

Counting the Faces of Randomly-Projected Hypercubes and Orthants, with Applications

David L. Donoho · Jared Tanner

Received: 25 July 2008 / Revised: 7 August 2009 / Accepted: 10 August 2009 /
Published online: 30 September 2009
© Springer Science+Business Media, LLC 2009

Abstract Let A be an $n \times N$ real-valued matrix with $n < N$; we count the number of k -faces $f_k(AQ)$ when Q is either the standard N -dimensional hypercube I^N or else the positive orthant \mathbb{R}_+^N . To state results simply, consider a proportional-growth asymptotic, where for fixed δ, ρ in $(0, 1)$, we have a sequence of matrices A_{n, N_n} and of integers k_n with $n/N_n \rightarrow \delta$ and $k_n/n \rightarrow \rho$ as $n \rightarrow \infty$. If each matrix A_{n, N_n} has its columns in general position, then $f_k(AI^N)/f_k(I^N)$ tends to zero or one depending on whether $\rho > \min(0, 2 - \delta^{-1})$ or $\rho < \min(0, 2 - \delta^{-1})$. Also, if each A_{n, N_n} is a random draw from a distribution which is invariant under right multiplication by signed permutations, then $f_k(A\mathbb{R}_+^N)/f_k(\mathbb{R}_+^N)$ tends almost surely to zero or one depending on whether $\rho > \min(0, 2 - \delta^{-1})$ or $\rho < \min(0, 2 - \delta^{-1})$. We make a variety of contrasts to related work on projections of the simplex and/or cross-polytope.

These geometric face-counting results have implications for signal processing, information theory, inverse problems, and optimization. Indeed, face counting is related to conditions for uniqueness of solutions of underdetermined systems of linear equations. Below, let A be a fixed $n \times N$ matrix, $n < N$, with columns in general position.

The authors would like to thank the Isaac Newton Mathematical Institute at Cambridge University for hosting the programme “Statistical Challenges of High Dimensional Data” in 2008 and Professor D.M. Titterton for organizing this programme.

D.L. Donoho acknowledges support from NSF DMS 05-05303 and a Rothschild Visiting Professorship at the University of Cambridge.

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

D.L. Donoho
Department of Statistics, Stanford University, Stanford, CA, USA
e-mail: donoho@stanford.edu

J. Tanner (✉)
School of Mathematics, University of Edinburgh, Edinburgh, UK
e-mail: jared.tanner@ed.ac.uk

- (a) Call a vector in \mathbb{R}_+^N *k-sparse* if it has at most k nonzeros. For such a k -sparse vector x_0 , $b = Ax_0$ generates an underdetermined system $b = Ax$ having k -sparse solution. Among inequality-constrained systems $Ax = b$, $x \geq 0$, having k -sparse solutions, the fraction having a unique nonnegative solution is $f_k(A\mathbb{R}_+^N)/f_k(\mathbb{R}_+^N)$.
- (b) Call a vector in the hypercube I^N *k-simple* if all entries except at most k are at the bounds 0 or 1. For such a k -simple vector x_0 , $b = Ax_0$ generates an underdetermined system $b = Ax$ with k -simple solution. Among inequality-constrained systems $Ax = b$, $x \in I^N$, having k -simple solutions, the fraction having a unique hypercube-constrained solution is $f_k(AI^N)/f_k(I^N)$.

Keywords Zonotope · Random polytopes · Random cones · Random matrices · Wendel’s theorem · Winder–Cover theorem · Arrangements of hyperplanes · Compressed sensing · Threshold phenomena · Unique solution of underdetermined systems of linear equations

1 Introduction

There are three fundamental *regular* polytopes in \mathbb{R}^N , $N \geq 5$: the hypercube I^N , the cross-polytope C^N , and the simplex T^{N-1} . For each of these, projecting the vertices into \mathbb{R}^n , $n < N$, yields the vertices of a new polytope; in fact, up to translation and dilation, every polytope in \mathbb{R}^n is obtained by rotating the simplex T^{N-1} and orthogonally projecting on the first n coordinates, for some choice of N and of N -dimensional rotation. Similarly, every centrosymmetric polytope can be generated by projecting the cross-polytope, and every zonotope by projecting the hypercube.

1.1 Random Polytopes

Choosing the projection A at random has become popular. Let A be a random orthogonal projection obtained by first applying a uniformly-distributed rotation to \mathbb{R}^N and then projecting on the first n coordinates. Let Q be a polytope in \mathbb{R}^N . Then AQ is a random polytope in \mathbb{R}^n . Taking Q in turn from each of the three families of regular polytopes, we get three arenas for scholarly study:

- Random polytopes of the form AT^{N-1} were first studied by Affentranger and Schneider [2] and by Vershik and Sporyshev [25];
- Random polytopes of the form AC^N were first studied extensively by Böröczky and Henk [6];
- The random zonotope AI^N was studied in passing in [6] and will be heavily studied in this paper; a literature on zonotopes can be found in [3, 5, 20, 23, 29].

Starting with [2, 25], interest has focused on the number $f_k(AQ)$ of k -faces of such random polytopes AQ ; in those papers, fundamental formulas were developed for the expected values $\mathcal{E} f_k(AQ)$. Deriving insights from these formulae in the high-dimensional case has also been an important theme; Böröczky and Henk [6] studied the expected number $f_k(AQ)$ for each of these families of random polytopes, focusing on the asymptotic framework where the small dimension n is held fixed while the large dimension $N \rightarrow \infty$ (this was previously done for $Q = T^{N-1}$ in [2]).

Vershik and Sporyshev [25] studied the case AT^{N-1} in an asymptotic framework with the dimensions N and n both proportionally large, and observed a phenomenon of *sharp thresholds*: random polytopes can have face lattices undergoing abrupt changes in properties as dimensions change relatively slightly. Our own previous work considered both AT^{N-1} and AC^N [10, 12, 14, 16] and gave precise information about several such threshold phenomena.

To make precise the notion of “threshold phenomenon,” consider the following *proportional-dimensional* asymptotic framework. A *dimension specifier* is a triple of integers (k, n, N) , representing a “face” dimension k , a “small” dimension n , and a “large” dimension N ; $k < n < N$. For fixed $\delta, \rho \in (0, 1)$, consider sequences of dimension specifiers, indexed by n , and obeying

$$k_n/n \rightarrow \rho \quad \text{and} \quad n/N_n \rightarrow \delta \quad \text{as } n \rightarrow \infty. \tag{1.1}$$

For such sequences, the small dimension n is held proportional to the large dimension N as both dimensions grow. We omit subscripts on k_n and N_n when possible. For $Q = T^{N-1}, C^N$, the papers [10, 12, 14, 16] exhibited thresholds $\rho_W(\delta; Q)$ for the ratio between the expected number of faces of the low-dimensional polytope AQ and the number of faces of the high-dimensional polytope Q :

$$\lim_{n \rightarrow \infty} \frac{\mathcal{E} f_k(AQ)}{f_k(Q)} \begin{cases} = 1, & \rho < \rho_W(\delta; Q), \\ < 1, & \rho > \rho_W(\delta; Q). \end{cases} \tag{1.2}$$

(In this relation, we take a limit as $n \rightarrow \infty$ along some sequence obeying the proportional-dimensional constraint (1.1).) In words, the random object AQ has roughly as many k -faces as its generator Q for k below a threshold and has noticeably fewer k -faces than Q for k above the threshold. The threshold functions are defined in terms of Gaussian integrals and other special functions, and can be calculated numerically.

1.2 Random Zonotopes

Missing from the above picture is information about the third family of regular polytopes, the hypercube. Böröczky and Henk [6] mentioned in passing the case of the projected hypercube, in the case of A a random orthogonal projection. Böröczky and Henk largely worked in the asymptotic framework n fixed and $N \rightarrow \infty$. In that framework the threshold phenomenon of interest here is not visible. In this paper, we adopt the proportional-dimensional framework (1.1) and prove the following.

