here are due to a referee.

Lemma 2.1. *Let A be an arbitrary linear transformation. Let $P = AC$ have the same number of vertices and $(k-1)$-faces as C: $f_\ell(P) = f_\ell(C)$, $\ell = 0, k-1$. Then:*

- *The $(k-1)$-skeleton of P is isomorphic to the $(k-1)$-skeleton of C; in particular, $f_\ell(P) = f_\ell(C)$, and every ℓ-face of P is an ℓ-simplex, for $\ell = 0, \ldots, k-1$.*
- *P is centrally k-neighborly.*

Proof. Let Q be a centrosymmetric polytope with $2n$ vertices in $\mathbf{R}^d$ ($n \geq d$). Let $\mathcal{T}_k$ be the set of all k-tuples of vertices of Q not containing an antipodal pair. Then $\#\mathcal{T}_k = 2^k \binom{n}{k}$. Let $F \in \mathcal{F}_{k-1}(Q)$. The face F is the convex hull of suitable vertices of Q. Among these vertices, there are k affinely independent ones, and there is no antipodal pair. Hence, with each $F \in \mathcal{F}_{k-1}(Q)$ we can associate at least one k-tuple in $\mathcal{T}_k$ which affinely spans the affine hull of F. The k-tuples associated with different $(k-1)$-faces are distinct. It follows that $f_{k-1}(Q) \leq 2^k \binom{n}{k}$. If equality holds here, then the vertex set of each $F \in \mathcal{F}_{k-1}(Q)$ contains exactly one affinely independent k-tuple, hence each such F is a simplex, and every k-tuple of $\mathcal{T}_k$ is the vertex set of a $(k-1)$-face of Q. For $\ell < k$, every ℓ-tuple of vertices of Q containing no antipodal pair is contained in some k-tuple of $\mathcal{T}_k$ and hence determines an $(\ell-1)$-simplex which is a face of Q. It is now clear that any bijection between the vertices of Q and the vertices of C preserving antipodal pairs establishes a combinatorial isomorphism between the $(k-1)$-skeleta of Q and C. This implies the assertions of the lemma. $\qquad\square$

3. Random Projections of Cross Polytopes

We now outline the proof of Theorem 1. Key lemmas and inequalities are justified in later sections.

3.1. *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 [23] and [2]) studied the properties of random projections $R = AQ$ where Q is an n-polytope and R is its d-dimensional orthogonal projection. They [1] derived the formula

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

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 $\gamma(G, Q)$ is the external angle of Q at face G; for definitions of these terms see, e.g. Chapter 14 of [14].

The slogan underlying the formula is that each face $F \in \mathcal{F}_k(Q)$ will either "survive" under projection, so that AF is a k-face of R, or it will get "swallowed up" inside R. The expected number of faces in R is thus the number of faces in Q minus the expected number of faces "swallowed up" in projection. The chance of a particular face's getting "swallowed up" is exactly the chance that the subspace spanned by columns of A^t in $\mathbf{R}^n$ intersects trivially with the cone of separating linear functionals associated to face $F \in Q$. The chance that a uniform random subspace hits a cone is precisely the so-called Grassmann angle as defined by Grünbaum [13]. Hence the expected number of faces $f_k(R)$ involves a sum of Grassmann angles, one for each k-face F of Q, evaluating the probability that AF is a k-face of R. McMullen [16] developed nonlinear angle-sum relations which are used to decompose these Grassmann angles into the above sums involving internal and external angles.

Specializing to the case where $Q = C$, the n-dimensional cross polytope, we write

$$Ef_k(P) = f_k(C) - \Delta(k, d, n) \tag{3.1}$$

with

$$\Delta(k, d, n) = 2\sum_{s \geq 0} \sum_{F \in \mathcal{F}_k(C)} \sum_{G \in \mathcal{F}_{d+1+2s}(C)} \beta(F, G)\gamma(G, C). \tag{3.2}$$

3.2. *Exact Equality from Expectation*

Because of (2.1) we view (3.1) as showing that on average $f_k(P)$ is about the same as $f_k(C)$, except for a nonnegative 'discrepancy' Δ. We will show that under the stated conditions on k,d, and n, for some $\varepsilon > 0$,

$$\Delta(k, d, n) \leq n \exp(-n\varepsilon), \qquad n > n_0(\rho, \delta, \varepsilon). \tag{3.3}$$

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

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

The valid range for k in (3.3) includes k as in the statement of the theorem. Thus with overwhelming probability we get equality of $f_k(P)$ with $f_k(C)$, as claimed in

the theorem. To extend this into the needed simultaneous result—that $f_\ell(P) = f_\ell(C)$, $\ell = 0, \ldots, k-1$—one defines events $E_\ell = \{f_\ell(P) \neq f_\ell(C)\}$ and notes that by Lemma 2.1, the event $(E_0 \cup E_k)^c$ implies $(E_0 \cup \ldots \cup E_k)^c$. Hence by Boole's inequality

$$Prob\left(\bigcup_0^k E_\ell\right) \leq Prob(E_0) + Prob(E_k) \leq \Delta(0, d, n) + \Delta(k, d, n).$$

The analog of (3.3) also holds for $\Delta(0, d, n)$. Hence by establishing (3.3) we get

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

as is to be proved.

To establish (3.3), we rewrite (3.2) as

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

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

$$D_s = 2 \cdot \sum_{F \in \mathcal{F}_k(C)} \sum_{G \in \mathcal{F}_\ell(C)} \beta(F, G)\gamma(G, C).$$

We will show that, for a sequence $(k, d) = (k_n, d_n)$ with $k_n/d_n \to \rho$, $d_n/k_n \to \delta$, where $\rho < \rho_N(\delta)$ (the function ρ_N still to be defined) and for sufficiently small $\varepsilon > 0$, then for $n > n_0(\rho, \delta; \varepsilon)$,

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

This implies (3.3) and hence our main result follows.

3.3. *Decay and Growth Exponents*

Böröczky and Henk [2] studied exactly the setting $P = AC$ with C the cross polytope—though for a different range of k, d, n (they considered k, d fixed and $n \to \infty$), and also used a different formula for $Ef_k(P)$, so they did not directly study the term $\Delta(k, d, n)$. They did, however, make the following useful observations:

- There are $2^{k+1}\binom{n}{k+1}$ k-faces of C.
- For $\ell > k$, there are $2^{\ell-k}\binom{n-k-1}{\ell-k}$ ℓ-faces of C containing a given k-face of C.
- The faces of C are all simplices, and for faces $F \in \mathcal{F}_k(C)$, $G \in \mathcal{F}_\ell(C)$, the internal angle $\beta(F, G) = \beta(T^k, T^\ell)$, where T^d denotes the standard d-simplex.
- The external angle $\gamma(G, C)$ is the same for all ℓ-faces G of C; it has a closed form integral expression very similar to $\gamma(T^\ell, T^n)$.

Thus we can write

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

$$= \alpha_s \, \beta(T^k, T^\ell)\gamma(F^\ell, C),$$

say, with α_s the combinatorial prefactor.

We now plan to 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. This viewpoint turns our problem into the study of exponents of exponentially growing and decaying terms. To get a first taste, consider 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 [23], we also remark that

$$n^{-1}\log\binom{n}{\lfloor pn\rfloor} \to H(p), \qquad p \in [0,1], \quad n \to \infty, \qquad (3.4)$$

so that $H(p)$ measures the exponential growth of combinatorial terms. Relations of this kind can be developed for α_s and for the internal and external angles. We introduce the new variable $\nu = \ell/n \geq \delta$, and we suppose that $k/d \approx \rho$ and $d/n \approx \delta$. Operating formally, we obtain

$$n^{-1}\log(\alpha_s) = \nu\log(2) + H(\rho\delta) + H\left(\frac{\nu - \rho\delta}{1 - \rho\delta}\right)(1 - \rho\delta) + R_1 \qquad (3.5)$$

with remainder $R_1 = R_1(s, k, d, n, \rho, \delta)$. Next, we introduce the *growth* exponent

$$\Psi_{\text{com}}(\nu; \rho\delta) \equiv \nu\log(2) + H(\rho\delta) + H\left(\frac{\nu - \rho\delta}{1 - \rho\delta}\right)(1 - \rho\delta),$$

and claim that it describes the exponential growth of the combinatorial factor α_s. To make this notion precise, we describe bounds on R_1 uniform in various parameters. By refining the analysis behind (3.4) one can easily show that, for fixed $\varepsilon > 0$, $\rho \in (0, 1)$ and $\delta \in (0, 1)$, there is $\eta > 0$ and $n_1(\rho, \delta; \varepsilon, \eta)$ so that

$$n^{-1}\log(\alpha_s) \leq \Psi_{\text{com}}(\nu; \rho\delta) + \varepsilon,$$

uniformly in $n > n_1(\rho, \delta; \varepsilon, \eta)$, $|k/d - \rho| < \eta$, $|d/n - \delta| < \eta$ and $\ell = d + 1 + 2s$, $s = 0, 1, 2, \ldots, \lfloor(n - d - 1)/2\rfloor$.

Section 4.1 below defines a so-called *decay* exponent $\Psi_{\text{ext}}(\nu)$. Section 5 shows that $\gamma(F^\ell, C)$ decays exponentially at least at the rate $\Psi_{\text{ext}}(\nu)$; for each $\varepsilon > 0$,

$$n^{-1}\log(\gamma(F^\ell, C)) \leq -\Psi_{\text{ext}}(\nu) + \varepsilon,$$

uniformly in $\ell \geq \delta n$, $n \geq n_2(\delta; \varepsilon)$. The graph of Ψ_{ext} is depicted in Fig. 4.1.

Similarly, Section 4.2 below defines a decay exponent $\Psi_{\text{int}}(\nu; \rho\delta)$. Section 6 below shows that the internal angle $\beta(T^k, T^\ell)$ indeed decays with this exponent; thus

$$n^{-1}\log(\beta(T^k, T^\ell)) \leq -\Psi_{\text{int}}(\nu; \rho\delta) + \varepsilon,$$

uniformly in $n > n_3(\rho, \delta; \varepsilon, \eta)$, $|k/d - \rho| < \eta$, $|d/n - \delta| < \eta$ and $\ell \geq \delta n$.

