Neighborly Polytopes and Sparse Solution of Underdetermined Linear Equations

David L. Donoho

December 2004

Abstract

Consider a $d \times n$ matrix A, with $d < n$. The problem of solving for x in $y = Ax$ is underdetermined, and has many possible solutions (if there are any). In several fields it is of interest to find the sparsest solution – the one with fewest nonzeros – but in general this involves *combinatorial* optimization.

Let a_i denote the i-th column of A, $1 \leq i \leq n$. Associate to A the *quotient polytope* P formed by taking the convex hull of the $2n$ points $(\pm a_i)$ in $\mathbf{R}^d$. P is centrosymmetric and is called (centrally) *k-neighborly* if every subset of $k+1$ elements $(\pm_{i_l} a_{i_l})_{l=1}^{k+1}$ are the vertices of a face of P.

We show that if P is k-neighborly, then if a system $y = Ax$ has a solution with at most k nonzeros, that solution is also the unique solution of the *convex* optimization problem $\min \|x\|_1$ subject to $y = Ax$; the converse holds as well.

This complete equivalence between the study of sparse solution by ℓ^1 minimization and neighborliness of convex polytopes immediately gives new results in each field. On the one hand, we get new families of neighborly centrosymmetric polytopes, by exploiting known results about sparsity of ℓ^1 minimization; on the other, we get new limits on the ability of ℓ^1 minimization to find sparse solutions, by exploiting known limits on neighborliness of centrally symmetric polytopes.

Weaker notions of equivalence between ℓ^1 and sparse optimization have also been studied recently. These are equivalent to other interesting properties of the quotient polytope. Thus, suppose the columns of A are in general position. Consider the vectors having $k < d/2$ nonzeros that are simultaneously the sparsest solution of $y = Ax$ and the minimal ℓ^1 solution. These make up a fraction $1 - \epsilon$ of all vectors with k nonzeros if and only if the quotient polytope P has at least $1 - \epsilon$ as many k-dimensional faces as the regular cross polytope C^n. Combining this with recent work on face numbers of randomly-projected cross-polytopes, we learn that for large d, the overwhelming majority of systems of linear equations with d equations and $4d/3$ unknowns have the following property: *if there is a solution with fewer than $.49d$ nonzeros, it is the unique minimum ℓ^1 solution.*

A stylized application in digital communication is sketched; for large n, it is possible to transmit $n/4$ pieces of information using a codeword of length n with immunity to $.49n$ gross errors in the received codeword, if the signs and sites of the gross errors are random, and with immunity to $.11n$ gross errors chosen by a malicious opponent. The receiver uses ℓ^1 minimization.

Key Words and Phrases: Centrosymetric Polytopes. Centrally-Neighborly Polytopes. Underdetermined Systems of Linear Equations. ℓ^1 optimization. Combinatorial Optimization. Signal Recovery with Gross Errors. Breakdown Point.

1 Introduction

Consider an underdetermined system of linear equations $y = Ax$, where $y \in \mathbf{R}^d$, $x \in \mathbf{R}^n$, A is an $n \times d$ matrix, $d < n$, y is known and x is unknown. If there are any solutions to this system, there are many. However, in numerous fields, driven by parsimony, one wants to find the *sparsest* solution – the one with fewest nonzeros. Formally, one wants to solve

$$(P_0) \qquad \min \|x\|_0 \text{ subject to } y = Ax.$$

where the 0-'norm' $\|x\|_0$ counts the number of nonzeros. Because of the extreme non-convexity of the zero-'norm', (P_0) is NP-hard in general; it fact it contains various combinatorial optimization problems (knapsack, satisfiability) as special cases. However, as we will see below, there has recently been a great deal of interest spurred by the realization that for many matrices A, the sparsest solution – if it is sufficiently sparse – is available by ℓ^1 minimization, namely by solving

$$(P_1) \qquad \min \|x\|_1 \text{ subject to } y = Ax.$$

This is a convex optimization problem and can be considered tractable. The phenomenon of interest is that for certain matrices A, whenever the solution to (P_0) is sufficiently sparse, it is also the solution of (P_1). As a general label, we call this phenomenon *ℓ^1/ℓ^0 equivalence*. For literature, see [10, 9, 12, 14, 17, 23, 4].

This paper develops an understanding of this equivalence phenomenon based on ideas from the theory of convex polytopes; the books of Grünbaum [18] and Ziegler [26] are typical starting points. Specifically, we associate to the matrix A the convex polytope $P = AC$ where $C \subset \mathbf{R}^n$ is the n-dimensional cross-polytope, characterized equally as the convex hull of the signed unit basis vectors $\pm e_i$ with $i = 1, \dots n$ and as the ℓ^1 ball in $\mathbf{R}^n$:

$$\|x\|_1 \leq 1.$$

The polytope P is centrosymmetric, and is called *k-neighborly* if, whenever we take $k+1$ vertices not including an antipodal pair, the resulting vertices span a face of P. Thus, all sensible ways of combining vertices create valid faces.

(Note: the notion of neighborliness we discuss here is the one appropriate to centrally-symmetric polytopes – hence the proviso 'not including an antipodal pair'. It is discussed in, for example, [21, 22, 3]. For asymmetric polytopes one can use the simpler notion of neighborliness without this proviso; see, for example [15, 16, 18]. The notion we consider is sometimes called 'centrally neighborly' e.g. [19]. Unless we say otherwise, only the 'centrosymmetric interpretation' is intended.)

In Section 3 we connect neighborliness to the question of ℓ^1/ℓ^0 equivalence.

Theorem 1 *Let A be a $d \times n$ matrix, $d < n$. These two properties of A are equivalent:*

- *The quotient polytope P has $2n$ vertices and is k-neighborly,*
- *Whenever $y = Ax_0$ has a solution x_0 having at most k nonzeros, x_0 is the unique optimal solution of the optimization problem (P_1).*

1.1 Corollaries

In short, two areas of scholarly work – ℓ^1/ℓ^0 equivalence and neighborliness of polytopes – are tightly connected. In Section 4, we use Theorem 1 to immediately transfer several results from one area to the other one. To cite a very simple example, we use known results on ℓ^1/ℓ^0 equivalence in so-called *incoherent dictionaries* to get

Corollary 1.1 *Let a_i denote the i-th column of A. Suppose that $\|a_i\|_2 = 1$, $1 \le i \le n$ and*

$$|\langle a_i, a_j \rangle| \le 1/(2k-1), \qquad i \neq j;$$

then $P = AC$ is k-neighborly.

There are many examples of matrices with $d < n$ satisfying this condition; a simple example is the concatenation of the d by d identity matrix with a d by d Hadamard matrix, making a $d \times n$ matrix $A = [IH]$ with $n = 2d$. The corollary implies that P is $\sqrt{d}/2$-neighborly. Actually, the quotient polytope is roughly $.9\sqrt{d}$-neighborly, as we discuss in Section 4 below.

It is of course more interesting to consider polytopes which are even more highly neighborly, perhaps proportional to d. To do this, we apply known results on ℓ^1/ℓ^0 equivalence with large random A with d and n. In this result, d and n are increasing to ∞ together in a proportional way.

Corollary 1.2 *Let $0 < \delta < 1$, and let n tend to infinity along with $d = d_n = \lfloor \delta n \rfloor$, and let $A = A^{d,n}$ be a random d by n matrix whose columns are iid random points on S^{d-1}. There is $\rho = \rho(\delta) > 0$ so that, with overwhelming probability for large n, $P = AC$ is ρd-neighborly.*

In this corollary, P is the convex hull of a random set of points together with their antipodes. The points are uniformly-distributed on the sphere. This is a natural model of 'uniform random centrosymmetric polytope' in dimension d with $2n$ vertices. From that perspective, Corollary 1.2 shows that 'most' centrosymmetric polytopes have neighborliness 'proportional to dimension'. Prior to this result, there appears to be no work in the polytope literature implying neighborliness proportional to dimension when n is some multiple of d. For example, from Schneider's work [22], we can get k-neighborliness for $k \approx .23d$, but only for $n = d + O(1)$, d large.