Theorem 1.1 (“Weak” Threshold for Hypercube) *Define*

$$\rho_W(\delta; I) := \max(0, 2 - \delta^{-1}), \quad 0 < \delta < 1. \tag{1.3}$$

For ρ, δ in $(0, 1)$, consider a sequence of dimension specifiers (k, n, N) obeying (1.1). Consider a sequence of real-valued $n \times N$ matrices $A = A_{n,N}$, each one with columns in general position in \mathbb{R}^n . Then

$$\lim_{n \rightarrow \infty} \frac{f_k(AI^N)}{f_k(I^N)} = \begin{cases} 1, & \rho < \rho_W(\delta, I), \\ 0, & \rho > \rho_W(\delta, I). \end{cases} \tag{1.4}$$

Remarks

- Use of the modifier “weak” and the subscript W on ρ matches corresponding usage with T^{N-1} and C^N .
- The result shows a sharp discontinuity in the behavior of the face lattices of random zonotopes; the location of the threshold is precisely identified. Such discontinuity is also observed empirically for the other two polytopes (1.2) above; to our knowledge, a proof of discontinuity has not yet been published in that setting.
- The result is *universal* across matrices; only general position is required. Universality of threshold effects across a range of matrix ensembles has been observed empirically for the other two regular polytopes [17]. However, theoretical results [1] for other polytopes do not yet match empirical facts. This result gives a rigorous universality result for one family of regular polytopes; this may inspire studies to see if parallel results exist for the others.

We briefly discuss the ideas leading to this result. Böröczky and Henk [6] applied a fundamental identity of Affentranger and Schneider [2] on general projected polytopes and gave the explicit expression

$$\mathcal{E} f_k(AI^N) = 2 \binom{N}{k} \sum_{\ell=N-n}^{N-k-1} \binom{N-k-1}{\ell}, \tag{1.5}$$

valid where A is a uniformly-distributed random orthoprojector. In a previous version of this manuscript [15], the authors proved that the same formula holds much more generally, in fact under the assumption that A has an orthant-symmetric nullspace in general position. One of our referees pointed out that even more is true: for any A in general position, $f_k(AI^N)$ is the *fixed number*

$$f_k(AI^N) = 2 \binom{N}{k} \sum_{\ell=N-n}^{N-k-1} \binom{N-k-1}{\ell}. \tag{1.6}$$

This fact follows from Theorem 1.7, [27], on partitions of n -space by hyperplanes, as we show below in Sect. 2.1. (1.6) appears to be known to workers on oriented matroids ([4, p. 220]) but may not seem evident to workers on convex polytopes. The recent survey article *What is known about unit cubes* states that “no good bound for... $[f_k(AI^N)]$ is known,” [29]. However, see [21, p. 410a].

1.3 Random Cones

Convex *cones* provide another family of fundamental polyhedral sets. Amongst these, the simplest and most natural is surely the positive orthant $P = \mathbb{R}_+^N$. The image $K = AP$ of a cone under projection $A: \mathbb{R}^N \rightarrow \mathbb{R}^n$ is again a cone. Such a cone may be expected to have $f_0(K) = 1$ vertex (at 0), and as many as $f_1(K) = N$ extreme rays, etc. In fact, every pointed cone in \mathbb{R}^n can be generated as a (nonorthogonal) projection of the positive orthant under an appropriate projection from an appropriate \mathbb{R}^N .

There seems to be relatively little prior research on random projections of the positive orthant, except for the special case $k = n$, which was studied by Buchta [8].

As with the polytope models, surprising threshold phenomena can arise when the projector is random, and we work in the proportional-dimensional framework. The following result makes use of the notion of a random matrix with centrosymmetric exchangeable columns; for details, see Sect. 2.2 below.

Theorem 1.2 (“Weak” Threshold for Orthant) *Let A be a random matrix with centrosymmetric exchangeable columns which are in general position almost surely. In the proportional-dimensional framework (1.1), we have*

$$\lim_{n \rightarrow \infty} \frac{\mathcal{E} f_k(A\mathbb{R}_+^N)}{f_k(\mathbb{R}_+^N)} = \begin{cases} 1, & \rho < \rho_W(\delta; \mathbb{R}_+) \\ 0, & \rho > \rho_W(\delta; \mathbb{R}_+) \end{cases} \tag{1.7}$$

with $\rho_W(\delta; \mathbb{R}_+) \equiv \rho_W(\delta; I)$ as defined in (1.3).

Here the threshold for the orthant is at precisely the same place as it was for the hypercube.

1.4 Exact Equality in the Number of Faces

Our focus in Sects. 1.1–1.3 was on “weak” agreement of $\mathcal{E} f_k(AQ)$ with $f_k(Q)$; in the proportional-dimensional framework, for ρ below threshold $\rho_W(\delta; Q)$, we have limiting relative equality:

$$\frac{\mathcal{E} f_k(A\mathbb{R}_+^N)}{f_k(\mathbb{R}_+^N)} \rightarrow 1, \quad n \rightarrow \infty.$$

We now focus on the “strong” agreement; it turns out that in the proportional dimensional framework, for ρ below a somewhat lower threshold $\rho_S(\delta; Q)$, we actually have exact equality with overwhelming probability:

$$\text{Prob}\{f_k(Q) = f_k(AQ)\} \rightarrow 1, \quad n \rightarrow \infty. \tag{1.8}$$

The existence of such “strong” thresholds for $Q = T^{N-1}$ and $Q = C^N$ was proven in [10, 12], which exhibited thresholds $\rho_S(\delta; Q)$ below which (1.8) occurs. These “strong thresholds” and the previously mentioned “weak thresholds” (1.2) are depicted in Fig. 3. A similar strong threshold exists for the projected orthant.

Theorem 1.3 (“Strong” Threshold for Orthant) *Let*

$$H(\gamma) := \gamma \log(1/\gamma) - (1 - \gamma) \log(1 - \gamma) \tag{1.9}$$

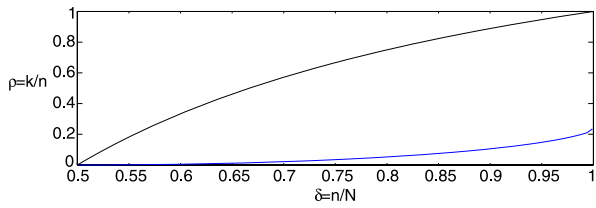
denote the usual (base- e) Shannon Entropy. Let

$$\psi_S^{\mathbb{R}_+}(\delta, \rho) := H(\delta) + \delta H(\rho) - (1 - \rho\delta) \log 2. \tag{1.10}$$

For $\delta \geq 1/2$, let $\rho_S(\delta; \mathbb{R}_+)$ denote the zero crossing of $\psi_S^{\mathbb{R}_+}(\delta, \rho)$. In the proportional-dimensional framework (1.1) with $\rho < \rho_S(\delta; \mathbb{R}_+)$,

$$\text{Prob}\{f_k(A\mathbb{R}_+^N) = f_k(\mathbb{R}_+^N)\} \rightarrow 1 \quad \text{as } n \rightarrow \infty. \tag{1.11}$$

Fig. 1 The “weak” thresholds, $\rho_W(\delta; I)$ and $\rho_W(\delta; \mathbb{R}_+)$ (black), and a lower bound on the strong threshold for the positive orthant, $\rho_S(\delta; \mathbb{R}_+)$ (blue)



The threshold $\rho_W(\delta; Q)$ for $Q = \mathbb{R}_+^N$ and I^N , and $\rho_S(\delta; \mathbb{R}_+)$ are depicted in Fig. 1. In contrast, the hypercube offers no phenomenon like (1.8).