(Remark: these upper bounds on the α_s's β's and γ's possess matching lower bounds, but we have no need for them in this article.)

Hence for any fixed choice of ρ, δ, for $\varepsilon > 0$, and for $n \geq n_0(\rho, \delta; \varepsilon)$ we have the inequality

$$n^{-1}\log(D_s) \leq \Psi_{\text{com}}(\nu; \rho\delta) - \Psi_{\text{int}}(\nu; \rho\delta) - \Psi_{\text{ext}}(\nu) + 3\varepsilon, \qquad (3.6)$$

valid uniformly in s.

3.4. *Defining ρ_N*

Now define 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. We will see 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

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

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

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

Combine this with (3.6). Then for all $s = 0, 2, \ldots, \lfloor (n - d - 1)/2 \rfloor$ and all $n > n_0(\delta, \rho; \varepsilon)$,

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

This implies (3.3) and our main result follows.

4.   Properties of Exponents

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

4.1. *Exponent for External Angle*

Let G denote the cumulative distribution function of a half-normal $HN(0, \frac{1}{2})$ random variable, i.e. a random variable $X = |Z|$ where $Z \sim N(0, \frac{1}{2})$, and $G(x) = Prob\{X \leq x\}$. It has density $g(x) = 2/\sqrt{\pi} \exp(-x^2)$. Writing this out,

$$G(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-y^2) \, dy; \tag{4.1}$$

so G is just the classical error function *erf*. For $\nu \in (0, 1]$, define x_ν as the solution of
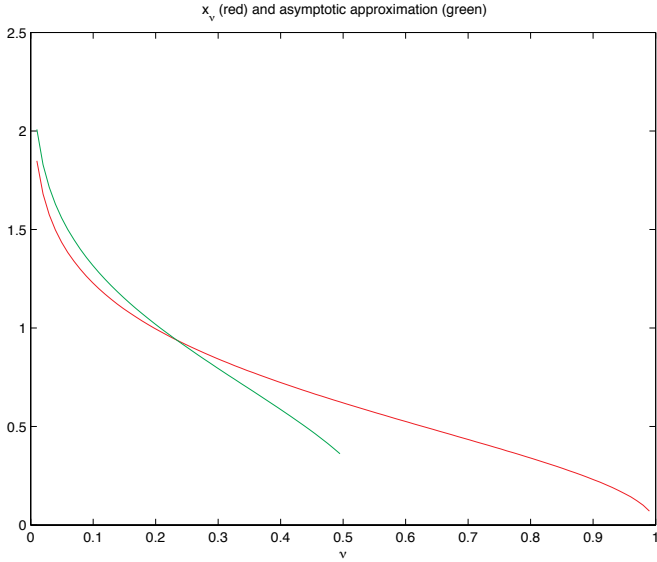
$$\frac{2xG(x)}{g(x)} = \frac{1 - \nu}{\nu}. \tag{4.2}$$

Since xG is a smooth strictly increasing function on $[0, \infty)$ with value 0 at $x = 0$ and $g(x)$ is strictly decreasing, the function $2xG(x)/g(x)$ is one–one on the positive axis. So x_ν is well-defined, and a smooth, decreasing function of ν. See Fig. 4.1 for a depiction.
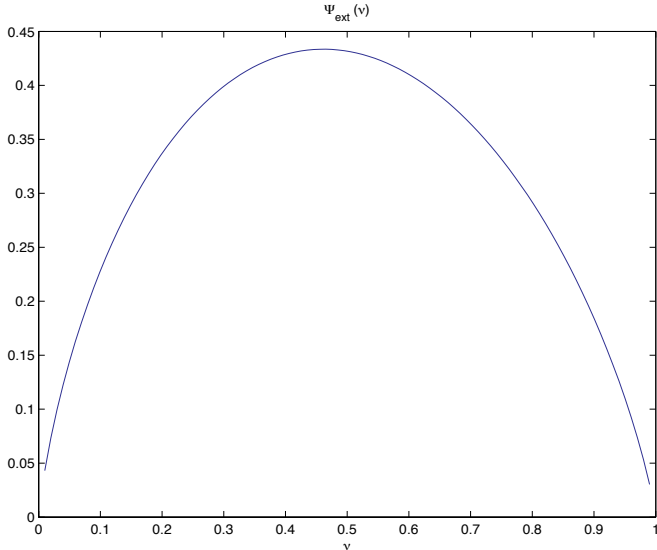
This has limiting behavior $x_\nu \to 0$ as $\nu \to 1$ and $x_\nu \sim \sqrt{\log((1 - \nu)/(2\nu))}$ as $\nu \to 0$. Define now

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

This is depicted in Fig. 4.1.

(a)

(b)

Fig. 4.1. (a) The minimizer x_ν of ψ_ν (see (5.3)), as a function of ν (red) and the asymptotic approximation $\sqrt{\log(1/\sqrt{\pi}\nu)}$ (green). (b) The exponent Ψ_{ext}, a function of ν.

This function is smooth on the interior of $(0, 1)$ and concave, with endpoints $\Psi_{\text{ext}}(1) = 0$, $\Psi_{\text{ext}}(0) = 0$. A useful fine point is the asymptotic

$$\Psi_{\text{ext}}(\nu) \sim \nu \log\left(\frac{1}{\nu}\right) - \frac{1}{2}\nu \log\left(\log\left(\frac{1}{\nu}\right)\right) + O(\nu), \qquad \nu \to 0. \qquad (4.3)$$

4.2. *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 [6])

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

This is smooth and convex on $(0, \infty)$, strictly positive except at $\mu = EY = \sqrt{2/\pi}$. More details are provided in Section 6. See Fig. 4.2.

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

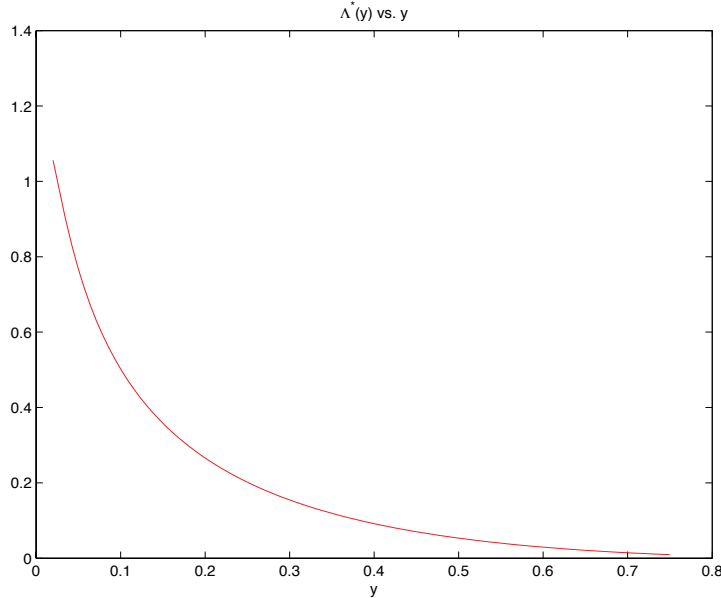$$\xi_\gamma(y) = \frac{1 - \gamma}{\gamma}\frac{y^2}{2} + \Lambda^*(y).$$

Fig. 4.2. $\Lambda^*(y)$, rate function for half-normal distribution; only the 'left half' $0 < y < \mu$ is depicted. The function diverges at zero.

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

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

This is depicted in Fig. 4.3. For fixed ρ, δ, Ψ_{int} is continuous in $\nu \geq \delta$. Most importantly, in Section 6.4 below we get the asymptotic formula

$$\xi_\gamma(y_\gamma) \sim \tfrac{1}{2} \cdot \log\left(\frac{1 - \gamma}{\gamma}\right), \qquad \gamma \to 0. \tag{4.4}$$

Since $\gamma = \rho\delta/\nu \leq \rho$, (4.4) implies that for given $\eta > 0$ and small ρ,

$$\Psi_{\text{int}}(\nu; \rho\delta) \geq \left(\tfrac{1}{2} \cdot \log\left(\frac{1 - \rho}{\rho}\right)(1 - \eta) + \log(2)\right)(\nu - \rho\delta), \qquad \nu \in [\delta, 1]. \tag{4.5}$$

4.3. *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} is the sum of a linear function in ν, and a scaled, shifted version of the Shannon entropy, which is a symmetric, roughly parabolic shaped function. This is the exponent of a growing function which must be outweighed by the sum $\Psi_{\text{ext}} + \Psi_{\text{int}}$. It is depicted in Fig. 4.3.
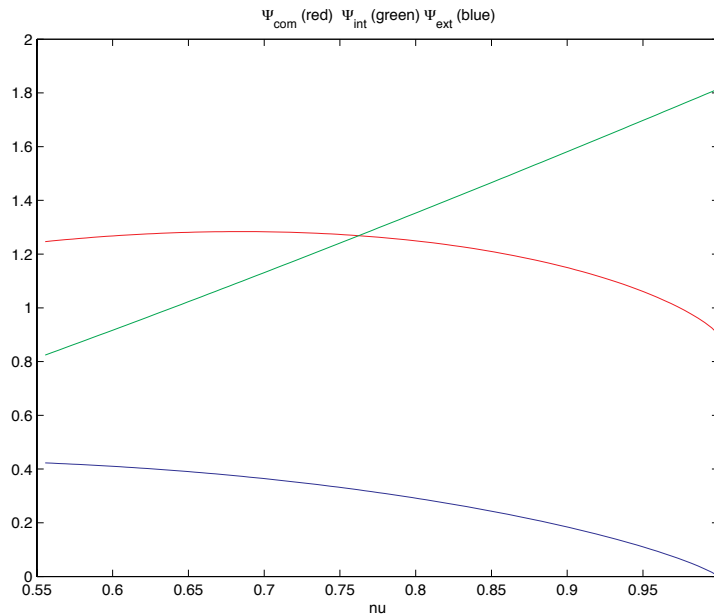