One can go in the other direction, getting results on sparse solution from results on neighborliness. In Section 4 we point out that bounds on neighborliness of centrosymmetric polytopes by McMullen and Shephard [21] imply the following

Corollary 1.3 *Let $n - 2 \ge d > 2$. Then if, with the matrix A, ℓ^1 minimization correctly finds all sparse solutions having $\le k$ nonzeros,*

$$k \le \lfloor (d+1)/3 \rfloor. \tag{1.1}$$

This establishes what seems, to the author, a surprising limit on the sparsity level at which ℓ^1 minimization can solve ℓ^0 problems. While the upper bound $k < d/2$ can be easily seen without knowledge of polytopes, (1.1) is much stronger and unlikely to have been stumbled across.

Thus the connection we establish with Theorem 1 opens new insights in each area.

1.2 Other Notions of Equivalence

Students of ℓ^1/ℓ^0 equivalence have also used weaker notions of correspondence between solutions of the two problems – correspondence between the solutions for *most* x with $\le k$ nonzeros, rather than for all x with $\le k$ nonzeros. Below we describe the two main such notions which have been proposed, and show that each one is equivalent to an interesting property of the quotient polytope.

One of these weaker notions - *local equivalence* - asks that, for a given $I \subset \{1, \dots, n\}$, every x supported in I is both the sparsest solution to $y = Ax$ and the minimal ℓ^1 solution. This turns out equivalent to saying that the section of P by the linear span of $(\pm_i a_i : i \in I)$, is

$|I|$-neighborly. In consequence, *most* sets I of size k are sets of local equivalence if and only if *most* k-dimensional 'intrinsic' sections are k-neighborly.

The other notion - *individual equivalence* - asks that, for a fraction $1 - \epsilon$ of vectors x with k nonzeros, x is both the sparsest solution to $y = Ax$ and the minimal ℓ^1 solution. If the columns of A are in general position, this turns out equivalent to saying that P has at least $1 - \epsilon$ times as many $(k-1)$-faces as C.

Each of these weaker ℓ^1/ℓ^0 equivalences suggests notions of weak neighborliness. Existing results immediately yield examples of centrosymmetric polytopes with low degree of neighborliness but high degree of weak neighborliness. Thus for example, applying results of [10, 12] with the breakthrough of Candès-Romberg-Tao [4] we conclude:

Corollary 1.4 *Let d be a perfect square. The quotient polytope P generated by $A = [I\ F]$ with I the $d \times d$ identity and F the real $d \times d$ Fourier matrix is not $\sqrt{d}$-neighborly. But most k-dimensional intrinsic sections are k-neighborly for $k \approx cd/\log(d)$, for a certain $c > 0$.*

In the other direction, the above equivalences allow methods for counting faces of quotient polytopes to give new insights into sparsity properties of ℓ^1 optimization. In [8], the author refines tools developed to count faces of randomly-projected polytopes and develops a series of results implying precise quantitative bounds on the strong and weak neighborliness of $P = AC$ when A is a random orthoprojector, and d is proportional to n. These results from counting polytope faces imply the strongest known results on the combinations k, n, d for which ℓ^1/ℓ^0 equivalence is highly likely. As a simple example:

Corollary 1.5 *There is $\delta_0 < 1$ with the following property. Let n and d_n tend to ∞ together so that $d_n = \lfloor \delta \cdot n \rfloor$, where $\delta_0 < \delta < 1$. Let $k = k(n)$ obey $k/d \leq .49$.*

Let $y = Ax_0$, where x_0 contains nonzeros at k sites selected uniformly at random, with signs chosen uniformly at random (amplitudes can have any distribution), and where A is a uniform random orthoprojector from $\mathbf{R}^n$ to $\mathbf{R}^d$.

With overwhelming probability for large n, the minimum ℓ^1 norm solution to $y = Ax$ is also the sparsest solution, and is precisely x_0.

In this result .49 can be replaced by any other number $< 1/2$. Numerical evidence provided in [8] shows that $\delta_0 \approx .7$. This justifies the claim in the abstract that, for large n, among all underdetermined systems with n unknowns and $3n/4$ equations and which admit a solution with $\leq .49n$ nonzeros, the overwhelming majority have a minimal ℓ^1 solution which is unique and which is also the sparsest solution. Section 7 gives further discussion, and numerical evidence.

Section 8 describes a stylized application: transmitting $n/4$ pieces of information over a channel subject to $.49n$ randomly-triggered gross errors or $.11n$ maliciously-chosen gross errors, with perfect recovery by ℓ^1 minimization.

1.3 Contents

Sections 2 and 3 prove Theorem 1, while Section 4 explains how Corollaries 1.1-1.3 follow from Theorem 1 and existing results. Section 5 defines local equivalence between ℓ^1 and ℓ^0 optimization, proves equivalence to sectional neighborliness, and relates this to Corollary 1.4. Section 6 defines individual equivalence between ℓ^1 and ℓ^0 optimization, and proves that (when the columns of A are in general position) this equivalence is prevalent if and only if P has nearly as many k-faces as C. Section 7 discusses random projection and other random linear systems and proves Corollary 1.5. Section 8 discusses applications to 'perfect recovery' of signals corrupted by overwhelming, impulsive noise.

2 Preliminaries

Before beginning the proof, we make a few obvious observations connecting (P_1) with the polytope P. Note that the value function of (P_1) is a function of $y \in R^d$; we call this the *quotient norm*:

$$Q(y) = val(P_1) = \inf \|x\|_1 \text{ subject to } y = Ax.$$

The quotient polytope $P = AC$ is also the unit ball for this norm:

$$P = \{y : Q(y) \le 1\}.$$

Indeed, the cross-polytope C is the set of x with ℓ^1 norm bounded by 1. $Q(y) \le 1$ just in case y is the image Ax of such an x; but this means $Q(y) \le 1$ exactly in $AC = P$.

The unit ball for this norm can also be characterized as the set of y realizable as convex combinations of signed columns of A. The reader will find it instructive to prove the following.

Lemma 2.1 *Consider the problem of representing $y \in \mathbf{R}^d$ as a convex combination of the signed columns $(a_1, -a_1, \dots, a_n, -a_n)$. This problem has a solution if and only if $val(P_1) \le 1$. If this problem has a unique solution then (P_1) has a unique solution for this y.*

We now fix notation concerning convex polytopes; see [18] for more details. In discussing the (closed, convex) polytope P, we commonly refer to its vertices $v \in \mathrm{vert}(P)$ and k-dimensional faces $F \in \mathcal{F}_k(P)$. $v \in P$ will be called a vertex of P if there is a linear functional λ_v separating v from $P \backslash \{v\}$, 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 $P = \mathrm{conv}(\mathrm{vert}(P))$. Vertices are just 0-dimensional faces, and a k-dimensional face is a 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 $\mathrm{vert}(F) \subset \mathrm{vert}(P)$; thus if F is a face, $F = \mathrm{conv}(\mathrm{vert}(F))$. A k-dimensional face will be called a *k-simplex* if it has $k + 1$ vertices. Important for us will be the fact that for k-neighborly polytopes, all the low-dimensional faces are simplices.

It is standard to define the face numbers $f_k(P) = \#\mathcal{F}_k(P)$. We also need the simple observation that

$$\mathrm{vert}(AC) \subset A\,\mathrm{vert}(C), \tag{2.1}$$

which implies

$$\mathcal{F}_\ell(AC) \subset A\mathcal{F}_\ell(C), \qquad 0 \le \ell < d; \tag{2.2}$$

and so the numbers of vertices obey

$$f_0(AC) \le f_0(C). \tag{2.3}$$

3 Equivalence

We now turn to the proof of Theorem 1.

3.1 Basic Insights

In our opinion there are two insights, which we record as lemmas. The first is the importance and convenience of having simplicial faces of P.