Theorem 1.4 (Zonotope Vertices) *Let A be an $n \times N$ matrix with $n < N$. Then*

$$f_k(AI^N) < f_k(I^N), \quad k = 0, 1, 2, \dots, n.$$

Proof of Theorem 1.4 $f_k(AI^N)$ obtains its maximum when A is in general position, and in this case Theorem 1.8 gives the exact value of $f_k(I^N) - f_k(AI^N)$, a value which is strictly positive when $n < N$. □

1.5 Exact Nonasymptotic Results

We have so far emphasized the Vershik–Sporyshev proportional-dimensional asymptotic framework; this makes for the most natural comparisons between results for the three families of regular polytopes. However, for the positive orthant and hypercube, much more can be said than for the other two polytopes as there are simple exact expressions for finite N . Moreover, these expressions can be derived from two beautiful results in geometric probability, Wendel’s Theorem and Theorem 1.7.

Theorem 1.5 (Wendel, [26]) *Let M points in \mathbb{R}^m be drawn i.i.d. from a centrosymmetric distribution such that the points are in general position. Then the probability that all the points fall in some half space is*

$$P_{m,M} = 2^{-M+1} \sum_{\ell=0}^{m-1} \binom{M-1}{\ell}. \tag{1.12}$$

Wendel’s elegant result is often known as simply a piece of recreational mathematics. Our original submission [15] obtained from it a simple proof of the following identity.

Theorem 1.6 *Let A be an $n \times N$ random matrix with centrosymmetric exchangeable columns in general position almost surely. Then*

$$\frac{\mathcal{E} f_k(A\mathbb{R}_+^N)}{f_k(\mathbb{R}_+^N)} = 1 - P_{N-n, N-k}. \tag{1.13}$$

In this revision, the result derives from Theorem 1.7.¹

Theorem 1.7 (Winder [27]; Cover [9]) *A set of M hyperplanes in general position in \mathbb{R}^m , all passing through some common point, divides the space into $2^M P_{m,M}$ regions.*

This shows that $f_k(AI^N)$ satisfies the same formula as $\mathcal{E}f_k(A\mathbb{R}_+^N)$, but without the expectation.

Theorem 1.8 *Let A be an $n \times N$ matrix with columns in general position in \mathbf{R}^n . Then*

$$\frac{f_k(AI^N)}{f_k(I^N)} = 1 - P_{N-n, N-k}. \quad (1.14)$$

Formula (1.14) coincides with Böröczky and Henk's formula (1.6), [6]; but whereas (1.6) was proven for the case where A is a uniformly-distributed random orthoprojector, Theorem 1.8 holds for any A in general position. Theorem 1.8 is proven in Sect. 2.1. Theorem 1.6 is proven in Sect. 2.2, where it is derived from Theorem 1.8 by symmetrization.

1.6 Contents

Proofs of the above results are given in Sect. 2. The hypercube is contrasted with the other regular polytopes in Sect. 3; the cone and hypercube are contrasted in Sect. 4, where we also present additional results for specially constructed matrices. These phenomena, described here from the viewpoint of combinatorial geometry, have surprising consequences in probability theory, information theory, and signal processing; see [11, 13, 16] and Sect. 5.

2 Proofs of Main Results

We start with the key nonasymptotic exact identities (1.13) and (1.14) and then derive from (1.13) Theorems 1.2 and 1.3 by asymptotic analysis of the probabilities $P_{m,M}$. Throughout the paper we write $\mathcal{N}(A)$ for the nullspace of A .

2.1 Proof of Theorem 1.8

For convenience, in this section we let I^N denote the hypercube $[-1, 1]^N$. Each k -face F of I^N is a set of vectors with $N - k$ particular coordinates taking fixed,

¹This formula appears to have been derived by multiple authors independently at about the same time; in the discrete geometry literature, Winder's paper is often cited [27]; in the Machine Learning and Information Theory literature, Cover's paper [9] is typically cited instead; see Cover [9] for a history of early related results and of the method of proof dating back to Schläfli [22].

specific values, namely for each particular coordinate a specific choice from the endpoints $\{-1, 1\}^{N-k}$ applies for every member in the face. Within each face, the remaining k coordinate values may vary throughout the range $[-1, 1]^k$.

Let Q be a polyhedron (polytope or polyhedral cone) in \mathbb{R}^N and $x_0 \in Q$. The vector v is a *feasible direction* for Q at x_0 if $x_0 + tv \in Q$ for all sufficiently small $t > 0$. Let $\text{Feas}_{x_0}(Q)$ denote the cone of all feasible directions for Q at x_0 .

Lemma 2.1 *Let F be a k -face of the polytope or polyhedral cone Q , and let x_0 be a vector in $\text{relint}(F)$. For an $n \times N$ matrix A in general position, the following are equivalent:*

- (Survive(A, F, Q)): AF is a k -face of AQ ,
- (Transverse(A, x_0, Q)): $\mathcal{N}(A) \cap \text{Feas}_{x_0}(Q) = \{0\}$.

A proof of Lemma 2.1 is given in [24, p. 329].

Each face F of I^N can be identified with its centroid x_F ; this is a vector in \mathbb{R}^N with k of its coordinates = 0 and $N - k$ entries taking the value $\sigma_F \in \{-1, 1\}^{N-k}$. We speak of $\text{supp}(F)$, the *support* of F ; it is the set of indices of coordinates which vary among members of the face and σ_F , the *sign pattern* of F as the common sign pattern of the coordinates which do not vary among members of the face and so are outside the support. Thus, for example, if F is the set of all vectors with $-1 \leq x(1), \dots, x(k) \leq 1$ and $x(k + 1) = \dots = x(N) = 1$, then $\text{supp}(F) = \{1, \dots, k\}$ and $\sigma_F = (1, \dots, 1)$.

For each whole number m , let $[m] := \{1, \dots, m\}$. For an index set $J \subset [N]$ of cardinality k , let $\mathcal{F}(J)$ denote the collection of all k -faces F with $\text{supp}(F) = J$. There are of course 2^{N-k} such faces; they differ in the choice of σ_F .

Lemma 2.2 *Let A be an $n \times N$ matrix with $n < N$ whose columns are in general position in \mathbb{R}^n . Then*

$$\text{Card}_{F \in \mathcal{F}([k])} \{(\text{Survive}(A, F, I^N)) \text{ does not hold}\} = 2^{N-k} P_{N-n, N-k}.$$

Proof Let $\text{Feas}_{x_F}(I^N)$ denote the cone of feasible directions for I^N at x_F . The collection of such cones associated to a common support is a cover of \mathbb{R}^N :

$$\bigcup_{F \in \mathcal{F}([k])} \text{Feas}_{x_F}(I^N) = \mathbb{R}^N; \tag{2.1}$$

moreover, the terms appearing in the union have pairwise disjoint interiors. If F, G are distinct members of $\mathcal{F}([k])$, then

$$\text{int}(\text{Feas}_{x_F}(I^N)) \cap \text{int}(\text{Feas}_{x_G}(I^N)) = \emptyset; \tag{2.2}$$

roughly speaking, the collection of feasible cones associated to $\mathcal{F}([k])$ forms a partition of the space \mathbb{R}^N .

Define hyperplanes $H_j = \{x : x(j) = 0\}$; the hyperplanes $\{H_j : j = k + 1, \dots, N\}$, where the index avoids the support set $[k]$, also induce a partition of \mathbb{R}^N ; it is the same as the one induced by the above cones.

Set now $m = N - n$; by general position, $\mathcal{N}(A) \cong \mathbf{R}^m$. Set $M = N - k$ and define

$$\mathcal{H}_j = H_{k+j} \cap \mathcal{N}(A), \quad j = 1, \dots, M.$$

Since $\mathcal{N}(A)$ is in general position, these are relative hyperplanes of $\mathcal{N}(A) \cong \mathbf{R}^m$. Thus, up to linear isomorphism, $\{\mathcal{H}_j : j = 1, \dots, M\}$ is a collection of hyperplanes in general position in \mathbf{R}^m ; these hyperplanes intersect in the common point 0. Theorem 1.7, tells us that \mathbf{R}^m is partitioned by M hyperplanes into $2^M P_{m,M}$ regions.