Fig. 4.3. The exponents $\Psi_{\text{com}}(\nu; \rho\delta)$ (red) and $\Psi_{\text{int}}(\nu; \rho\delta)$ (green), for $\rho = 0.095$, $\delta = 0.5555$. For comparison, Ψ_{ext} is displayed in blue.
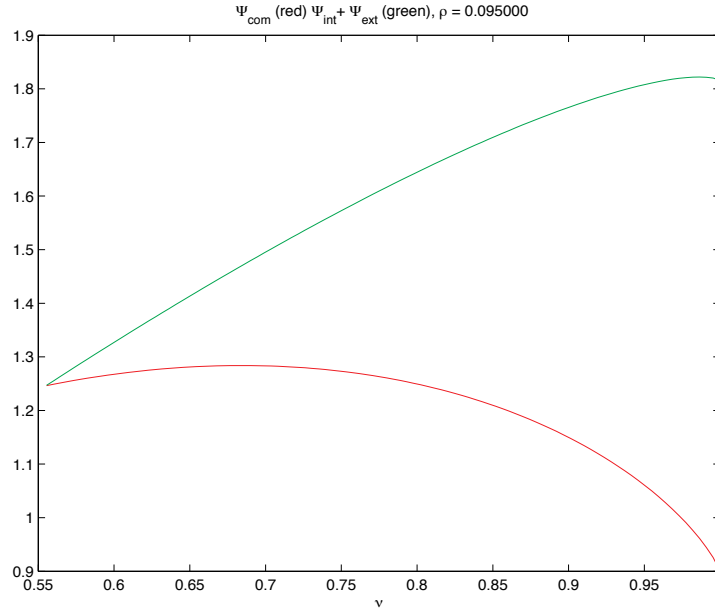
Fig. 4.4. The exponents $\Psi_{\text{com}}(\nu; \rho\delta)$ and $\Psi_{\text{int}}(\nu; \rho\delta) + \Psi_{\text{ext}}(\nu)$, for $\rho = 0.095$, $\delta = 0.5555$. The graph of Ψ_{com} (red) falls below that of $\Psi_{\text{int}} + \Psi_{\text{ext}}$ (green) and so $\Psi_{\text{net}} < 0$.

Figure 4.4 shows both Ψ_{com} and $\Psi_{\text{ext}} + \Psi_{\text{int}}$ with $\delta = 0.5555$ and $\rho = 0.095$. The desired condition $\Psi_{\text{net}} < 0$ is the same as $\Psi_{\text{com}} < \Psi_{\text{ext}} + \Psi_{\text{int}}$, and this is distinctly obeyed except near $\nu = \delta$, where the two curves are close. We have $\rho_N(\delta) \approx 0.095$.

4.4.   *Properties of ρ_N*

The asymptotic relations (4.5) and (4.3) allow us to see two key facts about ρ_N, both proved in the Appendix. Firstly, the concept is nontrivial:

Lemma 4.1.

$$\rho_N(\delta) > 0, \qquad \delta \in (0, 1). \tag{4.6}$$

This result was to be expected. Exploiting [9], [8] and [5] it could have previously been inferred that, for some $\rho = \rho(\delta) > 0$, such random polytopes are, with overwhelming probability, $\lfloor \rho d \rfloor$-neighborly. Effectively, (4.6) shows that the techniques of this paper are at least as strong as those of [8] and [5].

Secondly, one can show that, although $\rho_N(\delta) \to 0$ as $\delta \to 0$, it goes to zero slowly. We prove the following in the Appendix.

Lemma 4.2. *For $\eta > 0$,*

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

Again this result could have been anticipated. Exploiting the connection with studies of ℓ^1 optimization [9], we could have inferred from [7] and [5], that in the case $n \sim d^\gamma$, $\gamma > 1$, AC is centrally k-neighborly with $k \geq cd/\log(n)$.

5. Bounds on the External Angle

We now justify the use of Ψ_{ext}.

Lemma 5.1. *Fix $\delta, \varepsilon > 0$.*

$$n^{-1}\log(\gamma(F^\ell, C)) \leq -\Psi_{\text{ext}}(\ell/n) + \varepsilon, \tag{5.1}$$

uniformly in $\ell \geq \delta n$, $n \geq n_0(\delta, \varepsilon)$.

We start from an exact identity. Böröczky and Henk [2], building on work of Vershik and Sporyshev [22] and ultimately of Ruben [18], give the integral formula

$$\gamma(F^\ell, C) = \sqrt{\frac{\ell+1}{\pi}} \int_0^\infty \exp\{-(\ell+1)x^2\} \left(\frac{2}{\sqrt{\pi}} \int_0^x \exp(-y^2)\,dy\right)^{n-\ell-1} dx.$$

We recognize the term in the large parentheses as the error function G from (4.1). Set $v_{\ell,n} = (\ell+1)/n$. The integral formula can be rewritten as

$$\sqrt{\frac{nv_{\ell,n}}{\pi}} \int_0^\infty \exp\{-nv_{\ell,n}x^2 + n(1-v_{\ell,n})\log G(x)\}\,dx. \tag{5.2}$$

The appearance of n in the exponent suggests using Laplace's method, i.e. using the fact that such an integral behaves in n roughly like the integrand at its maximum. This is now often called Varadhan's lemma in the large deviations literature [6]. (It is sometimes also called the saddlepoint approximation, for example in quantum field theory, but this is misleading as it really represents only the "tip of the iceberg" of the full saddlepoint method). We define, for v fixed,

$$f_{v,n}(y) = \exp\{-n\psi_v(y)\} \cdot \sqrt{\frac{nv}{\pi}}$$

with

$$\psi_v(y) \equiv vy^2 - (1-v)\log G(y). \tag{5.3}$$

We note that ψ_v is smooth and convex and (in the Appendix) develop expressions for its second and third derivatives. Applying Laplace's method to ψ_v in the usual way, but taking care about regularity conditions and remainders, gives a result with the uniformity in v, which is crucial for us.

Lemma 5.2. *For $v \in (0, 1)$ let x_v denote the minimizer of ψ_v. Then*

$$\int_0^\infty f_{v,n}(x)\,dx \leq \exp\{-n\psi_v(x_v)\}(1 + R_n(v)),$$

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

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

Of course the minimizer x_v mentioned in this lemma is the same x_v defined earlier in (4.2) in terms of the error function, and the minimum value identified in this lemma as driving the exponential rate is the same as our exponent Ψ_{ext}:

$$\Psi_{\text{ext}}(v) = \psi_v(x_v). \tag{5.4}$$

In fact Lemma 5.2 easily leads to Lemma 5.1. We first note that $\Psi_{\text{ext}}(v) \to 0$ as $v \to 1$. For given $\varepsilon > 0$ in the statement of the lemma, there is (a largest) $v_\varepsilon < 1$ with $\Psi_{\text{ext}}(v_\varepsilon) = \varepsilon$. Note that $\gamma(F^\ell, C) \leq 1$, so that for $\ell > v_\varepsilon n$,

$$n^{-1}\log(\gamma) \leq 0 < -\Psi_{\text{ext}}(v) + \varepsilon,$$

for $n \geq 1$. Consider now $\ell \in [\delta n, v_\varepsilon n)$. Taking into account (5.2), we now have

$$\gamma(F^\ell, C) = \int_0^\infty f_{v_{\ell,n}}(x)\,dx.$$

Applying the uniformity in v given in Lemma 5.2, we conclude

$$n^{-1}\log(\gamma(F^\ell, C)) = \psi_{v_{\ell,n}}(x_{v_{\ell,n}}) + o(1), \qquad \ell \geq \delta n.$$

Then invoking the identity (5.4) and the uniform continuity of ψ_v in x and of x_v in $v \in [\delta, 1]$, we get

$$n^{-1}\log(\gamma(F^\ell, C)) \leq -\Psi_{\text{ext}}(\ell/n) + o(1).$$

Lemma 5.1 follows.

6. Bounds on the Internal Angle

In this section we justify

Lemma 6.1. *For $\varepsilon > 0$, and $n > n_0(\varepsilon, \delta, \rho)$,*

$$n^{-1}\log(\beta(T^k, T^\ell)) \leq -\Psi_{\text{int}}(\ell/n; k/n) + \varepsilon,$$

uniformly in $\ell \geq \delta n, k \leq \rho\delta n$.

6.1. *Background*

By definition, the internal angle $\beta(F, G)$ is the fraction of the vector space $span(G - x^F)$ taken up by the positive cone $pos(G - x^F)$ where x^F is the barycenter of face F. Let now V_{m-1} denote $(m - 1)$-dimensional surface measure on the sphere S^{m-1}, while $\Sigma_{m-1}(\alpha)$ denotes the regular spherical simplex (a generalization of triangle) with m vertices on the sphere and sides of angle α. Böröczky and Henk [2] give the formula

$$\beta(T^k, T^\ell) = \frac{V_{\ell-k}(\Sigma_{\ell-k}(1/(k+2)))}{V_{\ell-k}(S^{\ell-k})}; \tag{6.1}$$

see also [1] and [23].

Defining

$$B(\alpha, m) = \frac{V_{m-1}(\Sigma_{m-1}(\alpha))}{V_{m-1}(S^{m-1})}$$

it is thus of interest to evaluate/bound $B(1/(k+2), \ell - k + 1)$. By [2],

$$B(\alpha, m) = \theta^{(m-1)/2}\sqrt{(m-1)\alpha + 1}\,\pi^{-m/2}\alpha^{-1/2}J(m, \theta), \tag{6.2}$$

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

$$J(m, \theta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left(\int_0^{\infty} \exp(-\theta v^2 + 2iv\lambda)\, dv \right)^m \exp(-\lambda^2)\, d\lambda. \tag{6.3}$$

Vershik and Sporyshev [23] also considered $\beta(T^k, T^\ell)$ in the proportional-dimensional setting $k \sim \rho n$ and $\ell \sim \rho n$. They analyzed a seemingly different integral expression based on contour integration. They sketched an approach to the asymptotics of that integral by the saddlepoint method. We pursue here a probabilistic approach, and later correlate our findings to theirs.

6.2. *Probabilistic Analysis*

Seemingly taking a different tack from previous authors, we recognize in the expression for $J(m, \theta)$ a convolution of $m + 1$ probability densities being expressed in the Fourier domain. This leads to the following probabilistic interpretation, proved in the Appendix.