Lemma 3.1 (Unique Representation). *Consider a k-face $F \in \mathcal{F}_k(P)$ and suppose that F is a k-simplex. Let $x \in F$. Then*

[a] *x has a unique representation as a convex combination of vertices of P.*

[b] *This representation places nonzero weight only on vertices of F.*

Conversely, suppose that F is a k-dimensional closed convex subset of P with properties [a] and [b] for every $x \in F$. Then F is a k-simplex and a k-face of P.

Proof. Let $v_1, \ldots, v_N$ be the vertices of P. As F is a simplex, it has $k+1$ vertices; without loss of generality take these as $v_1, \ldots, v_{k+1}$. The vertices of F are affinely independent; hence there is a unique representation of $x \in F$ as a convex combination of the vertices.

The fact that F is a face implies the existence of a linear functional λ and constant c so that $\lambda(v_1) = \ldots \lambda(v_{k+1}) = c$ while for other vertices $\lambda(v_i) < c$, $i > k+1$. Suppose there were a convex combination $x = \sum_i \beta_i v_i$ placing nonzero weight on $i > k+1$. Then

$$\begin{aligned} \lambda(x) &= \sum_{i=1}^{k+1} \beta_i \lambda(v_i) + \sum_{i=k+2}^{N} \beta_i \lambda(v_i) \\ &< \sum_{i=1}^{k+1} \beta_i c + \sum_{i=k+2}^{N} \beta_i c = c, \end{aligned}$$

but $\lambda(x) < c$ contradicts $x \in F$. Hence there is no such convex combination.

For the converse direction, assumption [a] implies that F does not meet the interior of P, since throughout the interior, the representation by convex combination is nonunique. Hence F is a subset of some k-face of P - G, say. Now the uniqueness property [a] can hold on a k-dimensional subset of G only if G is a simplex, for we need the vertices of G to be affinely independent. Similarly, the weighting property [b] can hold on the k-dimensional set F only if F contains at least the vertices of G. As $F \subset G$ but $\mathrm{vert}(G) \subset F$ we conclude $F = G$, i.e. F is a k-simplex and a k-face of P. □

The second insight: neighborliness can be thought of as saying that the $k-1$-faces of P are simply images under A of the faces of C – nothing more complicated for the face lattice is allowed, even though for non-neighborly polytopes it would be expected.

Lemma 3.2 (Alternate Form of Neighborliness). *Suppose the centrosymmetric polytope $P = AC$ has $2n$ vertices and is k-neighborly. Then*

$$\forall \ell = 0, \ldots, k-1, \quad \forall\, F \in \mathcal{F}_\ell(C), \quad AF \in \mathcal{F}_\ell(AC). \tag{3.1}$$

Conversely, suppose that (3.1) holds; then $P = AC$ has $2n$ vertices and is k-neighborly.

Proof. Since P has $2n$ vertices, (2.1)-(2.3) tell us that these must be exactly the $\pm a_i$.

Now each set of k disjoint indices $i_1, \ldots i_k$ $1 \le i_\ell \le n$, and k signs $\sigma_\ell \in \{+1, -1\}$ $\ell = 1, \ldots k$ determines a face F of the cross-polytope C. Let e_i be the i-th canonical unit basis vector in R^n. Then

$$F = \mathrm{conv}\{\sigma_1 e_{i_1}, \ldots, \sigma_k e_{i_k}\}.$$

Let a_i denote the i-th column of A. Then

$$AF = \mathrm{conv}\{\sigma_1 a_{i_1}, \ldots, \sigma_k a_{i_k}\}.$$

We have seen that the $\sigma_\ell a_{i_\ell}$ are vertices of P, and by the neighborliness assumption, their convex hull makes a face of P. But this convex hull is just AF which therefore makes a face of AC.

In the converse direction, note that (3.1) for $\ell = 0$ exactly says that AC has $2n$ vertices.

Note that for $\ell = k-1$ (3.1) tells us that, for each set of k disjoint indices $i_1, \ldots i_k$, $1 \le i_\ell \le n$, and k signs $\sigma_\ell \in \{+1, -1\}$, $\mathrm{conv}(\{\sigma_\ell a_{i_l} : 1 \le \ell \le k\})$ is a face of P. This is exactly the definition of k-neighborly. □

3.2 Main Result, Forward Direction

At last we turn to the proof of Theorem 1, in the forward direction. We suppose that P is k-neighborly, that x_0 has at most k nonzeros, and show that the minimal ℓ^1-norm solution is precisely x_0. We assume without loss of generality that the problem is scaled so that $\|x_0\|_1 = 1$.

Now since x_0 has at most k nonzeros, it belongs to a k-dimensional face F of the cross-polytope: $F \in \mathcal{F}_k(C)$. Hence y belongs to AF, which, by neighborliness and Lemma 3.2, is a k-dimensional face of P. Now, by Lemma 3.1, y has a unique representation by the vertices of P, which is a representation by the vertices of AF only, and which is unique. But x_0 already provides such a representation. It follows that x_0 is the unique representation for y obeying

$$\|x\|_1 \leq 1.$$

Hence it is the unique solution of (P_1). □

3.3 Main Result, Converse Direction

We now go in the converse direction, and suppose that A has the property that, if $y = Ax_0$ with x_0 having fewer than k nonzeros, then x_0 is the unique solution to the instance of (P_1) generated by y. We then derive that P has $2n$ vertices and is k-neighborly.

By considering the case $k = 1$ with every $x_i = \pm e_i$, we learn that in each case the corresponding $y_i = Ax_i$ belongs to P and is uniquely representable among convex combinations of $(\pm_j a_j)_j$ by $\pm a_i$. This implies by Lemma 3.1 above that each y_i is a vertex of P, so P has at least $2n$ vertices. Since by (2.3) the number of vertices of $P = AC$ is at most the number of vertices of C, we see that P has exactly $2n$ vertices.

Consider now a collection of k disjoint indices $i_1, \dots i_k$, $1 \leq i_\ell \leq n$, and k signs $\sigma_\ell \in \{+1, -1\}$. By hypothesis, for every x_0 which can be generated

$$x_0 = \sum_\ell \alpha_\ell \sigma_\ell e_{i_\ell},$$

with $\alpha_\ell \geq 0$ and $\sum_\ell \alpha_\ell = 1$ the corresponding problem (P_1) based on $y = Ax_0$ has a unique solution, equal to x_0. Since this latter problem has a unique solution, there is (by Lemma 2.1) a unique solution to the problem of representing each such y as a convex combination of signed columns of A, and that solution is provided by the corresponding x_0. By the converse part of Lemma 3.1, AF is a face in $\mathcal{F}_k(AC)$.

Combining the last two paragraphs with the converse part of Lemma 3.2, we conclude that P has $2n$ vertices and is k-neighborly. □

4 Corollaries

We now explain how Theorem 1 yields Corollaries 1-3. To simplify our discussion, we introduce terminology:

Definition 1 *The* **Equivalence Breakdown Point** *of the matrix A, $EBP(A)$, is the maximal number N such that, for every x_0 with fewer than N nonzeros, the corresponding vector $y = Ax_0$ yields problems (P_1) and (P_0) with identical unique solutions, both equal to x_0.*

With this terminology, Theorem 1 can be restated so:

$$P \text{ is } k\text{-neighborly iff } EBP(A) > k. \tag{4.1}$$

Corollary 1 flows from the following result, proved recently by several different authors.

Theorem 4.1 *[10, 9, 12, 14, 17, 23] Suppose that* $\|a_i\|_2 = 1$, $1 \le i \le n$ *and*

$$|\langle a_i, a_j \rangle| \le M, \qquad i \ne j;$$

then $EBP(A) > (M^{-1} + 1)/2$.

Corollary 1.1 follows by setting $M = 1/(2k-1)$. For an example, let H be a $d \times d$ Hadamard matrix. Then $A = [IH]$ yields $M = \frac{1}{\sqrt{d}}$, and we get that $P = AC$ is k-neighborly with $k = \lceil \sqrt{d}/2 \rceil$.

The dependence of $EBP(A)$ on d in Theorem 4.1 is partially disappointing; something proportional to d would be more interesting. In that direction, the following result implies the 'neighborliness-proportional-to-dimension' of Corollary 2.