Correspondingly $\mathcal{N}(A)$ is partitioned into $2^{N-k} P_{N-n,N-k}$ regions. The relative interior of each such region in $\mathcal{N}(A)$ belongs to the interior of exactly one cone $\text{Feas}_{x_F}(I^N) \subset \mathbf{R}^N$ (by (2.2) and (2.1)). That cone specifies exactly one k -face F for which $(\text{Transverse}(A, x_F, I^N))$ does not hold. Equivalently, $(\text{Survive}(A, F, I^N))$ does not hold. □

Theorem 1.8 follows from Lemmas 2.1 and 2.2 by noting that the set of all k -faces of I^N can be partitioned cleanly by specifying one of the $\binom{N}{k}$ k -element subsets $J \subset [N]$, $\text{card}(J) = k$, and then considering $\mathcal{F}(J)$. In combinatorics one denotes by $\binom{N}{k}$ the collection of different k -element subsets of $[N]$. Thus we have the disjoint union

$$\mathcal{F}_k(I^N) = \bigsqcup \left\{ \mathcal{F}(J) : J \in \binom{N}{k} \right\}.$$

Hence,

$$\begin{aligned} f_k(I^N) - f_k(AI^N) &:= \text{Card}_{F \in \mathcal{F}_k(I^N)} \{ (\text{Survive}(A, F, I^N)) \text{ does not hold} \} \\ &= \sum_{J \in \binom{N}{k}} \text{Card}_{F \in \mathcal{F}(J)} \{ (\text{Survive}(A, F, I^N)) \text{ does not hold} \} \\ &= \binom{N}{k} \cdot 2^{N-k} P_{N-n,N-k}. \end{aligned} \tag{2.3}$$

2.2 Proof of Theorem 1.6

In the original submission of this manuscript, we proved (1.13) using Wendel’s Theorem and then derived (1.14) from it, by an averaging argument. Prompted by a referee, in this revision, we go in the opposite direction, having first proved Theorem 1.8 using Theorem 1.7, we now derive (1.13) from Theorem 1.8 by symmetrization.

We start with the following observation on the expected number of k -faces of \mathbb{R}_+^N :

$$\frac{\mathcal{E} f_k(A\mathbb{R}_+^N)}{f_k(\mathbb{R}_+^N)} = \text{Ave}_F [\text{Prob}\{AF \text{ is a } k\text{-face of } A\mathbb{R}_+^N\}]. \tag{2.4}$$

Here Ave_F denotes “the arithmetic mean over all k -faces of $A\mathbb{R}_+^N$.”

In this section, it is convenient to let $I^N = [0, 1]^N$. This choice does not affect face counts. With this representation of I^N , it becomes true that, the “lower k -faces” of I^N are in one-one correspondence with the k -faces of \mathbb{R}_+^N . Namely, if, in the previous subsection’s notation, F is a k -face of I^N with $\sigma_F > 0$, then the cone $\text{pos}(F)$ is a k -face of \mathbb{R}_+^N . Adopt now the notational convention that within the proof, for a lower face F of I^N , $\tilde{F} = \text{pos}(F)$ denotes the corresponding face of \mathbb{R}_+^N . We observe that for a vector x_0 with nonnegative coordinates all strictly less than 1, we have

$$\text{Feas}_{x_0}(\mathbb{R}_+^N) = \text{Feas}_{x_0}(I^N). \tag{2.5}$$

Combining this with Lemma 2.1 applied once to I^N and once to \mathbb{R}_+^N , we see that if A is a matrix with columns in general position, and F is a lower face of I^N ,

$$\text{Prob}\{A\tilde{F} \text{ is a } k\text{-face of } A\mathbb{R}_+^N\} = \text{Prob}\{AF \text{ is a } k\text{-face of } AI^N\}.$$

Now the face counts for the projected hypercube obey:

$$\frac{\mathcal{E} f_k(A\mathbb{R}_+^N)}{f_k(\mathbb{R}_+^N)} = \text{Ave}_F[\text{Prob}\{AF \text{ is a } k\text{-face of } AI^N\}]. \tag{2.6}$$

Here Ave_F denotes “the arithmetic mean over all k -faces of AI^N .”

Definition 2.3 (Centrosymmetric Exchangeable Columns) Let A be a random n by N matrix such that, for each signed permutation matrix Π and for every measurable set Ω ,

$$\text{Prob}\{A \in \Omega\} = \text{Prob}\{A\Pi \in \Omega\}.$$

Then we say that A has *centrosymmetric exchangeable columns*.

Below we assume without loss of generality that A has centrosymmetric exchangeable columns. Then all k -faces of \mathbb{R}_+^N become statistically equivalent:

$$\text{Prob}\{AF \text{ is a } k\text{-face of } A\mathbb{R}_+^N\} = \text{Prob}\{AG \text{ is a } k\text{-face of } A\mathbb{R}_+^N\}$$

for each distinct F, G in $\mathcal{F}_k(\mathbb{R}_+^N)$; indeed, there is always a permutation Π for which G is the image of F under Π : $G = \Pi F$, and the probabilities are Π -invariant. Then (2.4) becomes: let F be a fixed k -face of \mathbb{R}_+^N ; then

$$\frac{\mathcal{E} f_k(A\mathbb{R}_+^N)}{f_k(\mathbb{R}_+^N)} = \text{Prob}\{AF \text{ is a } k\text{-face of } A\mathbb{R}_+^N\}. \tag{2.7}$$

Similarly, all k -faces of I^N become statistically equivalent; indeed, there is always a signed permutation Π for which G is the image of F under Π : $G = \Pi F$, and the probabilities are Π -invariant. Hence (2.6) becomes: let F be a fixed k -face of I^N ; then

$$\frac{\mathcal{E} f_k(AI^N)}{f_k(I^N)} = \text{Prob}\{AF \text{ is a } k\text{-face of } AI^N\}. \tag{2.8}$$

Combining these displays with (1.14) implies (1.13).

2.3 Some Generalities About Binomial Probabilities

The probability $P_{m,M}$ has a classical interpretation: it gives the probability of at most $m - 1$ heads in $M - 1$ tosses of a fair coin. The usual Normal approximation to the binomial tells us that

$$P_{m,M} \approx \Phi\left(\frac{(m-1) - (M-1)/2}{\sqrt{(M-1)/4}}\right),$$

with Φ the usual standard normal distribution function $\Phi(x) = \int_{-\infty}^x e^{-y^2/2} dy / \sqrt{2\pi}$; here the approximation symbol \approx can be made precise using standard limit theorems, e.g., appropriate for small or large deviations. In this expression, the approximating normal has mean $(M-1)/2$ and standard deviation $\sqrt{(M-1)/4}$. There are three regimes of interest, for large m, M , and three behaviors for $P_{m,M}$.

- Lower Tail: $m \ll M/2 - \sqrt{M/4}$. $P_{m,M} \approx 0$.
- Middle: $m \approx M/2$. $P_{m,M} \in (0, 1)$.
- Upper Tail: $m \gg M/2 + \sqrt{M/4}$. $P_{m,M} \approx 1$.

2.4 Proof of Theorem 1.2

Using the correspondence $N - n \leftrightarrow m$, $N - k \leftrightarrow M$, and the connection to Wendel's theorem, we have three regimes of interest:

- $N - n \ll (N - k)/2$.
- $N - n \approx (N - k)/2$.
- $N - n \gg (N - k)/2$.