Lemma 6.2. *Put $\theta = (1 - \alpha)/\alpha$. Let T be a random variable with the $N(0, \frac{1}{2})$ distribution, and let W_m be a sum of i.i.d. half-normals $U_i \sim HN(0, 1/2\theta)$. Let T and W_m be stochastically independent, and let g_{T+W_m} denote the probability density function of the random variable $T + W_m$. Then*

$$B(\alpha, m) = \sqrt{\frac{\alpha(m - 1) + 1}{1 - \alpha}} \cdot 2^{1-m} \cdot \pi^{1/2} \cdot g_{T+W_m}(0). \tag{6.4}$$

Using the probabilistic interpretation, and applying large deviations techniques, we obtain effective bounds on $g_{T+W_m}(0)$. By the convolution formula, symmetry of the

standard normal, support properties of the half-normal and integration by parts we have

$$g_{T+W_m}(0) = \int_{-\infty}^{\infty} g_T(0-v) g_{W_m}(v)\, dv = \int_0^{\infty} g_T(v) g_{W_m}(v)\, dv$$

$$= \int_0^{\infty} (-g_T'(v)) G_{W_m}(v)\, dv,$$

where G_{W_m} denotes the cumulative distribution function of the random variable W_m. Since $g_T'(t) = (-2t)g_T(t)$,

$$g_{T+W_m}(0) = 2 \int_0^{\infty} v g_T(v) G_{W_m}(v)\, dv.$$

Let $\mu_m = EW_m$ denote the mean of W_m. The part of the integral occurring above μ_m obeys

$$2 \int_{\mu_m}^{\infty} v g_T(v) G_{W_m}(v)\, dv \le 2 \int_{\mu_m}^{\infty} v g_T(v)\, dv = g_T(\mu_m) \le \exp(-\mu_m^2) \cdot 2/\sqrt{\pi}. \quad (6.5)$$

Since $\mu_m = m \cdot \sqrt{1/\pi\theta}$, it turns out that this part of the integral is typically negligible. The part of the integral below the mean can be controlled by the basic inequalities of large-deviations theory [6]; for $v \le \mu_m \equiv E(W_m)$,

$$G_{W_m}(v) \le \exp(-\Lambda_{W_m}^*(v)). \quad (6.6)$$

Here $\Lambda_{W_m}^*$ is the rate function of the random variable W_m (i.e. the Fenchel–Legendre dual of the cumulant generating function Λ_{W_m}).

Now as $W_m = U_1 + \cdots + U_m$ with the U_i i.i.d., then $\Lambda_{W_m}^*(v) = m\Lambda_{U_1}^*(v/m)$; also $\Lambda_{aU_1}^*(av) = \Lambda_{U_1}^*(v)$. Combining these, we have that if Y denotes a standard half-normal random variable $Y \sim HN(0,1)$, and if Λ^* denotes the rate function associated to Y, then from $U_1 = Y/\sqrt{2\theta}$ we get

$$g_{T+W_m}(0) \le \frac{2}{\sqrt{\pi}} \cdot \int_0^{\mu_m} v \exp\left(-v^2 - m\Lambda^*\left(\frac{\sqrt{2\theta}}{m}v\right)\right) dv + \exp(-\mu_m^2) \cdot \frac{2}{\sqrt{\pi}}$$

$$= I_m + II_m, \quad \text{say.}$$

We have already argued that II_m is negligible. We now change variables $y = (\sqrt{2\theta}/m)v$, getting

$$I_m \le \frac{2}{\sqrt{\pi}} \cdot \frac{m^2}{2\theta} \cdot \int_0^{\sqrt{2/\pi}} y \exp\left(-m\left(\frac{m}{2\theta}\right)y^2 - m\Lambda^*(y)\right) dy. \quad (6.7)$$

6.3. *Laplace's Method*

The m in the exponent of (6.7) was defined by $m = \ell - k + 1$ which we think of as $(\nu - \rho\delta)n$. Thus m is growing with n, suggesting again a recourse to Laplace's

method. We recognize that essentially the function ξ_γ appears in the exponent, with $\gamma = \theta/(m + \theta - 1)$. Recalling that $\theta = (1 - \alpha)/\alpha = k + 1$, and $m = \ell - k + 1$ this then has the form $\gamma_{k,\ell} = (k + 1)/(\ell + 2)$; for $k \sim \rho\delta n$ and $\ell \sim \nu n$ this is essentially the constant $\gamma = \rho\delta/\nu$. So define the integral

$$f_{\gamma,m} = \int_0^\infty y \exp(-m\xi_\gamma(y)) \, dy.$$

Again applying Laplace's method with careful attention to uniformity gives

Lemma 6.3. *For $\gamma \in (0, 1)$, let $y_\gamma \in (0, 1)$ denote the minimizer of ξ_γ. Then*

$$\int_0^\infty f_{\gamma,m}(x) \, dx \leq \exp(-m\xi_\gamma(y_\gamma)) \cdot R_m(\gamma),$$

where, for $\eta > 0$,

$$m^{-1} \sup_{\gamma \in [\eta, 1]} \log(R_m(\gamma)) = o(1) \qquad as \quad m \to \infty.$$

The proof is very similar to that of Lemma 5.2 and we omit it.

We conclude that

$$g_{T+W_m}(0) \leq \exp(-m\xi_\gamma(y_\gamma)) R_m(\gamma),$$

where $n^{-1} \log(R_m(\gamma_{k,\ell})) = o(1)$. Applying (6.2) we get

$$n^{-1} \log(\beta(T^k, T^\ell)) \leq -(\xi_\gamma(y_\gamma) + \log(2))(\nu - \rho\delta) + o(1),$$

where the $o(1)$ is uniform over a range of k and ℓ. Arguing much as in the section on Internal Angles, we get Lemma 6.1.

6.4. *Properties of ξ_γ*

We now briefly review the properties of ξ_γ relevant to computing Ψ_{int}.

By standard convex duality, associated to the cumulant generating function $\Lambda(s)$ and its dual Λ^*, we have the duality relations

$$y = \Lambda'(s), \qquad s = (\Lambda^*)'(y), \tag{6.8}$$

defining a one–one relationship $s = s(y)$ and $y = y(s)$ between $s < 0$ and $0 < y < \sqrt{2/\pi}$.

In particular, from $\Lambda(s) = s^2/2 + \zeta_0(s)$, $\zeta_0(s) = \log(2\Phi(s))$,

$$y(s) = s + \zeta_1(s), \qquad \zeta_1 = \frac{d}{ds}\zeta_0 = \frac{\varphi}{\Phi}, \tag{6.9}$$

where again φ and Φ are the normal density and cumulative. It is common culture in probability and statistics that the normal density and cumulative are closely related, with

$\Phi(s)$ nearly $\varphi(s)/|s|$ for s large negative. To refine this insight, we define a variant of "Mills' Ratio" [19]; a function $M(s)$ with domain $s < 0$ having $0 < M(s) < 1$ and $M(s) \to 1$ as $s \to -\infty$ defined by

$$\Phi(s) = M(s) \cdot (\varphi(s)/|s|), \qquad s < 0.$$

Ruben studies series expansions for M which we use farther below; there are convenient rational approximations to $M(s)$ in Bryc's article [3] which we found essential for numerical work. It follows that $\zeta_1(s) = |s|/M(s)$. Using (6.9) we conclude

$$y(s) = s \cdot (1 - 1/M(s)), \qquad s < 0. \tag{6.10}$$

We can also obtain a useful identity for ζ_0:

$$\zeta_0(s) = \log(2\Phi(s)) = \log(2 \cdot M(s) \cdot (\varphi(s)/|s|)) = \log(M(s)) + \log(2/\pi)/2 - s^2/2 - \log(|s|).$$

Substituting in $\Lambda^*(y) = ys - \Lambda(s) = ys - s^2/2 - \zeta_0(s)$, $s = s(y)$, gives

$$\Lambda^*(y(s)) = s^2(1 - 1/M(s)) - \log(M(s)) - \log(2/\pi)/2 + \log(|s|). \tag{6.11}$$

Applying duality (6.8) to $(d/dy)\xi_\gamma(y) = 0$, we see that the minimizer y_γ of ξ_γ obeys

$$\frac{1-\gamma}{\gamma} y_\gamma = -s_\gamma. \tag{6.12}$$

Inserting this in (6.10) gives the convenient characterization

$$M(s_\gamma) = 1 - \gamma.$$

Combining with (6.11) we obtain the exact formula

$$\Lambda^*(y_\gamma) = -y_\gamma^2 \frac{1-\gamma}{\gamma} - \frac{\log(2/\pi)}{2} + \log\left(\frac{y_\gamma}{\gamma}\right),$$

and the identity

$$\xi_\gamma(y_\gamma) = -\frac{1}{2} y_\gamma^2 \frac{1-\gamma}{\gamma} - \frac{\log(2/\pi)}{2} + \log\left(\frac{y_\gamma}{\gamma}\right). \tag{6.13}$$

We now consider asymptotics as $\gamma \to 0$. We have

$$E\exp(sY) \sim \sqrt{\frac{2}{\pi}} \cdot \frac{-1}{s}, \qquad s \to -\infty.$$

Hence $\Lambda(s) \sim -\log(|s|) + \log(2/\pi)/2$ as $s \to -\infty$. Strengthening this with remainders shows $\Lambda'(s) \sim -1/s$. Duality then implies the crude asymptotics

$$y(s) \sim -\frac{1}{s}, \quad s \to -\infty; \qquad s(y) \sim -\frac{1}{y}, \quad y \to 0,$$

as well as

$$y_\gamma \sim \sqrt{\frac{\gamma}{1-\gamma}}, \qquad s_\gamma \sim -\sqrt{\frac{1-\gamma}{\gamma}}, \qquad \gamma \to 0. \tag{6.14}$$

From this the leading-order asymptotics of $\xi_\gamma(y_\gamma)$ given in (4.4) follow immediately. Compare also Lemma A.3.