Theorem 4.2 *[7] Let* $A_{d,n}$ *have its columns sampled iid from the uniform distribution on the sphere* S^{d-1}. *Fix* $\delta \in (0,1)$. *There is* $\rho = \rho(\delta) > 0$ *so that for* $d \ge \delta n$,

$$Prob\{EBP(A_{d,n}) > \rho d\} \to 1, \qquad n \to \infty.$$

Corollary 1.2 follows on noting (4.1). The proof in [7] gives only very crude estimates of $\rho(\delta)$, which are absurdly small. Emmanuel Candès has used different methods (see Theorem 7.1 below) to estimate ρ but again with a very small value as the result. In contrast – see Section 7 below – polytope methods; allow to show that for $\delta = .5$, $\rho \ge .089$.

Now we turn in the other direction, and establish Corollary 1.3. We invoke the following result of McMullen and Shephard [21]:

Theorem 4.3 *Let* P *be a centrosymmetric* d*-polytope with* $d \ge 2$ *and* $n \ge d + 2$. *If* P *is* k*-neighborly, we have*

$$k \le \lfloor \frac{d+1}{3} \rfloor.$$

Corollary 1.3 follows on noting (4.1). We believe that while most students of ℓ^1/ℓ^0 equivalence would expect that getting k close to 1/2 could be challenging, none would have claimed that it's actually impossible, even in small-scale cases. Yet that's the necessary implication.

5 Local Equivalence and Sectional Neighborliness

The notion of equivalence between ℓ^1 and ℓ^0 discussed above is very strong; it holds for *any* x_0 with at most k nonzeros. Two weaker notions of ℓ^1/ℓ^0 equivalence make sense, where we consider *most* rather than *all* x_0 of a certain form. Each is equivalent to an interesting property of the quotient polytope P. In the next sections we discuss each of these in turn. We start with a notion equivalent to neighborliness of low-dimensional sections.

5.1 Sets of Local Equivalence

Fix the number of nonzeros k in x_0; the support I of x_0 can be any of the $\binom{n}{k}$ possible choices.

Definition 2 *We say there is* **local equivalence** *(between* ℓ^1 *and* ℓ^0 *optimization)* **at a given support** $I \subset \{1, \ldots, n\}$ *if, for every* x_0 *supported in* I, *the corresponding problems* (P_0) *and* (P_1) *generated by* $y = Ax_0$ *both have* x_0 *as their unique solution. A set* I *having this property is called a* **set of local equivalence**.

The adjective 'local' emphasizes the possibility that ℓ^1/ℓ^0 equivalence can hold for some support sets while failing on others. Note if $EBP(A) > k$, then *every* support set of size $\leq k$ is a set of local equivalence. Hence the EBP notion is seen to be 'global' rather than local.

Since the ℓ^1/ℓ^0 equivalence phenomenon was first identified, it has been recognized that global equivalence can be far stronger than local equivalence [10]. Indeed, there is the possibility that for a given matrix A, $EBP(A) \ll k$ while 'most' support sets I of size k exhibit local equivalence [10, Final Section].

A simple and important example is offered by the d by $2d$ matrix

$$A = [I \ F]$$

where I is the d by d identity and F is the d by d real Fourier transform matrix. This example arises in attempting to represent a signal as a sum of 'spikes' and sinusoids. Although there is no unique way to do this – the system $y = Ax$ is underdetermined – [10] showed that, if a signal y is made from fewer than $\sqrt{d}/2$ spikes and sinusoids, then (P_1) will precisely recover those terms.

The $\sqrt{d}/2$ bound can be of the right order for global equivalence. [10] shows that if d is a perfect square, $EBP(A) \leq \sqrt{d}$; [12, 13] shows that $EBP(A) \approx .9\sqrt{d}$. Hence, (P_1) is only guaranteed to recover the sparsest solution in the Identity/Fourier case if it has somewhat fewer than $\sqrt{d}$ nonzeros.

In a breakthrough, Candès, Romberg, and Tao [4] showed that *most* support sets I of size $k \sim cd/\log(d)$ are sets of local equivalence; here c is an absolute constant. Thus the size of subsets under which we can typically have local equivalence may be dramatically larger than the size under which we have global equivalence, at least for this specific class of matrix.

The result of [4] can be interpreted probabilistically, as saying that if a set of at most k spikes and sinusoids is chosen uniformly at random, and if $k \leq cd/\log(d)$, then – even though a representation by sums of spikes and sinuosids is nonunique in general, with overwhelming probability ℓ^1 minimization will recover the specific combination of spikes and sinusoids generating the signal, no matter what the amplitudes and/or polarities in the combination might be.

Definition 3 *We say that* **sets of size k are typically sets of local equivalence***, with typicality coefficient* $(1-\epsilon)$*, when a fraction* $\geq (1-\epsilon)$ *of all possible support sets* I *of size* k *are sets of local equivalence.*

Thus for $A = [IF]$ and d a perfect square, by [10] some sets of size $\sqrt{d}$ are not sets of local equivalence, while by [4], for each $\epsilon > 0$, and for large d, sets of size $k \approx cd/\log(n)$ are typically sets of local equivalence.

5.2 Neighborliness of Intrinsic Sections

Local equivalence can be related to properties of *intrinsic sections* of the quotient polytope $P = AC$, defined as follows. Suppose that P has $2n$ vertices. Pick a set K of k vertices which does not contain an antipodal pair. This set spans a k-dimensional linear subspace V_K of R^d. Consider the section

$$P_K = P \cap V_K;$$

this is a polytope with vertices chosen from among the members of K and their antipodes. We call this an *intrinsic section* because, in distinction to arbitrary or random sections, it involves an intrinsic property of the polytope – the subspaces spanned by vertices; this is independent of the coordinate system. There are exactly as many intrinsic sections as there are subspaces V_K.

We say that such a section P_K is a cross-polytope if it has $2|K|$ vertices and is centrally $|K|$-neighborly. The terminology makes sense, as it forces P_K to be equivalent to a one-one affine transformation of the $|K|$-dimensional regular cross-polytope. We say that a given intrinsic section P_K is *neighborly* if it is a cross-polytope.

The reader may check that if $P = AC$ has $2n$ vertices and every intrinsic section is neighborly, then P is neighborly. Hence behavior uniform across all intrinsic sections offers nothing new; but typical behavior across most sections can be different than uniform behavior. Indeed,

Theorem 2 *Suppose that $P = AC$ has $2n$ vertices. These two properties are equivalent:*

- *Among the k-dimensional intrinsic sections of the quotient polytope P, a fraction $1-\epsilon$ are neighborly.*
- *Among sets of size k, a fraction $1-\epsilon$ are sets of local equivalence.*

In words, when supports of size k are typically sets of local equivalence, then k-dimensional intrinsic sections are typically k-neighborly.

Proof. Observe that, if K is a set of $|K|$ vertices of P not including an antipodal pair, it can be written as $(\pm_i a_i)_{i\in I}$ for some subset I. The subspace V_K actually depends only on the indices $i \in I$ and not on the signs $\pm_i$. By abuse of terminology, we write this subspace as V_I and the corresponding section as P_I. Hence, there is a one-to-one correspondence between the possible k-dimensional intrinsic sections of P and the possible support sets I of size k.

We now apply Lemma 5.1, showing that for each specific subset I there is complete equivalence between local equivalence at I and neighborly section by V_I. □

Lemma 5.1 *I is a set of local equivalence if and only if the corresponding $|I|$-dimensional intrinsic section P_I is a cross-polytope.*

Proof. In one direction, suppose I is a set of local equivalence; we show that P_I is a cross-polytope.