In the proportional-dimensional framework, the above discussion translates into three separate regimes, and separate behaviors we expect to be true:

- Case 1: $\rho < \rho_W(\delta; \mathbb{R}_+)$. $P_{N-n, N-k} \rightarrow 0$.
- Case 2: $\rho = \rho_W(\delta; \mathbb{R}_+)$. $P_{N-n, N-k} \in (0, 1)$.
- Case 3: $\rho > \rho_W(\delta; \mathbb{R}_+)$. $P_{N-n, N-k} \rightarrow 1$.

Case 2 is trivially true, but it has no role in the statement of Theorem 1.2. Cases 1 and 3 correspond exactly to the two parts of (1.7) that we must prove.

To prove Cases 1 and 3, we need an upper bound deriving from standard large-deviations analysis of the lower tail of the binomial.

Lemma 2.4 *Let $N - n < (N - k)/2$. Then*

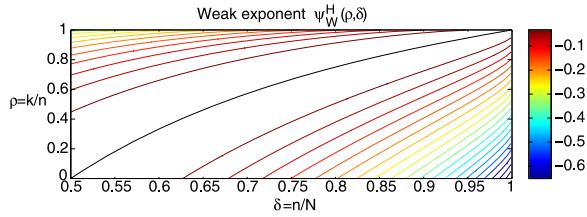
$$P_{N-n, N-k} \leq n^{3/2} \exp\left(N \psi_W^{\mathbb{R}_+}\left(\frac{n}{N}, \frac{k}{n}\right)\right), \quad (2.9)$$

where the exponent is defined as

$$\psi_W^{\mathbb{R}_+}(\delta, \rho) := H(\delta) + \delta H(\rho) - H(\rho\delta) - (1 - \rho\delta) \log 2 \quad (2.10)$$

with $H(\cdot)$ the Shannon Entropy (1.9), see Fig. 2.

Fig. 2 Exponent for the weak phase transition, $\psi_{W^+}^{\mathbb{R}^+}(\rho, \delta)$, (2.10), which has its zero level curve at $\rho_W(\delta; \mathbb{R}_+)$, (1.3). The projected hypercube has the same weak phase transition and exponent $\psi_W^I \equiv \psi_W^{\mathbb{R}^+}$



Proof Upperbounding the sum in $P_{N-n, N-k}$ by $N - n - 1$ times $\binom{N-k-1}{N-n}$, we arrive at

$$P_{N-n, N-k} \leq 2^{N-k-1} \frac{n-k}{N-k} \cdot (N-k+1) \binom{N}{n} \binom{n}{k} \binom{N}{k}^{-1}. \tag{2.11}$$

We can bound $\binom{m}{\gamma \cdot m}$ for $\gamma < 1$ using the Shannon entropy (1.9):

$$c_1 n^{-1/2} e^{mH(\gamma)} \leq \binom{m}{\gamma \cdot m} \leq c_2 e^{mH(\gamma)}, \tag{2.12}$$

where $c_1 := \frac{16}{25} \sqrt{2/\pi}$, $c_2 := 5/4 \sqrt{2\pi}$. Recalling the definition of $\psi_W^{\mathbb{R}^+}$, we obtain (2.9). □

We will now consider Cases 1 and 3, and prove the corresponding conclusion.

Case 1: $\rho < \rho_W(\delta; \mathbb{R}_+)$. The threshold function $\rho_W(\delta; \mathbb{R}_+)$ is the location of the lowest zero crossing $\psi_W^{\mathbb{R}^+}(\delta, \rho)$ as a function of ρ for δ fixed; i.e.,

$$\rho_W(\delta; \mathbb{R}_+) = \inf\{\rho : \psi_W^{\mathbb{R}^+}(\delta, \rho) \geq 0\}.$$

Thus, for any ρ strictly below $\rho_W(\delta; \mathbb{R}_+)$, the exponent $\psi_W^{\mathbb{R}^+}(\delta, \rho)$ is strictly negative. Lemma 2.4 thus implies that $P_{N-n, N-k_n} \rightarrow 0$ as $n \rightarrow \infty$.

Case 3: $\rho > \rho_W(\delta; \mathbb{R}_+)$. Binomial probabilities have a standard symmetry (re-label every “head” outcome as a “tail”, and vice versa). It follows that $P_{m, M} = 1 - P_{M-m, M}$. We have $P_{N-k, N-n} = 1 - P_{N-k, n-k}$. In this case $N - n > (N - k)/2$, so Lemma 2.4 tells us that $P_{N-k, n-k} \rightarrow 0$ as $n \rightarrow \infty$; we conclude $P_{N-k, N-n} \rightarrow 1$ as $n \rightarrow \infty$.

2.5 Proof of Theorem 1.3

$P_{N-n, N-k}$ is the probability that one fixed k -dimensional face F of \mathbb{R}_+^N generates a k -face AF of $A\mathbb{R}_+^N$. The probability that some k -dimensional face generates a k -face can be upperbounded, using Boole’s inequality, by $f_k(\mathbb{R}_+^N) \cdot P_{N-n, N-k}$.

From (2.12), (2.9), and $f_k(\mathbb{R}_+^N) = \binom{N}{k}$ we have

$$f_k(\mathbb{R}_+^N) \cdot P_{N-n, N-k} \leq n^{3/2} \exp(N\psi_S^{\mathbb{R}^+}(\delta_n, \rho_n)),$$

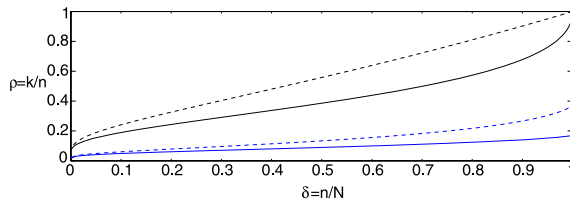


Fig. 3 Weak thresholds for the simplex, $\rho_W(\delta; T)$ (black-dash), and cross-polytope, $\rho_W(\delta; C)$ (black-solid). Consider sequences obeying the proportional-dimensional asymptotic with parameters δ, ρ . For (δ, ρ) below these curves and for large n , each projected polytope has nearly as many k -faces as its generator; above these curves the projected polytope has noticeably fewer. Strong thresholds for the simplex, $\rho_S(\delta; T)$ (blue-dash), and cross-polytope, $\rho_S(\delta; C)$ (blue-solid). For (δ, ρ) below these curves and for large n , each projected polytope and its generator typically have exactly the same number of k -faces

where $\psi_S^{\mathbb{R}_+}$ was defined earlier in (1.10) as

$$\psi_S^{\mathbb{R}_+}(\delta, \rho) := H(\delta) + \delta H(\rho) - (1 - \rho\delta) \log 2. \tag{2.13}$$

Recall that for $\delta \geq 1/2$, $\rho_S(\delta; \mathbb{R}_+)$ is the location of the lowest zero crossing of $\psi_S^{\mathbb{R}_+}$ as a function of ρ for δ fixed; i.e.,

$$\psi_S^{\mathbb{R}_+}(\delta; \mathbb{R}_+) = \inf\{\rho : \psi_S^{\mathbb{R}_+}(\delta, \rho) \geq 0\}.$$

For any $\rho < \rho_S(\delta; \mathbb{R}_+)$, we have $\psi_S^{\mathbb{R}_+}(\delta, \rho) < 0$, and as a result (1.11) follows.

3 Contrasting the Hypercube with Other Polytopes

The theorems in Sect. 1 contrast strongly with existing results for other polytopes.