6.5. *Reconciliation with Vershik and Sporyshev*

As we have pointed out, Vershik and Sporyshev, in pioneering work [23], considered (in our notation) the internal angle $\beta(T^k, T^\ell)$ with $k \sim \rho\delta n, \ell \sim \nu n, n \to \infty$. (To compare our papers, use the dictionary: $\nu \leftrightarrow \beta, \rho \leftrightarrow \varepsilon, \delta \leftrightarrow \alpha$ and $s_\gamma \leftrightarrow t_\varepsilon$, our paper $\leftrightarrow$ [23].) They work with an expression apparently different from (6.3) defined by contour integration and sketch a saddlepoint procedure. Their results are not stated in terms of large-deviations properties of half-normal random variables. From the viewpoint of this paper, we understand their results as implicitly working with what we call the dual variables s_γ. It appears that their formula can be rewritten using our notation as

$$n^{-1} \log \beta(T^k, T^\ell) \leq \frac{\rho\delta s_\gamma^2}{2} + \log\left(\frac{1-\gamma}{\sqrt{2\pi}|s_\gamma|}\right)(\nu - \rho\delta) + o(1).$$

Our results can be re-expressed, using (6.11) and (6.12), as follows:

$$
\begin{aligned}
n^{-1} \log \beta(T^k, T^\ell) + o(1) &\leq -\Psi_{\text{int}}\left(\frac{\ell}{n}, \frac{k}{n}\right) \\
&= -(\xi_\gamma(y_\gamma) + \log(2))(\nu - \rho\delta) \\
&= -\left(-\frac{\gamma}{1-\gamma}\frac{s_\gamma^2}{2} + \log\left(\sqrt{\frac{\pi}{2}} \cdot \frac{|s_\gamma|}{1-\gamma}\right) + \log(2)\right)(\nu - \rho\delta) \\
&= \frac{\gamma}{1-\gamma}\frac{s_\gamma^2}{2}(\nu - \rho\delta) - \log\left(\frac{\sqrt{2\pi}|s_\gamma|}{1-\gamma}\right)(\nu - \rho\delta) \\
&= \frac{\rho\delta s_\gamma^2}{2} + \log\left(\frac{1-\gamma}{\sqrt{2\pi}|s_\gamma|}\right)(\nu - \rho\delta). \quad (6.15)
\end{aligned}
$$

Thus the two answers agree—as of course they would if both approaches were equally precise in determining exponential order. Our systematic approach, based on a probabilistic interpretation, provides a sound motivation for the correctness of this answer. It also provides detailed information useful in other ways, for example in proving Lemmas 4.1 and 4.2.

7. Theorems 2 and 3

Given our lengthy proof of Theorem 1, it may be a welcome relief to learn that no serious work is required to get the other results announced in the Introduction.

7.1. *Proof of Theorem* 2

Observe that $f_{k-1}(C) = 2^k\binom{n}{k}$; this combinatorial factor has exponential growth with n according to an exponent $\Psi_{\text{face}}(\rho\delta) \equiv \rho\delta \log(2) + H(\rho\delta)$; thus, if $k = k(n) \sim \rho\delta n$,

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

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

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

$$\Psi_{\text{net}}(v; \rho, \delta) < \Psi_{\text{face}}(\rho\delta), \qquad v \in [\delta, 1). \tag{7.1}$$

Recall Section 3's definition $\Delta(k, d, n) = f_{k-1}(C) - f_{k-1}(AC) \geq 0$. The proof of Theorem 2 is based on observing that (7.1) implies

$$\Delta(k, d, n) = o(f_{k-1}(C)). \tag{7.2}$$

We immediately get (1.2). Showing that (7.1) implies (7.2) requires no new ideas; one proceeds as in Section 3 almost line-by-line; we omit the exercise. $\square$

7.2.   *Proof of Theorem* 3

We retrace slightly the discussion of sectional neighborliness in [9]. With probability one, the projected cross polytope $P = AC$ has $2n$ vertices. Every subset K of k of these vertices which does not contain an antipodal pair defines a k-dimensional subspace V_K of $\mathbf{R}^d$. Every such V_K defines a so-called *intrinsic section* $P_K = V_K \cap P$. There are $\binom{n}{k}$ different intrinsic sections of P. Define the exponent $\Psi_{sect} = H(\rho\delta)$. Then for $k \sim \rho\delta n$, we have

$$n^{-1} \log(\#\{\text{intrinsic } k\text{-sections}\}) \to \Psi_{sect}(\rho\delta), \qquad n \to \infty.$$

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

Definition 3.   Let $\delta \in (0, 1]$. $\rho_S(\delta)$ is the supremum of $\rho \in [0, 1]$ obeying

$$\Psi_{\text{net}}(v; \rho, \delta) < \Psi_{sect}(\rho\delta), \qquad v \in [\delta, 1). \tag{7.3}$$

Recall again the definition $\Delta(k, d, n) = f_{k-1}(C) - f_{k-1}(AC) \geq 0$. The proof of Theorem 3 is based on the observation that (7.3) implies, for $k \sim \rho\delta n$,

$$\Delta(k, d, n) = o(\#\{\text{intrinsic } k\text{-sections}\}). \tag{7.4}$$

We have with probability one that the columns of A are in general position. As in Section 6 of [9], it follows that all the k-faces of P are k-simplices, $k < d/2$. In order for a given section P_K *not* to be a k-dimensional cross polytope, it must have

$$f_{k-1}(P_K) < f_{k-1}(C^k),$$

where C^k denotes the k-dimensional cross polytope. In words, to not have a cross polytope, a section must "lose at least one $(k - 1)$-face." However, as there are $\binom{n}{k}$ intrinsic sections P_K, and the condition (7.4) permits us to "lose" only $o(1)$ faces per section, we conclude that there are relatively few sections which can "lose" any faces. Hence the overwhelming majority are intact cross polytopes. $\square$

Acknowledgements

Appendix

Proof of Lemma 4.1

Fix $\delta > 0$. Apply (4.5) with $\eta = \frac{1}{2}$. For small enough ρ, the lower bound forces $\Psi_{\text{int}}(\nu; \rho\delta)$ to be quite big, uniformly in $\nu \in [\delta, 1]$ and $\delta < \delta_0$, simply by picking ρ small.

For $\delta < \delta_0$ small positive, the terms Ψ_{com} and Ψ_{ext} can be bounded independently of $\rho < \rho_0$ and $\nu \geq \delta$. Hence, for ρ small, the large size of Ψ_{int} can be big enough to force the net exponent Ψ_{net} negative, uniformly in ν. $\qquad\square$

Proof of Lemma 4.2

Proof. We will show that, with $\rho(\delta) = \log(1/\delta)^{-1-\eta}$ and $\delta_0 > 0$ to be chosen below,

$$\Psi_{\text{net}}(\nu; \rho(\delta), \delta) < -\delta, \qquad \delta < \delta_0, \quad \nu \in [\delta, 1). \tag{A.1}$$

Hence $\rho_N(\delta) \geq \rho(\delta)$ for $\delta < \delta_0$.

Define $\Omega(\nu) = H(\nu) - \Psi_{\text{ext}}(\nu)$. This is concave; see Fig. A.1.

Based on $H(\nu) \sim \nu \log(1/\nu)$ as $\nu \to 0$ and (4.3), we have the following asymptotic as $\nu \to 0$:

$$\Omega(\nu) \sim \frac{1}{2} \cdot \log(\log(1/\nu))\nu, \qquad \nu \to 0. \tag{A.2}$$

Also define $K(\nu; \rho, \delta) = \Omega(\nu) - \xi_\gamma(y_\gamma)(\nu - \rho\delta)$. Our proof of (A.1) will be reduced to showing, via (A.2), that $K < 0$ for small ν.
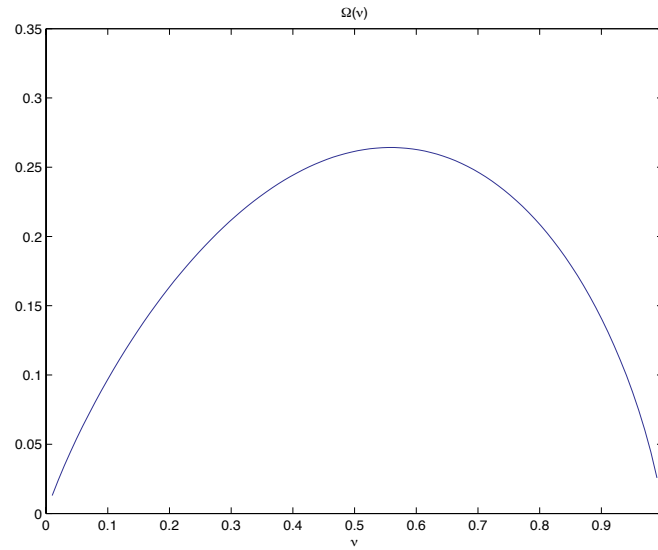
Note the combinatorial identity

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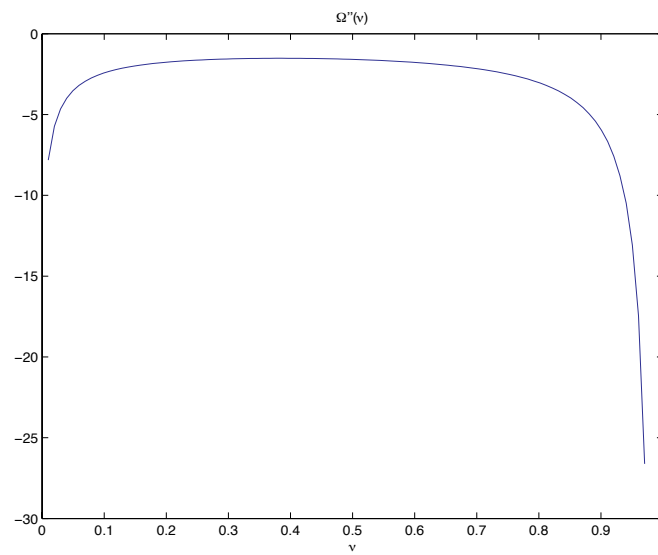
This implies

$$H(\nu) + H\left(\frac{\rho\delta}{\nu}\right)\nu = H(\rho\delta) + H\left(\frac{\nu - \rho\delta}{1 - \rho\delta}\right)(1 - \rho\delta). \tag{A.3}$$

Use this to rewrite the net exponent as

$$\Psi_{\text{net}}(\nu; \rho, \delta) = K(\nu; \rho, \delta) + \rho\delta \log(2) + H(\rho\delta/\nu)\nu.$$

(a)

(b)

Fig. A.1. (a) $\Omega(\nu) = H(\nu) - \Psi_{\text{ext}}(\nu)$ is concave; (b) $\Omega'' < 0$.