By assumption, any y formed by linear combinations of the columns $(a_i)_{i\in I}$ in fact has a unique representation by such terms; this is obtained by using (P_1). Consider the set Q_I of all such vectors y which can be formed by linear combinations of columns $(a_i)_{i\in I}$ with sums of absolute coefficients at most 1. This is simply the affine image of an I-dimensional cross-polytope. That image must be one-to-one because the elements of Q_I by assumption have unique representations as convex combinations of the columns. Hence Q_I is a cross-polytope. We claim that $Q_I = P_I$, thus proving that P_I is a cross-polytope, and so the section by V_I is neighborly. To establish our claim, note that $Q_I \subset P_I$. Indeed $Q_I \subset V_I$, as Q_I is in the convex hull of the $(\pm_i a_i)_{i\in I}$ At the same time, Q_I is in the convex hull of all the $(\pm_i a_i : 1 \le i \le n)$, placing it inside P. Hence $Q_I \subset V_I \cap P = P_I$. However, $P_I \subset Q_I$; indeed the vertices of P_I are among the signed columns $(a_i)_{i\in I}$, so $\mathrm{vert}(P_I) \subset Q_I$, so $P_I = \mathrm{conv}(\mathrm{vert}(P_I)) \subset \mathrm{conv}(Q_I) = Q_I$.

In the other direction, suppose that P_I is a cross-polytope. We take any x_0 supported in I, generate $y = Ax_0$ and show that the minimum ℓ^1 solution to $y = Ax$ is precisely x_0.

Without loss of generality, assume $\|x_0\|_1 = 1$. Then y is in P_I. In fact, it is on the boundary of P_I, so it lies in some $(|I|-1)$-dimensional face F of P_I. As P_I is a cross-polytope, F is simplicial. Applying now Lemma 3.1, we conclude that y is uniquely representable, among all convex combinations of all signed columns of A, which of course must be as $y = Ax_0$. Hence x_0 must be the unique ℓ^1 solution. □

5.3 Discussion

To compress the above discussion, we propose additional terminology (compare [24]):

Definition 4 *The* **Local Equivalence Breakdown Point** *of the matrix A, $LEBP_\epsilon(A)$, is the largest N so that a fraction $\geq 1-\epsilon$ of all sets I of size $< N$ are sets of local equivalence.*

Clearly $LEBP_0(A) = EBP(A)$.

Definition 5 *We say that the polytope P is (k,ϵ)*-**sectionally neighborly** *if all but a fraction $\leq \epsilon$ of intrinsic k-dimensional sections are k-neighborly.*

Evidently, $(k,0)$-sectionally neighborly is the same as k-neighborly. With this additional terminology, we have the correspondence:

$$LEBP_\epsilon(A) > k \;\Leftrightarrow\; P = AC \text{ is } (k,\epsilon)\text{-sectionally neighborly;}$$

this contains (4.1) as a special case by setting $\epsilon = 0$.

Summarizing the discussion concerning the quotient polytope based on $A = [IF]$ when d is a perfect square, we have, for $\epsilon > 0$ and large d,

$$EBP(A) < \sqrt{d}, \qquad \text{and} \qquad LEBP_\epsilon(A) \geq cd/\log(d).$$

Alternatively, $P = AC$ is not $\sqrt{d}$-neighborly, but it is overwhelmingly likely to be $(cd/\log(d),\epsilon)$-sectionally neighborly for large d. This proves Corollary 1.4.

Section 7 considers the case where A is a high-dimensional random projection, and gives numerical information about neighborliness and sectional neighborliness. In that case also we observe marked differences between neighborliness and sectional neighborliness, although both can be proportional to d.

6 Individual Equivalence and Face Numbers

6.1 Individual Equivalence

We say there is *individual equivalence* (between ℓ^1 and ℓ^0 optimization) *at a specific x_0* when, for that x_0, the result $y = Ax_0$ generates instances of (P_1) and (P_0) which both have x_0 as the unique solution. In such a case we say that x_0 is a *point of individual equivalence.*

While in general the situation of describing such points may be very complicated, under a simplifying assumption a good deal can be said intelligibly. We say that *the columns of A are in general position* in $\mathbf{R}^d$ if there is no nontrivial linear relation $Ax = 0$ where x has fewer than d nonzeros. While this sounds innocuous, it is not universal; when d is a perfect square, it fails for the example $A = [IF]$ discussed above. On the other hand it succeeds for random A.

Under this assumption, the face structure of the quotient polytope is very simple.

Lemma 6.1 *Suppose the columns of A are in general position. Then for $k < d-2$, the k-dimensional faces of $P = AC$ are all simplicial.*

Proof. Suppose, to the contrary, that there were some k-dimensional face of P with more than $k+1$ vertices. As the vertices of P are among the signed columns of A, there would then be $k+2$ signed columns of A lying in some k-flat. If $k < d-2$ this violates general position. Hence there is no such k-dimensional face. Hence the k-dimensional faces have $k+1$ or fewer vertices.

But we can only have a k-dimensional face, if we have at least $k+1$ vertices; hence the face must have precisely $k+1$ vertices, i.e. it is simplicial. □

General position allows equivalence to be checked empirically.

Lemma 6.2 *Suppose that the columns of A are in general position. Suppose that (P_1) has a solution x_1 with fewer than $d/2$ nonzeros; then x_1 is a point of individual equivalence.*

The practical meaning of this lemma: if we know A to be in general position and we find ourselves in a situation where (P_1) has a solution with fewer than $d/2$ nonzeros, then we have actually found the solution of (P_0).

Proof. Indeed, without loss of generality, suppose $\|x_1\|_1 = 1$, otherwise simply rescale. Then $y = Ax_1$ sits on the boundary of P. Let $k+1$ be the number of nonzeros in x_1. The point x_1 sits in the relative interior of a k-face F of the cross-polytope, Lemma 6.3 below shows that AF is a simplicial face of P. Then Lemma 3.1 implies that x_1 must be the unique solution to (P_1).

We claim that x_1 must also be the unique solution to (P_0). Indeed, suppose there were a different solution x_0 to (P_0), also with $k < d/2$ nonzeros. Then we would have $A(x_1 - x_0) = 0$. This would supply a set of fewer than d columns of A obeying a nontrivial linear relation, contradicting general position. □

Lemma 6.3 (Dichotomy) *Suppose the columns of A are in general position. Let F be a k-dimensional face of the cross-polytope $C \subset \mathbf{R}^n$, $k < d/2$. Then*

- *If the relative interior of AF meets the boundary of P, then AF is a simplicial face of P.*
- *If the relative interior of AF meets the interior of $P = AC$ then the relative interior of AF does not meet the boundary of P, and AF is not part of any face of P.*

Proof. This is a consequence of a still simpler fact. Any line L intersects the interior of the polytope P in a relatively open line segment S, say, which may be empty. The closure of this line segment intersects the boundary of P in the relative boundary of S. Such an S cannot contain a relative interior point which is not also an interior point of P.

To apply this, note that general position of the columns of A implies that AF is a simplex. Consider a point y relatively interior to AF. Lemma 6.4 below shows that every relatively interior point z of the simplex AF not equal to y can be represented as a convex combination between y and at most k vertices of F, with positive weights on y and on at least some vertices of F. We can rewrite this representation as

$$z = \alpha y + (1-\alpha)\sum_{i=1}^{k} \beta_i v_i$$

where the v_i are vertices of AF and $\alpha \in (0,1)$, $\beta_i \geq 0$, $\sum_i \beta_i = 1$. This gives us a representation of z as a relatively interior point of the *line segment* from y to a boundary point $b = \sum_{i=1}^{k} \beta_i v_i$ on some $(k-1)$-face of the simplex. Since y is in the relative interior of AF, and the line segment lies in AF, it has a continuation beyond the endpoint y which still lies in $relint(AF)$. Let S denote the maximal such continuation. z and y are both relatively interior points of S. Hence by the previous paragraph, either they are both interior points of P or neither is an interior point of P.

Suppose now that AF contains an interior point of P, call this point y; it is in the relative interior of AF. The above observation shows that all other relative interior points of AF are also interior points. Similarly, if AF contains a relatively interior point y which is not interior to P then no other relative interior points of AF are interior to P. □

Lemma 6.4 *Given a point y in the relative interior of a k-simplex P, any other relatively interior point z can be represented as a convex combination of y with at most k vertices of P, with positive weights on y and on at least some vertices of P.*

Geometrically, the lemma says that we may dissect $P = \text{conv}(v_1, \ldots, v_{k+1})$ into $k+1$ simplices $P_i = \text{conv}(v_1, \ldots, v_{i-1}, y, v_{i+1}, \ldots, v_{k+1})$, with y as a vertex for each one. This is a standard idea in simplicial subdivision. We omit the details.