3.1 Nonexistence of Weak Thresholds at $\delta < 1/2$

Theorem 1.1 identifies a region of the phase diagram ($\delta = \frac{n}{N}, \rho = \frac{k}{N}$) where the typical random zonotope has nearly as many k -faces as its generating hypercube; in particular, if $n < N/2$, it has many fewer k -faces than the hypercube, for every k . This behavior at $\delta = n/N < 1/2$ is quite different from the behavior seen for random projections of the simplex and the cross-polytope at small δ . Those polytopes have $f_k(AQ) \approx f_k(Q)$ for quite a large range of k even at relatively small values of δ ; [16] shows that we can have $k \sim n/(2 \log(\delta^{-1}))$ for small δ at both of those polytopes, while we could not have even $k = 1$ at the hypercube for such small δ ; see also the visual evidence in Fig. 3.

3.2 Nonexistence of Strong Thresholds for Hypercube

Lemma 1.4 shows that projected zonotopes always have strictly fewer k -faces than their generators $f_k(AI^N) < f_k(I^N)$, for every $n < N$. This is again quite different from the situation with the simplex and the cross-polytope, where we can

even have $n \ll N$ and still find k for which $f_k(AQ) = f_k(Q)$, [16], roughly $k \sim n/(2e \log(\delta^{-1}))$; again see visual evidence in Fig. 3.

3.3 Universality of Weak Phase Transitions

Theorem 1.1 holds for any A in general position.

In proving weak and strong threshold results for the simplex and cross-polytope, we required A to either be a random ortho-projector or to have Gaussian i.i.d. entries. Thus, what we proved for those families of regular polytopes applies to a much more limited range of matrix ensembles than for hypercubes.

4 Contrasting the Cone with the Hypercube

4.1 Universality of Weak Phase Transitions

For Theorem 1.2, A can be sampled from any ensemble of random matrices invariant under right multiplication by signed permutations. The result is thus *universal* across a wide class of matrix ensembles.

In proving weak and strong threshold results for the simplex and cross-polytope, we required A to either be a random ortho-projector or to have Gaussian i.i.d. entries. Thus, what we proved for those families of regular polytopes applies to a much more limited range of matrix ensembles than for the orthant.

Our empirical studies suggest that the same ensembles of matrices which “work” for the orthant weak threshold also “work” for the simplex and cross-polytope thresholds. It seems to us that the universality across matrix ensembles proven here may point to a much larger phenomenon, valid also for other polytope families; however, this universality class is far more restrictive than is the case for the hypercube. For our empirical studies, see [17].

The weak threshold for the orthant depends very much more delicately on details about A than do the hypercube thresholds; unlike $f_k(I^N)$, $f_k(A\mathbb{R}_+^N)$ is *not* the same number for all A in general position. It makes a substantial difference to the results if the matrix A is not “zero-mean.”

4.2 The Low-Frequency Partial Fourier Matrix

Consider the special partial Fourier matrix made only of the n lowest frequency entries.

Corollary 4.1 *Assume n is odd and, for $j = 1, 2, \dots, N$, let*

$$\Omega_{ij} = \begin{cases} \cos\left(\frac{\pi(j-1)(i-1)}{N}\right), & i = 1, 3, 5, \dots, n, \\ \sin\left(\frac{\pi(j-1)i}{N}\right), & i = 2, 4, 6, \dots, n-1. \end{cases} \tag{4.1}$$

Then

$$f_k(\Omega\mathbb{R}_+^N) = f_k(\mathbb{R}_+^N), \quad k = 0, 1, \dots, \frac{1}{2}(n-1).$$

The result is a corollary of [18, Theorem 3, p. 56]. The key steps of the proof are included in an extended technical report [15].

This behavior is dramatically different than the case for random A of the type considered so far, and in some sense dramatically better.

Corollary 4.1 is closely connected with the classical question of *neighborliness*. There are famous polytopes which can be generated by projections AT^{N-1} and have exactly as many k -faces as T^{N-1} for $k \leq \lfloor n/2 \rfloor$. A standard example is provided by the matrix Ω defined in (4.1); it obeys $f_k(\Omega T^{N-1}) = f_k(T^{N-1})$, $0 \leq k \leq \lfloor n/2 \rfloor$. (There is a vast literature touching in some way on the phenomenon $f_k(\Omega T^{N-1}) = f_k(T^{N-1})$. In that literature, the polytope ΩT^{N-1} is usually called a *cyclic polytope*, and the columns of Ω are called points of the *trigonometric moment curve*; see standard references [21, 28].)

Hence the matrix Ω offers both $f_k(\Omega T^{N-1}) = f_k(T^{N-1})$ and $f_k(\Omega \mathbb{R}_+^N) = f_k(\mathbb{R}_+^N)$ for $0 \leq k \leq \lfloor n/2 \rfloor$. This is exceptional. For random A of the type discussed in earlier sections, there is a large disparity between the sets of triples (k, n, N) where $f_k(AT^{N-1}) = f_k(T^{N-1})$ —this happens for $k/n < \rho_S(n/N; T)$ —and those where $f_k(A\mathbb{R}_+^N) = f_k(\mathbb{R}_+^N)$ —this happens for $k/n < \rho_S(n/N; \mathbb{R}_+)$. These two strong thresholds are displayed in Figs. 3 and 1, respectively.

Even if we relax our notion of agreement of face counts to weak agreement, the collections of triples where $f_k(AT^{N-1}) \approx f_k(T^{N-1})$ and $f_k(A\mathbb{R}_+^N) \approx f_k(\mathbb{R}_+^N)$ are very different, because $\rho_W(n/N; T)$ and $\rho_W(n/N; \mathbb{R}_+)$ are so dramatically different, particularly at $n < N/2$.

4.3 Adjoining a Row of Ones to A

An important feature of the random matrices A studied earlier is orthant symmetry. In particular, the positive orthant plays no distinguished role with respect to these matrices. On the other hand, the partial Fourier matrix Ω constructed in the last subsection contains a row of ones, and thus the positive orthant has a distinguished role to play for this matrix. Moreover, this distinction is crucial; we find empirically that *removing* the row of ones from Ω causes the conclusion of Corollary 4.1 to fail drastically.

Conversely, consider the matrix \tilde{A} obtained by *adjoining* a row of N ones to some matrix A :

$$\tilde{A} = \begin{bmatrix} 1 \\ A \end{bmatrix}.$$

Adding this row of ones to a random matrix causes a drastic shift in the strong and weak thresholds.

Theorem 4.1 *Consider the proportional-dimensional asymptotic framework with parameters δ, ρ in $(0, 1)$. Let a random $(n - 1) \times N$ matrix A have i.i.d. standard normal entries. Let \tilde{A} denote the corresponding $n \times N$ matrix whose first row is all ones and whose remaining rows are identical to those of A . Then*

$$\lim_{n \rightarrow \infty} \frac{\mathcal{E} f_k(\tilde{A}\mathbb{R}_+^N)}{f_k(\mathbb{R}_+^N)} = \begin{cases} 1, & \rho < \rho_W(\delta, T), \\ < 1, & \rho > \rho_W(\delta, T); \end{cases} \tag{4.2}$$

$$\lim_{n \rightarrow \infty} P\{f_k(\tilde{A}\mathbb{R}_+^N) = f_k(\mathbb{R}_+^N)\} = \begin{cases} 1, & \rho < \rho_S(\delta, T), \\ 0, & \rho > \rho_S(\delta, T). \end{cases} \tag{4.3}$$

Note particularly the *mixed* form of this relationship. Although the conclusions concern the behavior of faces of the randomly-projected *orthant*, the thresholds are those that were previously obtained for the randomly-projected *simplex*.

Since there is such a dramatic difference between $\rho(\delta; T)$ and $\rho(\delta; \mathbb{R}_+)$, the single row of ones can fairly be said to have a huge effect. In particular, the region “below” the simplex weak phase transition $\rho_W(\delta; T)$ comprises ≈ 0.5634 of the (δ, ρ) parameter area, and the hypercube weak phase transition $\rho_W(\delta; I)$ comprises $1 - \log 2 \approx 0.3069$.

Theorem 4.1 is an immediate consequence of the following identity.