Lemma A.1 gives for $\delta < \delta_1$ that $H(\rho\delta/\nu)\nu \le H(\rho)\delta + 2\rho(\nu - \delta)$. Then

$$\Psi_{\text{net}} \le K(\nu; \rho, \delta) + \left[\rho\delta \log(2) + H(\rho)\delta\right] + 2\rho(\nu - \delta), \qquad (A.4)$$

say. Lemma A.2 establishes concavity of K in ν. Suppose that for $\delta < \delta_1$, $K'(\delta) \le 0$, then we would have

$$K(\nu; \rho, \delta) \le K(\delta), \qquad \nu \in [\delta, 1).$$

Additionally, suppose that for some $\varepsilon_2 > 0$, $K'(\delta) \le -\varepsilon_2 < 0$, then we would have for any $\varepsilon_1 > 0$,

$$K(\nu; \rho, \delta) + \varepsilon_1 + \varepsilon_2(\nu - \delta) \le K(\delta) + \varepsilon_1, \qquad \nu \ge \delta. \qquad (A.5)$$

Below we will show that for $\delta < \delta_2$,

$$K'(\delta; \rho, \delta) \le -\eta/4 \log(\log(1/\delta)), \qquad (A.6)$$

$$K(\delta; \rho, \delta) \le -\delta \cdot \eta/4 \cdot \log(\log(1/\delta)), \qquad (A.7)$$

where $K'(\nu; \rho, \delta) \equiv (\partial/\partial\nu)K(\nu; \rho, \delta)$. If follows, recalling (A.4) and setting $\varepsilon_1 = \rho\delta \log(2) + H(\rho)\delta$ and $\varepsilon_2 = 2\rho$, that

$$K'(\delta; \rho, \delta) \le -\varepsilon_2, \qquad \delta < \delta_3,$$

and so (A.5) gives, for $\delta < \delta_3$,

$$\Psi_{\text{net}}(\nu; \rho, \delta) \le K(\nu; \rho, \delta) + \varepsilon_1, \qquad \nu \in [\delta, 1).$$

Now (A.7) shows that $K(\delta)/\delta \to -\infty$ as $\delta \to 0$ while evidently $\varepsilon_1 = O(\delta)$, hence $K(\delta) + \varepsilon_1 < -\delta$ for $\delta < \delta_4$. Inequality (A.1) follows with $\delta_0 = \min(\delta_1, \delta_2, \delta_3, \delta_4)$.

It remains to verify (A.6)–(A.7). Writing

$$\Omega(\nu) = \tfrac{1}{2}\nu \log\left(\log\left(\frac{1}{\nu} + R(\nu)\right)\right) + (1 - \nu)\log\left(\frac{1}{(1 - \nu)}\right)$$
$$+ \nu \log\left(\frac{1 - \nu}{\sqrt{\pi}}\right) - \log\left(\frac{2}{\sqrt{\pi}}\right),$$

where $R(\nu) = -\log(x_\nu G(x_\nu))$, we compute that

$$\Omega'(\nu) \sim \tfrac{1}{2}\log\left(\log\left(\frac{1}{\nu}\right)\right), \qquad \nu \to 0.$$

We also recall (A.2). Now by (6.13)–(6.14), we have that as $\delta \to 0$,

$$\xi_\rho(y_\rho) \sim \frac{1}{2}\log\frac{1}{\rho} \sim \frac{1}{2}\log\left(\log\frac{1}{\delta}\right)(1 + \eta).$$

Hence for $\delta < \delta_{1,1}$,

$$\xi_\rho(y_\rho)(1 - \rho) \ge (1 + \eta/2)\Omega(\delta)/\delta, \qquad (A.8)$$

$$\xi_\rho(y_\rho)(1 - \rho) \ge (1 + \eta/2)\Omega'(\delta). \qquad (A.9)$$

This gives

$$K(\delta; \rho(\delta), \delta) < -(\eta/2) \log(\log(1/\delta))(1 + o(1));$$

and of course for $\delta < \delta_{1,2}$, the $(1 + o(1))$ factor on the right-hand side exceeds $\frac{1}{2}$. Hence (A.7) holds for $\delta_1 = \min(\delta_{1,1}, \delta_{1,2})$. $\qquad\square$

Lemma A.1. *For* $0 \leq \rho < \frac{1}{2}$,

$$H(\rho\delta/\nu)\nu \leq H(\rho)\delta + 2\rho(\nu - \delta), \qquad \nu \in [\delta, 1). \qquad (A.10)$$

Proof. One can compute that, with $\gamma = \rho\delta/\nu$,

$$\frac{\partial}{\partial \nu}[H(\gamma)\nu] = -\log(1 - \gamma).$$

Now for $0 \leq \rho < \frac{1}{2}$, $-\log(1 - \rho) \leq 2\rho$. Also $\gamma = \rho\delta/\nu \leq \rho$ for $\nu \geq \delta$. Hence

$$\frac{\partial}{\partial \nu}[H(\gamma)\nu] \leq 2\rho, \qquad \nu \in [\delta, 1).$$

Equation (A.10) follows. $\qquad\square$

Lemma A.2. *For* $\rho < \rho_0$ *and* $\delta < \delta_0$, $K(\nu; \rho, \delta)$ *is concave as a function of* ν.

Proof. Let

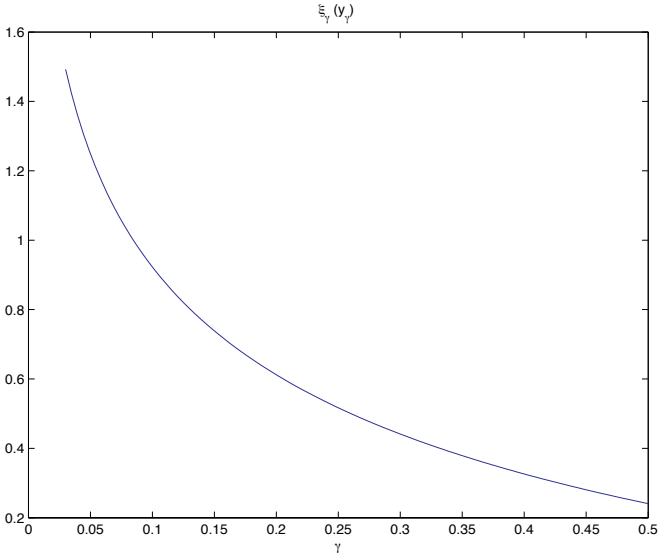$$\Xi(\nu; \rho, \delta) \equiv \xi_\gamma(y_\gamma)(\nu - \rho\delta).$$

Then $K = \Omega - \Xi$. Since Ω is concave, it is sufficient to show that Ξ is convex. This involves properties of $\xi_\gamma(y_\gamma)$; the reader may be helped by Fig. A.2.

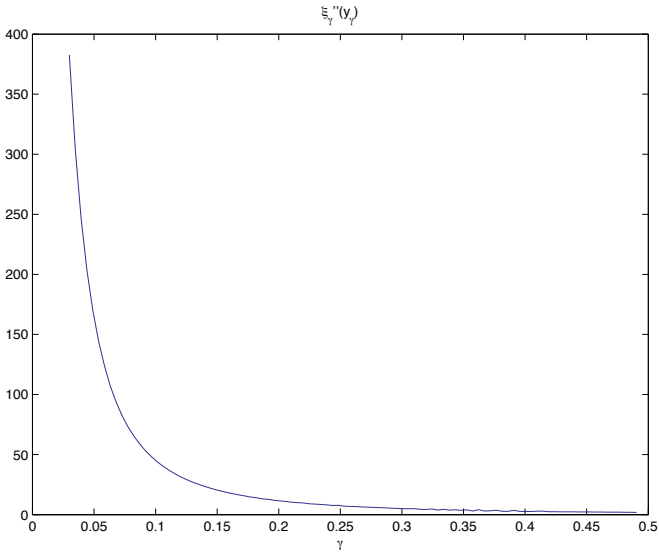Using $\partial\gamma/\partial\nu = -\gamma/\nu$ we write

$$\frac{\partial}{\partial \nu}\Xi(\nu; \rho, \delta) = \xi_\gamma(y_\gamma) + \frac{\partial}{\partial \gamma}\xi_\gamma(y_\gamma) \cdot \frac{\partial \gamma}{\partial \nu} \cdot (\nu - \rho\delta)$$

$$= \xi_\gamma(y_\gamma) - \frac{\partial}{\partial \gamma}\xi_\gamma(y_\gamma) \cdot \gamma(1 - \gamma).$$

Hence

$$\frac{\partial^2}{\partial \nu^2}\Xi(\nu; \rho, \delta) = \frac{\partial}{\partial \gamma}\xi_\gamma(y_\gamma)\frac{\partial \gamma}{\partial \nu} - \frac{\partial^2}{\partial \gamma^2}\xi_\gamma(y_\gamma) \cdot \frac{\partial \gamma}{\partial \nu} \cdot \gamma \cdot (1 - \gamma)$$

$$- \frac{\partial}{\partial \gamma}\xi_\gamma(y_\gamma) \cdot \frac{\partial \gamma}{\partial \nu} \cdot (1 - \gamma) - \frac{\partial}{\partial \gamma}\xi_\gamma(y_\gamma) \cdot \frac{\partial \gamma}{\partial \nu} \cdot (-\gamma)$$

$$= -\frac{\partial^2}{\partial \gamma^2}\xi_\gamma(y_\gamma) \cdot \frac{\partial \gamma}{\partial \nu} \cdot \gamma \cdot (1 - \gamma) + 2\frac{\partial}{\partial \gamma}\xi_\gamma(y_\gamma)\frac{\partial \gamma}{\partial \nu}\gamma$$

$$= \frac{\partial^2}{\partial \gamma^2}\xi_\gamma(y_\gamma) \cdot \gamma^2 \cdot \frac{1 - \gamma}{\nu} - 2\frac{\partial}{\partial \gamma}\xi_\gamma(y_\gamma)\frac{\gamma^2}{\nu}.$$

(a)

(b)

Fig. A.2. (a) $\xi_\gamma(y_\gamma)$ is convex; (b) $(d^2/d\gamma^2)\xi_\gamma(y_\gamma) > 0$.