6.2 Discreteness of Individual Equivalence

While one might imagine that success or failure of individual equivalence varies arbitrarily with the point under consideration, in fact it is much more constrained.

Lemma 6.5 *Suppose that x_0 has fewer than $d/2$ nonzeros. The property of individual equivalence depends only on the support of x_0 and on the sign of x_0 on its support.*

This result must have been implicitly observed by numerous authors, it has also been discussed explicitly – e.g. [10, 20]. We give here a proof illustrating the viewpoint of this article. The proof also builds up a viewpoint needed in the next result. It assumes that the columns of A are in general position.

Proof. Let $y = Ax_0$. Without loss of generality, let $\|x_0\|_1 = 1$. As x_0 has (say) $k+1$ nonzeros, it belongs to the relative interior of a k-face F of C defined by the support of x_0 and sign pattern of x_0 on its support. The Dichotomy Lemma 6.3 says that either AF is a simplicial face of P or $relint(AF)$ is interior to P.

If AF is a face of P, then because AF is simplicial, Lemma 3.1 shows that x_0 is the unique solution to the instance of (P_1) posed by y. As x_0 has fewer than $d/2$ nonzeros, by Lemma 6.2 it is a point of local equivalence.

If AF is not a face of P, the Dichotomy Lemma says that the relative interior of AF is interior to P. At any point interior to P, $val(P_1) < 1$. Hence x_0 is not the solution of (P_1).

These assertions depend only upon whether AF is a face of P or not. They are true simultaneously for all x belonging to the relative interior of a given face F. Hence these assertions only depend on the support and sign pattern of x_0. $\square$

In view of this discreteness, we can define a sensible notion of 'typicality ' for individual equivalence. Let I be a subset of $\{1, \ldots, n\}$ of size k and let $(\sigma_i)_{i \in I}$ be a sequence of signs ± 1. We call (I, σ) a *signed support.* There are $2^k \binom{n}{k}$ signed supports of size k.

Definition 6 *Given a $d \times n$ matrix A, we say that* **a fraction $\geq (1-\epsilon)$ of all vectors x with k nonzeros** *are points of local equivalence if individual equivalence holds for a fraction $\geq (1-\epsilon)$ of all signed supports of size k.*

A practical computer experiment can be conducted to test the typicality of individual equivalence, and has been carried out by numerous authors, including [10, 4, 24]. One generates a sparse vector x_0 with random signs on its support, creates $y = Ax_0$, and solves (P_1). Then one checks whether the solution of (P_1) is again x_0. ϵ can be estimated as the fraction of computer experiments where failure occurs. Experiments of this kind reveal that for $A = [IU]$ with U a random orthogonal $d \times d$ matrix, individual equivalence is typical for $k < .25d$ [10]. Similarly that for $A = [IF]$ with F the real Fourier $d \times d$ matrix, individual equivalence is typical for $k < .25d$ [4]. For random orthoprojectors A, with $n = 2d$ and d large, for $k < .3d$, individual equivalence is typical, while for $k > .35d$, individual equivalence is atypical [24].

It is also worth remarking that *local* equivalence says that for *any* choice of signs on a given support, ℓ^1/ℓ^0 equivalence will hold, while *individual* equivalence says that for a *specific* choice of signs, equivalence will hold. The requirement of uniformity over choices of sign makes local equivalence hard to verify, but also very powerful when it holds.

6.3 Individual Equivalence and Face Numbers

Typicality of individual equivalence can be posed using completely classical language.

Theorem 3 *Let the columns of A be in general position. These statements are equivalent for $k < d/2$:*

- *The face numbers of AC and C agree within a factor $1-\epsilon$:*

$$(1-\epsilon)f_{k-1}(C) \le f_{k-1}(AC) \le f_{k-1}(C).$$

- *A fraction $\ge (1-\epsilon)$ of all vectors with k nonzeros are points of individual equivalence.*

Proof. A given signed support of size k corresponds uniquely to a $k-1$ face F of C. Individual equivalence at the given signed support occurs if and only if AF is a face of P. By (2.2), the faces of P are a subset of the AF where F is a face of C. Hence the identity

$$\frac{\#(\text{signed supports giving equivalence})}{\#(\text{signed supports of size } k)} = \frac{f_{k-1}(AC)}{f_{k-1}(C)}.$$

□

Of course, counting faces of polytopes is an old story. This result points to a perhaps surprising *probabilistic* interpretation. Suppose that the columns of A are in general position. We randomly choose a vector x with $k < d/2$ nonzeros in such a way that all arrangements of the nonzeros are equally likely and all signs on the nonzero coefficients are equally likely. We then generate $y = Ax$. If the quotient polytope P has 99% as many k-faces as C, then there is a 99% chance that x is both the unique sparsest representation of y and also the minimal ℓ^1 representation of y. This is a quite simple and, it seems, surprising outcome from mere comparison of face counting.

However, we should note that counting faces is important for neighborliness also. The companion paper [8] proves that

$$P = AC \text{ is centrally } k\text{-neighborly iff } f_\ell(P) = f_\ell(C), \quad \ell = 0, 1, \dots, k-1.$$

Hence we have the appealing picture that our *strongest*, uniform notion of ℓ^1/ℓ^0-equivalence reduces to *exact equality* of $k-1$ dimensional face numbers while our *weakest*, average-case notion reduces to agreement of those same face numbers to within $(1 \pm \epsilon)$ factors.

6.4 Discussion

To compress the above discussion, we propose additional terminology (compare [24]):

Definition 7 *The* **Individual Equivalence Breakdown Point** *of the matrix A, $IEBP_\epsilon(A)$, is the largest N so that a fraction $\ge 1-\epsilon$ of all x with $< N$ nonzeros are points of local equivalence.*

Obviously $IEBP_0(A) = EBP(A)$.

Definition 8 *We say that the polytope P is* (k, ϵ)**-facially neighborly** *if all but a fraction $\leq \epsilon$ of k-sets of vertices span $k-1$-faces of P.*

Again $(k, 0)$-facial neighborliness is the same as k-neighborliness. With this additional terminology, we have the correspondence that, if $k < d/2$ and if the columns of A are in general position,

$$IEBP_\epsilon(A) > k \;\Leftrightarrow\; P = AC \text{ is } (k, \epsilon)\text{-facially neighborly;}$$

this contains (4.1) as a special case by setting $\epsilon = 0$. We obviously also have

$$EBP(A) \leq LEBP_\epsilon(A) \leq IEBP_\epsilon(A). \tag{6.1}$$

Applying known results in ℓ^1/ℓ^0 equivalence, we immediately get many examples of polytopes with nearly as many faces as the cross-polytope, but with varying degrees of sectional and/or ordinary neighborliness. Thus $A = [IF]$ yields quotient polytopes with $.9\sqrt{d}$-neighborliness and $(cd/\log(d), \epsilon)$-sectional neighborliness. Neither of these is 'proportional to dimension'. However, numerical experiments show that the facial-neighborliness is indeed 'proportional to dimension'; see [4, 24].

7 Randomly-Projected Cross-Polytopes

We now focus attention on the case where A is a uniformly-distributed random orthogonal projection from $\mathbf{R}^n$ into $\mathbf{R}^d$. We suppose throughout that $d = d_n = \lfloor \delta n \rfloor$, and consider the high-dimensional setting $n \to \infty$.