Lemma 4.2 *Suppose that the row vector 1 is not in the row span of A. Then*

$$f_k(\tilde{A}\mathbb{R}_+^N) = f_{k-1}(AT^{N-1}), \quad 0 < k < n.$$

Proof We observe that there is a natural bijection between k -faces of \mathbb{R}_+^N and the $(k - 1)$ -faces of T^{N-1} . The $(k - 1)$ -faces of T^{N-1} are in bijection with the corresponding support sets of cardinality k : i.e., we can identify with each k -face F the union I of all supports of all members of the face. Similarly, to each support set I of cardinality k , there is a unique k -face \tilde{F} of \mathbb{R}_+^N consisting of all points in \mathbb{R}_+^N whose support lies in I . Composing bijections $F \leftrightarrow I \leftrightarrow \tilde{F}$, we have the bijection $F \leftrightarrow \tilde{F}$.

Concretely, let x_0 be a point in the relative interior of some $(k - 1)$ -face F of T^{N-1} . Then x_0 has k nonzeros. x_0 is also in the relative interior of the k -face \tilde{F} of \mathbb{R}_+^N . Conversely, let y_0 be a point in the relative interior of some k -face of \mathbb{R}_+^N ; then $x_0 = (1' y_0)^{-1} y_0$ is a point in the relative interior of a $(k - 1)$ -face of T^{N-1} .

The last two paragraphs show that for each pair of corresponding faces (F, \tilde{F}) , we may find a point x_0 in both the relative interior of \tilde{F} and also of the relative interior of F . For such x_0 ,

$$\text{Feas}_{x_0}(\mathbb{R}_+^N) = \text{Feas}_{x_0}(T^{N-1}) + \text{lin}(x_0).$$

Clearly $\mathcal{N}(\tilde{A}) \cap \text{lin}(x_0) = \{0\}$, because $1' x_0 > 0$. We conclude that the following are equivalent:

$$(\text{Transverse}(A, x_0, T^{N-1})): \quad \mathcal{N}(A) \cap \text{Feas}_{x_0}(T^{N-1}) = \{0\}.$$

$$(\text{Transverse}(\tilde{A}, x_0, \mathbb{R}_+^N)): \quad \mathcal{N}(\tilde{A}) \cap \text{Feas}_{x_0}(\mathbb{R}_+^N) = \{0\}.$$

Rephrasing [13], the following are equivalent for x_0 a point in the relative interior of F :

$$(\text{Survive}(A, F, T^{N-1})): \quad AF \text{ is a } (k - 1)\text{-face of } AT^{N-1},$$

$$(\text{Transverse}(A, x_0, T^{N-1})): \quad \mathcal{N}(A) \cap \text{Feas}_{x_0}(T^{N-1}) = \{0\}.$$

We conclude that for two corresponding faces F, \tilde{F} , the following are equivalent:

$$\begin{aligned} (\text{Survive}(A, F, T^{N-1})): & \quad AF \text{ is a } (k-1)\text{-face of } AT^{N-1}, \\ (\text{Survive}(\tilde{A}, \tilde{F}, \mathbb{R}_+^N)): & \quad \tilde{A}\tilde{F} \text{ is a } k\text{-face of } \tilde{A}\mathbb{R}_+^N. \end{aligned}$$

Combining this with the natural bijection $F \leftrightarrow \tilde{F}$, the lemma is proved. \square

5 Application: Compressed Sensing

Our face counting results can all be reinterpreted as statements about “simple” solutions of underdetermined systems of linear equations. This reinterpretation allows us to make connections with numerous problems of current interest in signal processing, information theory, and probability. The reinterpretation follows from the two following lemmas, which are restatements of Lemma 2.1 for $Q = \mathbb{R}_+^N$ and $Q = I^N$, rephrasing the notion of $(\text{Transverse}(A, x_0, Q))$ with the all but linguistically equivalent $(\text{Unique}(A, x_0, Q))$.

Lemma 5.1 *Let x_0 be a vector in \mathbb{R}_+^N with exactly k nonzeros. Let F denote the associated k -face of \mathbb{R}_+^N . For an $n \times N$ matrix A , let AF denote the image of F under A and $b_0 = Ax_0$ the image of x_0 under A . The following are equivalent:*

$$\begin{aligned} (\text{Survive}(A, F, \mathbb{R}_+^N)): & \quad AF \text{ is a } k\text{-face of } A\mathbb{R}_+^N, \\ (\text{Unique}(A, x_0, \mathbb{R}_+^N)): & \quad \text{The system } b_0 = Ax \text{ has a unique solution in } \mathbb{R}_+^N. \end{aligned}$$

Lemma 5.2 *Let x_0 be a vector in I^N with exactly k entries strictly between the bounds $\{0, 1\}$. Let F denote the associated k -face of I^N . For an $n \times N$ matrix A , let AF denote the image of F under A and $b_0 = Ax_0$ the image of x_0 under A . The following are equivalent:*

$$\begin{aligned} (\text{Survive}(A, F, I^N)): & \quad AF \text{ is a } k\text{-face of } AI^N, \\ (\text{Unique}(A, x_0, I^N)): & \quad \text{The system } b_0 = Ax \text{ has a unique solution in } I^N. \end{aligned}$$

Note that the systems of linear equations referred to in these lemmas are underdetermined: $n < N$. Hence these lemmas identify conditions on underdetermined systems of linear equations such that, when the solutions are known to obey certain constraints, such seemingly weak prior knowledge in fact uniquely determines the solution. The first result can be paraphrased as saying that nonnegativity constraints can be very powerful if the object is known to have relatively few nonzeros; the second result says that upper and lower bounds can be very powerful, provided that many of those bounds are binding.

5.1 Reconstruction Exploiting Nonnegativity Constraints

We wish to reconstruct the unknown x , knowing only the linear measurements $b = Ax$, the matrix A , and the constraint $x \in \mathbb{R}_+^N$.

Let $J(x)$ be some function of x . Consider the positivity-constrained variational problem

$$(\text{Pos}_J) \quad \min J(x) \quad \text{subject to } b = Ax, \quad x \in \mathbb{R}_+^N.$$

Let $\text{pos}_J(b, A)$ denote any solution of the problem instance (Pos_J) defined by data b and matrix A .

We conclude the following:

Corollary 5.1 *Suppose that*

$$f_k(A\mathbb{R}_+^N) = f_k(\mathbb{R}_+^N).$$

Let $x_0 \geq 0$ and $\|x_0\|_{\ell^0} \leq k$. For the problem instance defined by $b = Ax_0$,

$$\text{pos}_J(b, A) = x_0.$$

In words: under the given conditions on the face numbers, *any* variational prescription which imposes nonnegativity constraints will correctly recover the k -sparse solution in *any* problem instance where such a k -sparse solution exists.

Corresponding to this “strong” statement is a “weak” statement. Consider the following probability measure on k -sparse problem instances.

- Choose a random subset L of size k from $\{1, \dots, N\}$, by k simple random draws without replacement.
- Set the entries of x_0 not in the selected subset to zero.
- Choose the entries of x_0 in the selected set L from some fixed joint distribution ψ_L supported in $(0, 1)^k$.
- Generate the problem instance $b = Ax_0$.

We speak of drawing a k -sparse random problem instance at random.

Corollary 5.2 *Suppose that for some $\epsilon \in (0, 1)$,*

$$f_k(A\mathbb{R}_+^N) \geq (1 - \epsilon) \cdot f_k(\mathbb{R}_+^N).$$

For (b, A) a problem instance drawn at random, as above:

$$\text{Prob}\{\text{pos}_J(b, A) = x_0\} \geq (1 - \epsilon).$$

In words: under the given conditions on the face lattice, *any* variational prescription which imposes nonnegativity constraints will correctly succeed to recover the k -sparse solution in at least a fraction $(1 - \epsilon)$ of all k -sparse problem instances.