We next show that the first term is positive for all small γ; it will also emerge that it dominates the second, and the claimed convexity follows. The key to evaluating the required derivatives is to express $\xi_\gamma(y_\gamma)$ in terms of the dual variables s_γ, as in (6.15). We then have

$$\xi_\gamma(y_\gamma) = \log(|s_\gamma|) + R(\gamma), \tag{A.11}$$

where the remainder $R(\gamma)$ involves other terms, also explicitly expressed in terms of γ and s_γ:

$$R(\gamma) = \log(\sqrt{2\pi}/(1-\gamma)) - s_\gamma^2/2 \cdot (\gamma/(1-\gamma)).$$

The term $\log(|s_\gamma|)$ gives rise to large derivatives. Thus, using (A.11) and Lemma A.3,

$$\frac{\partial}{\partial\gamma}\log(|s_\gamma|) = s_\gamma^{-1}\frac{d}{d\gamma}s_\gamma \sim (-\gamma^{1/2})(\tfrac{1}{2}\gamma^{-3/2}) \sim -\frac{\gamma^{-1}}{2}, \qquad \gamma \to 0, \tag{A.12}$$

while

$$\frac{\partial^2}{\partial\gamma^2}\log(|s_\gamma|) = \frac{((d/d\gamma)s_\gamma)^2}{s_\gamma^2} + \frac{(d^2/d\gamma^2)s_\gamma}{s_\gamma} \sim \frac{\gamma^{-4}}{4}, \qquad \gamma \to 0. \tag{A.13}$$

The remainder R turns out to have relatively negligible influence—principally because $s_\gamma^2\gamma = O(1)$ does not have such large derivatives in γ; indeed,

$$\frac{\partial^2}{\partial\gamma^2}s_\gamma^2\gamma = 2s_\gamma''s_\gamma\gamma + 2(s_\gamma')^2\gamma + 4s_\gamma's_\gamma;$$

applying Lemma A.3 below, this is $O(\gamma^{-2}) \ll O(\gamma^{-4})$. Hence we conclude that for some $\gamma_0 > 0$ sufficiently small, $(\partial^2/\partial\gamma^2)\xi_\gamma(y_\gamma)$ is positive uniformly in $\gamma \le \gamma_0$, and, since $\gamma \le \rho$, uniformly in $\rho \le \rho_0 \equiv \gamma_0$. The dominance of this term flows from the different asymptotic orders of (A.12)–(A.13). $\qquad\square$

Lemma A.3. *As $\gamma \to 0$,*

$$s_\gamma \sim -\gamma^{-1/2},$$

$$\frac{d}{d\gamma}s_\gamma \sim \tfrac{1}{2}\gamma^{-3/2},$$

$$\frac{d^2}{d\gamma^2}s_\gamma \sim -\tfrac{3}{4}\gamma^{-5/2}.$$

Proof. We use the classical asymptotic series

$$R_x \equiv \frac{\Phi(-x)}{\varphi(x)} \sim \frac{1}{x} - \frac{1}{x^3} + \frac{1\cdot 3}{x^5} - \frac{1\cdot 3\cdot 5}{x^7} + \cdots, \qquad x > 0;$$

Ruben [19] attributes this to Laplace. Since $M(s) = |s|R_{|s|}$, we have

$$M(-x) \sim 1 - \frac{1}{x^2} + \frac{1\cdot 3}{x^4} - \frac{1\cdot 3\cdot 5}{x^6} + \cdots, \qquad x > 0.$$

Because there is no $1/x$ term in this formula, the solution s_γ of

$$M(s_\gamma) = 1 - \gamma \tag{A.14}$$

obeys

$$s_\gamma^{-2} \sim \gamma, \qquad \gamma \to 0,$$

i.e. $s_\gamma \sim -\gamma^{-1/2}$. To take derivatives of s_γ we use the inverse function theorem:

$$\frac{d}{d\gamma} s_\gamma = -\frac{1}{M'(s_\gamma)}; \qquad \frac{d^2}{d\gamma^2} s_\gamma = \frac{M''(s_\gamma)}{(M'(s_\gamma))^3}.$$

We can explicitly compute the derivatives of $N(x) = M(-x)$, getting

$$N' = N/x + x + xN; \qquad N'' = 1 + (N-1)(3 + x^2).$$

Using this and the asymptotic series for M given above, we obtain asymptotic expressions for $M'(s_\gamma)$ and $M''(s_\gamma)$, which give the desired result. $\qquad \square$

Proof of Lemma 5.1

The uniform Laplace method ultimately boils down to this statement:

Lemma A.4. *Let ψ be convex and let C^2 be on an interval I and suppose that it takes its minimum at an interior point $x_0 \in I$ where $\psi'' > 0$ and that in a vicinity $(x_0 - \varepsilon, x_0 + \varepsilon)$ of x_0,*

$$|\psi''(x) - \psi''(x_0)| \leq C|\psi''(x_0)||x - x_0|. \tag{A.15}$$

Let $\bar{\psi}$ be the quadratic approximation $\psi(x_0) + \psi''(x_0)(x - x_0)^2/2$. Then

$$\int_I \exp(-n\psi(x))\, dx \leq \int_{-\infty}^{\infty} \exp(-n\bar{\psi}(x))\, dx \cdot (S_{1,n} + S_{2,n}),$$

where

$$S_{1,n} = \exp(n\psi''(x_0)C\varepsilon^2/6),$$
$$S_{2,n} = 2/(n\varepsilon(|\psi''(0)|2\pi)^{1/2}(1 - C\varepsilon^2/2)).$$

We defer the proof to the next subsection.

The constant C referred to in the lemma is effectively a (scaled) third derivative, since if ψ is C^3, we may take in (A.15)

$$C = C(\varepsilon) = \sup_{(x_0-\varepsilon, x_0+\varepsilon)} \psi^{(3)}(x)/\psi''(x).$$

The lemma has the following significance for Lemma 5.2. Pick $\varepsilon_n = n^{-2/5}$. Pick $n_0 = n_0(\psi''(x_0), C)$ so that

$$\psi''(x_0)Cn^{-1/5}/6 < 1, \qquad Cn^{-2/5}/3 < \tfrac{1}{2}.$$

Then for $n > n_0$, $\exp(u) \le 1 + (e - 1)u$ gives

$$S_{1,n} \le 1 + (e - 1)(\psi''(x_0)C/6) \cdot n^{-1/5}$$

while $(1 - Cn^{-2/5}/3) > \frac{1}{2}$ gives

$$S_{2,n} \le 2/\psi''(0)^{1/2}n^{-3/5}.$$

Hence

$$S_{1,n} + S_{2,n} \le 1 + o(1).$$

We conclude that

$$\int_I \exp(-n\psi(x))\,dx \le \int_{-\infty}^{\infty} \exp(-n\bar{\psi}(x))\,dx \cdot (1 + o(1)). \tag{A.16}$$

Here the $o(1)$ is uniform over any collection of convex functions with prescribed bounds on $\psi''(x_0)$ and C. Now we consider the collection of convex functions ψ_ν defined in Section 5. Recall $\eta > 0$ and $\delta > 0$ in the statement of Lemma 5.2. Over the interval $\nu \in [\delta, 1 - \eta)$, they obey uniform bounds on ψ_ν'' and ψ_ν''' of the type needed for Lemma A.4. We record this in the next lemma. Lemma 5.2 then follows from this and the uniformity in (A.16). $\qquad\square$

Lemma A.5. *The function ψ_ν is C^∞ with the second derivative at the minimum*

$$\psi_\nu''(x_\nu) = 2\nu\left(1 + x_\nu^2 \frac{4\nu}{1 - \nu}\right) \tag{A.17}$$

and the third derivative at the minimum

$$\psi_\nu^{(3)}(x_\nu) = (1 - \nu)((2 - 4x_\nu^2)z + 6x_\nu z^2 + 2z^2), \tag{A.18}$$

where $z = z_\nu = 2\nu x_\nu/(1 - \nu)$. We have

$$0 < 2\delta \le \inf_{\nu \in (\delta, 1]} \psi_\nu''(x_\nu).$$

For $\varepsilon < \min(\delta/2, \eta/2)$, the ratio

$$C(\varepsilon; \delta, \eta) = \sup_{\nu \in (\delta, 1-\eta]} \sup_{|x - x_\nu| < \varepsilon} \psi_\nu^{(3)}(x)/\psi_\nu''(x)$$

is finite.

Proof. We calculate that

$$\psi'(x) = -(1 - \nu)g/G + 2\nu x; \qquad \psi''(x) = -(1 - \nu)(g'/G - g^2/G^2) + 2\nu;$$

$$\psi^{(3)}(x) = -(1 - \nu)(g''/G - 3g'g/G^2 - 2g^3/G^3).$$

We also remark that $g' = (-2x)g$ and $g'' = (-2 + 4x^2)g$. At the point x_ν, we have

$$g(x_\nu)/G(x_\nu) = 2\nu x_\nu/(1 - \nu) = z_\nu, \quad \text{say.}$$

We combine these to get (A.17) and (A.18).

The formula for $\psi_\nu''(x_\nu)$ immediately gives $\psi_\nu''(x_\nu) \geq 2\nu$, so it is bounded away from zero on any interval $\nu \in [\delta, 1], \delta > 0$.

Now as for $\psi^{(3)}$, we note that clearly z_ν and x_ν are continuous functions on $(0, 1)$. Note that as $\nu \to 1$,

$$z_\nu \sim \sqrt{2\nu/(1 - \nu)},$$

which diverges, and as $\nu \to 0$,

$$x_\nu \sim \sqrt{\log((1 - \nu)/2\nu)},$$

which also diverges. However, both are bounded on any interval $\nu \in [\delta, 1 - \eta]$. As a polynomial in ν, x_ν and z_ν, $\psi_\nu'''(x_\nu)$ is also bounded. Boundedness also holds, by inspection if we consider the behavior at x near to x_ν. $\qquad\square$

Proof of Lemma A.4

We start with a preliminary lemma.