7.1 Relevant Results About ℓ^1/ℓ^0 equivalence

Solution of equations $y = Ax$ with A a random projection has been studied by Candès and Tao, who proved the following:

Theorem 7.1 *[5] Let $A_{d,n}$ be a uniform random projection. Fix $\delta \in (0, 1)$. There is $\rho = \rho(\delta) > 0$ so that for $d \geq \delta n$,*

$$Prob\{EBP(A_{d,n}) > \rho d\} \to 1, \qquad n \to \infty.$$

In short, ℓ^1/ℓ^0 equivalence holds up to a threshold which is proportional to dimension. By Theorem 1, this implies that randomly-projected cross-polytopes are neighborly proportional to dimension.

This result should be compared to Theorem 4.2 above, from [7]; the two results may be regarded as basically interchangeable. The bounds to emerge from the proofs [5, 7] of these results are both quite small. Candès claims to be able to show $\rho > 1/1000$ for moderate δ. Donoho did not even bother to mention to others an estimate deriving from his proof. As we will see, polytope methods can do much better.

In contrast, computational experiments in ℓ^1/ℓ^0 equivalence, described in [24], paint a far rosier picture; for example, they show most x with $\leq .3d$ nonzeros are points of individual equivalence. Moreover, these numerical experiments agree with the theoretical evidence from polytope methods.

7.2 Results about Projections of Polytopes

Over the years, several authors have considered the properties of randomly-projected polytopes, in particular their face numbers. [25] and [1] considered randomly-projected simplices, and [2] considered randomly-projected cross-polytopes. The different authors used different assumptions on n and d, and developed a variety of useful tools and representations. For example, Affentranger and Schneider [1] and Böröczky and Henk [2] considered d fixed, $n \to \infty$ while Vershik and Sporyshev pioneered the n-proportional-to-d case.

(Aside: the author of *this* article considers it remarkable that *anyone* was interested in random projections of cross-polytopes or in the high-dimensional n proportional to d case prior to this article. One of the main purposes here is to explain why study of projected cross-polytopes might be important. Apparently there are far-seeing researchers who need no such motivation!)

In [8], Donoho studied A which are uniformly-distributed random orthogonal projections from $\mathbf{R}^n$ into $\mathbf{R}^d$, in the proportional-to-dimension case where $k \sim \rho d$ and $d \sim \delta n$. He made various asymptotic estimates of face numbers of $P = AC$. In some sense this is a cross-breeding of the problem posed by Böröczky and Henk, with the high-dimensional asymptotic of Vershik and Sporyshev.

Fix $\epsilon > 0$. [8] describes a lower bound $\rho_N(\delta)$ on the neighborliness threshold. For $\rho < \rho_N(\delta)$. there is overwhelming probability for large n that $P = AC$ is $\lfloor \rho d \rfloor$-neighborly. It also describes a lower bound $\rho_S(\delta)$ on the sectional neighborliness threshold. For $\rho < \rho_S(\delta)$. with overwhelming probability for large n, most ρd-sections are ρd-neighborly. The bound $\rho_S(\delta)$ is substantially larger than $\rho_N(\delta)$, signalling that for these random polytopes, *typical* sectional neighborliness may be substantially better than *uniform* sectional neighborliness.

The paper also studies the face numbers $f_k(AC)$, asking for the range of k so that these face numbers are approximately the same as those of C. [8] derives a threshold $\rho_F(\delta) > 0$ so that for $\rho < \rho_F(\delta)$, the $\lfloor \rho d \rfloor$-dimensional face numbers of AC are the same as those of C, to within a factor $(1 + o(1))$. Informally, this describes the fraction k/d at which 'phase transition' occurs from (k, ϵ)-facial neighborliness to non-neighborliness.

Because of (6.1), we know that $\rho_N(\delta) \le \rho_S(\delta) \le \rho_F(\delta)$. Fixing some small $\epsilon > 0$, we have with overwhelming probability for large d that

$$\begin{aligned} P = AC \quad \text{is} \quad & (\tilde{\rho}_N \cdot d)\text{-neighborly, and} \\ & (\tilde{\rho}_S \cdot d, \epsilon)\text{-sectionally neighborly, and} \\ & (\tilde{\rho}_F \cdot d, \epsilon)\text{-facially neighborly;} \end{aligned}$$

here $\tilde{\rho}_N \equiv \rho_N(\delta) - \epsilon$, $\tilde{\rho}_S \equiv \rho_S(\delta) - \epsilon$, and $\tilde{\rho}_F \equiv \rho_F(\delta) - \epsilon$ obey

$$0 < \tilde{\rho}_N \approx \rho_N(\delta) < \tilde{\rho}_S \approx \rho_S(\delta) < \tilde{\rho}_F \approx \rho_F(\delta).$$

Some numerical information is provided in Table 1 below. Two key points emerge:

- ρ_N, the smallest of the three, is still fairly large, perhaps surprisingly so. While it tends to zero as $\delta \to 0$, it does so only at a logarithmic rate; and for moderate δ it is on the other of .1. This justifies our labelling prior results, like $\rho = 1/1000$ in Theorem 7.1 'small'.

- ρ_F, the largest of the three, is substantially larger than either ρ_N or ρ_S. The fact that it 'crosses the line $\rho = 1/2$' for δ near .75 is noteworthy; this is the source of Corollary 1.5.

Table 1: Numerical Results Using Methods from [8].

	$\delta = .1$	$\delta = .25$	$\delta = .5$	$\delta = .75$	$\delta = .9$
ρ_N	.049	.065	.089	.117	.140
ρ_S	.10	.14	.19	.24	.29
ρ_F	.18	.25	.38	.52	.66

7.3 Probabilistic $\ell^1 - \ell^0$ Equivalence

We now interpret the thresholds ρ_F, ρ_C and ρ_N through the lens of ℓ^1/ℓ^0 equivalence.

To interpret ρ_F, fix $\epsilon > 0$ and suppose that $k = k(n)$ obeys $k/d < \min(\rho_F(\delta), 1/2) - \epsilon$. Let x_0 have k nonzeros at sites chosen uniformly at random, and also with signs chosen uniformly at random. We generate the corresponding $y = Ax_0$. With overwhelming probability for large n, the minimal ℓ^1-norm solution of $y = Ax$ is also the sparsest solution, and both are equal to x_0. Since $\rho_F(\delta) > 1/2$ for $\delta > \delta_0$, and we may take $\epsilon < .01$, this proves Corollary 1.5.

To interpret ρ_S, we now assume the more restrictive condition that $k/d < \rho_S(\delta) - \epsilon$. Let x_0 have k nonzeros at sites chosen uniformly at random; the signs need not be chosen at random. We again have ℓ^1-ℓ^0 equivalence with overwhelming probability for large n. However, the sparsity requirement is higher than before, as $\rho_S(\delta) < \min(\rho_F(\delta), 1/2)$.

Turning now to ρ_N, we suppose that $k/d < \rho_N(\delta) - \epsilon$. Let x_0 have k sites which need not be chosen at random; the amplitudes and signs of the nonzeros need not be chosen at random. We again have ℓ^1-ℓ^0 equivalence with overwhelming probability for large n. This time, the sparsity requirement is even higher, as $\rho_N(\delta) < \rho_S(\delta)$.

Hence the difference between ρ_F, ρ_S, and ρ_N indicates the 'price we pay' for not having random signs and/or locations for the nonzeros in x. In this sense, all three natural measures of neighborliness of polytopes appear to have their own unique and interesting implications for finding sparse solutions of linear equations.

8 Perfect Signal Recovery Despite Malicious Errors

So far our discussion has been abstract and general. We now provide a brief example which may help motivate the interest in, and applicability of such results, and the appearance of numerous papers about ℓ^1/ℓ^0 equivalence in Information Theory and Signal Processing journals.

For n large and divisible by 4, generate a random orthogonal n by n matrix U. Form a d by n matrix A with $d = \lfloor 3n/4 \rfloor$ by taking the first d rows of U. At the same time generate its m by n orthocomplement B by taking the last $m = n - d$ rows of U.

The matrix B can be used as part of a stylized communications scheme which is highly robust against gross errors. Suppose we wish to communicate a block of m pieces of information (numbers) to a remote receiver. We represent this information as an m-vector α, and we transmit the noiseless signal $S = B^T \alpha$. The receiver gets a distorted signal $R = B^T \alpha + z$. In digital communications terminology, B defines a 'constellation', but this time in the space $\mathbf{R}^n$ of n-blocks rather than in the so-called I, Q-plane.