For more discussion, including potential applications, see [7, 13, 18, 19].

5.2 Reconstruction Exploiting Box Constraints

Consider again the problem of reconstruction from measurements $b = Ax$, but this time assuming that the object x obeys box-constraints: $0 \leq x(j) \leq 1, 1 \leq j \leq N$.

Define the box-constrained variational problem

$$(\text{Box}_J) \quad \min J(x) \quad \text{subject to } b = Ax, \quad 0 \leq x(j) \leq 1, \quad j = 1, \dots, N.$$

Let $\text{box}_J(b, A)$ denote any solution of the problem instance (Box_J) defined by data b and matrix A . In this setting, the notion corresponding to “sparse” is “simple”. We say that a vector x is k -simple if at most k of its entries differ from the bounds $\{0, 1\}$.

Consider the following probability measure on problem instances having k -simple solutions. Recall that k -simple vectors have all entries equal to 0 or 1 except at k exceptional locations.

- Choose the subset L of k exceptional entries uniformly at random from the set $\{1, \dots, N\}$ without replacement.
- Choose the nonexceptional entries to be either 0 or 1 based on tossing a fair coin.
- Choose the values of the exceptional k entries according to a joint probability measure ψ_L supported in $(0, 1)^k$.
- Define the problem instance $b = Ax_0$.

Corollary 5.3 *Suppose that for some $\epsilon \in (0, 1)$,*

$$f_k(AI^N) \geq (1 - \epsilon) \cdot f_k(I^N).$$

Randomly sample a problem instance (b, A) using the method just described. Then

$$P\{\text{box}_J(b, A) = x_0\} \geq (1 - \epsilon).$$

In words: under the given conditions on the face lattice, *any* variational prescription which imposes box constraints will correctly recover at least a fraction $(1 - \epsilon)$ of all underdetermined systems generated by the matrix A which have k -simple solutions.

In the hypercube case there is no phenomenon comparable to that which arose in the positive orthant with the special constructions Ω and \tilde{A} ; $f_k(AI^N)$ is a fixed number if A is in general position and decreases if A is not in general position. Consequently, the hypercube weak threshold is the best general result on the ability to undersample by exploiting box constraints. In particular, the difference between the weak simplex threshold and the weak hypercube threshold has this interpretation:

A given degree k of sparsity of a nonnegative object is much more powerful than that same degree of simplicity of a box-constrained object.

Specifically, *we should not expect to be able to undersample a typical box-constrained object by more than a factor of 2* and then reconstruct it using some garden-variety variational prescription. In comparison, the last section showed that we can severely undersample very sparse nonnegative objects. Moreover, when $n < N$, there is no region where $f_k(AI^N) = f_k(I^N)$, and consequently box constraints are never enough to ensure $\text{box}_J(b, A) = x_0$ for all k -simple problem instances.

Acknowledgements Art Owen suggested that we pay attention to Wendel’s Theorem. We also thank Goodman, Pollack, and Schneider for providing scholarly background and the anonymous referees for the most useful and instructive referee reports we have ever seen.

References

1. Adamczak, R., Litvak, A.E., Pajor, A., Tomczak-Jaegermann, N.: Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling. [arXiv:0904.4723v1](https://arxiv.org/abs/0904.4723v1) (2009)
2. Affentranger, F., Schneider, R.: Random projections of regular simplices. *Discrete Comput. Geom.* **7**(3), 219–226 (1992)
3. Baryshnikov, Y.M.: Gaussian samplesm, regular simplices, and exchangeability. *Discrete Comput. Geom.* **17**(3), 257–261 (1997)
4. Björner, A., Las Vergns, M., Sturmfels, B., White, N., Ziegler, G.: Oriented Matroids. *Encyclopedia of Mathematics and Its Applications*, vol. 46. Cambridge University Press, Cambridge (1999)
5. Bolker, E.D.: A class of convex bodies. *Trans. Am. Math. Soc.* **145**, 323–345 (1969)
6. Böröczky, K., Jr., Henk, M.: Random projections of regular polytopes. *Arch. Math. (Basel)* **73**(6), 465–473 (1999)
7. Bruckstein, A.M., Elad, M., Zibulevsky, M.: On the uniqueness of non-negative sparse and redundant representations. In: ICASSP 2008. Special session on Compressed Sensing, Las Vegas, Nevada (2008)
8. Buchta, C.: On nonnegative solutions of random systems of linear inequalities. *Discrete Comput. Geom.* **2**(1), 85–95 (1987)
9. Cover, T.M.: Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Electron. Comput.* **EC-14**(3), 326–334 (1965)
10. Donoho, D.L.: High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput. Geom.* **35**(4), 617–652 (2006)
11. Donoho, D.L.: Neighborly polytopes and sparse solutions of underdetermined linear equations. Technical Report, Stanford University (2006)
12. Donoho, D.L., Tanner, J.: Neighborliness of randomly-projected simplices in high dimensions. *Proc. Natl. Acad. Sci. USA* **102**(27), 9452–9457 (2005)
13. Donoho, D.L., Tanner, J.: Sparse nonnegative solutions of underdetermined linear equations by linear programming. *Proc. Natl. Acad. Sci. USA* **102**(27), 9446–9451 (2005)
14. Donoho, D.L., Tanner, J.: Exponential bounds implying construction of neighborly polytopes, error-correcting codes and compressed sensing matrices by random sampling. Preprint (2007)
15. Donoho, D.L., Tanner, J.: Counting the faces of randomly-projected hypercubes and orthants, with applications. [arXiv:0807.3590v1](https://arxiv.org/abs/0807.3590v1) (2008)
16. Donoho, D.L., Tanner, J.: Counting faces of randomly-projected polytopes when the projection radically lowers dimension. *J. AMS* **22**(1), 1–53 (2009)
17. Donoho, D.L., Tanner, J.: Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos. Trans. A* (2009)
18. Donoho, D.L., Johnstone, I.M., Hoch, J.C., Stern, A.S.: Maximum entropy and the nearly black object. *J. R. Stat. Soc., Ser. B (Methodological)* **54**(1), 41–81 (1992)
19. Fuchs, J.-J.: On sparse representations in arbitrary redundant bases. *IEEE Trans. Inf. Theory* **50**(6), 1341–1344 (2004)
20. Goodey, P., Weil, W.: Zonoids and generalisations. In: *Handbook of Convex Geometry*, vols. A, B, pp. 1297–1326. North-Holland, Amsterdam (1993)
21. Grünbaum, B.: *Convex Polytopes*, 2nd edn. Graduate Texts in Mathematics, vol. 221. Springer, New York (2003). Prepared and with a preface by Volker Kaibel, Victor Klee, and Günter M. Ziegler
22. Schläfli, L.: In: *Gesammelte Mathematische Abhandlungen*, vol. 1, pp. 209–212. Birkhäuser, Basel (1950)
23. Schneider, R., Weil, W.: Zonoids and related topics. In: *Convexity and Its Applications*, pp. 296–317. Birkhäuser, Basel (1983)
24. Schneider, R., Weil, W.: *Stochastic and Integral Geometry*. Springer, Berlin (2008)
25. Vershik, A.M., Sporyshev, P.V.: Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem. *Sel. Math. Sov.* **11**(2), 181–201 (1992)
26. Wendel, J.G.: A problem in geometric probability. *Math. Scand.* **11**, 109–111 (1962)
27. Winder, R.O.: Partitions of n -space by hyperplanes. *SIAM J. Appl. Math.* **14**(4), 811–818 (1966)
28. Ziegler, G.M.: *Lectures on Polytopes*. Graduate Texts in Mathematics, vol. 152. Springer, New York (1995)
29. Zong, C.: What is known about unit cubes. *Bull. AMS* **42**(2), 181–211 (2005)