Lemma A.6. *Let $\psi(x)$ be a C^2 function defined on $[-\varepsilon, \varepsilon]$ so that*

$$\psi(0) = \psi'(0) = 0, \qquad \psi''(0) > 0, \qquad |\psi''(x) - \psi''(0)| \leq C \cdot \psi''(0)|x|.$$

Then

$$\|\psi - \psi''(0)x^2/2\|_{L^\infty(-\varepsilon,\varepsilon)} \leq C\psi''(0)\varepsilon^3/6 \tag{A.19}$$

and

$$|\psi'(x)| \geq \psi''(0)|x|(1 - Cx^2/2), \qquad x \in [-\varepsilon, \varepsilon]. \tag{A.20}$$

This involves standard estimates, and we omit the proof.

To apply this lemma, we split the integral

$$\int_I \exp(-n\psi(x))\, dx = \int_{-\varepsilon}^{\varepsilon} + \int_{[-\varepsilon,\varepsilon]^c} = I + II.$$

Near x_0, we use (A.19), i.e. the fact that the quadratic approximation has ε^3-order accuracy:

$$\begin{aligned}
I &\leq \left(\int_{-\varepsilon}^{\varepsilon} \exp(-n\bar\psi(x))\, dx \right) \cdot \exp\{n\|\psi - \bar\psi\|_{L^\infty(-\varepsilon,\varepsilon)}\} \\
&\leq \left(\int_{-\infty}^{\infty} \exp(-n\bar\psi(x))\, dx \right) \cdot \exp\{n\psi''(0)C\varepsilon^3/6\}.
\end{aligned}$$

Away from x_0 we use convexity. To simplify notation, suppose $x_0 = 0$. Convexity gives

$$\psi(x) \geq \psi(\varepsilon) + \psi'(\varepsilon)(x - \varepsilon), \qquad x > \varepsilon.$$

Hence, using the formula $\int_0^\infty \exp(-n(a + bu))\, du = \exp(-na)/(nb)$,

$$\int_\varepsilon^\infty \exp(-n\psi(x))\, dx \leq \exp(-n\psi(0))/(n\psi'(\varepsilon)).$$

Now by (A.20), $\psi'(\varepsilon) \geq \psi''(0)\varepsilon(1 - C\varepsilon^2/2)$. Also,

$$\exp(-n\psi(0)) = \int \exp(-n\bar\psi(x))\, dx \cdot \sqrt{\psi''(0)/2\pi}$$

so that

$$\int_\varepsilon^\infty \exp(-n\psi(x))\, dx \leq \exp(-n\psi(0))/(n\varepsilon|\psi''(0)|^{1/2}(2\pi)^{1/2}(1 - C\varepsilon^2/2)).$$

The lemma follows.   □

Proof of Lemma 6.2

Define half-normal random variables $U_i = |Z_i|$ and $Z_i \sim_{\text{iid}} N(0, 1/2\theta)$, and the scaled normal $T \sim N(0, \frac{1}{2})$. Use the notation g_X for the probability density of random variable X and $\hat{g}_X$ for the characteristic function (Fourier transform) $\hat{g}_X(\lambda) = \int \exp(i\lambda x)g_X(x)\, dx$. We remark that

$$g_{U_1}(u) = \begin{cases} \exp(-\theta u^2)\sqrt{\dfrac{2\theta}{\pi}}, & u \geq 0, \\ 0, & u < 0, \end{cases}$$

and

$$\hat{g}_{U_1}(\lambda) = 2\sqrt{\frac{\theta}{\pi}} \int_0^\infty \exp(-\theta u^2 + iu\lambda)\, du.$$

Hence

$$J(m, \theta) = 2^{-m}\left(\sqrt{\frac{\pi}{\theta}}\right)^m \frac{1}{\sqrt{\pi}} \int_{-\infty}^\infty \hat{g}_T(\lambda)(\hat{g}_{U_1}(2\lambda))^m\, d\lambda.$$

Now we define $W_m = 2\sum U_i$ and note that the convolution theorem gives

$$\hat{g}_{W_m}(\lambda) = \hat{g}_{U_1}(2\lambda)^m.$$

Hence

$$J(m, \theta) = 2^{-m}(\pi/\theta)^{m/2}\pi^{-1/2} \int_{-\infty}^\infty \hat{g}_T(\lambda)\hat{g}_{W_m}(\lambda)\, d\lambda.$$

Again by the convolution theorem $\hat{g}_{T+W_m} = \hat{g}_T \hat{g}_{W_m}$, and from the Fourier inversion formula $g_X(0) = (2\pi)^{-1} \int \hat{g}_X(\lambda) \, d\lambda$ we recognize that

$$\int_{-\infty}^{\infty} \hat{g}_T(\lambda) \hat{g}_{W_m}(\lambda) \, d\lambda = 2\pi g_{T+W_m}(0),$$

giving

$$J(m, \theta) = 2^{-m} (\pi/\theta)^{-m/2} 2\sqrt{\pi} g_{T+W_m}(0). \qquad \square$$

Proof of Theorem 4

Using (A.3) in the same way as in the proof of Lemma 4.2, we have

$$\Psi_{\text{net}} - \Psi_{\text{face}} = H\left(\frac{\nu - \rho\delta}{1 - \rho\delta}\right)(1 - \rho\delta) - \xi_\gamma(y_\gamma) \cdot (\nu - \rho\delta) - \psi_\nu(x_\nu).$$

We make three elementary observations:

$$0 = \lim_{\delta \to 1} \sup_{\nu \in [\delta, 1)} H\left(\frac{\nu - \rho\delta}{1 - \rho\delta}\right)(1 - \rho\delta), \qquad (A.21)$$

$$0 = \lim_{\delta \to 1} \sup_{\nu \in [\delta, 1)} \psi_\nu(x_\nu), \qquad (A.22)$$

$$\xi_\gamma(y_\gamma) \cdot (\nu - \rho\delta) \geq \xi_\rho(y_\rho) \cdot (1 - \rho)\delta, \qquad \nu \geq \delta. \qquad (A.23)$$

At the same time $\xi_\rho(y_\rho) > 0$ for $\rho < 1$; see Fig. A.2. Now fix $\rho < 1$ and $0 < \varepsilon < \xi_\rho(y_\rho)(1 - \rho)/2$. Combining (A.21)–(A.23), we have for sufficiently large $\delta < 1$,

$$\Psi_{\text{net}}(\nu; \rho, \delta) - \Psi_{\text{face}}(\nu; \rho, \delta) < \varepsilon - \xi_\rho(y_\rho)(1 - \rho)\delta$$
$$< -\xi_\rho(y_\rho)(1 - \rho)/2 < 0, \qquad \nu \in [\delta, 1). (A.24)$$

We conclude from the definition of ρ_N that $\rho_N(\delta) > \rho$. The desired relation (1.3) follows.

It follows in particular from (A.24) that for some $\delta < 1$, there is $\rho > \frac{1}{2}$ with

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

taking δ_0 to be the infimum of all such δ gives (1.4). $\qquad \square$

References

1. Fernando Affentranger and Rolf Schneider. Random projections of regular simplices. *Discrete Comput. Geom.*, 7(3):219–226, 1992.
2. Károly Böröczky, Jr. and Martin Henk. Random projections of regular polytopes. *Arch. Math.* (*Basel*), 73(6):465–473, 1999.
3. Wlodzimierz Bryc. A uniform approximation to the right normal tail integral. *Appl. Math. Comput.*, 127(2–3):365–374, 2002.

4. Geoffrey R. Burton. The nonneighbourliness of centrally symmetric convex polytopes having many vertices. *J. Combin. Theory Ser. A*, 58(2):321–322, 1991.

5. Emmanuel J. Candès and Terence Tao. Near optimal signal recovery from random projections and universal encoding strategies. Technical report, Applied and Computational Mathematics, California Institute of Technology, 2004.

6. Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*, second edition. Volume 38 of Applications of Mathematics (New York). Springer-Verlag, New York, 1998.

7. David L. Donoho. Compressed sensing. Technical report, Department of Statistics, Stanford University, 2004. Accepted *IEEE Trans. Inform. Theory*, pending revision.

8. 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. Accepted *Comm. Pure Appl. Math.*, pending revision.

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

10. David L. Donoho and Jared Tanner. Neighborliness of randomly-projected simplices in high dimensions. *Proc. Natl. Acad. Sci. USA*, 102:9452–9457, 2005.

11. David L. Donoho and Jared Tanner. Sparse nonnegative solutions of underdetermined linear equations by linear programming. *Proc. Natl. Acad. Sci. USA*, 102:9446–9451, 2005.

12. David Gale. Neighboring vertices on a convex polyhedron. In *Linear Inequalities and Related Systems*, pages 255–263. Annals of Mathematics Studies, no. 38. Princeton University Press, Princeton, NJ, 1956.

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

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

15. Nathan Linial and Isabella Novik. How neighborly can a centrally symmetric polytope be? Technical report, Department of Mathematics, Hebrew University, Jerusalem, 2005. Placed on ArXiv July 14, 2005.

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

17. Peter McMullen and Geoffrey C. Shephard. Diagrams for centrally symmetric polytopes. *Mathematika*, 15:123–138, 1968.

18. 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.

19. Harold Ruben. A convergent asymptotic expansion for Mill's ratio and the normal probability integral in terms of rational functions. *Math. Ann.*, 151:355–364, 1963.

20. Mark Rudelson and Roman Vershynin. Geometric approach to error-correcting codes and reconstruction of signals. Technical report, Department of Mathematics, University of California, Davis, 2005. Placed on ArXiv February 15, 2005.

21. Rolf Schneider. Neighbourliness of centrally symmetric polytopes in high dimensions. *Mathematika*, 22(2):176–181, 1975.

22. Anatoly Vershik and Piotr Sporyshev. An asymptotic estimate for the average number of steps in the parametric simplex method. *Zh. Vychisl. Mat. Mat. Fiz.*, 26(6):813–826, 958, 1986.

23. Anatoly Vershik and Piotr Sporyshev. Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem. *Selecta Math. Soviet.*, 11(2):181–201, 1992.