The receiver solves the minimum ℓ^1 problem

$$(L_1) \qquad \min_a \|R - B^T a\|_1.$$

Let $\hat{\alpha}$ denote any minimizer.

It turns out that if a majority of the distortions in z are zero then typically, there is *perfect recovery*: $\hat{\alpha} = \alpha$, with no errors, despite the fact that there is no constraint on the size of the distortion z; in particular it can be arbitrarily more energetic than the undistorted signal. This reflects the strong nonlinearity of (L_1); it is able to preserve weak signals in the presence of strong distortion.

Corollary 8.1 *Let the distortion z have at most $k_{rss} = .49n$ nonzeros, and suppose that the positions and signs of those nonzeros are random and uniformly distributed (the amplitudes of the nonzeros can be chosen in a signal dependent, in fact malicious way). With overwhelming probability for large n*

$$\hat{\alpha} = \alpha.$$

Here the probability refers to the joint distribution on matrices B, and signed support of the distortion. Note that the amplitudes of the nonzeros can depend on the signal and in fact can be much larger than the transmitted signal. Informally, this communications scheme is robust against $.49n$ overwhelming errors with random sites and polarities. The subscript rss reminds us of the random sites/signs constraint.

Suppose now that the distortion is not random, but is chosen by a malicious opponent with knowledge of the constellation matrix B, and even of the transmitted signal S. However, the distortion must be nonzero in a controlled number k_{mal} of sites.

Corollary 8.2 *With overwhelming probability for large n, the random matrix B has the property that, for every disturbance z containing at most $k_{mal} = .11n$ nonzeros.*

$$\hat{\alpha} = \alpha.$$

Informally, this communications scheme is robust against $.11n$ maliciously-chosen gross errors. Here we mean that both the sites of the errors, and even the values at those sites can be chosen in a signal-dependent way – for example, they can occur in bursts.

Candès recently informed the author of results with Tao [6] along the lines of Corollary 8.2; however, seemingly without comparably strong information about the size of the allowable degree of malicious contamination k_{mal}. The above Corollaries complement [6] by giving precise and perhaps unexpectedly strong quantitative information and by describing behavior with both randomly-sited and maliciously-sited distortions. An earlier set of 'perfect recovery' results using a matrix B of sinusoids rather than a random matrix B, allowed about n/π nonzeros in the distortion but required that the nonzeros obey stronger pattern restrictions; see [11].

These Corollaries follow easily from the polytope neighborliness results discussed in the previous section. The constants $m = n/4$ and $k = .11n$ actually derive from Table 1.

To see this, define $y = AR$ and consider the problem (P_1). $y \in \mathbf{R}^d$ and $x \in \mathbf{R}^n$ as usual.

Lemma 8.1 *The solution sets of (L_1) and (P_1) are in one-to-one correspondence. If x_1 is a solution to (P_1) then $\hat{\alpha} = B(R - x_1)$ is a solution to (L_1) and vice-versa.*

Proof. The matrices A and B are complementary orthoprojectors. Hence a given vector R in $\mathbf{R}^n$ is uniquely determined by the pair AR, BR. Now if x is feasible for (P_1), $Ax = y = AR$. Rewriting $a = B(R - x)$ as $BR = a + Bx$, and noting that $AB^T = 0$ we have

$$R = A^T AR + B^T BR = A^T Ax + B^T (Ba + Bx) = B^T a + x.$$

So an $x \in \mathbf{R}^n$ feasible for (P_1) is in one-to-one correspondence with an $a \in \mathbf{R}^m$ so that $R = B^T a + x$. In short we can make an identification of variables $z = x$; noting that the objective of (L_1) is in fact $\|z\|_1$ while that of (P_1) is $\|x\|_1$, we see that the two problems are in complete correspondence, both as regarding the set optimized over and the value of the objective on the feasible set. $\Box$

We now apply our earlier discussion of (P_1). Suppose that the received signal $R = B^T\alpha + z$ suffered a disturbance z with k nonzeros. Applying Lemma 8.1, $y = AR$ is then representable using an x with k nonzeros; where actually the representing $x = z$. If $\|z\|_0 = k < EBP(A)$, then the minimal ℓ^1 solution to $y = Ax$ is unique and equal to the sparsest solution, which has $x_1 = z$. But then the solution $\hat{\alpha}$ to (L_1) is unique and equal to $B^T(R - x_1) = B^T(R - z) = \alpha$. In short there is perfect recovery.

Turning now to Table 1, we read off that, for $\delta = .75$, $\rho_N(\delta) \approx .014$. Thus with overwhelming probability for large n, a $d \times n$ random orthoprojector with $d = d_n$ and $d/n \sim 3/4$ yields a $.11n$-neighborly quotient polytope for large n. Applying Theorem 1, there is overwhelming probability for large n that $EBP(A) \geq .11n$. Corollary 8.2 follows.

A similar argument holds for Corollary 8.1. We simply look in Table 1, note that $\rho_F(\delta) \approx .52 > .5$, and apply the appropriate Theorems and Lemmas above.

Other results are possible for other combinations of (d, n, m) by choosing other ratios $\delta = d/n$ and looking up the corresponding result either in Table 1 if it is listed here, or else in [8] if it is not.

Observe the connection between polytope theory and signal recovery. Facial neighborliness translates into immunity to gross errors at positions and signs chosen at random, while central neighborliness translates into immunity to gross errors even when the positions and signs are chosen maliciously.

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

[4] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. Technical report, Applied and Computational Mathematics, California Institute of Tecnology, 2004.

[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 Tecnology, 2004.

[6] Emmanuel J. Candès and Terence Tao. Decoding via linear programming. Technical report, Applied and Computational Mathematics, California Institute of Tecnology, in progress.

[7] 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.

[8] David L. Donoho. High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension. Technical report, Department of Statistics, Stanford University, 2004.

[9] David L. Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization. *Proc. Natl. Acad. Sci. USA*, 100(5):2197–2202 (electronic), 2003.

[10] David L. Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory*, 47(7):2845–2862, 2001.

[11] David L. Donoho and Benjamin F. Logan. Signal recovery and the large sieve. *SIAM J. Appl. Math.*, 52(2):577–591, 1992.

[12] Michael Elad and Alfred M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Trans. Inform. Theory*, 48(9):2558–2567, 2002.

[13] Aryeh Feuer and Arkady Nemirovsky. On sparse representation in pairs of bases. *IEEE Trans. Inform. Theory*, 49(6):1579–1581, 2003.

[14] Jean-Jacques Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Trans. Inform. Theory*, 50(6):1341–1344, 2004.

[15] David Gale. Neighboring vertices on a convex polyhedron. In *Linear inequalities and related system*, Annals of Mathematics Studies, no. 38, pages 255–263. Princeton University Press, Princeton, N.J., 1956.

[16] David Gale. Neighborly and cyclic polytopes. In *Proc. Sympos. Pure Math., Vol. VII*, pages 225–232. Amer. Math. Soc., Providence, R.I., 1963.

[17] Rémi Gribonval and Morten Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, 49(12):3320–3325, 2003.

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

[19] Frank H. Lutz. Triangulated manifolds with few vertices: Centrally symmetric spheres and products of spheres. Technical report, Technische Universität Berlin, 2004.

[20] Dmitry M. Malioutov, Müjdat Çetin, and Alan S. Willsky. Optimal sparse representations in general overcomplete bases. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, May 2004, Montreal, Canada.*, 2004.

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

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

[23] Joel A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, 2004.

[24] Yaakov Tsaig and David L. Donoho. Breakdown of local equivalence between sparse solution and ℓ^1 minimization. Technical report, Department of Statistics, Stanford University, 2004.

[25] A. M. Vershik and P. V. Sporyshev. Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem. *Selecta Math. Soviet.*, 11(2):181–201, 1992. Selected translations.

[26] Günter M. Ziegler. *Lectures on polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.